OMBRES ET LUMIERES DE L'ECONOMIE POLITIQUE

Paul Jael

L'économie politique, science ou philosophie ?

AVERTISSEMENT CONCERNANT LA NOTATION

L'ouvrage comportant pas mal de graphiques et d'équations, quelques précisions sur la notation s'imposent. Généralement, les significations des variables et courbes sont expliquées dans le texte. Toutefois, certaines abréviations reviennent couramment :
Pour la recette, le coût et la prodictivité, la deuxième lettre est « T » pour *total(e)*, « M » pour *moyen(ne)* et « m » pour *marginal(e)*. Ainsi, RT = recette totale, PM = productivité moyenne et Cm = coût marginal. Les lettres « O » et « D » accompagneront généralement les courbes d'*offre* et de *demande*. Les lettres « Q » et « P » sur l'abscisse et l'ordonnée signifient *quantité* et *prix*.

Les *vecteurs* sont représentés par une lettre minuscule en caractère gras, les *matrices* par une lettre majuscule en caractère gras.

Dans les systèmes d'équations, pour ne pas écrire toutes les équations, dont le nombre est d'ailleurs indéterminé en pratique sinon en théorie, une ligne « :: » entre deux équations représente l'ensemble des équations du système qui ne sont pas indiquées.

Il serait souhaitable de pouvoir désigner une même variable par une même lettre tout au long de l'ouvrage. C'est malheureusement impossible. Beaucoup de variables sont déclinées de multiples façons par les différents économistes. Il faudrait beaucoup plus de 26 lettres pour établir une codification unique de ces variables et elle serait peu intuitive. J'ai tenté dans la mesure du possible de garder les codifications originales des auteurs à la source, sauf quand elles me semblaient pouvoir prêter à confusion.

Pour les distinguer clairement du texte, les variables qui y apparaissent sont en italique, lorsqu'elles consistent en une lettre latine. Lorsqu'elle sont affectées d'un indice ou lorsqu'elles n'appartiennent pas à l'alphabet latin, elles ne sont pas en italique.

TITRE I- INTRODUCTION

Ici commence un récit de voyage. Un voyage de plus de dix ans. Un voyage dans une contrée qui n'est bien connue que de ses habitants. Ce pays étrange, c'est l'économie politique. On ne connaît bien un pays qu'en le visitant. Pendant plus de dix ans, mon hobby a été de lire les grands classiques écrits par des économistes de la fin du XVIIIe siècle à la fin du XXème.

Je n'ai pas regretté cette immersion. J'ai découvert un monde ; un monde assez dépaysant pour qui ne connaît que ce côté-ci de la frontière. Indéniablement, j'ai été surpris, ce qui est toujours fascinant. Cette surprise est à l'origine du présent ouvrage.

L'apprentissage s'accompagne souvent d'un certain étonnement, mais ici j'avais l'impression de rencontrer un problème inattendu. Les économistes ont construit en pensée un monde abstrait qui repose sur un assez petit nombre de principes, dont l'équivalence avec le monde connu est assez floue. Je me permets de considérer cette dichotomie comme une anomalie. Il m'est donc venu l'idée peut-être iconoclaste de mettre le sujet sur la place publique. Vu le prestige que confère le statut de science à l'économie politique et même le réel pouvoir social qui en découle, cette question me paraît d'intérêt général.

Suis-je entré dans la peau de l'enfant qui dit « le roi est nu » ? Dans la peau d'un lanceur d'alerte ? « Le roi est nu » : que vaut cette petite phrase dans le cas qui nous occupe ? Le prestige de la science économique dans l'opinion publique évoque la force royale. Sans doute ce roi n'est-il pas complètement nu, mais sa garde-robe comporte quelques habits bizarres. L'économie politique est toute empreinte de rationalité et de logique. La capacité des économistes de mettre n'importe quel problème en équations est impressionnante. Mais le rapport de certains principes avec la réalité paraît douteux.

J'imagine que l'économie politique, comme d'autres savoirs, connaît peu d'incursions extérieures de ce type. Les économistes arpentent leur science, les autres restent à l'extérieur et soit en ignorent tout soit se contentent d'une photo aérienne. Grâce à la palissade qui l'entoure, l'économie politique peut faire passer à l'extérieur l'image harmonieuse qui lui convient.

Certes, l'histoire de la science économique a connu pas mal de débats et il y en a encore. Mais depuis plus d'un siècle a émergé une orthodoxie. Celle-ci fut confrontée à plusieurs remises en cause, mais elle est suffisamment implantée pour garder les controverses sous un couvercle et les empêcher de développer leur potentiel. Cette affirmation sera plus claire dans les chapitres qui suivent. Le face-à-face macroéconomique entre keynésiens et néoclassiques fait figure d'exception. Depuis 1936, la macroéconomie est divisée en deux camps qui s'affrontent sans relâche. Cette opposition fondamentale fait maintenant partie du paysage. On pourrait parler d'une double orthodoxie[1]. Par contre, le triomphe de l'orthodoxie néoclassique a rendu la microéconomie plutôt monolithique, bien que toutes les questions soulevées n'aient pas reçu une réponse satisfaisante.

Cet ouvrage n'a pas l'ambition d'énoncer de nouvelles théories ou d'apporter de nouvelles démonstrations que telle théorie existante est erronée. Il a pour objectif de mettre sur la place publique certaines contradictions ; il en existe de deux types :

- Des controverses entre économistes qui ont été mal refermées, dont la clôture à laissé subsister des questions non résolues ;
- Des hypothèses de l'orthodoxie qui méritent d'être mieux connues. Certaines d'entre elles surprendront sans doute le lecteur autant que moi. Les économistes y sont tellement habitués qu'ils sont vaccinés contre la surprise. De plus, ils ont leurs propres raisons

[1] La doctrine keynésienne garde toutefois certaines caractéristiques d'une hétérodoxie, car elle est moins en phase avec l'orthodoxie microéconomique que sa rivale.

d'apprécier ces hypothèses : elles aident à la modélisation. Comme on le sait, la fin justifie les moyens.

Beaucoup a été écrit ou dit en relation avec un défaut souvent imputé à cette science : le peu de cas qu'elle ferait de l'humain. On lui a aussi reproché un caractère « antisocial » : elle servirait à justifier le système du tout-au-marché et les réformes des gouvernements qui mettent à mal certains droits sociaux. Une telle discussion pourrait être intéressante, mais ce n'est pas l'objet du présent ouvrage. C'est le réalisme économique comme tel qui est mis en question.

La critique n'est que l'une des facettes du présent ouvrage. La part quantitativement la plus importante (Titres II à IX) est constituée d'un exposé plus ou moins objectif des principales théories ainsi que des principaux débats. Ces chapitres exposent plus qu'ils ne commentent, mais ils participent néanmoins à la critique : les théories parlent d'elles-mêmes. Simplement, je voudrais inciter le lecteur à ne pas éteindre son sens critique face à plus savant que lui. S'il accepte ce rôle, il fera la moitié du travail de critique à ma place. Sa tâche sera facilitée par la place que j'accorde aux objections des autres économistes aux théories étudiées. En outre, je glisserai dans l'exposé quelques commentaires personnels qui ont un caractère ponctuel ; je réserve une critique plus générale pour la conclusion.

L'exposé des théories prend la forme d'une suite de résumés de livres et d'articles, organisée par écoles de pensée, par grands débats ou par grands thèmes, avec une dose d'ordre chronologique. Vu l'ampleur des œuvres analysées, la concision s'est imposée. Il a fallu départager les idées à approfondir de celles qui sont simplement survolées voire omises. J'ai éliminé des résumés les idées qui auraient fait figure de répétition ou qui, au contraire, seront répétées avec plus de force ultérieurement. De cette façon, l'emboitement des résumés constitue une espèce de manuel sans trop de redondance.

Quel type de lectorat serait intéressé par un tel ouvrage ? J'aurais souhaité qu'il puisse intéresser tous les intellectuels, mais la densité d'informations qu'il contient le rend indigeste à un public non averti. L'audience naturelle est donc celle des étudiants en sciences économiques. Mais pourquoi achèteraient-ils un livre écrit par un amateur alors que tant de docteurs publient sur le sujet ? L'atout de ce livre me paraît résider dans sa spécificité :

- L'amateur se sent moins engagé par le point de vue orthodoxe qui, inévitablement, régente un peu les usages dans la profession. Pour le découvreur extérieur, rien ne va de soi.
- Un condensé de résumés des principaux classiques de la littérature économique avec force citations, ce travail est quasiment sans concurrence d'après mes recherches sur les étalages des librairies et les étagères des bibliothèques.

Pourquoi avoir consacré autant de pages à résumer des théories compliquées, pourquoi ne pas s'être contenté d'une critique en règle de théories que le lecteur aura apprises ailleurs ? La première raison est tout simplement l'intérêt inhérent à ces résumés ; j'ose espérer qu'ils aideront les étudiants. Les trois autres raisons sont liées et concernent leur appui à la critique :

- Si l'ouvrage avait été conçu dans un sens plus vulgarisateur, son pouvoir critique eût été déforcé ; le lien entre chaque critique et sa cible doit être clair.
- Les critiques sont basées sur un socle de connaissance. La mise au point concernant son périmètre clarifie la perception par le lecteur.
- Je dois répéter que certaines hypothèses explicites ou implicites sont étonnantes. La meilleure façon de les mettre en évidence est d'exposer la théorie elle-même.

Depuis près d'un siècle, les mathématiques ont envahi la science économique. Les théories tendent à devenir des théorèmes. Elles en ont donc les trois composantes habituelles : les hypothèses, la thèse et la démonstration. La démonstration concentre les quatre cinquièmes du

développement mathématique. Le présent ouvrage a pour principe d'exposer les hypothèses et la thèse, généralement en notation mathématique, mais en escamotant la démonstration ; le cas échéant, les grandes lignes de la stratégie de démonstration sont mentionnées. Cette façon de procéder raccourcit et simplifie l'exposé. Elle se justifie par le fait que la démonstration est un exercice avant tout mathématique. L'absence de démonstration ne retire aucun contenu économique à l'ouvrage. La démonstration n'est pas non plus un élément crucial dans la critique de l'économie politique. On peut faire confiance aux économistes pour produire des démonstrations mathématiquement correctes. L'important, en ce qui nous concerne, est de savoir ce qui a été prouvé et sur base de quelles hypothèses.

*

Répétons-le, l'objet de ce livre n'est pas l'économie réelle, pas même les commentaires avisés que les économistes ont pu émettre sur l'économie réelle lorsque le politique ou la presse les interrogeaient. Comme toute science, l'économie politique élabore des concepts et énonce des lois qui les mettent en relation. C'est cet appareil théorique qui nous intéresse ici. Depuis plus d'un siècle, les économistes ont conscience de participer à une œuvre collective : la contribution de chacun est une brique dans une construction globale qui ne peut être posée sans penser à l'architecture générale.

De mes lectures transparaît une certaine fascination exercée par la physique sur les économistes. Ils lorgnent vers la reine des sciences avec envie et en font leur modèle. Pour les économistes, le statut de science de l'économie politique ne fait aucun doute. Cette question mérite une discussion ; elle serait prématurée dans la présente introduction et je la réserve pour la conclusion.

TITRE II- L'ECONOMIE POLITIQUE PRE-MODERNE

2.1- LES PRECURSEURS

Au XVIIe et surtout au XVIIIe siècle, apparaît une nouvelle espèce de théoriciens. On connaissait depuis l'antiquité les philosophes, depuis le moyen âge les théologiens, depuis la renaissance les physiciens. Voici que des intellectuels publient des ouvrages exclusivement centrés sur les problèmes économiques. Auparavant, les questions économiques étaient confinées dans quelques digressions perdues dans les ouvrages des philosophes et des théologiens, comme saint Thomas d'Aquin. Leur point de vue était uniquement normatif.

Ce n'est pas un hasard si une étude « professionnelle » et à vocation scientifique des problèmes économiques émerge à cette époque. Le temps est aux mutations. Ce n'est pas encore la révolution industrielle, mais celle-ci fut précédée de ce qu'on a appelé la révolution commerciale et la révolution agricole. Le moyen âge est déjà loin ; les campagnes ont changé ; l'agriculture est pénétrée par l'échange monétaire depuis longtemps. La bourgeoisie connaît un grand essor avec les manufactures et le commerce international. L'évolution technique s'accélère. L'horizon géographique du monde s'élargit. Au XVIIIe siècle, l'agriculture connaît des progrès importants ; nouvelles cultures (pomme de terre), intensification, nouveaux engrais. En Angleterre, elle subit une mutation sociale avec l'avènement de la grande propriété, qui jette sur le marché du travail une masse de petits paysans sans terre.

Parallèlement, les institutions politiques ont changé. La monarchie absolue limite le pouvoir de la noblesse[1] et favorise ainsi l'essor de la bourgeoisie. Elle y est d'ailleurs intéressée, car les Etats « absolutistes » sont toujours en guerre entre eux et les souverains pressentent l'avantage que leur procure dans cette guerre permanente la puissance économique. Les premiers économistes trouveront chez eux une oreille attentive.

2.1.1. Les mercantilistes

On les retrouve principalement en France et en Angleterre, accessoirement en Allemagne. En France, les principaux noms sont : Jean Bodin (1530-1596), Jean-Baptiste Colbert (1619-1683) et John Law (1671-1629) ; en Angleterre : Thomas Mun (GB :1571-1641), John Locke (GB :1632-1704) et William Petty (GB :1623-1687).

Leur thèse centrale est bien connue. On peut la résumer très simplement : la richesse d'une nation dépend de la quantité de monnaie qui y circule, ce qui à l'époque, revenait au même que le stock d'or et d'argent. Les Etats doivent donc mener les politiques qui ont pour effet que l'or afflue sur leur territoire et n'en sorte pas. La meilleure façon d'y parvenir est d'exporter beaucoup et d'importer peu, autrement dit d'avoir une balance commerciale excédentaire. On trouve dans l'arsenal de mesures préconisées : les barrières douanières à l'entrée de produits finis, l'encouragement des industries exportatrices, les mesures populationnistes qui maintiennent les salaires à bas niveau… Les colonies sont conçues comme de simples pourvoyeuses de matières premières privées du droit d'exporter des produits finis. La conception des rapports internationaux qui sous- tend ces mesures est forcément de nature conflictuelle : l'Etat qui aura accumulé le plus d'or acquerra une certaine maîtrise internationale.

Pareille politique heurte la conception contemporaine de la rationalité économique et les mercantilistes auront plutôt mauvaise presse auprès de leurs successeurs classiques et de tous les économistes favorables au laissez faire. Adam Smith, l'un de leurs principaux détracteurs,

[1] En dernier ressort, le pouvoir monarchique demeurait le garant des intérêts de la noblesse, mais elle s'est arrogé l'exclusivité de la gestion de cet intérêt collectif.

est l'inventeur du terme « mercantilisme ». Le mercantilisme n'est pourtant pas une véritable école de pensée ; les auteurs ne s'inscrivent pas dans une démarche collective ; ils expriment simplement des idées en phase avec l'évolution historique de leur époque, traduisant l'ascendant des classes des grands marchands et des hauts fonctionnaires.

Le mercantilisme est plus qu'une idée ; il fut effectivement pratiqué en France et en Angleterre aux XVIIe et XVIIIe siècles. Colbert, le ministre des finances de Louis XIV, pratiqua cette politique à tue tête en même temps qu'il la théorisa. En Angleterre, les résultats furent plutôt bons : le mercantilisme contribua à amasser l'argent qui financera par la suite la révolution industrielle. En France, il concourut à la décadence de l'ancien régime et à l'affaiblissement du pays. Le petit artisanat et l'agriculture furent délaissés au profit des grandes manufactures et du commerce international.

A côté de l'obsession de la balance des paiements, la science économique de l'époque balbutie déjà quelques idées promises à un meilleur avenir. Les mercantilistes, ou certains d'entre eux, comprennent que la valeur de la monnaie est inversement proportionnelle à la quantité qui circule et qu'une monnaie plus abondante entraîne un niveau des prix plus élevé. Bodin explique l'inflation en Europe au XVIe siècle par l'afflux de métaux précieux en provenance des colonies espagnoles d'Amérique. Cette compréhension est paradoxale, car elle entre en contradiction avec la politique préconisée. Une balance commerciale systématiquement excédentaire doit engendrer une pléthore monétaire, avec comme conséquence un niveau de prix plus élevé qu'à l'étranger, ce qui désavantagera les exportations et stimulera les importations. Bref, l'excès d'exportations secrète les conditions de sa correction. Les mercantilistes connaissent les prémisses de ce mécanisme mais ne peuvent aboutir à la conclusion.

Peut-être ce paradoxe s'explique-t-il par la croyance répandue parmi les mercantilistes, qu'un afflux d'or peut avoir un effet stimulant sur l'activité industrielle. Des prix plus élevés entraînent des profits plus élevés, du moins si les salaires sont stables ; en outre, l'abondance de l'argent pourrait réduire le coût de l'emprunt.

CANTILLON ET LA CIRCULATION MONETAIRE

Richard Cantillon (1680-1734), banquier parisien d'origine irlandaise, rédigea son « Essai sur la nature du commerce en général » aux environs de 1730 mais l'ouvrage ne fut publié qu'en 1755 à titre posthume. Pressman écrit à son sujet : « His *Essay* can legitimately be regarded as the first real economic treatise. It envisioned the economy as an interrelated system, and explained how this system worked. For this reason, Cantillon probably deserves to be regarded as the first real economist »[1]. Tant d'enthousiasme doit être tempéré, car si Cantillon fut très inspiré, son exposé restait généralement flou, voire contradictoire.

Cantillon a l'intuition d'un équilibre général de l'économie où les goûts des consommateurs règlent l'importance relative des différentes branches de production au sein de la société. Les entrepreneurs assurent la jonction entre l'offre et la demande et assument le risque inhérent à l'imprévisibilité de cette dernière. Lorsque l'offre d'un produit est adéquatement réglée sur sa demande, c'est-à-dire lorsqu'il n'y a ni pénurie ni pléthore accidentelles ou temporaires, son prix correspond à sa « valeur intrinsèque ». Cantillon conçoit celle-ci comme étant « la mesure de la quantité de terre & du travail qui entre dans sa production »[2]. Mais il estime que dans cette expression, le travail peut être réduit à la quantité de terre nécessaire pour faire vivre le

[1] Pressman [289] p. 10
[2] Cantillon [45] p. 36

travailleur et sa famille ; donc, en définitive, toute la valeur peut être estimée simplement en quantité de terre. Cantillon pense qu'« en général, (les) prix ne s'écartent pas beaucoup de la valeur intrinsèque »[1].

La réputation de Cantillon lui vient principalement de ses réflexions sur la monnaie. Il écrit : « La proportion de la quantité d'argent comptant nécessaire pour la circulation d'un Etat n'est pas une chose incompréhensible & cette quantité peut être plus grande ou plus petite dans les Etats suivant le train qu'on y suit & la vitesse des paiements »[2]. Autrement dit, elle dépend de la somme des paiements qu'on fait dans le pays, mais également de la vitesse de sa circulation, en ce sens qu'« une accélération ou une plus grande vitesse dans la circulation de l'argent du troc[3], vaut autant qu'une augmentation d'argent effectif, jusqu'à un certain degré ».

Le système bancaire permet l'augmentation de la vitesse de circulation de l'argent, car il empêche l'argent d'être oisif. Par contre la thésaurisation des particuliers « dans leurs poches ou dans leurs caisses, contre les cas imprévus et pour n'être point à sec » diminue cette vitesse.

La masse monétaire en circulation peut augmenter de diverses façons ; Cantillon mentionne :
- La découverte de mines d'or ou l'augmentation de la production des mines existantes
- La balance commerciale excédentaire avec l'étranger,
- Les dépenses des résidents ou des voyageurs étrangers,
- Les investissements et prêts venant de l'étranger.

« Je conclus qu'une augmentation d'argent effectif dans un Etat y introduit toujours une augmentation de consommation & l'habitude d'une plus grande dépense. Mais la cherté que cet argent cause, ne se répand pas également sur toutes les espèces de denrées & de marchandises, proportionnément à la quantité de cet argent, à moins que celui qui est introduit ne soit continué dans les mêmes canaux de circulation que l'argent primitif, c'est-à-dire à moins que ceux qui offroient aux marchés une once d'argent ne soient les mêmes & les seuls qui offrent maintenant deux onces depuis que l'argent est augmenté du double de poids dans la circulation, ce qui n'arrive guère »[4]. Le scénario décrit par Cantillon est celui-ci : la nouvelle monnaie circule de poche en poche, comme un flux qui se ramifie progressivement dans des ruisseaux de plus en plus nombreux mais de plus en plus étroits. Les premiers servis en nouvelle monnaie augmentent vraiment leur consommation ; leur demande croissante élève les prix demandés par leurs fournisseurs. Au fur et à mesure que progresse l'irrigation de l'économie par la nouvelle monnaie, de plus en plus de prix et de coûts s'élèvent. Les derniers servis seront perdants : peu gâtés par la manne monétaire, ils devront faire face à la hausse des prix touchant la plupart de leurs achats ; la production aura été détournée pour satisfaire prioritairement les besoins des premiers servis, qui seront les principaux bénéficiaires. On appelle aujourd'hui *effet Cantillon* la reconnaissance du fait que les modalités d'injection de monnaie additionnelle en affectent le résultat. L'explication en est que, selon son origine, la nouvelle monnaie afflue dans des groupes sociaux différents dont les comportements somptuaires diffèrent[5].

---

[1] Cantillon [45] p. 157
[2] Cantillon [45] pp. 171 et 212
[3] Sous la plume de Cantillon, le terme *troc* n'a pas son acception moderne, celle d'un échange non monétaire. Il s'agit ici simplement d'un synonyme pour échange commercial.
[4] Cantillon [45] p. 238. L'auteur prend ici l'exemple d'un doublement de la masse monétaire
[5] L'*effet Cantillon* sera largement ignoré par les théoriciens monétaires des générations suivantes qui ne retiendront que la hausse du niveau GENERAL des prix.

Comme expliqué ci-dessus, l'accroissement de la monnaie en circulation enrichit dans un premier temps le pays où elle se produit. Mais Cantillon considère comme inévitable une phase de reflux : « La trop grande abondance d'argent, qui fait, tandis qu'elle dure, la puissance des Etats, les rejette insensiblement mais naturellement dans l'indigence »[1]. Le renchérissement de la terre et du travail se répercute sur le prix de tous les produits. L'Etat et les habitants, habitués à dépenser plus, importeront les marchandises des pays étrangers, devenus plus compétitifs. Ce scénario pessimiste a plus de chances de se produire lorsque l'accroissement monétaire original provenait des mines d'or que si la cause première était la balance commerciale excédentaire. Cantillon cite en exemple l'Espagne, en déclin, qui ayant engrangé l'or des Amériques, n'a pu éviter sa fuite vers les puissances européennes montantes. Sur cette base, Cantillon défend une politique mercantiliste dont le but est de conserver un maximum de métal précieux à l'intérieur des frontières. Tout obnubilé comme beaucoup de penseurs de son temps par la nécessité dans laquelle se trouve l'Etat de triompher de ses rivaux, Cantillon ne se rend pas compte de l'aspect contradictoire de sa recommandation politique.

Le taux d'intérêt, quant à lui, se fonde « sur les profits que les entrepreneurs en peuvent faire ». Cantillon conteste ensuite l'idée répandue « que l'augmentation de la quantité de l'argent effectif dans un Etat y diminue le prix de l'intérêt, parce que quand l'argent abonde, il est plus facile d'en trouver à emprunter »[2]. Un surplus de monnaie n'abaissera l'intérêt que s'il échoit « aux gens qui prêtent … mais si elle vient par l'entremise des personnes qui dépensent, elle aura l'effet tout contraire »[3]. Cantillon vise ici le prince et la noblesse. Leurs dépenses supplémentaires vont obliger les entrepreneurs à accroître leur production et créeront un besoin de capitaux supplémentaires qui poussera l'intérêt vers le haut.

2.1.2. Les physiocrates

C'est une école purement française, née en réaction contre la politique économique de l'ancien régime qui laissait le fossé se creuser entre les groupes privilégiés minoritaires et le reste de la nation. La situation de l'agriculture, entièrement négligée, se dégradait au même moment que son homologue anglaise était en plein essor. Alors que la grande industrie exportatrice était favorisée, l'artisanat et l'agriculture stagnaient dans un état semi-féodal ; le commerce à l'intérieur de la France se heurtait à mille et une barrières tarifaires et autres.

Contrairement aux mercantilistes, les physiocrates avaient un véritable chef de file et maître à penser, le docteur François Quesnay. Certains de leurs comportements semblent étrangement sectaires : la vénération dont firent l'objet leur mentor et ses écrits, l'usage d'un langage mystique. Mais le fonds de la pensée est rationnel. Quesnay était dans la mouvance des lumières et contribua à l'encyclopédie de Diderot. Smith et Marx l'apprécieront comme un grand économiste.

Face à l'interventionnisme mercantiliste de l'Etat français, Quesnay préconise le *laissez faire*, la libération du commerce de ses entraves. Si l'Etat s'abstient d'imposer un ordre jugé idéal, alors émergera un « ordre naturel », qui est au cœur de la philosophie de Quesnay.

La notoriété de la physiocratie est en grande partie due à cet axiome : seule l'agriculture est productive car sa production dépasse la valeur des biens qu'elle absorbe pour sa production. L'artisanat et les services sont qualifiés de « stériles » en ce sens qu'ils permettent juste la

[1] Cantillon [45] p. 244
[2] Cantillon [45] p. 283
[3] Cantillon [45] p. 284

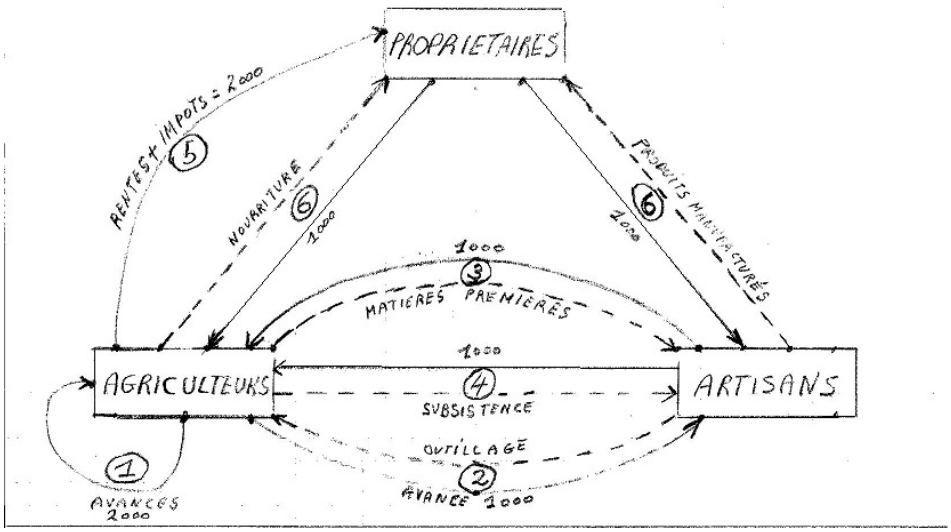
reproduction des biens entrants. Le qualificatif « stérile » est peu heureux car Quesnay reconnaît leur caractère utile et indispensable. Mais cette utilité ne se traduit pas par une création de richesse. La dignité exclusive de l'agriculture d'être productive n'est pas imputée à une spécificité du travail agricole mais aux bienfaits de la nature, à la fertilité de la terre.

LE TABLEAU ECONOMIQUE DE FRANÇOIS QUESNAY

Plusieurs versions de ce tableau économique ont été publiées. Nous retenons ici celle de 1766, intitulée « Analyse de la formule arithmétique du Tableau économique de la distribution des dépenses annuelles d'une nation agricole »[1]. Laissons parler l'auteur :

« La Nation est réduite à trois classes de Citoyens :
- La *classe productive* est celle qui fait naître par la culture du territoire les richesses annuelles de la Nation, qui fait les avances des dépenses des travaux de l'agriculture& qui paye annuellement les revenus des propriétaires de la terre (…)
- La *classe des propriétaires* comprend le Souverain, les Possesseurs des terres et les Décimateurs.
- La *classe stérile* est formée par les citoyens occupés à d'autres services et d'autres travaux que ceux de l'Agriculture ; & dont les dépenses sont payées par la classe productive & par la classe des propriétaires, qui eux-mêmes tirent leurs revenus de la classe productive. »

Graphique 2.1 : le Tableau Economique de Quesnay



Valeurs en millions. Quesnay ne précise pas l'unité.

Quesnay imagine une économie idéale, dont il brosse le tableau en quelques chiffres ronds. Le graphique ci-dessus (qui n'est pas de lui) résume son exemple. La classe productive fait des avances[2] à la production valant trois milliards, qui lui permettent de produire cinq milliards. Il y a donc un produit net de deux milliards. Ce revenu net est distribué aux propriétaires sous forme d'argent (rente foncière, impôt). A noter que la nourriture des agriculteurs est considérée comme un investissement nécessaire de la production agricole et ne fait donc pas partie du revenu net qui ne comprend que le revenu des propriétaires. Les propriétaires dépensent un milliard en produits agricoles et un milliard en produits manufacturés auprès des deux classes concernées. Avec le milliard obtenu des propriétaires, la classe stérile achète à la classe

[1] Cette version me paraît plus d'une exposition plus claire que le très fameux tableau en zigzag.
[2] L'*avance à la production* est un concept important de la science économique, introduit par Quesnay.

productive les subsistances qui lui sont nécessaires. Les trois milliards d'avances de la classe productive comprennent deux milliards de la production agricole qui sont recyclés à cette fin (semences et nourriture) et un milliard acheté à la classe stérile (outils…), qui elle-même fait un achat réciproque du même montant comme avance des matières premières à la production manufacturière.

Concernant les deux milliards composant les ventes de la classe stérile, Quesnay écrit : « il n'y a là que consommation et anéantissement de production & point de reproduction. Cette classe ne subsiste que du payement successif de la rétribution de son travail, qui est inséparable d'une dépense employée en subsistances, sans régénération de ce qui s'anéantit par la consommation »[1].

L'agriculture produit un revenu net, qui est entièrement capté par les propriétaires. L'industrie de Quesnay ne rapporte que la subsistance des artisans. Celle-ci, pas plus que celle des fermiers n'est conçue comme un revenu, parce que Quesnay les perçoit exclusivement dans l'optique des avances à la production. On pourrait prendre Quesnay à contre-pied sur son propre terrain, en considérant l'intérêt payé par les artisans aux capitalistes qui leur avancent le capital sur le même pied que la rente payée par les fermiers aux propriétaires fonciers. Mais Quesnay se garde bien d'introduire l'intérêt sur le capital manufacturier dans son tableau économique.

Quesnay se préoccupe de la maximisation du revenu de la société. Les dépenses improductives font l'objet de sa réprobation, notamment dans le chef de la paysannerie. Son souhait est de voir une paysannerie aisée, mais dépensant son argent en produits agricoles, dépense censée productive, plutôt qu'en produits manufacturiers stériles. Dans le même ordre d'idées, les moyens des paysans doivent être préservés des soustractions : « On ne pourroit rien soustraire à cette distribution de dépenses au désavantage de l'agriculture, ni rien soustraire des reprises du cultivateur, par quelque exaction ou par quelques entraves dans le commerce, qu'il n'arrivât du dépérissement dans la reproduction annuelle des richesses de la Nation & une diminution de la population facile à démontrer par le calcul ».

« La plus grande partie des dépenses des propriétaires sont de la même nature (NDLR : improductive) ; on n'en peut excepter que celle qu'ils font pour la conservation et l'amélioration de leurs biens & pour l'accroissement de leur culture. Mais comme ils sont de droit naturel chargés de ces dépenses, des soins de la régie & des réparations de leur patrimoine, ils ne peuvent être confondus avec la partie de la population qui forme la classe purement stérile »[2]. La classe des propriétaires se voit pardonner la stérilité de nombre de ses dépenses, car en tant que dépositaires de la terre généreuse et bienfaitrice, ils affectent une part de leur revenu à son entretien, ce qui représente une dépense productive très estimable.

« Les Propriétaires, le Souverain & toute la Nation ont un grand intérêt que l'impôt soit établi en entier sur les revenus des terres immédiatement ; car toute autre forme d'imposition seroit contre l'ordre naturel, parce qu'elle seroit préjudiciable à la reproduction & l'impôt & que l'impôt retomberoit sur l'impôt même… »[3]. Les diverses taxes devraient être remplacées par un impôt unique, qui ne frapperait que la rente. Ainsi, les avances à la production seraient préservées, la double imposition serait évitée et le commerce n'aurait pas à souffrir des entraves que constituent les taxes indirectes.

---

[1] Quesnay [290] pp. 15-16
[2] Quesnay [290] p. 27
[3] Quesnay [290] pp. 18-19

S'attaquant au mercantilisme, Quesnay écrit : « Nous n'avons point parlé de la masse d'argent monnoyé qui circule dans le commerce de chaque nation & que le vulgaire regarde comme la vraie richesse des Etats, parce que avec de l'argent on peut acheter, dit-on, tout ce dont on a besoin : mais on ne se demande pas avec quoi on peut se procurer de l'argent : cependant cette richesse ne se donne pas pour rien, elle coûte autant qu'elle vaut à celui qui l'achète »[1].

Le tableau économique influencera beaucoup d'économistes importants aux XIXe et XXe siècles. Sa grande contribution à la science économique est de faire apparaître pour la première fois l'économie comme un circuit. Les comptes de chaque classe sont équilibrés : elle dépense auprès des autres secteurs autant que le revenu qu'elle obtient. Après tous les transferts entre les classes et en leur sein, la société se retrouve prête à repartir pour un nouveau cycle. Pour autant qu'on ne l'entrave pas, le système économique s'auto-reproduit et peut même connaître une croissance s'il est géré de façon à conforter les avances à la production et l'amélioration des terres.

L'idée, que seule la classe agricole est productive, ne fera évidemment pas recette auprès des générations d'économistes suivantes. On chercherait d'ailleurs en vain un argument probant dans les explications de Quesnay ; le fait qu'il a réussi à équilibrer son circuit économique ne démontre pas l'improductivité du secteur non agricole. On aurait pu produire un modèle équilibré sur base d'autres hypothèses. Smith et Marx, paradoxalement, après avoir réfuté que seule la terre est productive, tenteront eux aussi de départager certaines activités qui seraient productives d'autres qui ne le sont pas, ce qui les engluera dans des raisonnements nébuleux et vains.

[1] Quesnay [290] p. 37

2.2- LIBERALISME ET PENSEE ECONOMIQUE

2.2.1. L'économie politique à la fin du siècle des lumières

Venons-en maintenant à deux personnages qui seront en contact avec Adam Smith et l'influenceront considérablement. L'Ecossais **David Hume** est un philosophe des lumières. Ses « Political Discourses » de 1752 expriment bien ses idées économiques libérales. Et d'abord, son anti-mercantilisme.

Contre la « groundless apprehension » qu'un Etat pourrait se vider de ses métaux précieux par le commerce international, il écrit : « and I should dread, that all our springs and rivers should be exhausted, as the money should abandon a kingdom where there are plenty of people and industry. Let us carefully preserve these latter advantages; and we need never be apprehensive of losing the former »[1].

Il s'intéresse tout particulièrement à la monnaie et son credo est celui-ci : « the greater or less plenty of money is of no consequence ; since the prices of commodities are always proportioned to the plenty of money »[2]. Ceci favorise les pays dotés de peu de monnaie dans le commerce international : leurs prix plus bas les rendront plus concurrentiels, ce qui corrigera la distribution des métaux précieux en leur faveur.

Hume modère toutefois ce propos : si la masse monétaire en circulation augmente, l'effet sur les prix ne se fera sentir que progressivement. En attendant que tous les prix se soient ajustés, il s'écoule un délai pendant lequel l'activité économique peut être stimulée par la circulation monétaire plus abondante. Hume en conclut : « it is of no manner of consequence with regard to domestic happiness of a state, whether money be in a greater or less quantity. The good policy of the magistrate consists only in keeping it, if possible, still encreasing, because, by that means, he keeps alive a spirit of industry in the nation and encreases the stock of labour, in which consists all real power and riches »[3].

La proposition suivant laquelle les prix sont réglés par la proportion entre les biens et la monnaie, n'est valide que dans la mesure où l'on a affaire à des biens échangés sur des marchés contre de la monnaie. Depuis la découverte de l'Amérique, la quantité de métaux précieux a augmenté en Europe dans une proportion justifiant une hausse des prix largement supérieure à celle qui eut effectivement lieu. Mais en même temps, l'échange monétaire s'est développé au détriment du troc et de l'autoconsommation, ce qui a altéré le rapport entre les volumes respectifs des paiements et de la monnaie.

« Nothing is esteemed a more certain sign of the flourishing condition of any nation than the lowness of interest. And with reason; though I believe the cause is somewhat different from what is commonly apprehended. Lowness of interest is generally ascribed to plenty of money. But money, however plentiful, has no other effect, if fixed, than to raise the price of labour.

Mais le taux d'intérêt ne dépend pas de la quantité de monnaie. « High interest arises from three circumstances: A great demand for borrowing; little riches to supply this demand; and great

[1] Hume [162] § 2.5.4
[2] Hume [162] § 2.3.1
[3] Hume [162] § 2.3.9

profits arising from commerce: And these circumstances are a clear proof of the small advance of commerce and industry, not of the scarcity of gold and silver »[1].

« For suppose that, by miracle, every man in Great Britain should have five pounds slipt into his pocket in one night; this would much more than double the whole money that is at present in the kingdom ; yet there would not next day, nor for some time, be any more lenders, nor any variation in the interest »[2]. Il n'en résulterait qu'une hausse des prix.

« It is needless, therefore, to enquire which of these circumstances, to wit, low interest or low profits, is the cause and which is the effect ? They both arise from an extensive commerce and mutually forward each other. No man will accept low profits where he can have high interest ; and no man will accept of low interest where he can have high profits. An extensive commerce, by producing large stocks, diminishes both interest and profits… »[3]. « Those who have asserted, that the plenty of money was the cause of low interest, seem to have taken a collateral effect for a cause »[4].

Jacques Turgot fut ministre des finances de Louis XVI. A ce titre, il entreprit de nombreuses réformes favorisant l'économie de marché aux dépens de la vieille économie corporatiste et semi-féodale. Il fut révoqué lorsque ses réformes déplurent à l'aristocratie. Inspiré par la physiocratie, sa contribution dépasse largement le cadre étriqué de celle-ci. Comme Quesnay, il contribua à l'Encyclopédie de Diderot. En 1766, paraissent ses « Réflexions sur la formation et la distribution des richesses », un ouvrage en avance sur son temps. Turgot s'y intéresse successivement au concept de *valeur*, à la monnaie et au capital.

Concernant la valeur d'échange des biens, Turgot insiste sur la différence entre l'échange isolé et l'échange sur un marché. Deux hommes sur une île déserte, porteurs chacun d'un bien différent, s'ils trouvent avantage à échanger, conviendront d'un rapport de prix, qui reflète leur situation particulière à tous deux ; on peut presque parler d'un rapport de prix aléatoire ; deux autres hommes auraient sans doute fixé un autre prix. Par contre, sur un marché, un prix unique s'impose du fait de la concurrence. Ce prix (ou ce rapport de prix s'il s'agit de troc), égalisant l'offre et la demande entre de nombreux intervenants, peut être considéré comme la véritable valeur du bien considéré.

« Le commerce donne à chaque marchandise une valeur courante, relativement à chaque autre marchandise, d'où il suit que toute marchandise est l'équivalent d'une certaine quantité de toute autre marchandise & peut être regardée comme un gage qui la représente. (…) Chaque marchandise peut servir d'échelle de mesure commune pour y comparer la valeur de toutes les autres »[5]. Toutefois, dans cette fonction de réserve de valeur d'échange et d'unité de mesure de la valeur, les marchandises dont la qualité est peu variable conviennent beaucoup mieux. La marchandise choisie pour remplir cette double fonction est ce qu'on appelle la monnaie. Les métaux précieux « y sont plus propres qu'aucune autre substance », notamment car « ils sont susceptibles de toutes les divisions imaginables » ; en outre ils sont inaltérables et ont « une très grande valeur sous un poids et un volume très peu considérable (…) Tout homme qui a une

[1] Hume [162] § 2.4.6
[2] Hume [162] § 2.4.9
[3] Hume [162] § 2.4.14
[4] Hume [162] § 2.4.15
[5] Turgot [369] pp. 24-25

denrée superflue & qui n'a pas, au moment, besoin d'une autre denrée d'usage, s'empressera donc de l'échanger contre de l'argent »[1]

Turgot ajoute : « On ne peut prendre pour commune mesure des valeurs, que ce qui a une valeur, ce qui est reçu dans le commerce en échange des autres valeurs : & il n'y a de gage universellement représentatif d'une valeur, qu'une valeur égale. Une monnoie de pure convention est donc une chose impossible ». Turgot montre qu'il a correctement compris la nature de la monnaie métallique. Par contre, il n'a pas été capable d'anticiper l'avènement d'un autre système monétaire.

Turgot en vient ensuite au capital. « Tous les genres de travaux de la culture, de l'industrie, du Commerce exigent des avances. Quand on laboureroit la terre avec les mains, il faudroit semer avant de recueillir : il faudroit vivre jusqu'après la récolte. Plus la culture se perfectionne & s'anime, plus les avances sont fortes »[2]. Turgot conçoit également que le capital est sous-tendu par un processus de circulation constant : « toutes les branches de commerce roulent sur une masse de capitaux ou de richesses mobiliaires accumulées qui, ayant été d'abord avancées par les Entrepreneurs, dans chacune de ces différentes classes de travaux, doivent leur rentrer chaque année avec un profit constant ; savoir le capital pour être reversé et avancé de nouveau dans la continuation des mêmes entreprises & le profit pour la subsistance plus ou moins aisée des Entrepreneurs »[3].

Il ne faut pas confondre l'argent et le capital. « (Leurs) évaluations sont indépendantes l'une de l'autre & sont réglées par des principes tout différens ». Turgot imagine un doublement de la masse d'argent en circulation distribué proportionnellement à la masse initiale. « L'argent, considéré comme masse de métal, diminuera certainement de prix, ou, ce qui est la même chose, les denrées seront payées plus cher (…) Mais il ne s'ensuivra nullement de-là que l'intérêt de l'argent baisse, si tout cet argent est porté au marché & employé aux dépenses courantes de ceux qui le possèdent, comme l'étoit, par la supposition, le premier million d'onces d'argent ; car l'intérêt de l'argent ne baisse qu'autant qu'il y ait plus d'argent à prêter, à proportion des besoins des emprunteurs, qu'il n'y en avoit auparavant. Or l'argent qu'on porte au marché n'est point à prêter ; c'est l'argent mis en réserve, ce sont les capitaux accumulés qu'on prête »[4]. Turgot est manifestement sur la même longueur d'onde que Hume, avec qui, d'ailleurs, il correspondait. Il partage avec lui la conviction qu'un intérêt plus bas est représentatif d'une plus grande épargne accumulée et est donc un signe de prospérité. Parallèlement, un taux d'intérêt plus élevé est défavorable à l'activité : « Si l'intérêt est à cinq pour cent, toute terre à défricher, dont les produits ne rapporteront pas cinq pour cent, outre le remplacement des avances & la récompense des soins du cultivateur, restera en friche »[5].

Turgot montre aussi que « dès que les profits résultant d'un emploi quelconque d'argent augmentent ou diminuent, les capitaux s'y versent ou s'en retirent en se versant vers les autres emplois ; ce qui change nécessairement dans chacun de ces emplois le rapport du capital au produit annuel »[6].

[1] Turgot [369] pp 29-30.
[2] Turgot [369] p.34
[3] Turgot [369] p.45
[4] Turgot [369] p.49
[5] Turgot [369] pp. 58-59
[6] Turgot [369] p.58

Un autre article de Turgot a accédé à la renommée, consacré à réfuter un mémoire écrit par un certain Saint-Palavy ; estimant le rapport de la production de la terre aux avances du fermier à cinq pour deux invariablement, celui-ci croit pouvoir appliquer ce ratio pour calculer la perte en production causée par un impôt indirect qui grève les avances. Contre ce simplisme, Turgot écrit « il ne faut qu'un peu de réflexion pour sentir qu'il ne peut y avoir de proportion constante entre les avances et les produits »[1]. On peut l'expliquer ainsi : « La semence jetée sur une terre naturellement fertile mais sans aucune préparation serait une avance presque entièrement perdue. Si on y joint un seul labour, le produit sera plus fort ; un second, un troisième labour pourront non pas simplement doubler et tripler, mais quadrupler et décupler le produit, qui augmentera ainsi dans une proportion beaucoup plus grande que les avances n'accroissent, et cela jusqu'à un certain point où le produit sera le plus grand qu'il soit possible, comparé aux avances. Passé ce point, si on augmente encore les avances, les produits augmenteront encore, mais moins, et toujours de moins en moins jusqu'à ce que, la fécondité de la terre étant épuisée et l'art n'y pouvant plus rien ajouter, un surcroît d'avance n'ajouterait absolument rien au produit »[2]. Turgot remarque que tant que les avances rapportent un excédent de production sur leur valeur, accrue de l'intérêt courant, il est avantageux d'y procéder. Le point où le produit est maximal par rapport aux avances n'est en rien un optimum.

2.2.2. Smith et la « main invisible » du marché

Adam Smith étudia puis enseigna la philosophie morale à l'Université de Glasgow. Son ouvrage « Theory of Moral Sentiments », publié en 1759 lui apporta la renommée et le succès. Au cours d'un voyage en France, il rencontra les philosophes et les économistes les plus éminents que comptait ce pays : Turgot, d'Alembert, Helvétius, Voltaire et Quesnay.

En 1776 paraît son œuvre maîtresse « An Inquiry into the Nature and Causes of the Wealth of Nations », l'ouvrage d'économie politique le plus complet jamais écrit à son époque. Pour la postérité, il est resté l'un des écrits fondateurs de la pensée libérale. Il fait partie de ces monuments intellectuels que peu de gens ont lus mais dont chacun connaît les lignes de force ; que cet ouvrage clame les vertus du marché est bien connu. Contrairement à beaucoup d'auteurs contemporains de la révolution industrielle, Smith se soucie du sort des masses laborieuses : « It is but equity, besides, that they who feed, clothe, and lodge the whole body of the people, should have such a share of the produce of their own labour as to be themselves tolerably fed, clothed, and lodged »[3]. Il pense également que « the difference of natural talents in different men is, in reality, much less than we are aware of »[4]; ces différences sont largement attribuables aux coutumes et à l'éducation. Smith considère les lois du marché comme le meilleur garant d'une prospérité bien répartie.

Le libéralisme de Smith est le mieux exprimé par sa métaphore bien connue de la *main invisible* : «by directing that industry in such a manner as its produce may be of the greatest value, every individual intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. (…) By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. »[5]. Cette idée est paraphrasée par cette autre citation : « it is not from the

[1] Turgot [368] p. 419
[2] Turgot [368] p. 421
[3] Smith [336] § 1.8.35. L'honnêteté commande préciser que les préjugés sociaux de la classe dominante sont présents dans d'autres passages de l'ouvrage.
[4] Smith [336] § 1.2.4
[5] Smith [336] § 4.2.9

benevolence of the butcher, the brewer or the baker that we expect our dinner, but from their regard to their own interest. »[1].

En vertu de ce principe, Smith tire trois conclusions :

1- L'intervention des pouvoirs publics (l'Etat et les pouvoirs locaux) dans la vie économique doit être limitée à quelques domaines que Smith énumère : la défense nationale, le maintien de la paix intérieure (justice et police), l'enseignement, ainsi que les ouvrages d'infrastructure (ponts, chaussées, canaux…) lorsqu'il n'est pas possible d'en tirer des revenus qui assurent leur entretien et leur gestion.
2- Les monopoles, ententes et corporations permettent à leurs membres de prélever des surprofits au détriment du reste de l'économie et nuisent à l'efficacité économique. Leurs privilèges et pouvoirs doivent être combattus.
3- Le libre échange doit prévaloir dans le commerce international. Smith condamne le protectionnisme et se révèle un critique intransigeant de la doctrine mercantiliste.

Smith attribue la progression de la prospérité au cours des siècles à la **division du travail**. Les producteurs spécialisés dans une production pointue atteindront un niveau de productivité hors d'atteinte de ceux pour qui la même production n'est qu'un produit ou une opération parmi une vaste gamme. Mais les entreprises ne se spécialiseront que s'il existe une demande suffisante pour des produits hautement spécifiques. Selon les mots de Smith, la division du travail est « limitée par l'extension du marché ». D'où l'avantage que l'échange ne soit pas confiné dans une aire géographique trop limitée. D'où l'importance de moyens de communication à bas prix qui rapprochent les agents économiques, jusque par delà les océans.

Abordant la problématique de la ***valeur***, Smith observe que ce mot a deux significations différentes :

- la *valeur d'usage* : c'est l'utilité du bien pour son possesseur ;
- la *valeur d'échange* que Smith définit ainsi : « the power of purchasing other goods which the possession of that object conveys »[2]. La valeur d'échange n'est donc rien d'autre que le prix, à une nuance près, qui apparaîtra d'ici peu.

Smith note le fameux paradoxe : les biens dont la valeur d'usage est très élevée ont souvent une faible valeur d'échange et vice versa ; il suffit de comparer des biens très utiles comme l'eau et l'air, dont la valeur d'échange est nulle ou quasi-nulle, avec le diamant, nettement moins utile mais plus cher. Ce n'est donc pas l'utilité qui détermine la valeur d'échange.

De ces deux notions de valeur, la plus importante aux yeux de l'économiste est la valeur d'échange. On appelle *théorie de la valeur*, toute explication des facteurs qui déterminent la valeur d'échange. Comme nous nous en rendrons compte tout au long de ce livre, la théorie de la valeur est l'un des thèmes qui préoccupera le plus les économistes de tous bords.

Smith opère une distinction, assez embrouillée- il est vrai-, entre la mesure de la valeur et la détermination de celle-ci.

Commençons avec la MESURE de la valeur. Le prix monétaire est un étalon insatisfaisant puisque la monnaie est un bien (du métal précieux en l'occurrence) dont la valeur est elle-même sujette à fluctuation. Smith prétend que ce problème d'instabilité ne se pose pas avec le travail pris comme étalon de mesure. Pourquoi le travail ? « The real price of every thing, what every thing costs to the man who wants to acquire it, is the toil and trouble of acquiring it.

[1] Smith [336] § 1.2.2
[2] Smith [336] § 1.4.13.

What every thing is really worth to the man who wants to exchange it for something else is the toil and trouble which he can save to himself and which he can impose upon other peoples. (…) (The goods we buy) indeed save us this toil. They contain the value of a certain quantity of labour which we exchange for what is supposed at the time to contain the value of an equal quantity. Labour was the first price, the original purchase-money that was paid for all things »[1]. En conséquence, « Labour alone, therefore, never varying in its own value, is alone the ultimate and real standard by which the value of all commodities can at all times and places be estimated and compared. It is their real price; money is their nominal price only »[2]. Comment concilier cette affirmation de l'invariance de la valeur du travail avec les variations du pouvoir d'achat du salaire ? Tout simplement, dit Smith, c'est la valeur des biens qui change et non celle du travail. Mais dans cette optique, la valeur d'échange n'est plus un rapport entre deux biens égaux dont chacun peut être pris comme étalon ; le travail est l'étalon par nature. On doit aussi en inférer que la valeur du travail et le salaire réel (et a fortiori le salaire nominal) sont deux choses différentes. Or il n'est pas évident que Smith aurait souscrit à cette conclusion. C'est donc dans une certaine confusion que Smith introduit la distinction très importante en économie entre la valeur REELLE et la valeur NOMINALE des biens.

Smith reconnaît que l'hétérogénéité du travail amène une difficulté dans la détermination de la valeur travail des biens, mais il la résout en réduisant les diverses variétés de travail en unités du travail le plus facile. Certaines tâches, soit parce qu'elles sont plus pénibles, soit parce qu'elles requièrent une compétence particulière, ont leur temps de travail compté avec un facteur multiplicatif. Pour déterminer ce facteur, Smith ne propose aucune formule ; il résulte du marchandage propre à la formation des prix.

Selon Smith, tout travail ne crée pas de valeur. C'est la fameuse et parfois fumeuse distinction entre *travail productif* et *travail improductif*. "Thus the labour of a manufacturer adds, generally, to the materials which he works upon, that of his own maintenance, and of his master's profit. The labour of a menial servant, on the contrary, adds to the value of nothing (…) But the labour of the manufacturer fixes and realises itself in some particular subject or vendible commodity (…) The labour of some of the most respectable orders in the society is, like that of menial servants, unproductive of any value (…) The sovereign, for example, with the officers both of justice and war who serve under him, the whole army and navy, are unproductive labourers. They are the servants of the public, and are maintained by a part of the annual produce of the industry of other people. Their service, how honourable, how useful, or how necessary soever, produces nothing for which an equal quantity of service can afterward be procured"[3]. Il ajoute à cette liste des improductifs : "churchmen, lawyers, physicians, men of letters of all kinds; players, buffoons, musicians, opera-singers, opera-dancers &c". A mon avis, il est impossible d'établir une classification des travaux productifs et improductifs entièrement cohérente et inattaquable.

Abordant ensuite la DETERMINATION de la valeur, Smith la considère différemment selon le type d'économie auquel on a affaire. Voyons d'abord – et rapidement- « the early and rude state of society which precedes both the accumulation of stock and the appropriation of land ». Dans la société PRIMITIVE, les valeurs des biens sont proportionnelles à la quantité de travail nécessaire pour les obtenir. L'exemple donné par Smith a été mille fois répété : s'il faut deux

[1] Smith [336] § 1.5.2
[2] Smith [336] § 1.5.8
[3] Smith [336] § 2.3.1 et 2.3.2

fois plus de travail pour chasser un castor qu'un daim, le castor s'échangera contre deux daims[1].

Dans la société AVANCEE, trois agents participent à la production : le travailleur, le propriétaire foncier et le capitaliste (une même personne peut, le cas échéant, cumuler deux de ces fonctions ou les trois). Chacun d'eux exigera sa part. La valeur d'échange est déterminée en additionnant ces parts qui ont pour noms le salaire, la rente et le profit. Lorsqu'on décompose le prix du pain en profits, salaires et rentes, on considère évidemment la rémunération non seulement du boulanger et de ses employés, mais aussi du meunier, du fermier et de tous ceux qui ont fourni des services accessoires tels que transports…

Il y a un salaire naturel, un profit naturel et une rente naturelle ; la valeur d'échange est le prix naturel résultant de l'addition des rémunérations à leur valeur naturelle[2]. Mais le PRIX DU MARCHE peut s'écarter temporairement de ce prix naturel, si des conditions accidentelles produisent pléthore ou pénurie (cf. Cantillon ci-dessus). Comme le dit Smith, le prix du marché gravite autour du prix naturel. Normalement, ces variations sont temporaires parce qu'un marché à demande insatisfaite attire les producteurs et qu'un marché à demande insuffisante les repousse. Il existe toutefois certaines situations où le producteur peut vendre durablement au-dessus de la valeur naturelle. Smith donne l'exemple de la manufacture qui bénéficie de secrets de production. Le monopole en est un autre exemple.

Smith analyse ensuite l'un après l'autre et en détail, chacun des trois facteurs de production. On s'attendrait à trouver dans cette analyse la formule déterminant leur rémunération naturelle, mais cette attente est déçue. Voyons brièvement les commentaires de Smith sur le marché du travail ainsi que sur le capital et le profit.

Le **salaire** est le résultat d'une négociation antagonique entre deux classes. Les employeurs étant moins nombreux que les salariés et mieux protégés par la loi, c'est un peu le pot de fer contre le pot de terre. Mais, contrairement à d'autres économistes, Smith n'en conclut pas que les salaires sont condamnés au minimum vital. Smith insiste également sur le surplus de productivité qui peut résulter d'une meilleure rémunération.

Qu'est-ce qui détermine la demande de travail et le volume de l'emploi ? « The demand of those who live by wages, it is evident, cannot increase but in proportion to the increase of the funds which are destined for the payment of wages »[3]. D'apparence tautologique, cette phrase est loin d'être anodine. Smith introduit ici la *théorie du fonds des salaires*, qui fera florès tout au long du XIXe siècle. Des salaires ne peuvent être payés pour telle production que parce que du capital a été constitué à cette fin, préalablement à la production. Cette accumulation préalable est indispensable, car la production ne nourrit pas elle-même ceux qui y sont employés. Le fonds des salaires est conçu comme ayant un volume assez rigide. Il sera plus élevé dans une société en croissance rapide que dans une société stationnaire ; la croissance du revenu social importe plus que son niveau absolu.

---

[1] Smith est ici dans un monde imaginaire ; ce n'est que la concurrence entre les chasseurs qui peut fixer les valeurs d'échange de cette manière. Un tel niveau de concurrence n'existait pas dans les sociétés primitives.

[2] L'articulation entre la mesure de la valeur selon la seule quantité de travail et sa détermination par l'addition des trois rétributions demeure obscure.

[3] Smith [336] § 1.8.18.

La rémunération du capitaliste est appelée ***profit*** lorsque l'entrepreneur travaille avec son propre capital et ***intérêt*** lorsqu'il travaille avec du capital emprunté[1]. Le taux d'intérêt se règle sur le taux de profit et Smith ne considère pas la distinction entre eux comme essentielle. En principe, des capitaux investis dans des branches différentes doivent rapporter un taux de profit identique ; si tel n'était pas le cas, l'équilibre serait rapidement rétabli du fait du déplacement de capitaux vers les branches les plus rentables. Ce mécanisme est toutefois assorti d'un autre facteur : plus une affaire est risquée, plus le taux de profit qui rémunère son capital doit être élevé.

Smith considère l'accumulation du capital comme un facteur de progrès des salaires ; selon lui, cette même cause entraîne une baisse tendancielle du taux de profit. L'explication est celle-ci : plus de capital signifie plus de concurrence entre les producteurs et celle-ci rabote leur profit.

Smith analyse le capital en profondeur. L'épargne en est selon lui la source unique. Il introduit une distinction entre deux parties du capital, qui se révélera très importante, comme le lecteur le constatera tout au long du présent ouvrage :

1- Le ***capital fixe*** : Ce sont les bâtiments, machines et outils, le charroi. Smith y inclut également les « improvements of land » très importants en agriculture et la formation des producteurs nécessaire à l'acquisition de talents.
2- le ***capital circulant*** : appelé ainsi, parce qu'il change de forme en un cycle continuellement répété : argent puis matières et travail, puis encours de production, puis produits finis puis à nouveau argent[2].

[1] Comme nous le verrons, la distinction théorique entre le profit et l'intérêt évoluera par la suite.
[2] Le capital circulant est un concept plus abstrait, plus difficile à saisir que le capital fixe. Du point de vue comptable, on retrouve le capital circulant dans les stocks de matières premières et de produits finis, mais surtout dans l'*encours de production*, c'est-à-dire les biens dont la production a déjà absorbé des dépenses (notamment de salaire) mais qui ne sont pas encore mûrs pour être commercialisés.
Si on se place au niveau non pas de la firme mais de l'économie en général, tous les biens intermédiaires, non consommables comme tels peuvent être considérés comme du capital circulant. Le tissu et le fil appelé à être tissé sont des vêtements en devenir, donc du capital circulant.

2.3- LE CLASSICISME RICARDIEN ET SON TEMPS

L'*Inquiry* de Smith eut un retentissement considérable, en Grande-Bretagne évidemment, mais en France également. Il suscita probablement plusieurs vocations d'économiste.

Au début du XIXe siècle, L'Angleterre connaissait ce qu'on a appelé un *décollage* industriel. Le capitalisme triomphait et s'implantait également dans l'agriculture : c'est la fameuse structure à trois classes tant commentée par les économistes de cette époque : des fermiers capitalistes louaient la terre à des propriétaires fonciers vivant de la rente, et tiraient du profit d'exploitations agricoles occupant des travailleurs salariés. Malgré l'avance économique du pays, les problèmes économiques ne manquaient pas : les guerres napoléoniennes coûtaient cher, stimulant l'émission de papier-monnaie ainsi que la dette publique. L'inflation régnait. Le paupérisme frappait les classes laborieuses dont beaucoup d'individus recouraient à l'assistance publique (secours paroissial dans le cadre des *poor laws*). En 1815, le gouvernement vota la *corn law* qui fixait un prix minimum du grain assez élevé ; cette mesure protectionniste aggravait encore la précarité du prolétariat.

C'est dans ce contexte que nombre de grands bourgeois commencèrent à se passionner pour l'économie politique qui n'était pas encore une discipline scientifique officielle. Ces gentlemen publièrent des pamphlets puis des traités, débattirent entre eux. Leur expertise fut suffisamment reconnue pour les voir informer des commissions parlementaires. Les deux plus réputés étaient David Ricardo et Thomas Robert Malthus, mais il y avait également le colonel Torrens, Henry Thornton, James Mill, John Ramsey Mc Culloch… Malthus devînt le premier professeur d'économie politique, qu'il enseigna à l'*East of India College* de 1804 à sa mort.

Malthus est surtout connu pour ses théories démographiques fort teintées de pessimisme, voire apocalyptiques. La population augmente plus vite que la capacité de la terre de la nourrir. Augmenter le revenu des pauvres a des effets négatifs plutôt que positifs, car cela tend à accroître la natalité et donc le nombre de personnes à aider. Le malthusianisme a aujourd'hui mauvaise réputation car l'erreur complète de ses prédictions s'est avérée. S'il ne fit pas l'unanimité à sa sortie en 1798, l'« Essay on Population » influença immédiatement la plupart des économistes et maintint cette influence pendant tout le XIXe siècle.

Après ce début mémorable de sa carrière intellectuelle, Malthus s'orienta vers l'économie politique. Il partagea avec Ricardo une amitié indéfectible, mais étrange car leurs avis discordaient sur presque tous les sujets économiques et ils ne se privaient pas de critiquer mutuellement leurs théories. Leur abondante correspondance a été conservée et est publiée. Leurs rares points d'accord sont la théorie du salaire déterminée par la conception malthusienne de la population, l'hostilité aux poor laws, la *théorie de la rente différentielle* (cf. infra) qu'ils découvrirent simultanément[1] et la conception d'une rente foncière appelée à augmenter dans le futur du fait de la mise en culture de terres de moins en moins fertiles, sous la pression démographique. Marx considère Malthus comme le représentant de l'aristocratie foncière alors que Ricardo serait l'idéologue de la bourgeoisie industrielle. Ce jugement peut se comprendre au vu de l'énergie mise par Malthus à justifier la rente foncière et par sa défense des corn laws (qui ne bénéficiaient qu'aux propriétaires fonciers) alors que Ricardo s'y oppose au nom du libre échange. Toutefois, Malthus épouse sur la plupart des sujets un point de vue proche de celui des capitalistes et des entrepreneurs.

[1] Par une coïncidence, l'année 1815 vit cette théorie énoncée par quatre auteurs différents : Ricardo, Malthus, Torrens et West.

2.3.1. Ricardo : le coût de production détermine la valeur

David Ricardo est l'un des plus grands économistes de l'histoire. La plupart des grands économistes ultérieurs sentiront le besoin de se situer par rapport à lui. Comme nous le verrons plus loin, un courant de pensée économique se réclame encore aujourd'hui de ses théories.

Sa vocation d'économiste ne fut pas précoce. Ricardo s'engagea d'abord dans les affaires où il fit fortune très rapidement comme agent de change à la bourse de Londres. A l'âge de vingt-sept ans, il découvrit la *Richesse des Nations* qui l'impressionna.

C'est dans les débats monétaires que Ricardo eut la première occasion de se distinguer comme économiste. Voyons le contexte. A cette époque, l'or était encore la forme monétaire la plus courante et les billets de banque, émis par différentes banques privées, des certificats dont le détenteur pouvait demander la conversion en or. Les craintes quant à un débarquement français en Angleterre suscitèrent un début de panique en 1797, ce qui décida le gouvernement britannique à décréter l'inconvertibilité des billets de la Banque d'Angleterre. La situation se calma ensuite, mais le gouvernement ne décida pas le retour à la convertibilité. S'engagea alors tout un débat connu sous le nom de *controverse bullionniste*[1] .

Deux thèses s'opposaient : d'une part les *bullionistes* réclamaient le retour à la convertibilité. Ils craignaient l'inflation entraînée par l'émission excessive de papier-monnnaie. La convertibilité des billets représente un mécanisme autorégulateur de leur quantité et donc un gage de la stabilité des prix. La thèse opposée, celle des *anti-bullionistes* était favorable au statu quo. C'était le camp des optimistes qui faisaient confiance dans l'autorégulation du marché.

Ricardo fut l'une des voix bullionistes les plus écoutées. Il publia en 1810 un pamphlet intitulé « The High Price of Bullion, a Proof of the Depreciation of Bank Notes » qui fut un franc succès. En 1819, le gouvernement décidera le retour à la convertibilité, ce dont la renommée de Ricardo profita beaucoup. En 1816, Ricardo avait publié un autre pamphlet demandant qu'on modifie le statut de la Banque d'Angleterre, pour en faire une banque centrale capable et désireuse de contrôler la masse monétaire suivant des critères objectifs plutôt que de se préoccuper exclusivement de son propre profit.

Mais Ricardo pense que « to determine the laws which regulate (the) distribution is the principal problem of Political Economy »[2]. La répartition du produit entre la rente, le profit et le salaire occupe donc l'essentiel de son œuvre majeure « On the Principles of Political Economy and Taxation », parue en 1817.

Ricardo prend comme point de départ la théorie de la valeur-travail initiée par Smith. Il apporte cette double mise au point très importante :
1-la valeur d'échange est indépendante de l'utilité, mais il n'y a pas de valeur d'échange sans utilité.
2-la théorie de la valeur travail ne s'applique qu'aux biens reproductibles en série dans le cadre de la concurrence. A propos des biens tels les vins de grand cru, les œuvres d'art etc, Ricardo note : « (their) value is determined by their scarcity alone. No labour can increase the quantity of such goods and therefore their value cannot be lowered by an increased supply (…) Their value is wholly independent of the quantity of labour originally necessary to produce them »[3].

---

[1] De l'anglais « bullion », qui signifie « lingot ».
[2] Ricardo [297] p. 5
[3] Ricardo [297] pp. 9-10

Comme Smith, Ricardo distingue l'état primitif et l'état avancé de l'économie. L'état primitif n'est pas a priori le cas le plus intéressant, mais l'analyse ricardienne de cas servira de base à la théorie marxiste de la valeur-travail. Dans l'état primitif, la valeur est réglée par la quantité de travail incorporée. Le capital rapportant de l'intérêt est censé ne pas encore exister ; néanmoins, le travailleur-producteur achète des outils et des matières premières, qui ont une valeur préexistante ; qu'en advient-il ? Ricardo écrit : « not only the labour applied immediately to commodities affect their value but the labour also which is bestowed on the implements, tools and buildings, which much labour is assisted »[1]. Illustrons par un exemple chiffré: un producteur produit 100 unités par an, correspondant à 2000 heures de travail. Il utilise un outil qui a une durée de vie de 5 ans et qui a coûté 4000 heures de travail. La valeur unitaire de sa production intégrera 28 heures de travail, soit la somme de 20 heures directes et 8 heures indirectes.

Ricardo délaisse ensuite cet état primitif pour s'intéresser au cas plus actuel d'une production où le capital joue un rôle important. Là, sa théorie diverge de celle de Smith, tout en partageant la même philosophie de la valeur déterminée par le coût de production.

Ricardo s'attache à la structure du capital. Le capital fixe peut avoir une valeur-travail plus ou moins élevée. Sa durée de vie productive peut être plus courte ou plus longue. Les proportions respectives du capital fixe et du capital circulant peuvent différer. « It is also to be observed that the circulating capital may circulate, or be returned to its employer, in very unequal times »[2]. C'est donc un tableau contrasté qu'offre le capital investi dans les différentes branches industrielles. Comment prendre en compte cette diversité dans la détermination de la valeur ? Pour y arriver, Ricardo introduit un nouveau concept promis à un bel avenir *: la période de production*. La formulation est encore vague, comme en témoigne ce bout de phrase qui semble faire office de définition : « the length of time which must elapse before the most valuable can be brought to market »[3].

De deux marchandises incorporant la même quantité de travail, celle dont la période de production est la plus longue aura une valeur plus élevée, car elle devra « payer » l'intérêt au capitaliste plus longtemps. Ricardo donne l'exemple d'une production *A* qui nécessite le travail de cent hommes pendant un an et une production *B* nécessitant le travail de cent hommes pendant deux ans. La valeur de *B* sera plus du double de la valeur de *A*. La valeur de la marchandise ne dépend donc plus de la seule quantité de travail, mais également de la période de production et du taux de profit.

Les commentateurs actuels de Ricardo résument parfois sa théorie de la valeur en recourant à des équations comme celle-ci :

$$V_a = \Sigma_1^T L_t.(1+r)^t \qquad (2.1)$$

où V_a est la valeur d'échange du bien *A*
L_t est le temps de travail dépensé au temps t
r est le taux de profit
T est la durée de la période de production

Le taux de profit, *r*, tend à être le même dans tous les secteurs de production, sinon, il se produirait un afflux de capital dans les secteurs où il est supérieur et cette abondance de capital

[1] Ricardo [297] p. 22
[2] Ricardo [297] p. 31
[3] Ricardo [297] p. 34

y ramènerait le taux au niveau normal ; l'inverse se produirait dans les secteurs où *r* est inférieur. Ce taux de profit n'est pas indépendant du taux de salaire. Le revenu national est comme un gâteau à répartir entre le profit et le salaire : « there can be no rise in the value of labour without a fall of profits »[1].

Le lecteur pourrait s'étonner qu'à la place de (2.1), nous n'ayons pas écrit la formule :

$$V_a = \Sigma_1^T L_t.w.(1+r)^t \qquad (2.2)$$

où *w* est le salaire réel

Elle serait plus correcte, mais la simplification nous fait éliminer les variables qui ne sont pas indispensables. Or Ricardo s'ingénie à montrer que le salaire n'a pas d'influence directe sur la valeur relative, car il subit le même processus d'uniformisation par la concurrence que le taux de profit. Certes, Ricardo avait soutenu la thèse de Smith sur le facteur multiplicatif affectant certains travaux plus qualifiés ou plus pénibles, qui a pour effet de différencier *w* ; mais il estime pouvoir simplifier.

Ayant éjecté le taux de salaire comme facteur DIRECT, Ricardo le rappelle comme facteur influençant la valeur INDIRECTEMENT. En effet, vu que le salaire et le profit sont les deux composantes du revenu et que celui-ci est considéré comme donné, on a r = f(w) où la fonction *f* est décroissante. La formule de la valeur devient :

$$V_a = \Sigma_1^T L_t.(1+f(w))^t \qquad (2.3)$$

L'explication de l'influence indirecte du taux de salaire sur la valeur est l'un des apports décisifs de Ricardo. Beaucoup d'économistes, dont Smith, ont affirmé que la hausse du salaire entraînerait un accroissement des prix de tous les biens. Ricardo montre que l'effet serait plus complexe. N'oublions pas que la hausse du salaire réel obère le profit. En cas de hausse du salaire réel, les biens dont la production est intense en travail connaîtraient une forte hausse de leur coût salarial accompagnée d'une faible baisse de leur coût du capital : leur valeur augmenterait. Celle des biens intenses en capital baisserait car la faible hausse du coût salarial serait plus que compensée par la baisse plus importante du coût du capital. Cet effet différencié est souvent appelé *effet Ricardo*.

L'effet Ricardo se marque dans la formule (2.3). : lorsque *t* est plus élevé, la baisse de *f(w)* voit son poids accru par rapport à la hausse de *w*, du fait de l'exposant.

L'effet Ricardo concerne les variations du salaire REEL. Une variation du salaire NOMINAL aurait évidemment un impact sur le prix nominal. Mais cet effet serait imputable à la baisse de la valeur de la monnaie envers l'ensemble des autres biens y compris le travail, et non à la hausse des salaires.

Considérons les trois variables qui interviennent dans l'équation (2.1), en l'occurrence *L*, *r* et *t*. Ricardo estime que c'est le temps de travail *L* qui est, de fait, responsable de la plus grande part des variations de la valeur. Il constate que *r* fluctue généralement à l'intérieur d'une marge assez étroite. Quant à *t*, il ne semble plus s'en soucier beaucoup. Il est déconcertant de voir comment Ricardo oublie par moments ce facteur qu'il a lui-même mis en évidence.

C'est l'aspect dynamique de la théorie de la valeur qui intéressait le plus Ricardo. « When commodities varied in relative value, it would be desirable to have the means of ascertaining which of them fell and which rose in real value, and this could be effected only by comparing

[1] Ricardo [297] p. 35

them one after another with some invariable standard measure of value, which should itself be subject to none of the fluctuations to which other commodities are exposed »[1]. Si la quantité de travail seule déterminait la valeur, ce ne serait pas trop difficile. Par exemple, si de t_1 à t_2, la valeur du kilo de café passait de 1 à 1,2 kilo de thé, il suffirait de voir si c'est le temps de travail pour produire le café qui a augmenté de 20% ou si c'est le temps de travail pour produire le thé qui a baissé de 16,7%. Mais les variations dans la répartition entre le profit et le salaire viennent compliquer l'affaire, comme l'indique l'équation (2.3).

Le souci de Ricardo est bien légitime, car en l'absence d'un étalon invariable, les comparaisons de prix dans le temps sont toujours biaisées. Cette quête de la *mesure invariante de la valeur* fut sa chimère. Assez arbitrairement et sans trop de conviction, Ricardo décréta que l'or serait la mesure invariante de la valeur car sa période de production serait assez moyenne et la quantité de travail requise par sa production serait assez stable. Ces considérations manquent manifestement de rigueur mais Ricardo semble se contenter d'une approximation, dont la justesse n'est pas évidente. Toutefois, sa correspondance semble indiquer que vers la fin de sa vie, il se mit à douter de l'existence d'une telle mesure invariante.

Concluons cet important chapitre sur la théorie de la valeur avec l'opinion de Ricardo sur l'idée très répandue que la valeur d'échange serait simplement déterminée par l'offre et la demande ? «This opinion has become almost an axiom in political economy and has been the source of much error in that science »[2]. « The proportion between supply and demand may, indeed, affect the market value of a commodity, until it is supplied in greater or less abundance, according as the demand may have increased or diminished; but this effect will be only of temporary duration »[3]. Les écarts du prix de marché par rapport au prix naturel sont possibles, mais ils sont précisément l'élément déclencheur des forces du marché qui ramèneront le prix à la valeur. L'accroissement de la demande élèvera le prix, ce qui fera hausser le profit, d'où proviendra l'accroissement de l'offre qui poussera le prix de marché à la baisse jusqu'à qu'il égale le prix naturel.

Abordons maintenant le modèle de Ricardo quant à la répartition du revenu. Un élément clé est le concept de *rente différentielle*, qui a largement contribué à sa notoriété. Alors que selon les physiocrates, la rente rémunérait la productivité de la terre, selon Ricardo, elle paye sa rareté, ou plus exactement la rareté des bonnes terres fertiles. Pour nourrir sa population, la société doit mettre en culture des terres de fertilité variable et Ricardo postule que les meilleures terres sont les premières à être cultivées. Nous savons que la valeur du blé dépend de la quantité de travail[4] nécessaire pour le produire, mais celle-ci diffère précisément suivant qu'on ait affaire à une meilleure ou à une moins bonne terre. Par exemple, 50 heures de travail produiront 100 qx de blé sur une terre fertile alors qu'il en faudra 80 sur la terre la moins fertile. Puisque tous les producteurs doivent atteindre le seuil de rentabilité, c'est la quantité de travail sur la plus mauvaise terre qui déterminera la valeur. Comme les salaires des ouvriers agricoles sont les mêmes, quelle que soit la fertilité du sol, les fermiers exploitant de meilleures terres devraient bénéficier d'une rentabilité supérieure et faire des superprofits proportionnels à la qualité du sol qu'ils cultivent. Il n'en sera pourtant rien. Le propriétaire foncier fera jouer les enchères et louera sa terre au fermier qui acceptera de payer la rente foncière la plus élevée ; c'est donc la rente foncière qui intégrera le surplus dû à la productivité plus grande et non le profit du fermier

[1] Ricardo [297] p. 43
[2] Ricardo [297] p.382
[3] Ricardo [297] p.382
[4] Ricardo pratique la simplification consistant à assimiler le coût de production, véritable déterminant de la valeur, à la quantité de travail incorporée.

qui suivra la règle générale d'égalisation des taux de profit par la concurrence. La rente d'une terre vaut la différence entre le produit qui en est retiré et le produit qu'une même quantité et qualité de capital et de travail obtiendrait sur la terre la moins fertile.

La rente est une « marge » rendue par le prix au propriétaire d'avantages rares. « Corn is not high because a rent is paid, but a rent is paid because corn is high »[1]. Ricardo reproche à Smith de ne pas l'avoir compris. Rappelons-nous que Smith ajoutait la rente au salaire et au profit pour obtenir le prix.

Il y a donc une nuance subtile entre rente et coût. Certes, dans les comptes du fermier, la rente apparaîtra comme un coût, car il doit la payer au propriétaire foncier. Mais du point de vue économique, il en va autrement. Le fait que la terre marginale ne rapporte pas de rente montre bien que celle-ci n'est pas le coût de la terre de la même façon que le salaire est le coût du travail. Ricardo envisage l'hypothèse où les terres les plus fertiles seraient disponibles à volonté : « no rent could be paid for such land, for the reason stated why nothing is given for the use of air and water or for any other of the gifts of nature which exists in boundless quantity »[2]. Le même argument vaut pour le prix d'acquisition de la terre. Dans un pays où il resterait de bonnes terres non appropriées, le prix de la terre serait nul.

Soit une terre fertile dont la mise en culture requiert un capital[3] de £1000 produisant 100 qx de blé ; à côté une deuxième terre moins fertile dont la culture requiert également un capital de £1000 avec un produit de 80 qx. Avec un capital de £ 2000, les deux terres seront cultivées et la plus fertile rendra une rente valant 20 qx de blé. Ceci est la première forme de la rente différentielle, que nous avons déjà expliquée ci-avant. On l'appelle souvent la forme *extensive*. Mais Ricardo montre la possibilité d'une seconde forme de rente foncière, souvent appelée *intensive* : la deuxième tranche de £1000 de capital ne serait plus dépensée sur une deuxième terre, mais toujours sur la première ; celle-ci bénéficierait d'un capital de £2000 pour elle seule. Voyons comment il peut en résulter de la rente.

L'investissement d'un deuxième capital de £1000 sur la terre la plus fertile fait hausser sa production mais moins qu'un doublement, car les doses successives se révéleront de moins en moins productives. Supposons que la production passe de 100 à 185 qx. Cet investissement sera préféré à celui sur la terre moins fertile (qui ne rapportait que 80) et cette dernière ne sera pas cultivée. Mais si le capitaliste accepte un retour de 85 sur la 2e tranche de capital, il acceptera ce même taux sur la première tanche, comme il avait accepté 80 quand les deux tranches de capital étaient investies sur des terres différentes ; il ne peut y avoir des taux de profit différents pour des investissements différents. Le capital entier sera rémunéré selon la productivité de sa tranche la moins productive ; l'investissement de £2000 rapportera 170 qx au fermier ainsi qu'une rente de 15 qx au propriétaire. Des applications successives de capital et travail sur une même terre ont donc le même effet que des applications successives sur des terres de moins en moins fertiles. Dans les faits, les applications intensives et extensives du capital se combinent.

Comme tout bien, le travail a une valeur naturelle que Ricardo définit ainsi : « that price which is necessary to enable the labourers to subsist and to perpetuate their race, without either increase or diminution »[4]. Il s'agit plutôt d'un minimum culturel que d'un minimum

[1] Ricardo [297] p. 74
[2] Ricardo [297] p. 69
[3] « Capital » est ici un raccourci simplificateur pour l'ensemble capital et travail, que Ricardo fond en un tout indistinct.
[4] Ricardo [297] p. 93

biologique, puisque Ricardo considère qu'il dépend de la coutume. Le niveau du salaire dépend essentiellement du prix des biens de consommation ouvrière. Le prix du blé est donc le facteur qui guide le salaire, car il représente la plus grande part de la dépense de la classe ouvrière.

Comme les autres biens, le travail peut être tantôt abondant tantôt rare, ce qui écarte son prix de marché du salaire naturel. Mais le processus malthusien de hausse de la natalité lorsque le revenu est élevé ou de dénatalité lorsque le salaire est bas ramène le salaire à son niveau naturel, après un délai qui peut, certes, être assez long.

D'après Ricardo, il existe une tendance historique à l'accroissement de la population, donc aussi à l'accroissement de la production agricole, donc aussi du coût de production du blé à cause de la nécessité de mettre en culture des terres de moins en moins fertiles, donc aussi à l'accroissement du salaire. Le salarié ne bénéficie d'ailleurs pas de cette hausse, puisqu'elle sert à compenser l'augmentation des prix agricoles, dont nous avons vu qu'elle profite exclusivement aux propriétaires fonciers.

Toujours sur base de cette conception malthusienne, Ricardo critique les velléités interventionnistes du législateur, qui pourrait être tenté de réglementer les salaires, par exemple en fixant un plancher. Le seul effet serait d'accroître la natalité et donc de réduire le bien-être ouvrier. A propos des « poor laws », Ricardo ne craint pas d'écrire : « every friend to the poor must ardently wish for their abolition »[1].

Le prix se compose du profit, du salaire et de la rente. Comme la rente est déterminée de façon exogène par les différences de fertilité des terres, on peut considérer les variations du profit et du salaire comme une opération à somme nulle. Puisque le salaire connaît une tendance historique à l'augmentation, le profit tend forcément à diminuer à long terme. La raison fondamentale de cette baisse est la hausse du coût de production du blé qu'entraîne la satisfaction de la demande d'une population croissante. Le tableau est maintenant complet : la tendance historique de la répartition du produit entre les classes sociales est celle-ci : hausse réelle de la rente foncière, hausse nominale (mais pas nécessairement réelle) du salaire, baisse du taux de profit. Il y a toutefois une limite à cette chute du profit : si le taux de profit décline trop, l'accumulation du capital sera découragée et la demande de travail baissera, et avec elle le salaire, la natalité, la population … et les prix agricoles.

Ricardo admet que l'amélioration de la productivité, tant dans l'industrie que dans l'agriculture, est favorable au profit, mais il ne prête pas à ce facteur la capacité de contrecarrer le déclin du taux de profit. Ricardo note pourtant lui-même que les ouvriers sont toujours en concurrence avec les machines et que l'accroissement des salaires stimule la mécanisation de la production qui favorise le profit.

Ricardo réfute qu'à part le renchérissement des biens agricoles, d'autres causes puissent entraîner une chute tendancielle du taux de profit. Ce qui l'amène à critiquer la vision de Smith, selon laquelle l'accroissement du stock de capital à long terme entraînera une concurrence plus acharnée, qui réduira le taux de profit. Sa réplique invoque un argument que nous approfondirons dans la section consacrée à J.B. Say : toute hausse de la production et de l'offre crée du revenu et s'accompagne d'une hausse de la demande ; tout investissement a donc nécessairement des débouchés.

[1] Ricardo [297] p. 106

2.3.2. Mill, la synthèse parfaite de l'économie classique

John Stuart Mill, le fils de James Mill, est le génial compilateur de l'économie classique. Il en fait la synthèse qu'il présente avec plus de clarté mais aussi plus de nuance que ses prédécesseurs. « Pendant toute la seconde moitié du XIXe siècle, les Principes d'Economie Politique de Mill furent la bible incontestée des économistes »[1]. Cet ouvrage date de 1848, mais sa lecture demeure intéressante au XXIe siècle.

Concernant la valeur d'échange, Mill fait une remarque préliminaire très importante. Il est toujours question du prix d'un bien comme si tous les acheteurs payaient le même prix pour une même marchandise. Mill estime que sur les marchés de gros, règne en effet un prix unique. Par contre, sur le marché du commerce de détail, la multitude des prix prévaut. La faute en est imputable aux consommateurs, qui ne se donnent pas la peine de toujours rechercher le meilleur prix. Si les consommateurs étaient plus rigoureux, la concurrence entraînerait l'avènement d'un prix unique. Donc toute théorie de la valeur n'est valide que moyennant l'hypothèse « supposing all parties to take care of their own interest »[2]. La conception selon laquelle la concurrence conduit à un prix unique, loin d'être propre à Mill, est largement partagée.

Comme les autres économistes classiques, Mill croyait à la détermination de la valeur par le coût de production. Mais il analyse plus à fond le lien entre cette loi et jeu de l'offre et la demande sur le marché.

Deux conditions doivent être réunies pour qu'un bien quelconque représente une valeur : il doit être utile et « there must be some difficulties in its attainment ». Ladite difficulté provient soit d'une limitation absolue de la disponibilité du bien, soit des efforts qu'il faut mettre en œuvre pour le produire. A partir de là, Mill distingue trois cas :

1- le bien est absolument limité en quantité. C'est le cas de vignobles de grand crû. ou des œuvres d'art.
2- le bien peut être produit à volonté moyennant le travail et l'investissement du capital nécessaire. C'est le cas le plus courant.
3- le cas intermédiaire : la production d'une plus grande quantité du bien, toujours possible, entraîne un accroissement de son coût.

Les biens du premier cas, en quantité limitée, voient leur prix déterminé exclusivement par la loi de l'offre et de la demande.

Outre des raisons naturelles, l'offre peut aussi être limitée artificiellement : c'est le monopole. Le monopoleur peut fixer le prix aussi haut qu'il le désire, mais il ne le peut qu'en limitant l'offre.

Dans le deuxième cas, lorsque les biens peuvent être produits sans limite, leur prix est également déterminé par l'offre et la demande. Mill imagine la situation où l'offre et la demande aboutissent à un prix inférieur au coût de production. Alors, le bien ne sera tout simplement pas mis en production. Les capitalistes n'admettront pas de supporter des pertes si ce n'est de façon tout à fait temporaire. Le coût de production est donc un minimum et ceci est également vrai pour les biens de la première catégorie. Pour les biens de la deuxième catégorie, ce minimum est en même temps un maximum, car des profits plus élevés attireraient de nouveaux offreurs, ce qui ramènera le prix au niveau du coût de production. L'offre et la

[1] Blaug [35] p. 209.
[2] Mill [258] § 3.1.5.

demande ne déterminent pas le prix, mais permettent à celui-ci de se fixer à son niveau naturel. Les écarts sont nécessairement temporaires.

Mill donne l'exemple d'un progrès technique abaissant le prix de revient. Tant que le prix de marché n'a rejoint le nouveau coût de production, les profits excédentaires attireront de nouveaux producteurs, ce qui accroîtra l'offre. Selon Mill, il n'est même pas nécessaire que ces nouveaux producteurs arrivent effectivement. Les offreurs en place auront intérêt à réduire leur prix de vente simplement pour dissuader les nouveaux entrants.

Dans le troisième cas, le prix est déterminé par le coût de production dans les circonstances les moins avantageuses. Il laissera une rente aux producteurs favorisés. On reconnaît ici l'empreinte de Ricardo.

La démonstration de Mill est magistrale. Il analyse séparément les trois cas, mais il montre que fondamentalement, c'est la même loi qui est à l'œuvre dans chacun d'eux.

Mill s'attache à un cas particulier de la loi de la valeur : *la production jointe*. Il s'agit d'un processus de production dont sortent différents biens, sans qu'il soit possible de produire l'un sans l'autre. Par exemple la viande de mouton et la laine ou la viande de bœuf et la peau à tanner. La viande de mouton et la laine ont un coût de production commun qui détermine la somme de leurs prix. Mais la façon dont cette somme se répartit entre les deux composants est fonction exclusivement de l'offre et de la demande de chacun d'eux.

Ne quittons pas la loi de l'offre et la demande sans mentionner deux remarques que Mill fait à son sujet :

- A propos de la demande, seule compte la *demande solvable* : « A beggar may desire a diamond ; but this desire, however great, will have no influence on the price »[1].
- Mill découvre un paradoxe : le prix dépend de l'offre et de la demande mais celles-ci dépendent du prix : « How to solve the paradox of two things each depending upon the other ? »[2]. Il s'escrime à apporter une explication, sans vraiment clarifier[3].

Dans le sillage de Ricardo, Mill revient sur le concept de *période de production*. Il donne l'exemple suivant. Imaginons que la quantité de travail et les matières premières pour produire un litre de vin ou un mètre carré de tissus soient parfaitement identiques. Pour obtenir un vin vendable, il faut le conserver en fut pendant cinq ans alors que le tissu est mis en vente beaucoup plus rapidement. Le litre de vin coûtera plus cher que le mètre carré de tissus car le prix doit couvrir cinq années de profit accumulé suivant la formule de l'intérêt composé. « All commodities made by machinery are assimilated, at least approximately, to the wine in the preceding example »[4]. Plus grande est la proportion du capital fixe dans le capital total, plus élevée sera la part des profits dans le prix. Une durée de vie (et d'amortissement) plus longue de ce capital aura le même effet.

---

[1] Mill [258] § 3.2.3.
[2] Mill [258] § 3.2.3.
[3] Il n'y a de paradoxe qu'en apparence. Mathématiquement, il y a deux fonctions $D=f(p)$ et $O=g(p)$ qui donnent respectivement les quantités demandées D et les quantités offertes O aux différents prix p. Lorsque je dis que l'offre et la demande dépendent du prix, je m'attache en fait aux dépendances fonctionnelles f et g. Lorsque je dis que le prix dépend de l'offre et la demande, il ne s'agit pas d'une dépendance fonctionnelle mais de la résolution du système d'équations dont le prix et la quantité sont les inconnues.
[4] Mill [258] § 3.4.5.

Mill consacre un important chapitre à la nature du capital, du profit et de l'intérêt. Le capital est toujours et nécessairement le résultat de l'épargne, celle-ci étant définie comme l'excédent de la production sur ce qui est consommé. Pour expliquer l'acte d'épargne, Mill reprend le concept d'abstinence qui a été avancé par son contemporain W.N. Senior. L'épargne est un renoncement à une consommation. A côté de l'acte volontaire d'épargne qui permet à l'individu d'accumuler du capital, Mill attire l'attention sur *l'épargne forcée* à laquelle sont voués les esclaves et qui accumule le capital dans d'autres mains que les leurs ; toutefois, même dans ce cas, le capital ne peut s'accumuler que si le capitaliste renonce volontairement à consommer le produit de l'épargne forcée qu'il a imposée à ses esclaves.

L'une des propositions clés du chapitre est « that industry is limited by capital »[1]. On peut rapporter cette phrase un peu mystérieuse à la *théorie du fonds des salaires* déjà présente chez Smith. Le capital crée l'emploi. Le capital fixe et le capital circulant sont mutuellement nécessaires l'un à l'autre ; le capital circulant avance les salaires. Mill affirme qu'en accroissant indéfiniment le capital, on peut accroître l'emploi sans autre limite que la main d'œuvre disponible. En conséquence, Mill dénonce cette opinion répandue mais complètement fausse, « that the unproductive expenditure of the rich is necessary to the employment of the poor »[2]. Au contraire, l'épargne des riches est nécessaire. Le capital circulant qui en résulte créera du pouvoir d'achat dans les classes laborieuses, ce qui maintiendra la demande des produits de consommation à un niveau suffisant pour absorber le produit de l'investissement.

Le profit est ce qui reste à l'entrepreneur lorsqu'il a payé les avances à la production : salaires, machines et matières. Manifestement, l'économie à laquelle s'intéresse Mill est celle d'entrepreneurs capitalistes qui dirigent eux-mêmes leur firme. Le profit rémunère trois contributions : l'abstinence, le risque et le travail de l'entrepreneur. La rémunération de l'abstinence sans plus s'appelle l'intérêt. Alors que le taux de profit peut varier d'une branche à l'autre du fait du risque, le taux d'intérêt tend à l'égalité entre les branches vu les transferts de capital. On voit que l'analyse de Mill n'est pas adaptée aux sociétés par actions.

Si l'on considère l'économie dans son ensemble, dans le produit final, tout ce qui n'est pas du profit est du salaire. « We thus arrive at the conclusion of Ricardo and others, that the rate of profit depends on wages; rising as wages fall and falling as wages rise »[3]. Mill distingue toutefois le *coût du travail* et le salaire. Des salaires élevés pour un travail efficace représentent un coût du travail moindre que des salaires bas pour un travail inefficace.

De quels facteurs dépendent le taux de profit et le taux d'intérêt ? « To popular apprehension it seems as if the profits of business depended upon prices. A producer or dealer seems to obtain his profits by selling his commodity for more than it cost him »[4]. Ceci n'est qu'une observation superficielle ; la cause réelle du profit est que le travail produit plus que ce qui est nécessaire à son entretien. Le profit naît dans la production et non dans l'échange. Cette idée deviendra plus tard l'un des leitmotivs de l'économie marxiste.

Mill estime que le déclin du taux de profit est la contrepartie inévitable de l'accroissement général du capital investi ; ce taux finira par atteindre un niveau insuffisant pour inciter à l'épargne ; c'est donc l'état stationnaire qui nous attend à longue échéance. Mill rend même « sa » chute du taux de profit plus probable que celle de Ricardo, car l'augmentation de la

[1] Mill [258] § 1.5.1.
[2] Mill [258] § 1.5.3.
[3] Mill [258] § 2.15.7.
[4] Mill [258] § 2.15.4.

population n'est plus une condition nécessaire ; à population constante, plus de capital signifie des salaires plus élevés – en vertu de la théorie du fonds des salaires- et donc moins de profit. Un tel enchaînement implique la condition que toutes autres choses restent égales ; Mill admet que certains phénomènes freinent cette évolution, comme l'innovation technologique qui réduit le coût des biens de consommation ouvrière, l'importation de biens à bas prix, l'exportation de capitaux ; mais –selon lui- à long terme ils se révéleront insuffisants pour la contrecarrer.

Mill s'intéresse aussi au progrès technique. Le remplacement du travail par du capital fixe, est-il source de chômage ? Beaucoup d'économistes apportent une réponse optimiste basée sur l'argument que le progrès technique diminue le coût de production, et qu'aux prix réduits qui s'ensuivent, la demande pourra absorber la production accrue. Mill est également assez optimiste, mais il reconnaît que l'argument ci-dessus ne suffit pas. D'abord, les travailleurs qui ont perdu leur emploi diminuent leur demande des produits. Ensuite et surtout, pour occuper autant de travailleurs qu'avant, il faut plus de capital : le capital fixe a augmenté du fait de l'acquisition des nouvelles machines ; mais le maintien de l'emploi implique que soit maintenu le capital circulant. D'où proviendra ce capital supplémentaire ? Ce qui justifie l'optimisme de Mill, c'est la vision globale d'une économie en croissance quasi permanente. Les firmes y font régulièrement du profit et une partie de ce profit est réinvestie pour financer les investissements ; le capital fixe a donc une source de financement permanente et assez stable, qui n'exige pas la réduction du capital circulant affecté aux salaires. Néanmoins, Mill n'exclut pas qu'à court terme, les innovations techniques jettent des travailleurs au chômage. Il réclame pour eux l'aide des pouvoirs publics. La vision de Mill sur le progrès technique et sa théorie de la chute du profit, ne forment-elles pas un ensemble un peu schizophrénique ?

L'ouvrage se termine par un chapitre dénonçant les MAUVAISES raisons poussant les Etats à intervenir dans l'économie, suivi d'une énumération des BONNES raisons. Voyons d'abord les mauvaises :

1- Les lois qui fixent un taux d'intérêt maximum. Si le taux du marché se monte à, par exemple, 6%, la limitation à 5% frustrera une partie de la demande de capitaux.
2- La fixation d'un prix maximum pour la nourriture. L'offre de produits alimentaires se réduira, à moins que des subventions ne soient accordées aux fermiers. Dans ce cas, le public paiera quand même le prix plein de sa nourriture, dont une partie via les impôts.
3- Les lois contre le monopole doivent se garder d'empêcher l'inventeur de profiter des surprofits générés par son invention.
4- Les lois contre les coalitions de travailleurs (généralisées au XIXe siècle). Mill ne croit pas que les associations ouvrières soient capables de maintenir durablement le salaire au-dessus de son prix de marché. Mill, à la suite de Smith rappelle l'importance du marchandage dans l'émergence du prix de marché. Les coalitions et les grèves sont une technique de marchandage qui permet simplement aux salariés de composer avec les employeurs sans y perdre. « Associations of labourers, of a nature similar to trades unions, far from being a hindrance to a free market for labour, are the necessary instrumentality of that free market »[1].

En vertu du fait que les intéressés sont les meilleurs juges quant aux moyens d'atteindre leurs fins, Mill conclut que « Laisser-faire, in short, should be the general practice: every departure from it, unless required by some great good, is a certain evil »[2]. Dans cette optique, Mill défend l'intervention de l'Etat dans les cas suivants :

[1] Mill [258] § 5.10.5.
[2] Mill [258] § 5.11.7.

1- L'instruction obligatoire gratuite.
2- Les services qui ont tendance à devenir naturellement des monopoles comme l'eau, le gaz…
3- Les clauses des contrats qui imposent à une partie un engagement irrévocable pour une durée illimitée doivent être proscrites.
4- « There are matters in which the interference of law is required, not to overrule the judgement of individuals respecting their own interest, but to give effect to that judgement: they being unable to give effect to it except by concert, which concert again cannot be effectual unless it receives validity and sanction from the law »[1]. Mill donne l'exemple de plusieurs agents en concurrence qui voudraient chacun limiter leur temps de travail à neuf heures par jour mais dont aucun n'ose en prendre l'initiative, du fait de la perte majeure qu'il encourrait s'il était seul à limiter son temps de travail. Le prescrit légal sera la seule issue à ce blocage paradoxal.
5- L'assistance aux pauvres. « If the poor are left to individual charity, a vast amount of mendacity is inevitable »[2].
6- Les expéditions scientifiques, si les perspectives bénéficiaires ne sont pas suffisantes pour motiver l'initiative privée.

2.3.3. Say et la loi des débouchés

Economiste, homme politique et homme d'affaire, Jean-Baptiste Say est l'un des pionniers du libéralisme économique et politique en France. Il popularisa dans son pays l'œuvre d'Adam Smith, avec qui il partage une vision économique optimiste. Il devint professeur d'économie au Conservatoire des arts et métiers en 1819 et fonda cette même année la toute-première école de commerce. Son « traité d'économie politique » paraît en 1803.

C'est le chapitre intitulé « Des débouchés » qui a rendu cet ouvrage célèbre, mais d'une réputation peu flatteuse, comme nous le verrons. Say commence avec cette idée qu'il croit très répandue : « Lorsque le placement de leurs produits est lent, pénible, peu avantageux, (les entrepreneurs) disent que *l'argent est rare* »[3]. C'est contre cette idée qu'il s'insurge. Ce n'est pas la rareté de l'argent mais celle des autres produits qui est la cause du problème. « Dans les lieux qui produisent beaucoup, se crée la substance avec laquelle seule on achète : je veux dire la *valeur*. L'argent ne remplit qu'un office passager dans ce double échange ; et les échanges terminés, il se trouve qu'on a payé des produits avec des produits »[4]. Alors se pose immanquablement la question : « Cela étant ainsi, d'où vient, demandera-t-on, cette quantité de marchandises qui, à certaines époques, encombrent la circulation, sans pouvoir trouver d'acheteurs ? Pourquoi ces marchandises ne s'achètent-elles pas les une les autres ? Je répondrai que ces marchandises qui ne se vendent pas, ou qui se vendent à perte, excèdent la somme des besoins qu'on a de ces marchandises, soit parce qu'on en a produit des quantités trop considérables, soit plutôt parce que d'autres productions ont souffert. Certains produits surabondent parce que d'autres sont venus à manquer »[5]. Clairement, une crise de surproduction ne peut pas être *générale.* On sait que Keynes sera l'un des principaux contempteurs de la loi des débouchés qu'il résumera par l'aphorisme « l'offre crée sa propre demande »[6].

[1] Mill [258] § 5.11.12.
[2] Mill [258] § 5.11.13.
[3] Say [328] p. 138.
[4] Say [328] p. 141.
[5] Say [328] p. 142.
[6] Comme nous le verrons, l'expression *loi de Say* se retrouvera abondamment dans les écrits macroéconomiques modernes.

Changeons de chapitre. Dans une société où l'échange est généralisé et où la plupart des producteurs n'ont aucun besoin du produit direct de leur travail, on imagine à quel point le troc serait peu pratique. « S'il existe dans la société une marchandise qui soit recherchée non à cause des services qu'on en peut tirer par elle-même, mais à cause de la facilité qu'on trouve à l'échanger contre tous les produits nécessaires à la consommation, une marchandise dont on puisse exactement proportionner la quantité qu'on en donne avec la valeur de ce qu'on veut avoir, c'est celle-là seulement que notre coutelier cherchera à se procurer en échange de ses couteaux, parce que l'expérience lui a appris qu'avec celle-là il se procurera facilement par un autre échange, du pain ou toute autre denrée dont il pourra avoir besoin. Cette marchandise est la monnaie »[1].

« La quantité de monnaie dont un pays a besoin est déterminée par la somme des échanges que les richesses de ce pays et l'activité de son industrie entraînent nécessairement »[2].

Say attaque ensuite l'idée que l'unité monétaire permet de mesurer la valeur d'un bien, comme le mètre mesure une longueur. « Quand on dit : cette maison vaut 20.000 francs et cette autre vaut 10.000 francs, la phrase au fond ne dit rien d'autre que : cette maison vaut deux fois autant que celle-là. (…) Qu'on appelle cela *mesurer*, j'y consens mais (…) on aura la même idée du rapport qui existe entre la valeur des deux maisons, lorsqu'on dira : l'une vaut mille hectolitres de froment et l'autre n'en vaut que cinq cents »[3]. Pire, la mesure de la valeur est sans espoir. « C'est ce qui rend impossible la comparaison qu'on a quelquefois tenté de faire des richesses entre deux époques ou de deux nations différentes. Ce parallèle est la quadrature du cercle de l'économie politique, parce qu'il n'y a point de mesure commune pour l'établir. L'argent, et même la monnaie, de quelque matière qu'elle soit composée, n'est qu'une marchandise dont la valeur est variable, comme celle de toute marchandise… »[4]. Une valeur est toujours une valeur ici et maintenant, jamais un absolu. Say critique Smith, dont il était pourtant un admirateur, pour avoir pensé que le temps de travail pouvait être l'unité de mesure générale. L'offre et la demande de travail ne fluctuent pas moins que celle des autres marchandises et avec elles la valeur du travail.

Say enchaîne avec la détermination de la valeur d'échange. Il analyse successivement la demande et l'offre. Concernant la demande, « le motif qui détermine les hommes à faire un sacrifice quelconque pour se rendre possesseurs d'un produit, est le besoin que ce produit peut satisfaire, la jouissance qui peut naître de son usage (…) Chaque individu ou chaque famille sont obligés de faire une sorte de classement de leurs besoins pour satisfaire ceux auxquels ils attachent plus d'importance, préférablement à ceux auxquels ils en attachent moins »[5]. Ici joue un facteur déterminant : le revenu. Pour une personne à revenu élevé, il faudra descendre bas dans le classement des préférences pour arriver aux biens désirés auxquels elle devra renoncer. Par contre, une personne à bas revenu devra restreindre sa demande et exclure la satisfaction d'un nombre important de besoins. Concernant l'offre, s'opère un processus plus ou moins similaire. « (La marchandise) qui, par ses frais de production, reviendrait plus cher que le cours ne sera pas offerte »[6].

---

[1] Say [328] p. 241
[2] Say [328] p. 248
[3] Say [328] p. 281
[4] Say [328] p. 283
[5] Say [328] pp. 316 et 317
[6] Say [328] p. 320

La synthèse de l'offre et la demande peut s'exprimer ainsi : « J'ai dit que le prix des produits s'établissait en chaque endroit au taux où les portent leurs frais de production, pourvu que l'utilité qu'on leur donne fasse naître le désir de les acquérir »[1]. Quant aux frais de production, en définitive, ils se résolvent entièrement dans la rétribution des services producteurs, que sont le capital, la terre et le travail. « Le prix du produit sera dès lors la somme nécessaire pour payer les services indispensables à sa création ». Cela ressemble à la théorie de Smith, mais, comme l'explique Say, il y a une grosse différence : les économistes anglais (les classiques) considèrent le prix des services producteurs comme donné. Si le prix d'un produit s'élève, les économistes anglais en déduisent un afflux de services producteurs dans la production de ce bien qui ramènera le prix à son niveau normal. Say, quant à lui, n'exclut pas qu'en cas de hausse de la demande d'un produit, la hausse de son prix puisse accroître la rétribution des services qui sont spécialisés dans sa production, du fait de caractéristiques qui les avantagent. Les économistes anglais considèrent ces avantages possibles comme des exceptions à leur règle, mais, écrit Say, « si l'on rejette dans les exceptions les avantages qu'un producteur retire de la supériorité de son jugement, de son talent ou bien des circonstances plus ou moins favorables dans lesquelles agissent ses terres et ses capitaux, alors les exceptions l'emporteront sur la règle »[2]. Quelle est la règle et quelle est l'exception ; nous reviendrons sur le fond de ce débat dans un chapitre ultérieur.

Le prix des produits est distribué entre les possesseurs des services qui concourent à leur formation. Telle est le fondement de ce qu'on appelle le revenu. « Quant aux richesses que les hommes acquièrent sans avoir concouru, directement ou indirectement, à une production quelconque, un homme n'en peut jouir qu'au détriment d'un autre ».

Say considère que la distribution des revenus se passe comme s'il y avait un échange entre les possesseurs des services producteurs et les consommateurs, où les entrepreneurs en seraient que de simples intermédiaires. Cette façon de voir deviendra courante chez beaucoup d'économistes néoclassiques.

Concernant l'intérêt ou le profit rapportés par un capital, Say fait cette mise au point pas nécessairement inutile : « Ce capital est un produit antérieur : le profit que j'en ai recueilli dans l'année, est un produit nouveau et tout à fait indépendant du travail qui a concouru à la formation du capital lui-même »[3].

Terminons par une question, qui hantera bien des économistes par la suite. « Si les profits des capitaux baissent à mesure qu'ils deviennent plus abondants, on peut se demander si, dans un pays éminemment industrieux et économe, les capitaux pourraient se multiplier au point que leurs profits se réduisent à rien. Il est difficile de croire ce cas possible ; car plus les profits diminuent, et plus diminuent aussi les motifs qui portent les hommes à l'épargne. Il est évident que l'homme qui pourrait épargner une somme sur ses revenus, la dépensera, si cette somme devient incapable d'être employée avec profit »[4].

[1] Say [328] p. 321
[2] Say [328] p. 323.
[3] Say [328] p. 401.
[4] Say [328] pp. 400-401.

2.3.4. Malthus et la controverse sur la surproduction générale

En 1820 paraissent les « Principles of Political Economy Considered with a View to Their Practical Application ». Thomas Robert Malthus y développe les mêmes thèmes que Smith, Ricardo et Say, mais avec un point de vue souvent original.

Malthus commence par poser la question : qu'est-ce que la richesse ? Il constate que la frontière des éléments à y inclure fait débat. : tout ce qui a de l'utilité ? Tout ce qui a une valeur dans l'échange ? Tout ce qui est destiné à l'échange ? Selon le critère choisi, on risque d'éliminer des éléments de façon contre-intuitive ou bien on risque le paradoxe que des unités d'un même type de bien soient les unes incluses, les autres exclues. Les talents constituent la principale pierre d'achoppement. Say les inclut dans la richesse. Mais comment les calculer, les valoriser ? Leur intégration sacrifie la précision, essentielle en économie. Au nom de la mesurabilité, Malthus ne retient finalement que les objets matériels ; on peut estimer un instrument de musique, pas la musique qu'il jouera.

Liée à la question de la richesse est celle du *travail productif* sur laquelle Smith s'était échiné. Malthus adhère à la conception de ce dernier, si ce n'est qu'il préfère substituer le terme *services personnels*, moins péjoratif, à *travail improductif*. Il dit considérer cette distinction comme la pierre angulaire de la construction smithienne et ne pas comprendre ceux qui prétendent l'admirer tout en rejetant cette distinction. Malthus estime que la définition du travail productif et de la richesse léguée par Smith est la meilleure possible, mais il avoue qu'elle n'est pas inattaquable.

Après la richesse, il est naturel d'enchaîner avec la valeur, plus précisément la valeur d'échange. Sur le marché émerge une valeur unique, même lorsque tous les participants attribuent à l'objet une valeur subjective différente. Chaque bien est une mesure de la valeur de tous les autres et un pouvoir d'achat sur eux. Mais la valeur ne suffit pas à assurer l'échange ; le troc se heurte au problème pratique de la réciprocité des besoins. Sa solution implique la circulation d'un medium dans l'échange. Le prix n'est autre que la valeur d'échange exprimée en unités de ce medium (métal précieux).

Contrairement à Smith et Ricardo, Malthus insiste sur le rôle essentiel de l'offre et de la demande dans la détermination de la valeur. La force du désir dont il est l'objet accroît la demande et donc la valeur ; la difficulté de le produire réduit son offre et accroît donc également la valeur. Evaluons un bien A et prenons le bien B comme unité. Supposons que un *A* vaille dix *B*. Ladite valeur de A peut changer INTRINSEQUEMENT en cas de modification du désir de *A* ou de la difficulté de le produire. Elle peut changer EXTRINSEQUEMENT si ce sont le désir et la difficulté de produire *B* qui varient. Si une amélioration technique touche également la production de tous les biens, les valeurs d'échange resteront constantes. La mesure standard de la valeur nécessiterait la constance de la difficulté de produire un large éventail de biens ainsi que la constance de goûts, une condition en total contradiction avec l'expérience. Malthus ne s'embarrasse donc pas de la problématique de Ricardo ; il sait devoir renoncer à la valeur intrinsèque.

Nous savons que pour Ricardo et Smith, le prix du marché établi selon l'offre et la demande est attiré par une espèce de prix permanent qui correspond au coût de production ou à la quantité de travail, ces deux critères ayant été mis en avant. Concernant le coût de production, Malthus s'accorde avec eux sur le principe, mais constate qu'empiriquement beaucoup de prix n'égalent pas le coût de production : les prix des secteurs monopolisés échappent toujours à cette détermination et à court terme tous les prix y échappent. Malthus apprécie la vision smithienne où sont additionnés le salaire nécessaire pour inciter à travailler, le profit nécessaire pour inciter à investir et la rente nécessaire pour offrir le sol. On obtient ainsi le prix minimum pour qu'un

bien soit offert, mais Malthus veut réserver l'appellation « valeur » à un concept qui intègre le désir des consommateurs autant que du coût. Malthus insiste sur le fait que les trois composantes précitées résultent du jeu de l'offre et de la demande comme n'importe quel prix.

Vis-à-vis de la valeur-travail, Malthus est plus critique. D'abord, il constate l'ambiguïté de Smith : tantôt, il semble considérer le travail incorporé dans l'objet, tantôt la quantité de travail que cet objet peut commander (contre laquelle il peut s'échanger). Il condamne la première interprétation, celle qu'adopta Ricardo, car elle ignore l'effet du capital et de la terre. L'*effet Ricardo* est d'ailleurs l'aveu implicite de la part de Ricardo que le profit influence la valeur. Malthus préfère largement la seconde interprétation, bien que, selon lui, elle ne mesure adéquatement que la valeur INTRINSEQUE.

La valeur de la monnaie est susceptible de fluctuer comme toutes les autres et Malthus critique la fiction de Ricardo considérant ces fluctuations comme négligeables. Si la facilité de l'exploitation minière s'accroît, le prix de toutes les marchandises (y compris le travail) tendra à croître ; un afflux d'or généré par la balance commerciale excédentaire aura le même effet. Par contre, la variation du taux de profit élèvera certains prix et en abaissera d'autres (effet Ricardo).

Venons-en à l'important chapitre sur la rente foncière. Malthus s'oppose à Smith, Say et Ricardo qui ne la considèrent que comme une rente de monopole (liée à l'appropriation du sol et à sa rareté) et donc un pur prélèvement sur le revenu des autres agents. Selon Malthus, elle est la digne rémunération d'une contribution ; trois conditions sont nécessaires pour qu'il y ait rente :

- La rareté des terres fertiles, qu'on ne peut certes pas nier
- La capacité de la terre d'entretenir plus de personnes que le nombre nécessaire pour la travailler. Malthus écrit: "rent is the natural result of a most inestimable quality in the soil, which God has bestowed on man- the quality of being able to maintain more persons than are necessary to work it"[1].
- La qualité des biens de subsistance (le produit de la terre) de créer leur propre demande. Plus de nourriture incite à l'accroissement de la population et donc à celui de la demande de nourriture.

En fait, Malthus s'ingénie, vainement, à trouver une différence essentielle entre la production agricole et les autres[2], une différence autre que le caractère monopolistique de l'accaparement des terres et qui serait liée à la fertilité. Cette chimère surprend moins si l'on se place du point de vue de l'homme de XIXe siècle, privé du recul que nous avons aujourd'hui pour observer l'évolution de l'agriculture devenue une industrie comme une autre où la surproduction menace continuellement malgré la part décroissante de la population active dans ce secteur et l'affectation d'une part croissante du sol à d'autres usages.

Concernant la nature de la rente, il importe peu qu'elle soit perçue par un landlord ou qu'elle s'ajoute au profit du fermier propriétaire. La rente n'est pas le profit ; elle obéit dans tous les cas à d'autres lois économiques. Malthus oppose la rente purement économique obtenue par les landlords anglais à la rente vorace due au souverain dans les empires asiatiques.

Débarrassée des jugements de valeurs et des considérations sociales, la théorie de la rente de Malthus est proche de celle de Ricardo. Toutefois, Ricardo rejette l'idée d'une rente autre que

[1] Malthus [248] p. 148. Cette invocation divine nous rappelle que parallèlement à son métier d'économiste, Malthus était également prêtre de l'Eglise anglicane.

[2] Je me permets de ne pas exposer ici ses multiples arguments en ce sens.

différentielle alors que Malthus pense que même le propriétaire de la terre la moins fertile peut en retirer une rente grâce à des occupations comme le pâturage, la collecte de bois de chauffage plutôt que de la laisser à l'abandon.

Lorsqu'une société connaît un progrès économique sur le long terme, il s'accompagne d'une augmentation à la fois absolue et relative de la rente.

- D'une part, parce que l'accroissement de la production entraîne celui de la population et mène donc à la mise en culture de terres moins fertiles ; lorsque l'éventail des niveaux de fertilité s'élargit, la rente différentielle totale s'accroit.
- D'autre part, parce que les fermiers investissent pour améliorer la terre, ce qui élargit également l'éventail des fertilités. Par la renégociation périodique des baux, conjuguée avec l'égalisation des taux de profit, le landlord récolte en définitive les fruits de ces améliorations. Malthus recommande aux propriétaires fonciers de ne pas adapter la rente trop rapidement de façon à sauvegarder l'incitation à investir dans le chef des fermiers.

Paradoxalement, Malthus tire argument de cette coïncidence entre le progrès économique et l'accroissement de la rente pour affirmer l'identité d'intérêt entre les landlords et la société dans son ensemble. Il doit toutefois admettre que la situation aux Etats-Unis caractérisée par une agriculture prospère avec une rente très basse ne conforte pas son opinion.

Après la rente, Malthus s'intéresse au salaire et au profit. Il distingue le salaire NOMINAL et le salaire REEL ce dernier tenant compte de l'évolution des prix. Leur évolution est appelée à diverger puisque le prix du blé, prépondérant dans les budgets ouvriers, doit augmenter pour les raisons expliquées ci-avant ; Malthus pense que le salaire nominal augmentera aussi mais moins rapidement que les prix ; le salaire réel tendra donc à baisser. A moins long terme, le salaire nominal varie suivant les fluctuations de l'offre et de la demande de travail. La demande dépend du *fonds de subsistance*, qui est la partie du capital circulant destinée aux salaires. L'offre de travail est fonction des habitudes de la classe ouvrière. L'analyse de l'offre de travail dans ce chapitre surprend un peu, car Malthus s'y écarte du catastrophisme qui le caractérise habituellement lorsqu'il aborde la question de la population. Certaines ouvertures laissent entrevoir la possibilité d'un réel progrès de la condition ouvrière.

Comme chez Ricardo, le salaire nominal et la rente augmentent… et le taux de profit diminue. Deux causes se conjuguent :

- Le taux de profit agricole se règle sur l'exploitation de la terre marginale. La baisse de la fertilité nécessite un accroissement de la quantité de travail pour obtenir un produit donné. A salaire constant, la part de la valeur produite disponible pour le profit doit donc baisser. Malthus fait ici peu de cas de l'accroissement des prix agricoles qu'il annonçait ailleurs. L'industrie, quant à elle, n'est pas sujette à la baisse de la fertilité du sol, mais elle en subit le contrecoup. La baisse du profit agricole entraîne le transfert des capitaux de l'agriculture vers l'industrie, ce qui y accroît la concurrence. Les prix des biens industriels doivent donc diminuer. En définitive, tous les profits baissent à cause de l'évolution de la fertilité.
- En outre, le salaire nominal ne restera pas constant mais devrait hausser. Malthus évoque une pression de la demande de travail sur l'offre, ce qui n'est pas cohérent avec le reste de sa théorie, mais on a vu que la hausse du prix du blé amènera ce même effet.

La deuxième et dernière partie de l'ouvrage est intitulée « On the Progress of Wealth ». Il s'agit d'une leçon de macroéconomie qui anticipe largement les conceptions modernes. L'analyse intègre la croissance économique et la problématique des crises de surproduction.

On sait que chaque produit fait l'objet d'une offre et d'une demande ; Malthus agrège l'offre et la demande au niveau de l'économie prise globalement. Pour favoriser une croissance

harmonieuse, il faut assurer la croissance de l'offre et de la demande globales dans des proportions adéquates.

Malthus examine successivement les facteurs favorables à l'offre globale et ceux favorables à la demande globale. Concernant l'offre, il cite :

1- ***L'accroissement de la population***. Si l'offre de travail s'accroît, le salaire tend à baisser, ce qui incitera à l'embauche. Malthus reconnaît que ce principe comporte en lui sa propre limite. Par ailleurs, ce facteur est également favorable à la demande puisqu'il accroît le nombre de consommateurs.
2- ***L'épargne et l'accumulation de capital***. L'épargne est à la fois nécessaire parce qu'elle crée le capital et nuisible parce qu'elle réduit la consommation. La difficulté est de la proportionner adéquatement. La circonstance la plus favorable est celle d'une hausse du revenu national, car l'épargne peut alors croître sans imposer la baisse de la consommation.
3- ***La fertilité du sol***
4- ***Les inventions***. Malthus fait preuve d'un optimisme modéré à leur égard. Si l'invention réduit le coût de production d'un bien pour lequel existe un fort potentiel de hausse de la consommation, la valeur produite totale de ce bien pourrait augmenter. Dans le cas contraire, le revenu libéré doit se reporter sur d'autres produits. L'expérience montre que cette adaptation ne se produit pas toujours immédiatement.

Les facteurs favorables à la demande sont :

1- ***La division de la propriété terrienn***e. Un nombre restreint de très grands propriétaires fonciers ne peuvent assurer une consommation suffisante. Il est souhaitable de constituer une classe moyenne aisée dans les campagnes. Mais une division de la propriété à l'excès serait contre-productive, car les revenus des petits propriétaires ne leur permettraient pas de participer à la consommation de luxe. Malthus regrette donc l'abrogation de la primogéniture, telle qu'elle fut décidée par le gouvernement révolutionnaire en France.
2- ***Le commerce national et international***. Il permet à chaque unité produite d'aller là où elle est le plus désirée, c'est-à-dire là où les demandeurs offrent le prix le plus élevé.
3- ***L'existence d'une classe de consommateurs nets***. Les capitalistes ont tendance à trop épargner, en vue de constituer une grosse fortune. Il faut donc un contrepoids, un groupe qui consomme plus qu'il ne produit. Il y a bien-sûr les landlords, mails il est important qu'ils soient entourés d'une multitude de prestataires de *services personnels* : domestiques, prêtres, médecins et fonctionnaires et soldats. Cependant, ces derniers sont payés à partir des impôts et non par un acte volontaire comme les domestiques ; il faut veiller à ce que leur coût n'implique pas une fiscalité qui découragerait l'investissement.

Revenons à la loi des débouchés de Say. L'impossibilité d'une surproduction GENERALE (par opposition à PARTIELLE) est une idée partagée par la grande majorité des économistes de l'époque et notamment par Ricardo et Mill. Malthus ainsi que l'économiste suisse Sismondi s'y opposent. Il s'ensuit tout un débat connu sous le nom de « general glut controversy ».

Concernant la loi de Say, Malthus écrit : "It is by no means true, as a matter of fact, that commodities are always exchanged for commodities. An immense mass of commodities is exchanged directly, either for productive labour, or personal services: and it is quite obvious, that this mass of commodities, compared with the labour with which it is to be exchanged, may fall in value from a glut just as anyone commodity falls in value from an excess of supply, compared either with labour or money"[1].

[1] Malthus [248] p. 316.

Malthus considère que l'épargne est une fuite dans le circuit économique. L'individu peut épargner autant qu'il le souhaite, mais il n'en va pas de même de la société, dont l'épargne globale doit être proportionnée à la demande finale, qui, seule, peut valider tous les investissements qui seront opérés. Les opportunités d'investir rentablement ne sont pas illimitées. "But if the conversion of revenue into capital pushed beyond a certain point must, by diminishing the effectual demand for produce, throw the labouring classes out of employment, it is obvious that the adoption of parsimonious habits beyond a certain point, may be accompanied by the most distressing effects at first, and by a marked depression of wealth and population afterwards"[1]. J'insiste sur l'expression "effectual demad" qui revient régulièrement dans le chapitre et a la même signification que l' « effective demand » de Keynes. Celui-ci louera le travail clairvoyant de Malthus.

Malthus émet à l'encontre la loi des débouchés ces trois arguments :

1- Si la propension à épargner s'élève, la diminution du besoin abaissera la valeur des produits, qui pourrait tomber sous le coût de production.
2- Si la capacité productive s'élève considérablement, au lieu d'avoir tout le monde qui achète plus à tout le monde, il se pourrait que la population préfère un surplus de temps libre.
3- Selon Ricardo, l'individu qui investit ne diminue pas la demande, car par ricochet le revenu investi se retrouvera dans les mains de salariés qui le dépenseront en consommation. Malthus rétorque que si les riches affectent une forte proportion de leur revenu en ce sens, les biens de consommation ouvrière se retrouveront vite en état de pléthore.

2.3.5. Les débats monétaires à l'époque classique

Revenons à la *Bullionist controversy* déjà évoquée. Le contexte monétaire de l'époque est certes devenu anachronique, mais les arguments échangés ne le sont pas pour autant. Le pamphlet « The high price of the bullion, a proof of the depreciation of bank notes » est une critique de la politique monétaire laxiste qui prévalait à partir de la suspension de la convertibilité des billets en 1797. Ricardo reprécisera ensuite ces idées dans les *Principles*.

Pour appréhender les arguments de **Ricardo**, il importe bien comprendre sa théorie de la valeur de la monnaie. Ici une première difficulté surgit : il a apparemment deux théories :

- Premièrement : « Gold and silver, like other commodities, are valuable only in proportion to the quantity of labour necessary to produce them »[2]. Cela vaut pour les lingots. Et pour les pièces de monnaie ? Egalement, lorsque le monnayage est libre mais uniquement dans ce cas. Quant aux billets de banque, leur valeur suit celle des pièces qu'ils représentent et dépend donc également du coût de production du métal précieux.
- Deuxièmement : La valeur de la monnaie varie en proportion inverse de son abondance relativement au volume des biens dont elle assure la circulation. Que cette abondance soit due à la découverte de mines d'or ou à l'émission de papier-monnaie ne change pas la donne. En quantité excessive, pièces et billets, dont les valeurs sont indissolublement liées, peuvent être dévalorisés par rapport au lingot. Inversement, leur valeur se fixera au cours légal, même lorsque celui-ci dépasse le coût de production, pour autant que l'Etat limite l'offre de monnaie, dont il a le monopole.

[1] Malthus [248] p. 326.
[2] Ricardo [297] p. 213. Mentionnons un argument avancé par Wicksell contre l'idée que la valeur de la monnaie est déterminée par le coût de production de l'or : la production actuelle de l'or ne représente qu'une faible fraction de la totalité de l'or détenu dans le monde ; trop faible pour être influente.

Entre ces deux théories, la contradiction n'est qu'apparente. Si la quantité de monnaie métallique est excessive, sa valeur baissera jusqu'en en dessous du coût de production du métal précieux. En conséquence, la production de celui-ci sera réduite et une partie plus grande sera gardée hors de la sphère monétaire. La valeur de la monnaie remontera alors jusqu'à ce qu'elle revienne au coût de production. Si elle dépasse ce coût, la production du métal précieux sera stimulée. Les deux conceptions se concilient donc sur le long terme. Lorsque ce sont les billets de banque qui sont en excès, ce processus autorégulateur ne peut pas jouer et la deuxième explication prend le dessus sur la première.

La controverse a essentiellement porté sur la convertibilité des billets, mais fondamentalement, reconnaît Ricardo, ce n'est pas la convertibilité qui compte, mais la quantité. Pourquoi a-t-on tant parlé de la convertibilité ? Parce qu'elle est un garde-fou contre une émission pléthorique. Supposons que trop de billets soient émis. La valeur de la Livre Sterling sera réduite. Les produits deviendront plus chers en Grande Bretagne que dans les autres pays. Les Britanniques auront avantage à exporter leurs pièces pour importer des produits étrangers[1]. A cette fin, ils devront retirer leurs pièces des banques. Les banquiers se sentiront alors forcés d'appliquer le seul remède effectif, la réduction du volume des billets. Un entêtement de leur part entraînerait une hémorragie d'or presque suicidaire, qui devrait être dissuasive.

D'après Ricardo, l'expérience montre que sans la contrainte de la convertibilité, les émetteurs de billets, tant publics que privés, tombent facilement dans l'abus et le laxisme consistant à émettre des billets au-delà de la quantité qui préserve le pouvoir d'achat de la monnaie. « Parliament, by restricting the Bank from paying in specie, have enabled the conductors of that concern to increase or decrease at pleasure the quantity and amount of their notes »[2]. Le risque d'une dépréciation illimitée existe.

La règle préconisée par Ricardo consiste à limiter l'émission de papier-monnaie de telle façon que la valeur de la monnaie reste ce qu'elle serait si la monnaie métallique représentait la totalité. Ce principe assez restrictif est appelé le *Currency principle*. Paradoxalement, Ricardo souhaite en même temps que les billets de banque se substituent entièrement aux pièces dans la circulation, afin d'en réduire le coût. Il écrit : « A currency is in its most perfect state when it consists wholly of paper money but of paper money of an equal value with the gold which it professes to represent »[3]. A cette fin, dans des écrits postérieurs *au high price of the bullion*, Ricardo prône la convertibilité des billets en lingots et non plus en pièces.

Ricardo rejette la théorie selon laquelle l'augmentation de la masse monétaire ferait chuter le taux d'intérêt, ce qui stimulerait l'industrie. Le taux d'intérêt dépend de l'abondance ou de la rareté du capital productif et celui-ci ne gonfle pas nécessairement à la suite d'une émission monétaire. Ricardo exclut même que cela puisse être le cas. En fait, le taux d'intérêt se règle sur le profit que le capital peut rapporter. On acceptera de payer cher un capital qui rapportera beaucoup. Au risque de sembler se contredire, Ricardo admet qu'une création importante de billets de banque par des opérations de crédit fera baisser le taux d'intérêt à court terme mais- ajoute-t-il- la hausse des prix qui résultera de cette émission « absorbera » ce surplus de liquidités, ce qui ramènera le taux d'intérêt à son niveau antérieur.

---

[1] L'interdiction de l'exportation des pièces était peu respectée et facile à contourner. Ricardo estime qu'il n'y a pas lieu de s'alarmer de la fuite des pièces à l'étranger, car en bon libéral, il ne croit pas que l'intérêt général puisse s'opposer à l'intérêt particulier.

[2] Ricardo [297] p. 276.

[3] Ricardo [297] p. 218.

La théorie monétaire de Ricardo se caractérise par un pessimisme qu'on ne retrouve pas chez Smith, qui fut pourtant l'une de ses principales influences. Quelles sont les principales idées de **Smith** concernant la monnaie ?

On s'en doute, la production et l'entretien de la monnaie métallique sont coûteux. D'après Smith, ces activités ne peuvent être comptées dans le revenu national net d'une société. La monnaie sert à acheter les biens ; si dans le revenu national, on additionne la monnaie et les biens, on compte deux fois la même chose. « The revenue of the society consists altogether in those goods and not in the wheel which circulates them »[1].

En conséquence, Smith défend l'émission de billets au-delà de leur couverture métallique, car cela réduit la quantité de métaux précieux nécessaire pour assurer la circulation, ce qui permet au pays d'affecter plus de moyens à la production. D'après Smith, ce n'est pas tant le rapport entre les billets et les métaux précieux qui importe que le rapport entre la somme des deux et la masse des biens qu'elle doit faire circuler. Il considère que les billets se substituent à l'or plus qu'ils ne s'y ajoutent. Sans les billets, la monnaie comporterait plus de métaux précieux.

C'est principalement par l'escompte de lettres de change (bills of exchange), que les banquiers émettaient leur papier monnaie à l'époque. Smith trouve cette pratique très saine. C'est sa fameuse doctrine des *real bills*: « when a bank discounts a real bill of exchange drawn by a real creditor upon a real debtor, and which, as soon as it becomes due, is really paid by the debtor, it only advances to him a part of the value which he would otherwise be obliged to keep unemployed and in ready money for answering occasional demands »[2]. Grâce à ce type de crédit, les entrepreneurs sont dispensés de devoir avancer eux-mêmes un capital plus élevé dont une certaine part resterait en permanence à l'état de monnaie souvent inactive. Tant que la création monétaire respecte cette condition, elle ne peut être excessive, estime Smith. Par contre, il reconnaît qu'elle ne doit pas servir à fournir la totalité ou même la plus grande part du capital d'une entreprise. Mais pense-t-il, des avances aussi larges seraient contraires à l'intérêt des banques, car le retour serait trop lointain.

On comprend que la doctrine de Smith, tout particulièrement celle des *real bills*, aura servi de matière première aux arguments des anti-bullionistes.

Le débat se calma avec le retour à la convertibilité en 1819 mais ne s'arrêta pas. Il rejaillit même dans les années 1840. En effet, en 1844, le gouvernement britannique institua une réforme profonde du système monétaire, largement inspirée par le *currency principle*. La Banque d'Angleterre obtint le monopole de l'émission de billets et une couverture métallique complète lui était imposée pour les billets qu'elle émettait au-delà d'un montant déterminé. La controverse fit rage entre la *Currency School*, héritière des bullionistes, et la *Banking School* adversaire du currency principle, dont le principal avocat était Thomas Tooke.

En 1844, **Tooke** publie son pamphlet « An inquiry into the currency principle ». Le currency principle considère la monnaie métallique comme la monnaie idéale et veut faire des billets de banque un simple substitut de la monnaie métallique. A cette fin, il se préoccupe particulièrement du volume des billets en circulation. Cette focalisation, selon Tooke, s'avère surannée alors que la circulation monétaire évolue vers des formes différentes ; Tooke note

[1] Smith [336] § 2.2.14.
[2] Smith [336] § 2.2.59.

principalement les chèques et les lettres de change qui selon lui ont éliminé les billets de banque des paiements entre commerçants[1].

Tooke s'élève contre l'opinion selon laquelle la banque d'Angleterre peut faire varier la masse des billets en circulation dans un sens ou dans l'autre suivant son bon plaisir : « in no case could the Bank, by its own volition, add one hundred thousand pounds, perhaps not one hundred pounds to the amount already in circulation among the public »[2]. Les billets créés à l'initiative exclusive de la Banque d'Angleterre lui reviendraient rapidement sous forme de dépôts ou dormiraient dans les coffres des autres banques ; dès après leur création, ils seraient condamnés à l'oisiveté. En fait, le maintien de ces billets dans la circulation effective ne pourrait se faire qu'au prix d'une baisse sensible du taux d'intérêt. L'opinion de Tooke ne signifie pas que la banque centrale est incapable de créer et faire circuler de la nouvelle monnaie, mais qu'elle ne peut forcer l'économie à se servir d'une monnaie additionnelle que celle-ci n'a pas demandé.

C'est un leitmotiv de Tooke que « the prices of commodities do not depend upon the quantity of money indicated by the amount of bank notes nor upon the amount of the whole of the circulating medium ; but, on the contrary, the amount of the circulating medium is the consequence of prices »[3]. En fait, son avis est moins tranché ; il distingue deux cas:

- Une émission de monnaie inconvertible à cours forcé et légal pour financer des dépenses publiques. Il s'agit d'une « fresh source of demand »[4]. Elle agit causalement sur les prix.
- Le réescompte de lettres de change. Les lettres de change suivent la circulation des biens et les billets sont repris par l'émetteur à l'échéance. Il n'y a pas d'effet inflationniste. C'est la doctrine des *real bills*, qui fit l'objet d'intenses débats déjà lors de la *Bullionist controversy*.

Aujourd'hui, l'*open market policy* est devenue de très loin le principal mode de création monétaire et elle ne correspond précisément à aucun des deux cas mentionnés par Tooke. Il est certain qu'elle ne peut fonctionner que moyennant la manipulation des taux d'intérêt.

Tooke s'attaque également à l'opinion convenue qu'un taux d'intérêt bas favorise l'inflation et qu'un taux élevé la freine. Le prix des biens est déterminé par leur coût de production dont l'intérêt fait partie : le taux d'intérêt élevé est donc cause d'inflation. A l'appui de sa thèse, Tooke avance des données historiques indiscutables montrant que le niveau des prix fluctue dans le même sens que le taux d'intérêt. Au sous-chapitre 4.3.2, nous verrons que Wicksell donnera à ce problème connu sous le nom de *paradoxe de Gibson* une solution conciliant ces données empiriques avec la conception traditionnelle selon laquelle un taux bas stimule l'inflation.

Si les prix ne sont influencés ni par la masse monétaire ni par le taux d'intérêt, à quoi peut-on attribuer leurs fluctuations ? « …the fluctuations of price which are over and above those that are necessary incidental to the nature of commodities, are attributable to the expansion and contraction of credit, under the influence of the opinion of dealers and speculators, more or less exaggerated, of the prospects of markets »[5]. Les banques manquent parfois de prudence et emboîtent le pas à des transactions spéculatives ou exagérément optimistes. Tooke estime que les soubresauts du crédit et du taux d'intérêt seraient encore pires si on appliquait la

[1] Ces documents font circuler la monnaie scripturale qui est composée des montants déposés sur les comptes à vue.
[2] Tooke [366] p. 60.
[3] Tooke [366] p.123.
[4] Tooke [366] p.71.
[5] Tooke [366] pp.86-87.

revendication du Currency Principle tendant à séparer strictement les activités d'émission de billets des autres activités bancaires, car l'imbrication et la similarité de ces deux types d'activité permet aux excès de l'un d'être éventuellement compensés par l'autre.

Un sujet largement débattu à l'époque classique, comme ultérieurement bien-sûr, est celui de la ***neutralité de la monnaie***. Dans son intéressant article « Nonneutrality of Money in Classical Monetary Thought », Humphrey montre que contrairement à une idée répandue, l'effet de la circulation monétaire sur l'activité économique est admis par une majorité d'économistes de cette époque ; il dénombre pas moins de neuf arguments qu'ils avancent pour expliquer cet *effet réel*. Ricardo est la principale voix discordante dans ce concert, avec son aphorisme « Money cannot call forth goods ». Les effets réels de la monnaie ont été mis en avant par Hume, Thornton, Torrens, Malthus, Mc Culloch…

Citons **Thornton**: "When the Bank of England enlarges its paper, it augments, in the same degree, as we must here suppose, its loans to individuals. These favoured persons immediately conceive, and not without reason, that they have obtained an additional though borrowed capital, by which they can push their own particular manufacture or branch of commerce…"[1]. Mais il ajoute que si dans un premier temps, ce surplus d'activité offre de l'emploi à des personnes inoccupées, ensuite, si l'accroissement monétaire se poursuit, les entrepreneurs dotés de nouveaux moyens monétaires débaucheront des salariés déjà employés ailleurs ; l'effet bénéfique de l'accroissement monétaire sera alors épuisé.

Un deuxième canal s'offre à cet effet réel : la pression exercée par les billets additionnels sur la demande de produits accroît leur prix. Les prix plus élevés stimulent la production, du moins si les salaires n'augmentent pas parallèlement aux prix des produits. Précisément, lorsque la hausse des prix est perçue comme temporaire, les salaires ne suivent généralement pas le mouvement. Comme l'écrit Thornton, « … the rate of wages, we know, is not so variable as the price of goods »[2].

Malthus, pour sa part, pense que l'augmentation du stock monétaire accroît le capital par un effet de redistribution des ressources des classes à revenus fixes vers les entrepreneurs. Il écrit: "The new notes go into the market, as so much additional capital, to purchase what is necessary for the conduct of the concern. But before the produce of the country has been increased, it is impossible for one person to have more of it, without diminishing the shares of some others. This diminution is effected by the rise of prices, occasioned by the competition of the new notes, which puts it out of the power of those who are only buyers, and not sellers, to purchase as much of the annual produce as before (…) But, on every fresh issue of notes, not only is the quantity of the circulating medium increased, but the distribution of the whole mass is altered. A larger proportion falls into the hands of those who consume and produce, and a smaller proportion into the hands of those who only consume"[3]. Ici, la hausse des prix causée par la monnaie additionnelle est transitoire ; elle équilibre le système en attendant que les nouveaux biens arrivent sur le marché.

*

**

Avant de conclure ce chapitre, il convient d'insister sur l'importance de Ricardo dans l'histoire de la science économique. En avance sur son temps, il concentre son attention non plus sur l'économie réelle mais un modèle simplifié de celle-ci. Le modèle comporte un nombre

[1] Thornton [358] p. 236.
[2] Thornton [358] p. 119.
[3] Malthus [247] p. 364.

restreint de variables ; le rôle de l'économiste est de mettre en lumière les relations entre elles. Les générations suivantes d'économistes mathématiseront ces relations ; Ricardo se contente encore de donner des exemples chiffrés. Blaug estime la pensée de Smith plus riche en idées que celle de Ricardo, mais il ajoute : « Mais si l'économie politique est essentiellement un outil d'analyse, une méthode de pensée plutôt qu'un ensemble de résultats, Ricardo inventa littéralement la technique de l'économie politique »[1].

[1] Blaug [35] p. 157.

2.4- DEUX DIRECTIONS OPPOSEES DE L'ECONOMIE POLITIQUE

C'est Marx qui est l'auteur de l'appellation « les économistes classiques » pour désigner Smith, Ricardo et Mill, ainsi que leurs disciples. Il se considère comme leur héritier. A l'opposé, Cournot anticipe les résultats d'une école qui ne se développera qu'à la fin du XIXe siècle.

2.4.1. Cournot : l'introduction du calcul différentiel en économie

Antoine Augustin Cournot était un mathématicien réputé. Ses « Recherches sur les principes mathématiques de la théorie des richesses » (1838) sont le premier véritable ouvrage d'économie mathématique. Pour la première fois, un ouvrage économique abonde d'équations et de graphiques. Cournot est un précurseur de la théorie néoclassique que nous étudierons aux titres III et IV; ce sont d'ailleurs Walras et Jevons qui permettront à l'ouvrage de Cournot de sortir de l'anonymat.

A propos du concept de **valeur**, Cournot émet cette intéressante remarque : « De même que nous ne pouvons assigner la situation d'un point que par rapport à d'autres points, ainsi nous ne pouvons assigner la valeur d'une denrée que par rapport à d'autres denrées. Il n'y a en ce sens que des valeurs relatives. Mais lorsque ces valeurs relatives viennent à changer, nous concevons clairement que la raison de cette variation peut se trouver dans le changement de l'un des termes du rapport, ou de l'autre ou de tous deux à la fois »[1]. Contrairement à Ricardo, Cournot croit sans espoir la recherche d'un étalon invariant.

La fonction de **demande** est centrale dans l'analyse de Cournot. Elle est représentée mathématiquement par la fonction D = F(p) où *p* est le prix du bien concerné. La fonction *F* est évidemment décroissante. Sa forme dépend de facteurs comme les goûts des consommateurs, la distribution des revenus… La fonction réciproque p = f(D) fera également quelques apparitions dans les développements mathématiques qui suivent.

Chez Cournot, la fonction de demande n'est pas confrontée à une fonction d'offre. La variable *D* représente donc aussi bien la quantité produite et la quantité échangée que la quantité demandée.

Cournot pose l'hypothèse que la fonction F(p) est continue, « c'est-à-dire une fonction qui ne passe pas soudainement d'une valeur à une autre, mais qui prend dans l'intervalle toutes les valeurs intermédiaires ».

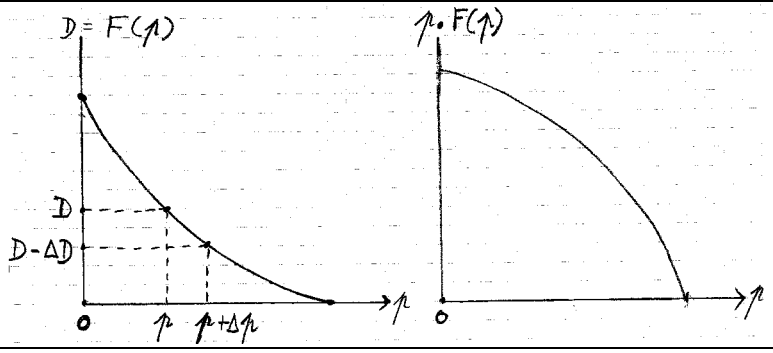
En multipliant la quantité échangée par le prix, on obtient évidemment le chiffre d'affaires total du marché, exprimé par la formule p.F(p), qui est également une fonction de *p*. Cournot s'intéresse à sa maximisation. Pour des quantités très faibles de même que pour des prix très faibles, le chiffre d'affaire sera peu élevé, mais entre ces extrêmes il existe une valeur de *p* qui maximise le chiffre d'affaires. On sait qu'une fonction atteint un maximum lorsque sa dérivée s'annule ; dans le cas de notre chiffre d'affaires, cette condition s'écrit[2] :

$$F(p) + p.F'(p) = 0 \qquad (2.4)$$

[1] Cournot [64] p. 16

[2] Pour avoir la certitude qu'il y ait un maximum unique, il faut que la dérivée seconde F" soit négative (concavité vers l'abscisse). Selon Cournot, même en dehors de cette condition, il est très improbable que dans un intervalle de variation de *p* réaliste, la fonction p.F(p) puisse avoir un minimum.

Graphique 2.2 : la loi de la demande selon Cournot



Supposons maintenant que le prix augmente de *p* à p+Δp ; il s'ensuivra une baisse de la demande de *D* à D-ΔD. Quel sera l'effet sur le chiffre d'affaires ? L'évolution de celui-ci dépendra du solde entre ces deux influences opposées, l'influence positive du prix et l'influence négative de la demande. Cournot montre que le chiffre d'affaires augmentera si ΔD/Δp < D/p et diminuera dans le cas contraire.

Après la demande, Cournot introduit le **coût de production**. Le coût total [1] de la production *D* est représenté par la fonction φ(D). Sa dérivée φ'(D) est également une fonction de *D* et représente la variation du coût total dû à une variation infinitésimale de la quantité *D*. Cournot juge la fonction φ' plus significative que la fonction φ. Ce concept sera appelé plus tard *coût marginal*. Quand *D* augmente, *φ'* peut soit croître soit décroître, mais il sera toujours positif. Cournot le voit croissant dans les activités d'exploitation du sol ou du sous-sol comme l'agriculture et les mines et décroissant dans l'activité manufacturière (au moins jusqu'à une certaine échelle où le renchérissement des matières premières et des salaires le ferait augmenter).

*

Pour analyser les lois du marché, Cournot passant successivement en revue les différentes formes de marché dans l'ordre croissant du nombre d'offreurs : le monopole, le duopole, l'oligopole et enfin la concurrence parfaite (que Cournot appelle la *concurrence indéfinie*). Pour chaque structure de marché, il détermine le prix, la quantité échangée et le profit des offreurs à l'équilibre, toujours à partir d'une même méthode : l'annulation de la dérivée de la fonction de profit.

Commençons par le **monopole.** La fonction de profit est la quantité multipliée par le profit unitaire, lui-même égal à la différence entre le prix et le coût unitaire :

$$F(p).[p - \varphi(D)/D] \qquad (2.5)$$

A partir de maintenant, nous simplifierons l'exposé en supposant un coût nul. Il faut donc maximiser p.F(p), problème dont la solution est donnée par l'équation (2.4).

Cournot enchaîne avec l'**oligopole** (concurrence entre un nombre restreint d'offreurs), dont le **duopole** est la forme élémentaire ; toute entente entre les producteurs est exclue par hypothèse, car elle nous ramènerait au cas précédent (monopole). *D* se décompose en $D_1 + D_2$, les parts de marché respectives des deux duopoleurs notés (1) et (2). Par hypothèse, les produits offerts sont identiques et c'est donc une même fonction F(p) qui gouverne la demande des deux offreurs. Le prix est quant à lui une fonction *f* de la quantité totale $D_1 + D_2$.

[1] Le coût comprend la rémunération de tous les facteurs, y compris le capital.

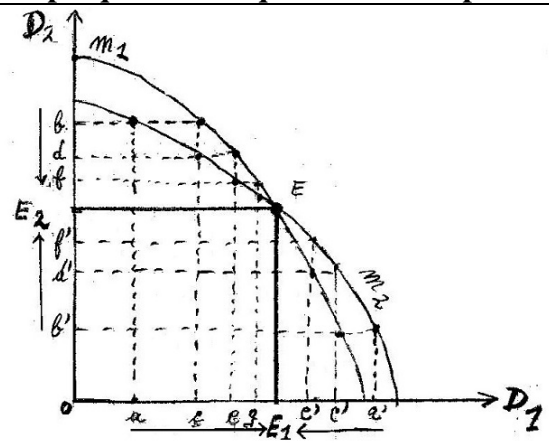
A coût nul, le profit du producteur (1) et celui du producteur (2) valent respectivement $D_1.f(D_1+D_2)$ et $D_2.f(D_1+D_2)$. Ce sont des fonctions à deux variables : D_1 et D_2. La dérivée à annuler est la dérivée partielle par rapport à la variable contrôlée : D_1 pour l'offreur (1) et D_2 pour l'offreur (2).

La maximisation du profit de l'offreur (1) et celle de l'offreur (2) sont données respectivement par les formules 2.6-A et 2.6-B qui forment ensemble un système d'équation simultanées permettant de déterminer p, D_1 et D_2 :

$$f(D_1+D_2)+ D_1.f'(D_1+D_2) = 0 \qquad (2.6\text{-A})$$
$$f(D_1+D_2)+ D_2.f'(D_1+D_2) = 0 \qquad (2.6\text{-B})$$

Comme le montrent les équations (2.6), D_1 est fonction de D_2 et vice versa. Le graphique 2.3 affiche deux courbes : la courbe m_1 représente les valeurs de D_1 décidées par le producteur (1) en fonction des valeurs de D_2 , de façon à respecter la règle de maximisation (2.6-A) ; la courbe m_2 représente les décisions du producteur (2) en fonction des valeurs de D_1. On peut les considérer comme des réactions d'un offreur aux décisions de l'autre. L'équilibre est donné par l'intersection de ces deux courbes et se caractérise notamment par l'égalité entre D_1 et D_2[1].

Graphique 2.3 : l'équilibre du duopole



Le graphique 2.3 expose un processus itératif où chaque producteur réagit à la décision précédente de son concurrent et se rapproche ainsi de sa position d'équilibre. Le producteur (1) commence et fixe D_1 à la valeur a. En réponse, le producteur (2) fixe D_2 à la valeur b ; en réponse, $D_1 = c$, puis $D_2 = d$; puis $D_1 = e$… La concurrence converge vers un équilibre E où $D_1 = E_1$ et $D_2 = E_2$. Si le producteur (1) avait fixé sa première quantité à D_1 = a', on aurait eu la suite a'-b'-c'-d'-e'… qui aurait convergé vers ce même équilibre. D'où qu'on vienne, on est attiré par ce point d'équilibre qui est donc stable. Certes, cette conclusion est dépendante de la forme des courbes ; si m_1 avait la forme de m_2 et vice versa, l'équilibre serait instable. Mais Cournot montre que lorsque $D_1 = 0$, l'équation (2.6-A) vaut nécessairement plus que (2.6-B).

En additionnant les équations 2.6-A et 2.6-B, on obtient l'équation du marché, qui après quelques manipulations algébriques devient :

[1] Ce résultat est lié à la double simplification : produit homogène et coûts non pris en compte. Dans des conditions plus réalistes, toutes autres choses restant égales, le producteur avec le coût marginal le plus faible aura la part de marché la plus grande.

$$D + 2p.\frac{dD}{dp} = 0 \qquad (2.7)$$

Alors que l'équation (2.4) du monopole peut être réécrite :

$$D + p.\frac{dD}{dp} = 0 \qquad (2.8)$$

Et que s'il y avait *n* offreurs, l'équation du marché prendrait la forme :

$$D + n.p.\frac{dD}{dp} = 0 \qquad (2.9)$$

Il apparaît clairement que plus il y a d'offreurs, plus la quantité échangée *D* est élevée et plus le prix *p* est bas. Cournot montre que le profit individuel décroît également.

Que *p* et *n* soient liés par une relation mathématique simple a de quoi surprendre même ceux qui intuitivement les imaginent varier en sens inverse. Ce résultat est lié à la conception implicite de la concurrence qui anime Cournot : une concurrence désincarnée, purement mécanique et inaboutie. Elle sera critiquée en 1883 par un autre économiste mathématicien français, Joseph **Bertrand** ; à propos de l'arrêt des hostilités entre les duopoleurs au point E, il écrit : « Une objection se présente : dans cette hypothèse, aucune solution n'est possible, la baisse n'aurait pas de limite ; quel que soit en effet le prix commun adopté, si l'un des concurrents abaisse seul le sien, il attire à lui, en négligeant des exceptions sans importance, la totalité de la vente, et il doublera sa recette si son concurrent le laisse faire. Si les formules de Cournot masquent ce résultat évident, c'est que, par une singulière inadvertance, il y introduit sous le nom de D_1 et D_2, les quantités vendues par les deux concurrents, et que les traitant comme des variables indépendantes, il suppose que, l'une venant à changer par la volonté d'un des propriétaires, l'autre pourra rester constante. Le contraire est de toute évidence »[1]. La solution de Bertrand implique la vente au prix coûtant et donc l'absence de profit au-delà de la rémunération du capital incluse dans le coût.

Le scénario de Bertrand est plus convaincant. Si les offreurs ne bride pas la concurrence entre eux, quel que soit leur nombre, celle-ci élimine le profit. Cournot ne pouvait l'appréhender car ses firmes se focalisent sur la quantité offerte plutôt que sur le prix.

En 1889, Edgeworth reprendra le scénario de Bertrand mais en ajoutant l'hypothèse que les duopoleurs font chacun face à une limite de leur capacité de production. Selon lui, dans ce cas, le prix sera indéterminé.

Lorsque *n* tend vers l'infini (mathématiquement parlant), on arrive à la limite de cette concurrence croissante, la **concurrence indéfinie**. Cournot la définit ainsi : « …chacune des productions partielles D_k est insensible, non seulement par rapport à la production totale D = F(p), mais aussi par rapport à la dérivée F'(p), en sorte que la production partielle D_k pourrait être retranchée de *D* sans qu'il en résultât de variation appréciable dans le prix de la denrée »[2].

Une fois de plus, suivant la méthode consistant à annuler la dérivée (mais cette fois en tenant compte du coût), la maximisation du profit du producteur *k* est donnée par (2.10) :

[1] Bertrand [32] p. 503.
[2] Cournot [64] p.69.

$$D_k + [p - \varphi'_k(D_k)].dD/dp = 0 \qquad (2.10)$$
$$p - \varphi'_k(D_k) = 0 \qquad (2.11)$$

Si on considère que D_k est négligeable, on doit déduire de cette équation que l'expression à l'intérieur de la parenthèse est nulle, ce qu'indique (2.11). La conclusion est qu'en concurrence indéfinie, le prix s'égalise avec le coût marginal.

Cournot montre que si la fonction φ' est décroissante, c'est-à-dire si le coût diminue quand augmente la quantité produite, les concurrents réaliseront immanquablement des pertes qu'un monopole pourrait éviter, ce qui tuera la concurrence. Par la suite, il ne retiendra plus que l'hypothèse des coûts croissants.

*

Cournot s'intéresse encore à ce qu'il appelle le *revenu social* (qu'on appelle aujourd'hui couramment le revenu national). Il analyse le cas suivant : soit un produit *A* qui renchérit à cause de l'instauration d'une taxe forfaitaire sur la consommation ou à la suite de conditions de production moins favorable. Comme nous l'avons vu plus haut, le chiffre d'affaires sur le marché *A*, en passant de $p_0.D_0$ à $p_1.D_1$, pourra soit croître soit décroître. La variation que connaîtra le revenu social (hausse ou baisse) sera approximativement égale à celle du revenu collectif des producteurs de *A*. En effet, la demande des autres biens subit deux effets inverses :

- baisse due à l'appauvrissement des producteurs de *A*
- hausse car les consommateurs renoncent à acheter du *A* devenu plus cher.

Cournot montre que ces écarts se compensent. Mais la variation du revenu social ainsi déterminée est une variation purement NOMINALE. Dans tous les cas, le renchérissement considéré entraînera une baisse REELLE du revenu social : la hausse du prix a réduit de D_0 à D_1 la production annuelle de la denrée *A*, et par cela seul anéanti annuellement une valeur égale à $p_0.(D_0 - D_1)$.

Supposons maintenant un bien qui était libre (l'eau par exemple) et qui cesserait de l'être, vu sa raréfaction ; le revenu des agents qui se lanceraient dans sa production s'accroîtrait forcément et le revenu social avec lui. « Mais ce qu'il y a de paradoxal dans cet énoncé disparaît, lorsque l'on a égard à la distinction que nous avons faite dans ce chapitre entre la valeur réelle et la valeur nominale »[1]. Ce serait seulement le revenu social nominal, qui augmenterait.

2.4.2. Marx : de la plus-value au profit

Karl Marx était à la fois philosophe, théoricien politique et économiste. Refusant d'envisager l'exploitation de l'homme par l'homme comme une question éthique, il a voulu donner le dernier mot à la science, en l'occurrence l'économie politique. L'économie de Marx fut largement influencée par les théories classiques, principalement celle de Ricardo que Marx admirait.

« Le Capital » est l'ouvrage le plus important que Marx a consacré à l'économie politique. Il l'a rédigé pendant toute la décennie 1861-1870 ainsi qu'en 1877 et 1878, après une interruption due à son état de santé. Le premier livre est paru en 1867 ; les deux suivants, inachevés, ont été publiés à titre posthume respectivement en 1885 et 1894 par son ami Frédéric Engels. Voyons en les grandes lignes.

[1] Cournot [64] pp.110-111.

Commençons avec la thèse marxiste, bien connue, selon laquelle la classe ouvrière est exploitée par celle des capitalistes. Dans la théorie marxiste, la démonstration de cette exploitation est totalement imbriquée avec la loi de la valeur-travail. Sur ce plan, Smith et Ricardo raisonnaient en deux temps : un stade primitif où la valeur est réglée par la quantité de travail suivi d'un stade industriel où les prix sont affectés par l'intervention du capital. Marx raisonne également en deux temps, mais les stades sont abstraits. Le premier est qualifié de « procès de production immédiat », sans que l'on comprenne sa spécificité, si ce n'est qu'il ne comporte pas toutes les propriétés du capitalisme moderne.

Ne nous appesantissons plus sur la loi de la valeur-travail, car Marx s'appuie sur le travail de Smith, Ricardo et Mill[1]. Citons quand-même ce passage qui illustre bien l'esprit d'un économiste philosophe : « Comme valeurs d'usage, les marchandises sont avant tout de qualité différente ; comme valeurs d'échange, elles ne peuvent être que de différente quantité. La valeur d'usage des marchandises une fois mise de côté, il ne leur reste plus qu'une qualité, celle d'être des produits du travail (…) Métamorphosés en sublimés identiques, échantillons du même travail indistinct, tous ces objets ne manifestent plus qu'une chose, c'est que dans leur production une force de travail humaine a été dépensée, que du travail humain y est accumulé. En tant que cristaux de cette substance sociale commune, ils sont réputés valeurs. »[2]

Là-dessus se greffe le concept de *plus-value*, c'est-à-dire qu'un agent économique (le capitaliste) peut accaparer un supplément de valeur, même quand toutes les marchandises y compris le travail sont échangées à leur valeur. Marx insiste sur le fait que l'échange des marchandises en tant que tel ne peut être source de plus-value. Il est courant qu'un commerçant revende une marchandise plus cher que son prix d'achat ; soit il l'a achetée en dessous de la valeur, soit il l'a vendue au-dessus. Mais aucune valeur supplémentaire n'a été créée.

Si ce n'est pas de la CIRCULATION des marchandises que le capital tire la plus-value, c'est de leur PRODUCTION[3]. Comment ? Il se fait qu'il existe une marchandise qui a la propriété particulière de pouvoir créer plus de valeur qu'elle n'en a elle-même : la *force de travail*. La valeur de la force de travail se mesure comme celle des autres marchandises par le temps de travail nécessaire pour la produire, c'est-à-dire pour produire les biens de consommation servant à l'entretien de la classe ouvrière. Mais le capitaliste devient propriétaire de la totalité de la valeur produite par le travail. Si donc il peut faire travailler le salarié plus d'heures qu'il n'en faut pour produire sa subsistance (le salaire), il obtient une plus-value que Marx assimile à du travail non payé. Marx utilise les termes *travail nécessaire* et *surtravail*, le premier reproduisant la valeur de la force de travail et le second produisant la plus-value captée par le capitaliste. La distinction entre le travail nécessaire et le surtravail est purement conceptuelle ; dans la pratique, les deux se confondent.

Marx écrit : « au milieu de notre société européenne où le travailleur n'achète la permission de travailler pour sa propre existence que moyennant surtravail, on se figure facilement que c'est une qualité innée du travail humain de fournir un produit net (…) Ricardo ne s'occupe jamais

[1] Rappelons quand-même que pour ces économistes, la quantité de travail (travail direct plus travail indirect) était une approximation de la valeur plus qu'un déterminant. Mais leur exposé était suffisamment équivoque pour permettre l'interprétation radicale de Marx.

[2] Marx [254] livre I, t1 pp. 53-54.

[3] La PRODUCTION doit être comprise au sens large. Elle inclut notamment le transport, le stockage et la manutention des marchandises. Marx s'engage dans une réflexion fastidieuse voire oiseuse, où Smith l'avait déjà précédé, pour départager les travaux productifs et improductifs de valeur d'échange. Sont notamment jugés improductifs les travaux servant à la circulation des marchandises comme le commerce et le crédit.

de la raison d'être de la plus-value. Il la considère comme une chose inhérente à la production capitaliste, qui pour lui est la forme naturelle de la production sociale »[1].

Dans la production, interviennent deux facteurs : les moyens matériels (bâtiments, outils, matières premières et consommables…) et le travail. Souvenons-nous que Ricardo avait analysé leur effet différencié sur la valeur travail : le travail a une double action : d'une part, il produit une valeur nouvelle ; d'autre part, il reproduit la valeur des moyens de production en l'intégrant dans la valeur des produits finis selon le nombre d'heures de travail nécessaire à leur production passée. Les moyens de production ne sont donc pas source de valeur nouvelle, comme le travail. Un exemple numérique facilitera la compréhension. Supposons qu'une heure de travail donne une unité de valeur. Les moyens matériels consommés dans telle production ont nécessité 1000 heures de travail dans une phase antérieure. La production présente absorbe 700 heures de travail « vivant », dont l'entretien (et donc le salaire) coûte 400 heures.

La valeur produite[2] pourra être décomposée comme suit :

$$1000\ (\mathrm{c}) + 400\ (\mathrm{v}) + 300\ (\mathrm{pl}) = 1700\ (\mathrm{val}) \qquad (2.12).$$

Marx appelle *capital constant* (noté *c*) la partie du capital servant à acheter les moyens de production et *capital variable* (noté *v*) la partie du capital qui avance le salaire. Cette appellation reflète le fait qu'à la fin du processus de production, *v* devient v+pl où *pl* est la plus-value. Si *c* passait de 1000 à 1100, la valeur du produit monterait à 1800. Si *v* passait à 500 par suite d'une hausse du salaire horaire (et non de la main d'œuvre employée), la valeur du produit resterait de 1700, mais la plus-value serait réduite à 200.

L'égalité (2.12) est l'une des équations fondamentales de l'économie marxiste[3]. En réalité, *c* représente, non pas le capital constant, mais la part de celui-ci qui est consommée dans le processus produisant les *x* unités de produit. L'exposé de Marx est imprécis, car les équations du type (2.12) sont toujours présentées comme si le capital constant et la consommation du capital constant se confondaient. C'est d'autant plus étonnant que, comme nous le verrons plus loin, Marx a analysé en profondeur le lien entre d'une part, le capital comme STOCK de valeur et d'autre part, le FLUX de valeur caractéristique de la production.

Mesurons l'exploitation de la force de travail à partir de l'équation (2.12). Marx exprime son niveau par le *taux de plus-value pl' = pl*/v, soit 75% dans notre exemple.
Il distingue deux manières de l'accroître :

- l'allongement la durée du travail à salaire égal (donc l'accroissement du *surtravail* en gardant le *travail nécessaire* fixe ; si *val* augmente alors que *c* et *v* sont constants, *pl* s'accroît).
- l'accroissement de la productivité dans les branches produisant les biens de consommation ouvrière, de façon à réduire *v*. C'est évidemment cette dernière méthode qui est la plus importante historiquement.

Le taux de plus-value dépend principalement du rapport de force entre employeurs et salariés.

Marx passe ensuite aux « formes concrètes (NDLR :prix et profit) auxquelles donne naissance le mouvement du capital considéré comme un tout ». Mais cette analyse ne vient qu'au livre III.

---

[1] Marx [254] livre I t2. pp188-189

[2] Il peut s'agir de la production annuelle d'une entreprise ou d'un secteur de production ; cette équation peut également être appliquée à la valeur d'une unité produite si on divise les variables par la quantité produite annuellement.

[3] Comme ses contemporains, Marx travaillait avec des exemples arithmétiques et non avec l'algèbre.

Après plus de mille pages rédigées comme si la valeur travail était le prix, soudain l'ouvrage bascule. Le chapitre « La transformation de la plus-value en profit et du taux de plus-value en taux de profit » éclate comme une révélation. On apprend que dans une économie capitaliste, les produits ne sont pas vendus à leur valeur, mais à ce que Marx appelle les *prix de production*. Comment ceux-ci se forment-ils ?

Le principe fondamental est celui de l'égalisation du taux de profit sous l'effet de la concurrence, déjà analysée par les autres auteurs classiques et que Marx appelle la *péréquation du taux de profit*. Le prix qui en résulte est égal au coût de production plus la marge bénéficiaire nécessaire pour que le capital investi dans sa production rapporte le profit moyen. Cette marge bénéficiaire n'est rien d'autre que la plus-value, analysée en long et en large au cours des mille pages précédentes, qui est redistribuée entre les industries en fonction de l'importance de leurs capitaux.

Marx donne l'exemple suivant, avec trois biens, qui sont chacun produits avec un taux de plus-value de 100%. :

$$\begin{aligned} &\text{A} : 80(\text{c}) + 20(\text{v}) + 20(\text{pl}) = \text{val } 120 \rightarrow 80(\text{c}) + 20(\text{v}) + 20(\text{p}) = \text{prix } 120 \quad \textbf{(2.13)} \\ &\text{B} : 90(\text{c}) + 10(\text{v}) + 10(\text{pl}) = \text{val } 110 \rightarrow 90(\text{c}) + 10(\text{v}) + 20(\text{p}) = \text{prix } 120 \\ &\text{C} : 70(\text{c}) + 30(\text{v}) + 30(\text{pl}) = \text{val } 130 \rightarrow 70(\text{c}) + 30(\text{v}) + 20(\text{p}) = \text{prix } 120 \end{aligned}$$

La plus-value (pl) totale était de 60. Elle reste à ce niveau mais devient du profit (p). Comme le capital total (c+v) est de 300, la plus-value extraite du travail rend possible un taux de profit *p'*de 20%. La formule du taux de profit est :

$$p' = pl\,/(c+v) \qquad (2.14)$$

Dans l'exemple ci-dessus, les trois branches ont un profit de 20, parce qu'il représente 20% d'un capital de 100 qui y est investi. Notons au passage, que les activités commerciales et bancaires, censées ne pas créer de valeur, participent malgré tout au partage de la plus-value via la péréquation du taux de profit. Tout capital doit amener le profit moyen à son propriétaire (sinon, il n'y serait pas investi). La péréquation du taux de profit résulte de la concurrence : si un secteur d'activité comporte des éléments monopolistiques, il ne participera pas ou pas entièrement à ce processus : le prix dépassera le *prix de production* et il y aura des surprofits. C'est le cas de l'agriculture, du fait de la rareté des bonnes terres. Le surprofit sera capté par la rente : une *rente absolue* s'ajoute à la *rente différentielle*, mais pour une raison différente de celle avancée par Malthus.

Marx insiste et réinsiste sur ce point : « La plus-value et le taux de plus-value sont, relativement, l'élément invisible et le point essentiel qu'il faut élucider, tandis que le taux de profit et donc la plus-value sous sa forme de profit sont des phénomènes qui n'apparaissent qu'à la surface »[1]. Le profit n'est que la plus-value redistribuée en pourcentage du capital investi. La démonstration marxiste de l'exploitation impose cette condition. Sinon, la théorie des prix dirait simplement : les prix se forment de façon à rapporter le profit moyen dans chaque branche ; les mots *exploitation* ou *plus-value* ne seraient même pas prononcés. L'artifice consiste à conserver les notions de plus-value et de valeur-travail alors que le prix est déconnecté de cette dernière, vidée de toute réalité.

N'y a-t-il pas ici une magistrale faute de logique ? Le raisonnement en deux phases est d'autant plus artificiel que la théorie de la plus-value a déjà elle-même pour objet un système capitaliste

[1] Marx [254] livre III t1 p. 62.

puisque la plus-value provient de l'achat de la force de travail par le capitaliste à un prix inférieur à la valeur qu'elle produira. Donc, dès le départ, il n'y avait pas de raison de retenir pour plus tard la théorie des prix de production basée sur le taux de profit moyen, puisque c'est ainsi que fonctionne le capital. Le détour par la valeur-travail et la plus-value est donc une pure construction de l'esprit.

*
* *

Après la théorie de la valeur travail, qui rendit Marx célèbre, nous pouvons aborder son analyse du capital, trop largement méconnue malgré quelques idées inspirées. Son analyse de la *circulation du capital* est importante : le capital a un cycle de vie dans l'exploitation de l'entreprise qui le fait passer par une succession de stades, qui se renouvelle sans cesse : 1°- argent (A), 2°- marchandises -matières premières et outils- (M) entrant dans la production (P), 3°- nouvelle marchandise -produits finis- (M') dont la valeur est supérieure à celle des biens qui entraient dans sa fabrication car la production produit une plus-value source de profit, 4°- retour à la forme monétaire (A'), pour repartir dans un nouveau cycle. Marx formalise ce processus ainsi :

$$A - M - \ldots[P]\ldots - M' - A'. \qquad \text{avec } M' > M \text{ et } A' > A \qquad (2.15)$$

Chaque cycle a une certaine durée, correspondant à la période de production introduite par Ricardo, que Marx appelle *période de rotation*. S'ensuit une analyse fouillée de la rotation du capital. De même que la fréquence peut être mesurée par la longueur d'un cycle ou par le nombre de cycles par unité de temps, de même la rotation du capital peut être mesurée par la durée du cycle (période de rotation) ou par le nombre de cycles par année (vitesse de rotation). Chaque élément du capital a un cycle propre mais on peut calculer la vitesse de rotation moyenne de l'ensemble du capital. Marx donne l'exemple suivant :

Soit un capital de £100.000 se composant ainsi :
-capital fixe : 80.000 à renouveler tous les 10 ans ; coût annuel : 8.000
- capital variable : 20.000 tourne cinq fois par an ; coût annuel : 100.000
Le coût annuel est de £108.000. Le capital de 100.000 accomplit 1,08 rotations par an ; il accomplit donc une rotation complète en 11,11 mois.

Il y a un lien direct entre la rotation du capital d'une part et les prix de production et le taux de profit d'autre part. Chaque cycle apporte son lot de plus-value d'une façon pouvant être représentée par une formule du type (2.15). Toutes autres choses restant constantes, il est mathématiquement évident que plus courts sont les cycles, plus élevé sera le taux de profit pour un taux de plus-value donné. Le capitaliste a donc intérêt à élever la vitesse de rotation de son capital, car il pourra atteindre le taux de profit moyen avec une marge bénéficiaire moindre.

Comme tant d'économistes avant et après lui, Marx s'est intéressé à l'intensité capitalistique de la production, en l'occurrence l'importance du facteur capital par rapport au facteur travail ; pour l'évaluer, il utilise deux outils : la *composition technique* du capital et sa *composition organique***.** La première est le rapport entre la masse des moyens de production et le nombre d'ouvriers ou d'heures de travail. Ce rapport est plutôt abstrait et difficile à mesurer. La composition organique du capital est le même rapport, mais exprimé en valeur, donné par le rapport c/v. Vu l'*accumulation du capital*, c'est à dire l'accroissement du stock de capital par le réinvestissement régulier d'une part des profits, la composition technique du capital tend à augmenter dans le temps. Marx a considéré que la composition organique du capital connaissait également cette tendance historique à l'accroissement, bien que moins rapidement.

Cette hausse de la composition organique du capital l'a amené à prophétiser la *baisse tendancielle du taux de profit*[1]. Comme le montre la formule (2.14), cette conclusion est mathématiquement imparable. Considérons que (c+v) égale le capital total investi[2]. Supposons que le taux de plus-value p/v reste constant. Vu la formule (2.14), *p'* ne peut que baisser lorsque c/v augmente. Dans la théorie marxiste, cette baisse tendancielle du taux de profit a des implications politiques, car elle est censée accentuer les contradictions du capitalisme qui doivent finalement mener à sa chute.

Malgré la logique de ce raisonnement, les études empiriques ne confirment pas la baisse du taux de profit. Comment l'expliquer ? Marx lui-même avait énuméré les causes qui contrecarrent sa loi. D'abord, la formule (2.14) montre que la hausse de la composition organique *c/v* peut être compensée par celle du taux d'exploitation *p/v*. Ensuite, le progrès technique dans l'industrie des biens de production peut mener à la baisse du prix des biens constituant *c*. La réalité est donc celle-ci : l'accroissement du capital physique par travailleur n'induit pas nécessairement une hausse de la composition organique du capital ; or celle-ci est la cause de la chute tendancielle du taux de profit. Marx avait tenu compte de cet élément mais en a sous-estimé l'effet. Nous sommes ici au cœur du problème le plus difficile de l'économie politique : la mesure de l'intensité capitalistique de la production et son rapport avec la rémunération des facteurs ; comme nous le verrons, des générations d'économistes y perdront leur latin et leurs cheveux.

Marx a également analysé les conditions d'un équilibre dynamique de l'économie. A cette fin, il divise l'économie en deux secteurs : le secteur I, qui produit les biens de production et le secteur II, qui produit les biens de consommation.

Marx présente le problème avec des exemples chiffrés basés sur sa fameuse équation (2.12). Deux situations sont analysées ; celle d'un équilibre stationnaire, que Marx appelle *reproduction simple*, où les inputs et l'output de la production demeurent constants ; ensuite, celle d'une économie en expansion, que Marx appelle *reproduction élargie*, qui accumule le capital, c'est-à-dire réalise des investissements nouveaux : les inputs et les outputs augmentent conjointement[3].

La reproduction simple

$$\begin{aligned} &\text{I} : 4000(\text{c}) + 1000(\text{v}) + 1000(\text{pl}) = 6000 \\ &\text{II} : 2000(\text{c}) + \ 500(\text{v}) + \ 500(\text{pl}) = 3000 \end{aligned} \qquad (2.16)$$

La condition d'équilibre est que chaque secteur fournisse à l'autre autant qu'il ne reçoit de lui, ce qu'indique l'équation :

$$\text{I}\ (\text{v}+\text{pl}) = \text{II}\ \text{c} \qquad (2.17)$$

Autrement dit : l'achat de biens de production de remplacement par le secteur II s'égalise avec l'achat de biens de consommation par les salariés et les capitalistes du secteur I. Dans l'état stationnaire, toute la production est consommée ; le capital n'augmente pas et au cycle suivant, les équations seront les mêmes.

[1] Smith et Ricardo avaient déjà fait cette même prévision, mais pour des raisons totalement différentes.

[2] La dualité entre le capital en tant que stock et le flux de sa consommation productive cesse d'être gênante si on prend l'hypothèse d'une vitesse de rotation égale à l'unité.

[3] Marx s'estime lui-même redevable envers Quesnay pour cette analyse. Marx était un grand admirateur du tableau économique.

La reproduction élargie

Stade initial :

$$\begin{aligned} &\text{I} && 5000(c)+ 1000(v)+ 1000(p)= 7000 && (2.18)\\ &\text{II} && 1430(c)+ 285(v)+ 285(p)= 2000 \end{aligned}$$

Stade 2

$$\begin{aligned} &\text{I} && 5417(c)+ 1083(v)+ 1083(p)= 7583\\ &\text{II} && 1583(c)+ 316(v)+ 316(p)= 2215 \end{aligned}$$

Stade 3

$$\begin{aligned} &\text{I} && 5869(c)+ 1173(v)+ 1173(p)= 8035\\ &\text{II} && 1715(c)+ 342(v)+ 342(p)= 2399 \end{aligned}$$

Cette fois, la condition d'équilibre est l'égalité :

$$\text{I}(v+pl) = \text{II}c + \text{II}\Delta c + \text{I}\Delta c \qquad (2.19)$$

Le revenu disponible en I doit compenser le remplacement des équipements en II plus les accroissements d'équipements en I et II. Par exemple, le revenu disponible en I au premier stade, valant 2000, servira à IIc pour 1430, à IΔc pour (5417-5000) et à IIΔc pour (1583-1430).

Le schéma de la reproduction élargie est une première tentative, héroïque, de définir les conditions d'une croissance équilibrée. Dans l'exemple numérique donné par Marx, le secteur I investit de façon autonome la moitié de sa plus-value ; l'investissement dans le secteur II, quant à lui, est calculé de façon à équilibrer les échanges entre les secteurs I et II. Il y a pas mal d'hypothèses simplificatrices : la totalité du capital est usée dans l'année (rotation unitaire), la composition organique du capital est constante, chaque secteur investit une partie de sa propre plus-value et il n'y a pas de transferts de capitaux entre eux… Au prix de conditions très strictes, Marx parvient à obtenir une croissance équilibrée. De période en période, le taux de plus-value reste constant tout comme les parts respectives des secteurs I et II dans le total.

TITRE III- L'ECOLE NEOCLASSIQUE

3.1- NAISSANCE D'UNE NOUVELLE ECOLE

On a utilisé l'expression « révolution marginaliste ». Il est vrai que dans les années 1870, une nouvelle école économique naît, dont les idées et les méthodes représentent une rupture presque totale par rapport à l'économie classique, qu'elle supplantera. La nouvelle économie efface et recommence en oubliant presque tout l'apport classique.

Certes, depuis longtemps, des économistes expriment des principes différents du paradigme classique, mettant en avant les notions d'utilité et de rareté dans l'explication de la valeur d'échange. Nous avons vu J-B. Say et A. Cournot, mais il y en eut d'autres, principalement en France, en Italie et en Allemagne et même en Grande Bretagne, la patrie de l'école classique. Toutefois, ces voix restaient isolées.

En 1871 paraissent « The Theory of Political economy » de William Stanley Jevons et les « Grundsätze der Volkwirtschaftlehre » de Carl Menger et en 1874 les « Eléments d'économie politique pure » de Léon Walras. La nouvelle école est sur les rails, mais il faudra plusieurs générations d'économistes pour que son corps théorique se façonne. Les trois ouvrages ci-dessus introduisent un concept neuf- *l'utilité marginale*- qui révolutionnera la science économique. Cette simultanéité est le fruit du hasard. Les trois auteurs ne se connaissaient pas avant la publication. Par la suite, Walras et Jevons collaboreront pour promouvoir leur nouvelle théorie au détriment sa rivale classique et favoriser l'usage des mathématiques. Par contre, Menger n'aura pas de contact avec eux et fera preuve de méfiance vis-à-vis de l'invasion de l'économie politique par les mathématiques.

Pour les classiques, la valeur était objective puisqu'elle correspondait au coût de production, qui avait une existence observable. La rareté et l'utilité étaient prises en compte mais confinées dans un rôle accessoire. Pour les marginalistes, la valeur est subjective. La valeur dépend de la désirabilité de l'objet, celle-ci étant évidemment fonction de l'abondance ou de la rareté [1] de l'objet désiré. Le concept d'*utilité marginale*, c'est à dire l'utilité propre à la dernière unité du produit à être consommée, conjugue l'utilité et la rareté.

L'explication marginaliste de la valeur implique une inversion de la relation causale entre la rémunération des facteurs et la valeur, telle que l'avaient définie les classiques : les marginalistes partent de la valeur déterminée par l'utilité marginale pour remonter à la répartition du revenu. C'est la théorie dite de *l'imputation*, très clairement exprimée par Menger et Walras[2].

3.1.1. Jevons : de l'utilitarisme à l'utilité marginale

A partir de la fin du XVIIIe siècle jusqu'à celle du XIXe, une théorie philosophico-politique fera florès auprès de la bourgeoisie intellectuelle en Angleterre : *l'utilitarisme*. Son fondateur et principal idéologue et apôtre est un certain Jeremy Bentham (1748-1832), personnage excentrique, qui influencera notamment Ricardo et Mill. William Stanley Jevons est également un disciple convaincu. L'utilitarisme vise à maximiser le bonheur total de la société, ce pour quoi il invite à une étonnante algèbre, entre plaisirs et peines, entre bonheur individuel et bonheur social. Le bonheur est conçu dans une optique hédoniste et assez matérialiste.

---

[1] Walras écrit : « …la rareté et l'abondance ne s'opposent pas l'une à l'autre : quelque abondante qu'elle soit, une chose est rare en économie politique dès qu'elle est utile et limitée en quantité, exactement comme un corps a de la vitesse, en mécanique, dès qu'il parcourt un certain espace en un certain temps » (Walras [380] p. 46).

[2] On avait déjà rencontré cette idée chez Say.

Dans l'introduction de sa *Theory*, Jevons se réfère abondamment au philosophe anglais qu'il cite : «Nature, he says, has placed mankind under the governance of two sovereign masters- pain and pleasure. It is for them alone to point out what we ought to do as well as to determine what we shall do (…) The principle of utility recognises this subjection and assumes it for the foundation of that system, the object of which is to rear the fabric of felicity by the hands of reason and law »[1]. Jevons en déduit que « the object of economics is to maximise happiness by purchasing pleasure, as it were, at the lowest cost of pain ». Il ajoute : « I have no hesitation in accepting the Utilitarian theory of morals which does uphold the effect upon happiness of mankind as the criterion of what is right and wrong »[2].

Jevons garde l'espoir qu'un jour le progrès scientifique permettra une mesure rigoureuse des plaisirs et peines ; l'analyse empirique de nos comportements en serait la base car nous ne pouvons faire un choix sans qu'il indique un excès de plaisir sur les peines dans sa direction.

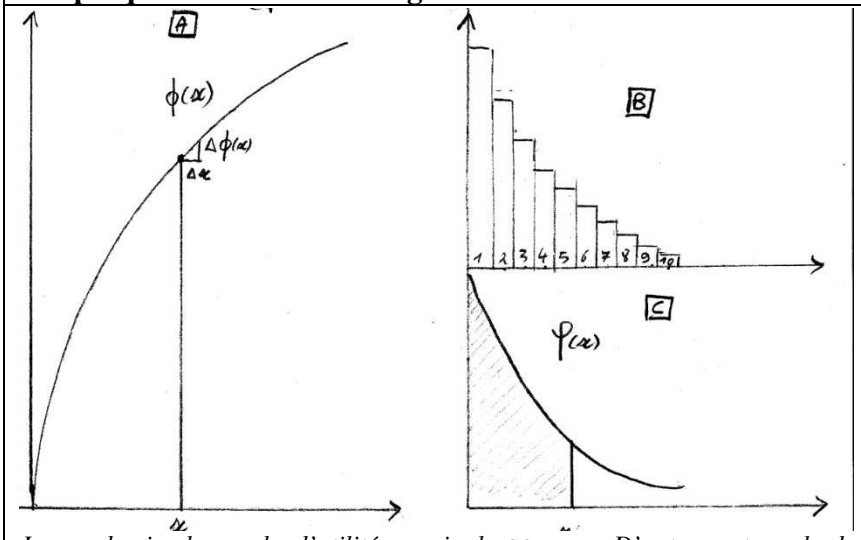
De l'utilitarisme à l'utilité et de l'utilité à l'**utilité marginale**, Jevons franchit le pas. « Whatever can produce pleasure and prevent pain may possess utility »[3]. Mais « we must carefully discriminate between the *total utility* arising from any commodity and the utility attaching to any particular portion of it (…) Let us imagine the whole quantity of food which a person consumes on an average during twenty-four hours to be divided into ten equal parts. If his food be reduced by the last part, he will suffer but little; if a second tenth part be deficient, he will feel want distinctly; the subtraction of a third tenth part will be decidedly injurious; with every subsequent subtraction of tenth part his sufferings will be more and more serious, until at length, he will be upon the verge of starvation. Now, if we call each of the tenth parts an *increment*, the meaning of these facts is that each increment of food is less necessary or possess less utility than the previous one »[4]. Voici formulée la fameuse loi de l'utilité marginale décroissante. « No commodity can be named which we continue to desire with the same force, whatever be the quantity already in use or possession ».

Les quantités considérées pour déterminer l'utilité totale ou l'utilité marginale sont des quantités consommées PAR UNITE DE TEMPS. On ne peut quantifier l'utilité d'un stock de bien comme tel.

Jevons met en graphique l'exemple des dix dixièmes de la ration de nourriture. Dans le graphique 3.1-B, chaque rectangle a sa base qui indique la quantité de nourriture représentée (égale pour tous) et sa hauteur l'utilité attachée à cette fraction. L'utilité totale des cinq premiers dixièmes est donnée par la surface totale des cinq premiers rectangles. Mais cette division en dix dixièmes est arbitraire et serait avantageusement remplacée par une l'analyse de variations infinitésimales de la quantité consommée. Dans ce cas, l'utilité marginale devient une fonction continue de celle-ci, comme le montre le graphique 3.1-C. A l'avenir, nous ne retiendrons que la fonction continue du graphique 3.1-C, l'analyse infinitésimale étant devenue la marque de fabrique de l'école néoclassique, d'où l'appellation « marginaliste » qui lui est souvent accolée.

[1] Jevons [165] § 1.30
[2] Jevons [165] § 1.29
[3] Jevons [165] § 3.1
[4] Jevons [165] § 3.15-16

Graphique 3.1 : l'utilité marginale et l'utilité totale



Jevons dessine la courbe d'utilité marginale CONVEXE. D'autres auteurs la dessinent concave. Les deux approches sont plausibles. Mais dans tous les cas, elle est décroissante. La courbe d'utilité totale, par contre, est nécessairement concave, du fait de l'utilité marginale décroissante.

Quel est le lien mathématique entre l'utilité totale et l'utilité marginale ? En accord avec Jevons mais plus clair que lui, Walras l'exprime ainsi : l'utilité totale et l'utilité marginale sont, toutes deux, fonction de la quantité consommée *x*. Soit φ(x) l'utilité marginale et ϕ(x) l'utilité totale. Les deux fonctions sont évidemment liées ; on a :

$$U_m = {d\Phi(x)}\Big/{dx} \qquad (3.1 - A)$$

$$U_T = \int_0^x \varphi(x).\,dx \qquad (3.1 - B)$$

L'utilité marginale apparaît donc sur le graphique de l'utilité totale comme étant la pente de la courbe ; l'utilité totale apparaît sur le graphique de l'utilité marginale comme étant la surface en dessous de la courbe entre l'origine et la quantité consommée.

Rappelons-nous le paradoxe de la valeur exprimé par Adam Smith, qui en avait conclu que l'utilité n'est pas le déterminant de la valeur. Jevons s'insurge contre cette vision : Smith n'envisageait que l'utilité totale et non l'utilité marginale. Toute l'eau dont nous disposons a en effet une utilité (totale) plus élevée que tout le diamant dont nous disposons. Mais la première est suffisamment abondante et le second suffisamment rare pour que l'utilité marginale du diamant soit supérieure à celle de l'eau.

Jevons peut maintenant s'attaquer à **l'échange**, le chapitre principal de son ouvrage. L'échange permet aux détenteurs de biens d'augmenter leur utilité totale en cédant des biens de faible utilité marginale pour en obtenir d'autres à forte utilité marginale.

Jevons préfère l'appellation RAPPORT D'ECHANGE entre le bien X et le bien Y plutôt que VALEUR D'ECHANGE, car elle révèle mieux que la valeur d'échange ne peut être que relative et circonstancielle. Le *rapport d'échange* est l'inverse du rapport entre les quantités échangées. Si un daim s'échange contre deux castors, cela revient à dire que le daim est deux fois plus cher que le castor. Jevons met en évidence ce qu'il appelle la *loi de l'indifférence* : « in the same open market, at any one moment, there cannot be two prices for the same kind of article. Such differences as may practically occur, arise from extraneous circumstances, such as the defective

credit of the purchasers, their imperfect knowledge of the market and so on »[1]. J.S. Mill avait déjà relevé cette loi.

Comment se détermine le rapport d'échange ? Jevons aborde d'abord le cas apparemment le plus simple : l'*échange bilatéral* ; l'agent 1, détenteur d'une quantité *a* du bien *X* en cède une quantité *x* en échange d'une quantité *y* du bien *Y* qu'il reçoit de l'agent 2, possesseur d'une quantité *b* de ce bien. Voyons d'abord comment l'affaire se présente pour l'agent 1, dont les utilités marginales des biens *X* et *Y* sont données respectivement par les fonctions φ_{x1} et φ_{y1}. « Exchange will thus go until each party has obtained all the benefit that is possible, and loss of utility would result if more were exchanged »[2]. Les premières unités de *X* cédées lui coûtent peu d'utilité et les premières unités obtenues de *Y* lui en rapportent beaucoup. Mais la cession de X se révèle de plus en plus pénible alors que l'acquisition de *Y* est de moins en moins avantageuse. Chaque avancée dans l'échange est moins profitable. L'agent 1 maximise évidemment son utilité totale en poussant l'échange jusqu'au moment où l'utilité de la dernière portion de *X* cédée égale celle de la dernière unité reçue de *Y*. Autrement dit, il faut égaliser $\varphi_{x1}(a\text{-}x).dx$ avec $\varphi_{y1.}(y)\ dy$. Ce qui donne : $\varphi_{x1}(a\text{-}x) / \varphi_{y1}(y) = dy/dx$. Or $dy/dx = y/x$. de par la loi d'indifférence. En appliquant le même principe pour l'agent 2, on obtient la double équation fondamentale de l'échange bilatéral :

$$\varphi_{x1}(a\text{-}x) / \varphi_{y1}(y) = y/x \qquad (3.2)[3]$$
$$\varphi_{x2}(x) / \varphi_{y2}(b\text{-}y) = y/x$$

« The two equations are sufficient to determine the results of exchange; for there are only two unknown quantities concerned, namely *x* and *y,* the quantities given and received »[4]. La règle consiste donc à échanger les quantités qui égalisent le rapport d'échange avec le ratio inverse des utilités marginales, dans le chef de chacun des participants.

Abordant ensuite l'*échange multilatéral,* Jevons prétend : « Exactly the same principles hold true, however numerous and complicated may be the conditions »[5]. Jevons se contente d'analyser deux cas :

- trois agents initialement possesseurs chacun d'un des trois biens X,Y et Z. Chacun des trois agents livre son bien aux deux autres : cela fait six quantités à déterminer : il y a six équations pour ces six inconnues.
- trois agents et deux biens dont l'un est donc offert par deux agents en concurrence : il y a quatre équations pour quatre inconnues.

Jevons reconnaît toutefois qu'il existe des cas où son modèle ne fonctionne pas : lorsque des biens intervenant dans l'échange sont indivisibles. Il donne l'exemple de la vente d'une maison. Le prix se déterminera en fonction de la capacité de marchandage des protagonistes.

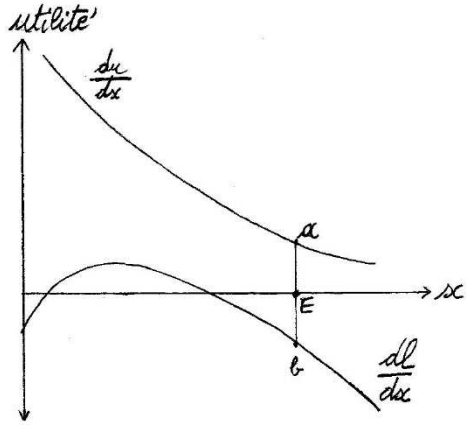
Jusqu'ici, nous nous sommes limités à l'*échange pur*. Introduisons maintenant la production dans le modèle. Jevons analyse successivement le travail, la terre et le capital. Voyons d'abord sa **théorie du travail**. En bon utilitariste, Jevons le conçoit sous l'angle des plaisirs et des

[1] Jevons [165] § 4.22.
[2] Jevons [165] § 4.30
[3] Les deux indices accolés à φ servent à différentier les fonctions d'utilité relatives à des biens différents (X et Y) ou à des agents différents (1 ou 2).
[4] Jevons [165] § 4.37. Le principe est que le système d'équations a une solution si le nombre d'équations égale le nombre d'inconnues. Nous y reviendrons plus d'une fois.
[5] Jevons [165] § 4.60

peines. Il formule le principe de la *désutilité marginale* croissante du travail : « Experience shows that as labour is prolonged the effort becomes as a general rule more and more painful. A few hours' work per day may be considered agreeable rather than otherwise; but as soon as the overflowing energy of the body is drained off, it becomes irksome to remain at work »[1].

Quelle quantité (autrement dit quelle durée) de travail, l'individu offrira-t-il ? Chaque heure de travail a une désutilité ainsi qu'une utilité, celle de la consommation rendue possible par le salaire. Les premières heures ont une faible désutilité et une forte utilité. Au fur et à mesure que la durée s'allonge, l'écart se réduit pour enfin s'annuler[2]. C'est à ce moment que le travailleur doit s'arrêter de travailler pour optimiser sa situation ; les heures qui suivraient feraient baisser l'utilité totale de l'individu. On voit que Jevons considère que l'utilité de la consommation et la désutilité du travail ont une même nature ; on peut les additionner ou les comparer ; seul le signe diffère. « He will cease to labour just at the point when the pain becomes equal to the corresponding pleasure gained »[3]. C'est ce que montre le graphique 3.2.

Graphique 3.2 : l'offre de travail selon Jevons



En dessous de l'abscisse, l'utilité est négative et devient de la désutilité. Au point E, les courbes de l'utilité marginale du produit et de la désutilité marginale pour le produire sont équidistantes de l'abscisse, donc égales.

Algébriquement, la formule d'optimisation est :

$$\frac{dl}{dt} = \frac{dx}{dt} \cdot \frac{du}{dx} \qquad (3.3)$$

Où x est la quantité consommée grâce au salaire, u l'utilité de cette consommation, l la désutilité subie pour la produire et t le temps de travail à optimiser.

Jevons estime que l'influence du salaire sur l'offre de travail conjugue deux tendances contradictoires :
- un salaire plus élevé rend l'allongement du temps de travail plus profitable
- le travailleur mieux payé pourrait être tenté de considérer plus rapidement qu'il a travaillé suffisamment d'heures et préférer les loisirs ou le repos

Selon les circonstances, l'une ou l'autre de ces deux tendances dominera : il n'y a pas de loi générale a priori[4].

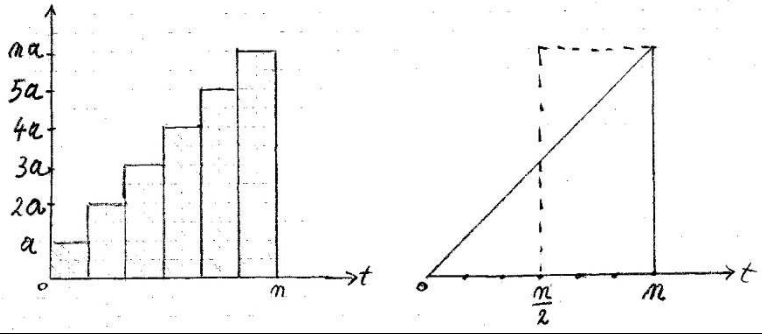
[1] Jevons [165] § 5.6
[2] L'utilité marginale décroissante du produit et la désutilité marginale du travail se conjuguent dans ce sens.
[3] Jevons [165] § 5.14
[4] Bien que la première tendance semblait la moins forte aux yeux de Jevons, les économistes dessinent généralement l'offre de travail croissant avec le salaire.

Concernant le **capital**, toujours dans son optique utilitariste, Jevons assimile l'*abstinence* affirmée par Senior et Mill à une forme de désutilité, ce qui permet un traitement symétrique du capital et du travail.

La grande avancée jevonienne en la matière est une idée qui fait de lui le précurseur de la théorie autrichienne du capital, que nous étudierons en profondeur dans le sous-chapitre 3.2.2. Il écrit : « The single and all important function of capital is to enable the labourer to await the result of any long lasting work ». Le capital est constitué des avances nécessaires pour payer les salaires (considérés ici comme les seuls frais de production) jusqu'au moment où la vente du produit amène de nouveaux fonds qui ne doivent donc plus être avancés par le capitaliste. Supposons qu'une production nécessite la dépense d'un salaire *a* pendant *n* années avant que le produit soit vendable. Le capital vaudra a+2a+3a+…+n.a. soit la valeur de produit très incomplet ayant subi la 1e année de travail + la valeur du produit déjà un peu plus avancé ayant subi les deux premières années de travail + … + le produit prêt à être vendu ayant subi les *n* années de travail. La somme de cette série vaut a.(n/2).(1+n) ; elle augmente proportionnellement au carré du temps.

Le graphique 3.3 illustre ce concept. Le *montant d'investissement du capital* est mesuré par l'aire totale du triangle formé par les bâtonnets. Si le temps est compté en continu, on obtient le triangle rectangle de la figure de droite, toujours dans l'hypothèse d'un investissement temporel constant. La surface du triangle est égale à celle du rectangle correspondant à la hauteur maximale multiplié par la moitié du temps *n*. On peut donc dire que le temps moyen d'investissement vaut dans ce cas la moitié de la durée totale de la production concernée.

Graphique 3.3 : le montant de l'investissement de capital selon Jevons



Jevons se rend compte que ce modèle ignore un facteur important : l'intérêt composé qui s'intègre progressivement dans le capital. Mais il estime, sans doute un peu légèrement, que « the consideration of compound interest would render the formulae very complex, and is not requisite for the purpose in view »[1].

Avant de se cristalliser en biens intermédiaires, le capital existe sous la forme de *capital disponible*, proche du *fonds des salaires* des classiques. C'est cet âge de jeunesse du capital qui intéresse Jevons : « The capital is not the railway but the food of those who made the railway »[2].

[1] Jevons [165] § 7.22. Sur le graphique 3.3, chaque bâtonnet a une hauteur valant celle du précédent plus *a*. Mais la valeur du capital intègre ses fruits tant qu'on ne les en retire pas. L'intérêt couru mais non perçu fait partie de l'avance de capital et s'accumule avec le principal. Les surfaces de la figure 3.3 devraient être augmentées de l'intérêt composé.

[2] Jevons [165] § 7.33

Le capital disponible peut s'investir dans une diversité de branches, mais comme les classiques, Jevons pense qu'il doit rapporter le même taux dans tous ses emplois.

Dans cette optique, il est logique de considérer la production comme une fonction continue *F* du temps d'attente *t*. Jevons donne pour l'intérêt (i) la formule suivante, équivalente au « rate of increase of the produce divided by the whole produce » :

$$i = \frac{F(t + \Delta t) - F(t)}{\Delta t} . \frac{1}{F(t)} \Leftrightarrow i = \frac{F'(t)}{F(t)} \qquad (3.4)$$

Jevons doute que l'accroissement de F(t) puisse à la longue rester suffisant pour garder le même rapport à F(t). Comme chez les classiques, le taux d'intérêt est promis à un déclin.

*

Jevons jouit encore aujourd'hui d'un réel prestige parmi les économistes néoclassiques, mais son œuvre est toutefois considérée comme dépassée. Deux lacunes importantes ont été pointées par deux de ses proches collègues :

1. Le modèle de l'échange bilatéral (équation (3.2) est inexact. Dix ans après la parution de la *Theory*, Edgeworth montrera qu'en échange bilatéral, le rapport d'échange est indéterminé et que c'est justement l'augmentation du nombre de participants à l'échange qui permet sa détermination. Nous examinerons la théorie de l'échange d'Edgeworth en détail. Dans sa démonstration, le tort de Jevons fut de recourir à la loi d'indifférence, alors qu'elle ne se justifie que par la concurrence parfaite, qui postule des intervenants infiniment plus nombreux[1].

2. L'utilité semble seule déterminer le prix. Mais quel est alors le rôle de l'offre et de la demande ? Jevons n'établit pas de connexion entre les deux principes.

3.1.2. Menger et la satisfaction des besoins

Pour fonder solidement sa théorie, Karl Menger veut l'attacher à des évidences d'un haut niveau de généralité. Il analyse donc (très) méticuleusement le concept de *besoin* et la relation qui s'établit entre un besoin et un objet qui peut le satisfaire. « Things that can be placed in causal connection with the satisfaction of human needs we term *useful things*. If, however, we both recognize this causal connection and have the power actually to direct the useful things to the satisfaction of our needs, we call them *goods* »[2]. Les *biens* incluent des actions humaines (par exemple le travail ou les services) au même titre que des objets matériels.

Menger appelle *biens de premier ordre* les biens de consommation et *biens d'ordre supérieur*, ceux qui servent à produire d'autres biens. On aurait par exemple l'enchaînement suivant : pain (premier ordre), farine (deuxième ordre), grain (troisième ordre) et terre (quatrième ordre). Parmi les biens d'ordre supérieur, il y a non seulement les équipement et matière première mais aussi les contributions personnelles : le travail, l'activité de l'entrepreneur, les services du capital (l'avance de fonds à la production) en tant que service distinct des biens matériels objets de l'avance … De premier ordre ou d'ordre supérieur, c'est toujours la satisfaction d'un besoin qui caractérise essentiellement un bien : avec les biens d'ordre supérieur, la seule différence est que cette relation est indirecte au lieu d'être directe.

---

[1] En fait, Edgeworth veut protéger son prédécesseur contre l'idée qu'il aurait commis une ERREUR. Son interprétation, pas totalement convaincante, est que Jevons aurait sorti deux agents *représentatifs* de la multitude se livrant concurrence, pour mettre en lumière le comportement général.

[2] Menger [256] p. 52

La production consiste en la transformation graduelle de biens d'ordre plus élevé en biens de consommation : ce processus implique l'écoulement de temps avant la satisfaction du besoin. Le progrès économique réside principalement dans le fait que l'homme peut se permettre d'affecter une part croissante du travail à la production de biens d'ordre plus élevé, dont l'utilisation précède la consommation d'un délai plus long.

« We shall call a person's *requirements* those quantities of goods that are necessary to satisfy his needs within the time period covered by his plans »[1]. Additionnons ces *demandes*[2] individuelles d'un bien donné *;* si la somme est supérieure à la quantité disponible, on a affaire à un *bien économique* ; dans le cas contraire, il sera dit *non économique*[3]. D'un bien économique, une part du besoin est donc toujours insatisfaite. C'est le cas de la grande majorité des biens. L'air que nous respirons est un bien non économique, car sa disponibilité suffit à assurer la satisfaction de la demande sans limite. Ce n'est pas sa nature qui rend tel bien économique, mais le contexte : des circonstances nouvelles (épuisement d'une carrière…) peuvent faire passer un bien du statut non économique au statut économique ou inversement. Avec le progrès de la civilisation, de plus en plus de biens deviennent économiques, car les besoins tendent à augmenter plus que les moyens de les satisfaire. La *richesse* d'un individu ou d'une collectivité est constituée par l'ensemble des biens économiques à sa disposition.

Jevons récusait le terme *valeur*, trop ambigu, et l'estimait superflu par rapport aux concepts d'*utilité* et de *rapport d'échange*. Menger, au contraire, lui réserve une place de choix et la définit ainsi : « the importance that individual goods or quantities of goods attain for us because we are conscious of being dependent on command of them for the satisfaction of our needs »[4]. Le concept « to have command of… » n'a pas de sens pour des biens qui s'offrent librement à l'homme. Seuls les biens économiques peuvent donc avoir de la valeur.

Il s'agit d'une conception totalement SUBJECTIVE de la valeur : « The value of goods arises from their relationship to our needs and is not inherent in the goods themselves. With changes in this relationship, value arises and disappears (…) value does not exist outside the consciousness of men »[5]. Il en résulte ces deux conséquences :

- Il n'y a pas de raisons pour qu'un bien ait la même valeur pour deux individus.
- La façon dont le bien a été obtenu ou produit n'influence pas sa valeur : un diamant trouvé à la même valeur que s'il avait fallu creuser pendant des jours.

Comment fonder la valeur, ainsi définie, en tant que GRANDEUR ? Menger estime qu'on peut graduer l'importance des satisfactions en accordant la première place à celles qui sont nécessaires à la survie, suivies par les autres dans l'ordre du degré de plaisir qu'elles apportent.

Considérons plusieurs biens, qui satisfont chacun un besoin différent. La consommation de ces biens est fractionnée en unités (homogènes). Chaque unité consommée apporte une certaine satisfaction à laquelle Menger attribue une cote, un nombre de points de 10 à 1. Il donne l'exemple suivant, comportant 10 biens (I à X) :

---

[1] Menger [256] p. 79
[2] Je traduis ici (un peu improprement) *requirements* par *demande*, faute de mieux.
[3] Les biens non économiques sont souvent appelés *biens libres*.
[4] Menger [256] p. 115
[5] Menger [256] pp. 120-21

Qnt	I	II	III	IV	V	VI	VII	VIII	IX	X
1	10	9	8	7	6	5	4	3	2	1
2	9	8	7	6	5	4	3	2	1	
3	8	7	6	5	4	3	2	1		
4	7	6	5	4	3	2	1			
5	6	5	4	3	2	1				
6	5	4	3	2	1					
7	4	3	2	1						
8	3	2	1							
9	2	1								
10	1									

La première unité du bien I apporte une utilité particulièrement élevée (10) : il s'agit d'un bien répondant à un besoin de première nécessité, par exemple la nourriture. Le bien VII satisfait un besoin plus superflu, par exemple le tabac. On constate toutefois que la 2e unité du bien VII rapporte plus d'utilité que la 9e unité du bien I. Le degré de satiété a donc plus d'importance que l'utilité intrinsèque du bien.

Le même principe régit le cas de biens dont la consommation peut être répartie en des usages variés. Menger prend l'exemple d'un fermier disposant d'un nombre donné de sacs de blé, dont les usages peuvent être classés en fonction décroissante du caractère fondamental du besoin satisfait : assurer sa survie (I), manger en suffisance (II), ressemer l'année suivante (III), nourrir les bêtes (IV),… produire de l'alcool (X) etc. Le tableau ci-dessus peut resservir.

Le fermier optimisera sa situation en satisfaisant par priorité les besoins les plus fondamentaux, ce qui aura pour effet d'apporter une satisfaction marginale égale aux différents besoins couverts. Par exemple, s'il a dix sacs de blé, il en affectera 4 à la satisfaction du besoin I, 3 au besoin II, 2 au besoin III et 1 au besoin IV. Tous ses besoins d'un niveau d'urgence supérieur ou égal à 7 seront couverts. La valeur d'un sac de blé sera donc de 7, car c'est le degré de satisfaction auquel il devrait renoncer s'il avait un sac en moins. Jevons avait exposé exactement le même principe.

Et comment se détermine la valeur des biens d'ordre supérieur ? Ces biens tirent leur valeur de celle des biens de premier ordre qu'ils contribuent à produire et non l'inverse. Prenons un ensemble de biens d'ordre supérieur qui sont complémentairement nécessaires pour produire du pain. Ensemble, ils représentent une valeur agrégée correspondant à la valeur prospective de tout le pain qu'ils produiront ensemble. Si une hausse de la valeur du pain est attendue, la valeur agrégée des biens d'ordre supérieur nécessaires à sa production augmentera. Mais comment cette valeur agrégée se répartit-elle entre les différents biens matériels, l'intérêt, la rente, le salaire et la rémunération de l'entrepreneur ? Comment se détermine la valeur d'un bien d'ordre supérieur pris individuellement ?

En fait, la règle est la même que pour les biens de consommation. Leur valeur dépend de la satisfaction dont la société se verrait privée si la dernière unité du bien était retirée. Prenons l'exemple d'un fermier qui réduit la surface qu'il cultive. Généralement, les facteurs autres que la terre ne seront pas mis au chômage à due concurrence, car une culture un peu plus intensive permettra peut-être d'utiliser tout le travail, les semences etc. Les biens d'ordre supérieur peuvent s'associer en proportions variables. Dans notre exemple, la culture plus intensive n'empêchera sans doute pas une certaine réduction de la production de pain (en valeur) : son ampleur détermine la valeur de la terre retirée à la culture.

Menger critique la théorie ricardienne de la rente foncière. Selon lui, « land occupies no exceptional place among goods »[1], alors que Ricardo conférait à la rente un statut qui la différenciait des autres rétributions. Pour Menger, la terre participe au partage de la valeur sur un pied d'égalité avec les autres biens d'ordre supérieur. La conclusion selon laquelle la terre la moins fertile ne peut rapporter de rente ne résiste pas à ce changement d'optique.

Le revenu incombant aux agents des différentes classes composant la société n'est pas arbitraire mais résulte de lois économiques régissant la valeur de leurs services. Menger en conclut : « The agitation of those who would like to see society allot a larger share of the available consumption goods to laborers than at present really constitutes a demand for nothing else than paying labor above its value »[2]. La fixation des salaires sur base d'une éthique sociale plutôt que des lois économiques, ajoute-t-il, impliquerait une transformation complète de l'ordre social, que-manifestement- il n'appelle pas de ses vœux.

Comme Jevons, Menger voit dans l'échange un moyen à la disposition des individus pour accroître la satisfaction de leurs besoins. C'est précisément parce que les agents attribuent des valeurs différentes aux biens qu'ils trouvent une incitation à échanger. Imaginons qu'un éleveur de chevaux assigne autant de valeur à un cheval qu'à 60 sacs de blé alors qu'un agriculteur estime sa valeur à 80 sacs de blé. Il est évident que tous deux auront intérêt à passer à l'échange. Comme Jevons, Menger situe l'équilibre de l'échange au point précis où les partenaires n'ont pas d'incitation à échanger plus et où le gain retiré de l'échange est maximal.

Selon Menger, le bien-être d'un individu ou d'une population dépend, non de la quantité de biens consommés, mais de la satisfaction totale des besoins. A partir du même stock de biens, l'échange opère une redistribution qui aboutit à une utilité totale supérieure. Le commerce est donc productif au même titre que l'agriculture ou l'industrie.

Le **prix** est une des caractéristiques de l'échange, une caractéristique plutôt mineure selon Menger. Comment se forme-t-il ? Menger distingue trois cas : le monopole bilatéral, le monopole des vendeurs et la concurrence.

- Concernant le monopole bilatéral (un vendeur unique face à un acheteur unique), Menger ne commet pas la même erreur que Jevons. Dans l'exemple ci-dessus, on peut dire que le prix s'établirait à un niveau qui est compris entre les deux bornes que sont 60 et 80 sacs de blé pour un cheval. Le niveau exact dépendra de la capacité de marchandage des deux protagonistes.

- Le monopole et la concurrence
  La thèse de Menger est qu'entre le monopole et la concurrence, il n'y a pas de différence qualitative mais seulement une différence quantitative. Dans tous les cas, le prix baisse quand la quantité offerte augmente, car les acheteurs « marginaux » cèdent une contrepartie de moins en moins consistante pour obtenir le bien. Or l'offre est nécessairement plus abondante en concurrence. En voici la raison : le monopoleur peut avoir intérêt à restreindre son offre pour accroître sa recette. En concurrence, un tel comportement est suicidaire, car son unique résultat serait d'offrir une part de marché aux concurrents, sans contrepartie. La concurrence fait donc baisser le prix parce qu'elle accroît la quantité.

[1] Menger [256] p. 164
[2] Menger [256] p. 174

Par rapport au monopole bilatéral, la concurrence conserve une marge d'indétermination du prix, mais elle se réduit. La concurrence entre les acheteurs tend à fixer un prix acheteurs supérieur au prix vendeurs. Une offre plus élevée signifie une moindre concurrence entre les acheteurs ; donc le surplus du prix acheteurs sur le prix vendeurs se réduira.

L'image de l'économie que Menger nous a présentée jusqu'ici est plutôt primitive, suggérant une énorme brocante entre particuliers. En réalité, la plus grande partie de l'offre dans une économie développée est le fait de vendeurs professionnels écoulant des biens produits dans le seul but d'être vendus, que les consommateurs payent avec de la monnaie. Vers la fin de l'ouvrage, Menger aborde enfin cet aspect de l'économie, mais il en a retardé la présentation à dessein. Selon lui, les forces fondamentales à l'œuvre restent celles décrites ci-avant : la différence n'est qu'en surface. Le concept de valeur décrit ci-dessus reste de rigueur, mais il existe sous deux formes : la *valeur d'usage*, déjà amplement étudiée, et la *valeur d'échange*, qui mérite quelques précisions.

Menger donne l'exemple d'une peau de bête que possède un trappeur. Elle peut servir DIRECTEMENT son besoin de se vêtir. Mais s'il la vend, elle satisfait INDIRECTEMENT une variété d'autres besoins grâce à la recette qu'elle apportera. Si la valeur d'usage de la peau est supérieure à celle des autres biens que le trappeur pourrait acquérir, il la conservera pour son usage propre, sinon il la vendra : sa valeur d'échange se transformera en valeur d'usage lors d'échanges ultérieurs. Les biens produits et détenus pour leur valeur d'échange sont appelés *marchandises*.

La théorie de la valeur d'échange de Menger pèche par la circularité du raisonnement, au moins partiellement. La valeur d'échange découle du prix de vente, car la quantité de valeur d'usage achetable ultérieurement en dépend. Mais inversement, la valeur est un facteur déterminant du prix.

L'une des propriétés permettant de caractériser les marchandises est la facilité avec laquelle elles circulent. Dans quelle mesure sont-elles acceptées dans l'échange, même en dehors de la satisfaction directe d'un besoin. Selon Menger, la **monnaie** a la même nature que les autres marchandises ; elle se distingue seulement par sa plus grande facilité de circulation, qui n'est qu'une question de degrés.

3.1.3. Walras : l'équilibre général

Léon Marie Esprit Walras est probablement le plus grand économiste néoclassique de tous les temps. N'étant pas prophète en son pays, il a fondé l'école de Lausanne. Comparés à ses très impressionnants « Eléments », les « Grundsätze » de Menger et la « Theory » de Jevons semblent assez légers.

Comme Jevons et Menger, Walras considère l'échange comme un moyen à la disposition de l'individu pour accroître son utilité totale. Mais il aborde la question différemment. Par hypothèse, les agents walrassiens sont ce que les économistes appellent des *price takers* ; en fonction de prix qu'ils considèrent comme donnés, ils déterminent les quantités des différents biens qu'ils désirent posséder. L'échange optimal est celui qui permet à notre agent d'obtenir – après échange- le panier de biens maximisant sa satisfaction à partir d'un budget déterminé. Quelle règle, l'agent applique-t-il inconsciemment pour déterminer ce panier optimal ? Walras le démontre mathématiquement, il faut que l'agent détienne de chaque marchandise la quantité telle que le rapport entre les utilités marginales et les prix soit égal pour toutes les marchandises. L'équation d'optimisation est :

$$\varphi_a(q_a + d_a) / p_a = \varphi_b(q_b + d_b) / p_b = \varphi_c(q_c + d_c) / p_c = ... \quad (3.5)$$

Les indices *a, b ,c* … désignent les biens. q_a est la quantité du bien *a* détenue avant l'échange, d_a en est la quantité échangée et p_a le prix. La somme $(q_a + d_a)$ est la quantité détenue après échange. $\varphi_a(q_a + d_a)$ est la fonction donnant l'utilité marginale du bien *a* après l'échange. Cette fonction est propre à un individu ; pour le bien *a*, les individus 1,2,3… ont respectivement les fonctions d'utilité marginale φ_{a1}, φ_{a2}, φ_{a3}…

En manipulant (3.5), on obtient pour toute paire de biens, l'égalité :

$$\varphi_a(q_a + d_a) / \varphi_b(q_b + d_b) = p_a / p_b \quad (3.6)$$

L'équation (3.6) de Walras correspond à l'équation (3.2) de Jevons, mais les deux auteurs ne lui attribuent pas la même signification. Jevons en déduisait directement l'équilibre de l'échange. Selon Walras, l'enchaînement est plus complexe : les équations individuelles (3.5) servent aux individus à déterminer leurs courbes d'offre et de demande personnelles des différents biens. L'addition des offres et demandes individuelles donne l'offre et le demande du marché. Le prix et la quantité échangée sont ceux qui égalisent l'offre et la demande du marché ; il en va ainsi sur chaque marché.

On dit d'un marché particulier qu'il est en équilibre lorsque l'offre et la demandent s'égalisent[1]. L'idée géniale de Walras est justement d'envisager l'équilibre des marchés dans leur ensemble. L'offre et la demande des différents biens étant déterminées sur base de la maximisation de son utilité par chacun des individus, Walras pose la question : ce processus, peut-il aboutir à un équilibre simultané de tous les marchés ? Cet équilibre simultané est appelé l'*équilibre général*.

Les marchés sont interdépendants vu la contrainte budgétaire, d'après laquelle on ne peut dépenser sur le marché *B* ce qui l'a été sur le marché *A*. Malgré cette interdépendance et surtout grâce à elle, Walras apporte la démonstration mathématique de l'existence de l'équilibre général[2]. Cet équilibre général est essentiellement constitué par un ensemble de prix (un par produit) ; les quantités demandées et offertes en découlent.

Le mode de démonstration de Walras consiste à mettre en équations les caractéristiques de l'équilibre. L'économie est représentée par un système d'équations linéaires simultanées ; il suffit de prouver que le système d'équations comporte une solution. Les inconnues du système d'équations sont les prix et les quantités échangées ; Walras pense que l'égalité entre le nombre d'équations et le nombre d'inconnues lui suffit à démontrer l'existence d'une solution. Il nous convie donc à un tel décompte et nous l'y suivrons, tout en simplifiant l'exposé très fastidieux qu'en donne Walras. Le lecteur rétif aux équations peut sauter cette section car la logique qui la sous-tend vient d'être exposée.

Avec le système d'équations simultanées, Walras évite l'écueil du raisonnement circulaire. Mais cette solution ne fut praticable que parce que Walras a dissocié la question de l'existence de l'équilibre de celle de la convergence de l'économie vers son équilibre. Menger, en essayant d'y répondre comme à une question unique, s'est égaré dans l'indétermination du prix. Pour

[1] L'égalité de l'offre et de la demande signifie qu'aucun acheteur ne peut dire « à ce prix là, j'aurais voulu acheter, mais je n'ai pas trouvé » et aucun vendeur ne peut dire « à ce prix là j'aurais vendu mais aucun acheteur ne s'est présenté ».

[2] Face à ce problème, Cournot, pourtant meilleur mathématicien que Walras, avait reculé : « Il semble donc que dans la solution complète et rigoureuse des problèmes relatifs à quelques parties du système économique, on ne puisse se dispenser d'embrasser le système tout entier. Or ceci surpasserait les forces de l'analyse mathématique et de nos méthodes pratiques de calcul… » (Cournot [64] p. 99).

éviter que l'indétermination soit reportée sur la question de la convergence, nous verrons à quelle astuce, diversement appréciée, Walras dut recourir.

Walras procède par étapes ; il analyse successivement des situations de complexité croissante. Il y a quatre étapes :
- l'*échange pur*, c'est-à-dire sans la production et sous la forme du troc.
- l'échange et la production *stationnaire* (avec une capacité de production constante)
- l'échange et la production progressive avec l'accumulation du capital qui augmente la capacité de production
- l'échange monétaire avec la production progressive. Cet aspect ne sera pas étudié ici mais au sous-chapitre 4.3.3.

L'ECHANGE PUR

Définissons d'abord le cadre. Il y a un certain nombre d'agents numérotés 1,2,3… Il y a *m* biens notés *A,B,C*…. Chaque agent détient au départ une quantité des différentes marchandises : q_{a1} est la quantité initiale de la marchandise *A* détenue par l'agent 1. x_{a1} est la quantité du bien *A* que l'agent un va échanger ; par convention, on a $x > 0$ s'il s'agit d'une demande et $x < 0$ s'il s'agit d'une offre.

La marchandise *A* sert de *numéraire*, c'est-à-dire que son prix est fixé arbitrairement à l'unité et qu'il sert d'unité de mesure pour les autres prix. Le prix p_a n'est donc plus une inconnue. Les prix à déterminer sont donc p_b, p_b, p_d,… et ils sont exprimés en unités de la marchandise *A*. Les prix d'équilibre sont cohérents et ne laissent pas d'opportunité d'arbitrage puisqu'il y a un seul prix par bien, quelle que soit la contrepartie contre laquelle il est échangé.

Avant d'aborder les équations montrant l'équilibre des marchés, voyons celles qui expliquent les comportements individuels, à partir desquelles les autres seront construites. Pour chaque agent, et nous prendrons l'agent *un* comme exemple, on peut construire un système d'équations indiquant qu'il applique la règle d'optimisation exposée ci-dessus (équation 3.5). Comment passons-nous de (3.5) au système (3.7) ?
- Si cette condition est satisfaite pour chaque bien avec le numéraire, elle est satisfaite pour tous les biens entre eux deux à deux. Nous passons donc à une équation par bien, sauf bien entendu qu'il n'y a pas d'équation pour le numéraire.
- La quantité échangée est représentée par la lettre *x* au lieu de *d* pour mettre en relief la possibilité qu'elle soit négative ou positive.
- Le prix p_a n'apparaît pas dans le membre de gauche, tout simplement parce qu'il vaut 1.

$$\varphi_{b1}(q_{b1} + x_{b1}) = p_b.\, \varphi_{a1}(q_{a1} + x_{a1}) \qquad (3.7)$$
$$\varphi_{c1}(q_{c1} + x_{c1}) = p_c.\, \varphi_{a1}(q_{a1} + x_{a1})$$
$$\vdots$$

A ces (m-1) équations s'en ajoute une qu'on peut qualifier de *contrainte budgétaire*, car à ce stade, l'épargne n'étant pas encore prise en compte, chaque agent doit offrir autant qu'il ne demande. En tenant compte que les *x* sont positifs en cas de demande et négatifs en cas d'offre, cette équation s'écrit ainsi :

$$x_{a1} + x_{b1}.p_b + x_{c1}.p_c + x_{d1}.p_d + \ldots = 0 \qquad (3.8)$$

Walras montre que du système (3.7) et de (3.8), on peut déduire le système d'équations (3.9) et (3.10)[1] :

$$x_{b1} = f_{b1}(p_b, p_c, p_d,...) \quad (3.9)$$
$$x_{c1} = f_{c1}(p_b, p_c, p_d,...)$$
$$::::::::::::::::::::::::::::::::$$

$$x_{a1} = -(x_{b1}.p_b + x_{c1}.p_c + x_{d1}.p_d + ...) \quad (3.10)$$

Par agent, pour chaque produit, nous avons donc une équation qui fait de son offre ou de sa demande une fonction non seulement de son propre prix mais aussi de tous les autres prix. Si le prix des autos double, je mangerai sans doute moins de chocolat.

Venons en maintenant aux équations du marché. Les quantités offertes ou demandées sur le marché (X_b, X_c…) ne que la somme des quantités offertes ou demandées individuellement.
Soient :

$$X_b = f_{b1}(p_b, p_c, p_d,...) + f_{b2}(p_b, p_c, p_d,...) + ... = F_b(p_b, p_c, p_d,...)^{2} \quad (3.11)$$
$$X_c = f_{c1}(p_b, p_c, p_d,...) + f_{c2}(p_b, p_c, p_d,...) + ... = F_c(p_b, p_c, p_d,...)$$
$$::$$

L'égalité entre l'offre et la demande sur chaque marché peut s'exprimer par le système d'équations :

$$X_b = F_b(p_b, p_c, p_d,...) = 0 \quad (3.12)$$
$$X_c = F_c(p_b, p_c, p_d,...) = 0$$
$$::::::::::::::::::::::::::::::::::::$$

Nous avons ici un système de (m-1) équations à (m-1) inconnues (p_b, p_c, p_d,...). Il a donc une solution. On peut prouver que si les (m-1) autres marchés sont en équilibre, celui de *A*, le numéraire, le sera nécessairement aussi :

Puisque $X_a = -(X_b.p_b + X_c.p_c + X_d.p_d + \ldots)$ et que X_b, X_c… valent 0, X_a doit valoir zéro.

Cette règle est d'ailleurs plus générale : il est impossible que tous les marchés soient en équilibre sauf un. Inversement, si les (m-1) marchés sont globalement en déséquilibre, le marché restant doit connaître un déséquilibre qui compense exactement le déséquilibre des autres.

Donc deux types de données suffisent à déterminer les prix d'équilibre lorsque les agents maximisent leur satisfaction : les quantités initialement détenues des différents biens par les agents et les fonctions d'utilité des différents intervenants. Nous avons d'ailleurs déterminé non seulement les prix d'équilibre mais aussi les quantités échangées, car dans les équations (3.11), si le prix est connu, on dispose de tous les éléments pour déterminer celles-ci.

Walras démontre que de ses équations, on peut déduire le théorème suivant : plus la quantité initialement disponible d'un bien est abondante, moins son prix d'équilibre sera élevé.

---

[1] Les équations précédentes forment un « système de *m* équations entre lesquelles on peut supposer successivement éliminées m-1 des inconnues x_{a1} x_{b1}… de sorte qu'il ne reste plus qu'une équation donnant la $m^{ième}$ en fonction des prix » (Walras [380] p. 179).
Walras tient compte de ce que les quantités ($q_{b1} + x_{b1}$), ($q_{c1} + x_{c1}$) … ne peuvent pas devenir négatives ; il montre que cette contrainte ne modifie pas le résultat.

[2] Par convention, Walras utilise la lettre minuscule pour l'agent individuel et la lettre majuscule pour le marché.

Walras rappelle que le numéraire n'est pas la monnaie ; il est une simple unité de compte alors que la monnaie est un instrument d'échange qui circule en sens inverse des marchandises à chaque transaction. Les prix en monnaie ne seront déterminés que lorsque nous aurons déterminé le prix de la monnaie en numéraire.

Walras écrit : « Lorsque je mesure une longueur donnée, par exemple celle d'une façade, il y a trois choses : la longueur de cette façade, la longueur de la dix millionième partie du quart du méridien terrestre (NDLR : le mètre) et le rapport de la première longueur à la seconde, qui est sa mesure (…) Pour que je pusse mesurer de même une valeur donnée, par exemple celle d'un hectolitre de blé, il faudrait qu'il y eût trois choses : la valeur de l'hectolitre de blé, la valeur du demi-décagramme d'argent au titre de 9/10 (NDLR : le franc) et le rapport de la première valeur à la seconde, qui serait sa mesure. Or de ces trois choses, deux n'existent pas, la première et la seconde. Il n'existe que la troisième (…) La valeur est une chose essentiellement relative (…) le mot *franc* est le nom d'une chose qui n'existe pas (…) Il s'ensuit que notre étalon doit être une certaine quantité d'une certaine marchandise et non la valeur de cette quantité de marchandise. »[1].

L'ECHANGE ET LA PRODUCTION

Dans le modèle ci-après, l'allocation initiale des agents (notion fondamentale dans l'économie walrassienne, rappelons-le) est une allocation de capacités productives et non plus de biens. Ils offrent des facteurs de production et, en retour, ils demandent des biens, avec une contrainte budgétaire qui impose l'égalité entre la valeur des services producteurs offerts et celle des biens demandés.

La conception des facteurs de production de Walras est assez novatrice. Il appelle *services producteurs* les inputs dans le système de production. Il y en a trois catégories :
-les services des différentes terres qui portent les symbôles *t,t',t"* …
-les services personnels, c'-à-d les différentes sortes de travail, notées *p,p',p"*…
-les services des capitaux, notés *k,k',k"*…

L'important n'est pas que la terre est un facteur de production mais qu'elle rend des services producteurs. De même pour le travail et le capital. La production est comme une machine dans laquelle entrent des services et de laquelle sortent des produits. La terre, les travailleurs et le capital n'entrent pas dans la machine, mais bien les services qu'ils rendent.

Il y a donc quatre intervenants : les détenteurs du capital, les propriétaires fonciers et les travailleurs ; face à eux, l'entrepreneur qui leur LOUE leurs services. Cette location à titre onéreux est fondamentale pour la bonne compréhension du système walrassien. L'unité de mesure du service est donc une certaine quantité d'une certaine qualité de service pendant un certain temps et c'est cette unité qui doit être évaluée sur le marché et donc faire l'objet d'un prix. Cette conception reflète bien la réalité quant aux services de la terre et du travail ; par contre, dans l'économie réelle, le capitaliste prête de l'argent à l'entrepreneur qui l'utilise pour acheter des outils. Fondamentalement, dit Walras, cela revient au même que si les capitalistes possédaient les outils et les donnaient en location à l'entrepreneur ; c'est sur cette représentation que seront bâties les équations.

L'entrepreneur rencontre les capitalistes, les travailleurs et les propriétaires fonciers une première fois sur le marché des services producteurs ; il les rencontre une deuxième fois sur le

[1] Walras [380] p. 224-225

marché des produits pour leur vendre le résultat de la production. En réalité, Walras ne s'intéresse pas beaucoup à l'entrepreneur. Dans le système walrassien, il est virtuellement absent. C'est comme si les consommateurs détenteurs des facteurs échangeaient directement les facteurs et les produits entre eux, avec au milieu une gigantesque machine à transformer les services en produits, dont le fonctionnement interne intéresserait plutôt l'ingénieur que l'économiste. Les *coefficients techniques* sont la caractéristique essentielle de la transformation des services en produits. Ils renseignent la quantité de chaque service devant entrer dans la production pour obtenir une unité d'un bien donné. Par exemple, a_t est la quantité de terre *t* nécessaire pour produire une unité du bien *A*. Dans les équations, les coefficients seront considérés comme fixes, pour la facilité. Mais Walras précise qu'ils sont la conséquence de choix technologiques. L'entrepreneur aura préalablement choisi la technologie qui minimise son coût.

Walras considère que l'équilibre général de l'échange avec production requiert trois conditions :

- l'offre de chaque produit égale sa demande.
- l'offre de chaque service producteur égale sa demande
- le prix de vente des produits égale leur prix de revient. C'est la fameuse thèse des classiques selon laquelle à long terme la concurrence efface le profit net. Mais Walras est adepte de la théorie de l'imputation ; selon lui, l'égalité du prix de revient et du prix de vente signifie que le second détermine le premier, à l'inverse de l'optique classique. Comme les prix des services producteurs et ceux des biens sont des inconnues d'un même système d'équations simultanées, le discours de Walras sur la causalité est en contradiction avec sa méthode mathématique.

Le profit chez les classiques et les néoclassiques : aspect terminologique

L'égalité entre le prix de revient et le prix du marché est une idée commune aux classiques et aux néoclassiques. Le langage pour exposer cette idée commune diverge. Chez les classiques, on avait un entrepreneur capitaliste, dont la rémunération était le profit, soumis à une loi d'égalisation entre les secteurs. Pour les néoclassiques, les fonctions de capitaliste et d'entrepreneur sont distinctes : la rémunération du capitaliste est l'intérêt (à prendre au sens large : il inclut la rémunération des actionnaires) et celle de l'entrepreneur est le profit[1]. Le profit des néoclassiques est ce qui reste quand les facteurs ont été correctement rémunérés, Y COMPRIS LE CAPITAL. Il peut donc être négatif (perte). A l'égalisation du profit (qui est un profit brut) mise en avant par les classiques correspond l'annulation du profit (qui est un profit net) chez les néoclassiques. Ce sont deux manières différentes d'exposer un même principe. La séparation des fonctions de capitaliste et d'entrepreneur est d'ailleurs formelle. Walras remarque qu'elles peuvent être cumulées sur une même tête.

La conception de l'offre des services mérite toute notre attention : les détenteurs des services en possèdent une dotation initiale et ils peuvent eux-mêmes directement en tirer une utilité. La terre peut servir de jardin, le travail peut consister à repeindre son living. Reconnaissons que cette optique convient moins bien pour les machines que pour les services de la terre et du travail. Les ménages prélèvent sur leur dotation de services une quantité qu'ils offriront aux entrepreneurs et ils achètent à ceux-ci une quantité des différents produits, de telle façon que l'utilité marginale de la quantité conservée de tous les services et celle de la quantité acquise de tous les produits sont égales, toujours en vertu de la même règle d'optimisation. L'offre des services se règle comme la demande des produits en un processus unique.

[1] Il ne faut pas confondre la présente distinction économique entre le profit et l'intérêt avec les formes concrètes de rémunération du capital (dividendes, plus-values…).

Venons en maintenant au système d'équations et à sa résolution. Les inconnues sont les prix des (m-1) biens autres que le numéraire *A* et des *n* services ainsi que les quantités échangées des *n* services et des *m* biens. Pour les biens, les quantités échangées sont les quantités produites.

On part avec les équations individuelles (3.13) à (3.15) :

$$o_t.p_t + o_p.p_p + o_k.p_k + ... = d_a + d_b.p_b + d_c.p_c + ... \qquad (3.13)$$[1]

$$\varphi_t(q_t - o_t) = p_t.\ \varphi_a(d_a) \qquad (3.14)$$
$$\varphi_p(q_p - o_p) = p_p.\ \varphi_a(d_a)$$
::::::::::::::::::::::::::::::::::::

$$\varphi_b(d_b) = p_b.\ \varphi_a(d_a) \qquad (3.15)$$
$$\varphi_c(d_c) = p_c.\ \varphi_a(d_a)$$
::::::::::::::::::::::::::::

L'équation (3.13) donne la contrainte budgétaire, compte tenu de ce que le revenu vient de l'offre des facteurs. L'équations (3.14) (il y en a une pour chacun des services *t, p, k*…) donne la règle d'optimisation de la quantité offerte des services producteurs ; elle stipule que la quantité après échange se règle de façon à égaliser son utilité marginale avec celle du numéraire et donc avec celles de tous les biens et facteurs après échange. Les équations (3.15) sont la nouvelle version des équations (3.7).

Suivant le principe déjà vu pour l'échange pur, on obtient par sommation, les équations de la demande collective des biens et de l'offre collective des facteurs :

$$O_t = F_t(p_t, p_p, p_k, ... p_b, p_c, p_d, ...) \qquad (3.16)$$
$$O_p = F_p(p_t, p_p, p_k, ... p_b, p_c, p_d, ...)$$
::

$$D_b = F_b(p_t, p_p, p_k, ... p_b, p_c, p_d, ...) \qquad (3.17)$$
$$D_c = F_c(p_t, p_p, p_k, ... p_b, p_c, p_d, ...)$$
::

$$D_a = O_t.p_t + O_p.p_p + O_k.p_k + ... - (D_b.p_b + D_c.p_c + ...) \qquad (3.18)$$

Il reste donc à ajouter *n* équations égalisant l'offre et la demande des services :

$$a_t.D_a + b_t.D_b + c_t.D_c + ... = O_t \qquad (3.19)$$
$$a_p.D_a + b_p.D_b + c_p.D_c + ... = O_p$$
::

Ainsi que *m* équations égalisant le prix des biens avec leur coût de production :

[1] L'agent offre des facteurs et demande des produits. La lettre *x* (quantité échangée en échange pur) fait place à la lettre *d* pour la demande de produits et la lettre *o* pour l'offre de facteurs.

$$a_t.p_t + a_p.p_p + a_k.p_k + \ldots = 1 \qquad (3.20)$$
$$b_t.p_t + b_p.p_p + b_k.p_k + \ldots = p_b$$
$$::$$

Le système (3.16 à 3.20) comporte 2(m+n) équations pour 2m+2n-1 inconnues, soit une de trop. Mais Walras montre que les équations ne sont pas linéairement indépendantes et que donc on peut en supprimer une. Le système a donc une solution. On peut s'étonner de l'absence d'équations stipulant l'égalité entre l'offre et la demande de produits. En fait cette égalité est impliquée par l'ensemble des équations.

INTEGRATION DE L'EPARGNE ET DES BIENS CAPITAUX.

Nous connaissons déjà les services des capitaux. Faisons connaissance avec les capitaux qui les fournissent. Dans le présent chapitre, Walras n'envisage que les biens durables servant de capital (capital fixe) qu'il appelle « les capitaux » et remet l'étude du capital circulant au chapitre ultérieur consacré à la monnaie (cf. sous-chapitre 4.3.3). Les capitaux sont notés K, K', K"…[1] et ont pour prix P_k,$P_{k'}$…[2] Les services des capitaux gardent leur rôle déjà connu ; le système est simplement complété par de nouvelles équations et de nouvelles inconnues.

Le prix p_k des services du capital *K* comporte deux parties bien distinctes :
-la prime d'amortissement et d'assurance contre la perte ou la destruction.
-le revenu net, l'intérêt en quelque sorte, noté π_k

L'origine du capital se trouve dans l'épargne. Pour qu'il y ait une offre et une demande de capitaux, la consommation doit être inférieure au revenu ; avec la différence, certains consommateurs deviennent capitalistes en acquérant les capitaux, dont ils louent les services aux entrepreneurs. En contrepartie, ceux-ci produisent non seulement des biens de consommation mais également des capitaux neufs, pour une valeur qui devra s'équilibrer avec celle de l'épargne. Le niveau du taux d'intérêt, conjointement avec l'ensemble du système des prix (comme toujours chez Walras), sert de régulateur pour équilibrer ces forces de l'épargne et de l'investissement.

Walras écrit : « Les capitaux neufs s'échangent contre l'excédent du revenu sur la consommation ». Comme on peut le voir, la théorie du capital de Walras est entièrement orientée vers une économie en croissance (puisque l'épargne accroît le stock de capital) ; on lui a reproché d'être ambiguë quant à ce qu'il adviendrait de l'intérêt dans une économie stationnaire.

Pour intégrer sa théorie du capital, Walras ajoute des équations et des inconnues à son système. D'abord, dans les équations relatives à l'équilibre du consommateur (3.13 et 3.15), l'épargne est ajoutée comme une marchandise supplémentaire, la marchandise *E*, dont le prix unitaire est P_e qui vaut 1/i où *i* est le taux d'intérêt. C'est la formule de la valeur d'une rente perpétuelle d'une unité de numéraire. Le prix P_e est somme toute un prix comme les autres ; il entre dans la série de prix qui déterminent l'offre des facteurs et la demande des biens (équations 3.16 et 3.17).

[1] Walras envisage également les capitaux T, T', T"… qui sont les variétés de terres et les capitaux P, P', P"… qui sont les capitaux personnels. La signification de ces derniers me paraît douteuse. A titre de simplification, le présent exposé se limite aux capitaux de type *K*.
[2] *P* majuscule pour le prix des capitaux ; *p* minuscule pour le prix de location de leurs services.

Pour déterminer la quantité produite des capitaux et leur prix, des équations spécifiques sont ajoutées : chaque capital a droit à deux équations expliquant son prix :
-la première stipule que le prix du capital sera égal à son coût de production (comme les autres biens)
-la deuxième égalise ce même prix égalera avec la capitalisation de son revenu net, au taux d'intérêt ; pour chaque *K*, on a :

$$P_K = \pi_k / i. \qquad (3.21)$$

Il s'ajoute encore deux équations centrées sur l'épargne.
- la première détermine l'épargne totale de la société ($E = D_e.P_e$) :

$$E = F_e(p_t...p_p...p_k,p_{k'},p_{k''}... p_b...p_d, i) \qquad (3.22)$$

- la deuxième assure l'égalité entre l'épargne et l'investissement.

$$E = D_k.P_k + D_{k'}.P_{k'} + D_{k''}.P_{k''} +... \qquad (3.23)$$

Ces équations supplémentaires sont en nombre pour déterminer les inconnues supplémentaires que sont : la quantité produite de chaque capital (les D_k), le prix de chaque capital (les P_k), le montant total de l'épargne de la société (E) et le taux d'intérêt (i).

DYNAMIQUE DE L'EQUILIBRE : LE TATONNEMENT

Comment les prix d'équilibre émergent-ils ? Le modèle de Walras implique que les agents opèrent comme de parfaits price takers par rapport à ces prix et que toutes les transactions se déroulent exactement au prix d'équilibre. Pour cette raison, Walras a conçu un mode de fonctionnement du marché qu'il a appelé le *processus de tâtonnement* ; il prévoit qu'aucun échange ne peut avoir lieu tant que tout le monde ne s'est pas mis d'accord sur tout. On fixe l'ensemble des prix qui doivent égaliser l'offre et la demande de tous les biens, avant de passer à la production et l'échange effectifs sur base de ces prix. Ainsi pas un seul bien ne sera acheté ou vendu à un autre prix que le prix d'équilibre.

Pour faciliter l'émergence des prix d'équilibre, Walras imagine même la présence d'un commissaire-priseur fictif. Pour utiliser l'expression de Walras, les prix sont « criés ». En fonction des prix criés, chaque intervenant exprime combien il achète ou vend. Le commissaire peut ainsi directement proposer un prix plus élevé si la demande était plus élevée que l'offre ou plus bas dans le cas inverse. En fait, les agents ne traitent pas bilatéralement mais avec le marché. Avec ce procédé, l'équilibre théorique ne peut manquer de se réaliser, éventuellement au prix de nombreux tâtonnements du commissaire.

Le fonctionnement réel de la plupart des marchés est assez éloigné du modèle du tâtonnement. On peut le concevoir comme une espèce de perfection vers laquelle tendraient les marchés réels. Assez naïvement, Walras semble assez confiant dans le réalisme de son hypothèse.

IMPLICATIONS SOCIALES ET POLITIQUES

Avec la solution de son système d'équations, Walras estime avoir démontré mathématiquement que la libre concurrence permet d'obtenir le maximum d'utilité et il n'en est pas peu fier : « C'est bien là, en somme, ce que les économistes ont déjà dit en préconisant le *laisser faire, laisser passer*. Malheureusement, il faut bien le dire : les économistes jusqu'ici ont moins démontré leur *laisser faire, laisser passer* qu'ils ne l'ont affirmé à l'encontre des socialistes, qui, de leur côté, affirment, sans la démontrer davantage, l'intervention de l'Etat »[1].

[1] Walras [380] p. 335

L'avantage de la démonstration scientifique d'un principe, continue Walras, est qu'on peut clairement délimiter l'aire où il s'applique de celle où il ne s'applique pas. Walras préconise donc le laissez faire ; dans ses « Etudes d'économie appliquée », il énumère néanmoins quelques exceptions :

1- L'Etat doit émettre la monnaie. En effet, la quantité de monnaie est indifférente au consommateur qui ne se préoccupe que de sa valeur, ce qui n'est pas compatible avec le fonctionnement concurrentiel.
2- « Une autre condition de la libre concurrence est que l'appréciation de l'utilité soit possible pour l'individu, ce qui n'est pas le cas en général pour les services publics. Donc il faut que l'Etat produise les services publics »[1].
3- L'Etat doit surveiller et règlementer certains services privés pour lesquels le consommateur n'est pas en mesure d'apprécier la qualité par lui-même. Walras ne donne pas d'exemple
4- La concurrence postule que la « multiplicité indéfinie des entreprises soit possible, ce qui n'est pas le cas des monopoles naturels et nécessaires tels que les mines, carrières, eaux minérales, chemins de fer etc ». L'Etat doit donc organiser les monopoles de façon à assurer l'égalité du prix de vente avec le prix de revient concurrentiel.

Dans le même ouvrage, Walras expose sa théorie de la justice sociale : « Il y a ainsi deux espèces de richesse sociale à répartir : les terres et les facultés personnelles, et il y a deux types sociaux entre lesquels cette richesse doit être répartie : l'Etat et l'individu. Eh bien, le principe de l'*inégalité des positions* exige que les *facultés personnelles* soient attribuées à l'*individu*, et le principe de l'*égalité des conditions* exige que les *terres* soient attribuées à l'*Etat* »[2]. Walras ne mentionne pas la richesse matérielle non foncière en tant que troisième espèce, car elle est incluse dans les deux autres : le propriétaire de la terre ou celui des facultés personnelles sont propriétaires des services rendus par ces richesses, donc des produits qu'ils obtiennent en échange de ces services et donc a fortiori de l'accumulation non consommée de ces produits.

En conséquence, Walras prône :
- la nationalisation de la terre et des richesses naturelles
- la suppression des impôts, car ils spolient les propriétaires des facultés personnelles.

Comme Ricardo, Walras prévoit que la rente est appelée à croître. D'après lui, elle suffira pour financer les dépenses de l'Etat en remplacement de l'impôt et l'emprunt.

3.1.4. Edgeworth et l'indétermination de l'échange

En 1881, dix ans après la « Theory of Political Economy » de Jevons, Francis Edgeworth publie les « Mathematical Psychics » où il aborde le problème de l'équilibre de l'échange avec un regard neuf. Comme le laisse supposer le titre de l'ouvrage, Edgeworth est lui aussi adepte de l'utilitarisme. L'auteur s'attache à montrer que « mathematical reasoning is not, as commonly supposed, limited to subjects where numerical data are attainable ». Il estime « remarkable that principal inquiries in Social Science may be viewed as *maximum*-problems »[3].

[1] Walras [379] p. 477
[2] Walras [379] p. 470. Les italiques sont de Walras.
[3] Edgeworth [85] pp. 2 et 6.

Edgeworth reprend le problème jevonnien de l'équilibre de l'échange. Il commence avec *l'échange bilatéral* où les protagonistes ne sont effectivement que deux, sans qu'on puisse interpréter cette situation comme un extrait représentatif du marché concurrentiel.

Les deux protagonistes sont Robinson Crusoé, le héros de Defoe, et le sauvage de son île, Vendredi ; le second offre son travail (bien *Y* en quantité *y*) contre le salaire (bien *X* en quantité *x*) offert par le premier. Chez Jevons, comme chez Walras, les fonctions d'utilité d'un individu, $\phi_x(x)$ et $\phi_y(y)$, étaient -de façon simpliste- indépendantes entre elles : ϕ_x donnait l'utilité de *X* et ne dépendait que de la quantité *x* ; ϕ_y donnait l'utilité de *Y* et ne dépendait que de la quantité *y*. L'utilité des deux biens ensemble valait $\phi_x(x) + \phi_y(y)$. Edgeworth introduit un progrès en recourant à des fonctions d'utilité à plusieurs variables (une variable par bien, donc deux dans l'exemple présent) ; l'utilité totale de Robinson est la fonction P(x,y) et celle de Vendredi la fonction $\prod(x,y)$; ce sont donc des COMBINAISONS de biens qui procurent l'utilité. L'utilité marginale de *X* et celle de *Y* sont données par les dérivées partielles dP/dx et dP/dy et dΠ/dx et dΠ/dy, respectivement pour Robinson et pour Vendredi. Si l'on prend la formule (3.2) de Jevons, qu'on la retraite suivant la présente méthode et qu'on supprime le terme intermédiaire (y/x), on obtient :

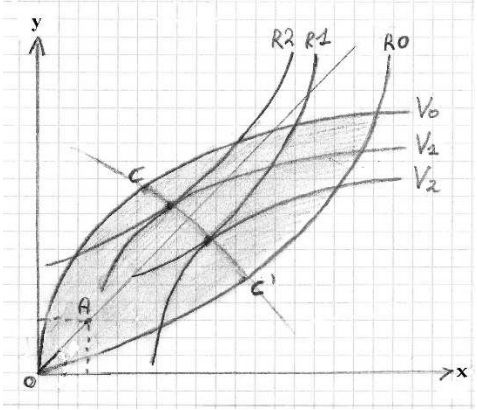
$$\frac{dP}{dx} \cdot \frac{\mathrm{d}\prod}{dy} = \frac{dP}{dy} \cdot \frac{\mathrm{d}\prod}{dx} \qquad (3.24)$$

Comment Robinson et Vendredi vont-ils conclure, sachant qu'il n'y a pas de commissaire-priseur comme chez Walras ? L'ensemble des couples (x,y), avec $x \geq 0$ et $y \geq 0$ constituent les échanges possibles, qu'Edgeworth appelle *contrats*,. Robinson préférera certains échanges à d'autres, car ils lui apportent plus d'utilité ; Vendredi de même. Dans l'ensemble de ces échanges, Edgeworth appelle *règlements* ceux pour lesquels il est impossible que Robinson et Vendredi, de commun accord, puissent leur en préférer un autre. Ce sont les contrats par rapport auxquels on ne peut améliorer l'utilité de Robinson qu'au détriment de Vendredi ou inversement.

Edgeworth conçoit une représentation graphique innovante du problème ; c'est la figure 3.4. Comme *x* et *y* sont des quantités échangées, l'origine indique l'échange zéro, le point de départ avant échange. Pour un point quelconque sur la carte des échanges, par exemple *A*, la pente de la droite le reliant à l'origine vaut y/x et exprime donc le *rapport d'échange*. Tous les points compris sur une droite partant de l'origine (par exemple celle passant par *A*) ont le même rapport d'échange (y/x). On peut donc tracer différentes droites passant par l'origine pour exprimer les différents *rapports d'échange* (autrement dit, les prix relatifs).

Pour chacun des deux échangeurs, on peut relier sur la carte les contrats lui apportant une utilité égale ; on forme ainsi des *courbes d'indifférence*. Ainsi, R0, R1 et R2 sont des courbes d'indifférence de Robinson par ordre croissant d'utilité ; V0, V1 et V2 de même pour Vendredi. Les courbes d'indifférence sont convexes par rapport à l'axe du bien offert, car pour chaque unité supplémentaire de *X*, Robinson exigera un incrément de *y* de plus en plus élevé , comme l'avait déjà montré Jevons ; pour chaque nouvelle unité de *y*, Vendredi exigera de même de plus en plus de *X* .

Graphique 3.4 : l'indétermination des contrats



Le nombre de courbes d'indifférence est infini, mais nous en avons dessiné trois pour Robinson et trois pour Vendredi. V0 et R0 sont les CI passant par l'origine. C-C' est la courbe de contrat.

Robinson et Vendredi vont négocier de pied ferme, puisque « the first principle of economics is that every agent is actuated only by self-interest »[1]. Ils feront propositions et contre-propositions dans le champ (x,y), Vendredi poussant vers le sud-est et Robinson vers le nord-ouest. Vers quel résultat se dirigent-ils ? Il est évident qu'ils n'envisageront des échanges que dans la surface (grisée) comprise entre les courbes d'indifférence passant par l'origine ; sinon l'un des protagonistes perdrait à l'échange.

Les points correspondant aux *règlements* sont évidemment ceux où les courbes d'indifférence de Robinson sont tangentes avec celles de Vendredi. Le graphique 3.4 montre que ces points sont multiples entre C et C'. Ils forment ensemble ce qu'Edgeworth appelle la *courbe de contrat*. La formule (3.24) est vérifiée pour chacun des points de cette courbe. Cette multiplicité des *règlements* possibles exprime le premier grand théorème d'Edgeworth : l'*indétermination du contrat*, qui prévaut en l'absence de concurrence parfaite (nécessitant un grand nombre d'acheteurs et de vendeurs). Turgot et Menger avaient déjà affirmé cette indétermination.

Selon Edgeworth, la présence à grande échelle de l'indétermination des contrats dans l'économie est un mal, car elle ouvre la porte de la bergerie au loup. Les plus rusés empocheront la plus grande part. Selon lui, cette situation appelle l'arbitrage. Un arbitrage basé sur les principes de l'utilitarisme. Il préconise que dans les cas où les forces du marché ne déterminent pas l'équilibre de l'échange, un arbitrage soit mis en œuvre, veillant à faire prévaloir le rapport d'échange permettant la maximisation de l'utilité globale de l'ensemble des membres de la société.

Voyons maintenant ce qu'on appelle la *conjecture d'Edgeworth*, son deuxième apport, corollaire au premier : lorsque le nombre d'intervenants augmente, la courbe de contrat se rétrécit, au point de n'être plus qu'un point lorsqu'il tend vers l'infini. Edgeworth ne démontre pas l'existence d'un équilibre unique en concurrence parfaite mais montre la dynamique du rétrécissement progressif de la courbe de contrat.

La méthode d'Edgeworth consiste à introduire successivement de nouveaux acteurs dans le champ d'échange, par couple : un clone de Robinson avec un clone de Vendredi. Cette extension de l'échange par réplication des agents a l'avantage qu'on peut conserver la présentation graphique car tous les Robinsons ont le même champ d'indifférence et la même

[1] Edgeworth [85] p.16.

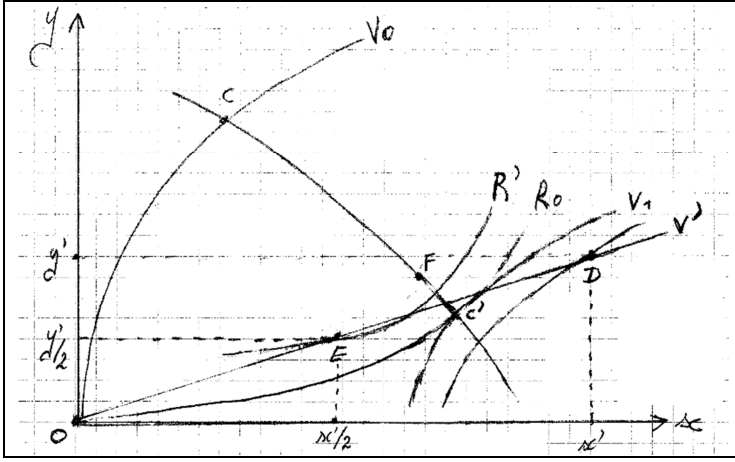
allocation de départ (du *X*) ; de même pour les Vendredis doués des mêmes goûts et de la même allocation de *Y*.

Edgeworth permet aux quatre protagonistes de passer ce qu'il appelle des « contrats », mais qui sont en réalité des projets de contrats dont il y a moyen de se désengager s'il s'avère possible d'en négocier un meilleur avec un ou plusieurs autres intervenants, ce qu'Edgeworth appelle *recontracter*. L'échange à quatre (ou plus) n'atteindra une position d'équilibre qu'à deux conditions :

- il se trouve sur la courbe de contrat, comme dans le cas « à deux »
- les quatre protagonistes se trouvent sur un point (x,y) unique. Un Robinson et un Vendredi situés en des points différents pourront toujours négocier des arrangements plus avantageux.

Supposons qu'à deux, le premier Robinson et le premier Vendredi aient conclu un *règlement* très défavorable à Robinson (infinitésimalement au-dessus de C' sur le graphique 3.5). Edgeworth montre justement que les règlements proches des extrémités de la courbe de contrat à deux ne pourront pas rester des *règlements* à quatre.

Graphique 3.5 : le rétrécissement de la courbe de contrat à quatre intervenants



Les deux Robinson pourraient proposer à un Vendredi un échange correspondant au point D = (x',y') hors de la courbe de contrat, laissant l'autre Vendredi sur le carreau. Ils lui fourniraient chacun x'/2 contre y'/2. Ce Vendredi monterait de la courbe d'indifférence V_1 à V', qui passe par le point D ; les deux Robinsons monteraient d'environ R_0 à R', qui passe par le point E. Ils ne cèdent que les unités de *X* les moins utiles et obtiennent les unités de *Y* les plus utiles, d'où le gain d'utilité. Le second Vendredi, pour ne pas rester à V_0, contre-proposera l'échange *F*, que les Robinsons s'empresseront d'accepter et auquel le premier Vendredi ne pourra pas s'opposer. L'extrémité sud-est de la courbe de contrat est donc rabotée et un raisonnement symétrique permet de raboter son autre extrémité.

Avec six échangeurs, les trois Robinsons pourraient proposer à deux Vendredi un échange (x",y") où ils fournissent chacun 2/3 de x" contre 2/3 de y". Le scénario serait similaire à celui du stade précédent et peut se reproduire à l'infini, avec une courbe de contrat qui rétrécit à chaque stade.

On constate que la réduction de l'indétermination des contrats obtenue par Edgeworth n'est pas liée à la diversité des goûts des agents puisqu'il n'y a que deux types de goûts : ceux de Robinson et ceux de Vendredi. C'est la multiplication des possibilités de recontracter qui joue.

Si des sous-groupes de Robinsons ou de Vendredi formaient des coalitions permanentes (cartels, syndicats par exemple), il en résulterait le même effet que s'ils étaient moins

nombreux : cela augmenterait l'indétermination et améliorerait leurs chances d'obtenir un règlement qui leur est favorable.

Edgeworth montre que sur le graphique 3.4, on pourrait déduire de leurs courbes d'indifférence, ce qu'il appelle une courbe de demande de Vendredi et une courbe de demande de Robinson. Ces deux courbes indiquent les quantités offertes et demandées de *X* contre *Y* ou *Y* contre *X* aux différents prix relatifs.

En concurrence parfaite avec une infinité d'acteurs, toutes ces courbes de demande se croisent sur la courbe de contrat, déterminant ensemble le point d'équilibre unique. En ce point, chaque échangeur est sur sa courbe de demande, ce qui signifie qu'au prix en vigueur il ne désire ni plus ni moins des produits que ce que lui laisse l'échange. Au contraire, en concurrence imparfaite, les échangeurs sont sur la courbe de contrat mais ne seraient sur leur courbe de demande que par hasard.

L'intersection des *courbes de demande* chez Edgeworth correspond à l'équilibre walrassien entre l'offre et la demande. L'équilibre général walrassien est concurrentiel. L'économie atteint son point d'équilibre, avec l'aide du commissaire-priseur. Edgeworth était plein d'admiration pour le système walrassien, sauf justement le processus de *tâtonnement* qu'il considérait comme parfaitement irréaliste. Avec la négociation et la renégociation des contrats, il entend apporter un modèle alternatif. Mais le modèle de négociation et renégociation des contrats par des agents formant des blocs, est-il une description du marché plus réaliste que l'intervention du commissaire priseur walrassien ?

3.2- LA CONSOLIDATION

La première phase de la révolution marginaliste s'était principalement attachée à l'échange et à la valeur. La deuxième privilégiera la production et la rémunération des facteurs de production.

Rappelons-nous la rente différentielle de Ricardo et plus précisément sa deuxième forme. Sur une terre donnée (autrement dit, en quantité fixe), l'investissement de doses successives de travail et capital, considérés ici comme formant ensemble un seul et même facteur homogène, apporte un supplément de produit, qui va décroissant. Comme toutes les doses de capital et travail doivent recevoir la même rémunération, celle-ci ne peut dépasser la productivité imputable à la dose la moins productive, c'est-à-dire la dernière. Tout le surplus produit par les unités précédentes par rapport à cette rémunération devient de la rente foncière.

Clark et Wicksteed vont « généraliser » cette loi, montrer que Ricardo a révélé un cas particulier d'une loi générale valant pour tous les facteurs de production aussi bien dans l'industrie que dans l'agriculture. Si à la place de faire varier la quantité de travail et capital sur une terre donnée, on prend n'importe quel facteur comme facteur variable en maintenant constante la quantité de tous les autres, les conclusions de Ricardo restent valables. La *productivité marginale* devient un concept central, défini comme la variation de la production causée par une variation infinitésimale du facteur variable. Elle est décroissante avec l'augmentation de la quantité du facteur. Si on considère que tout facteur de production est variable à un degré ou à un autre, la productivité marginale devient la norme de sa rémunération.

Comme toute loi, celle-ci a ses précurseurs. Johann von Thünen l'avait énoncée en 1850, mais elle passa inaperçue. Le lecteur aura reconnu que la décroissance de la productivité marginale est affirmée dans la dernière citation de Turgot au sous-chapitre 2.2.1. L'explication de la valeur des biens d'ordre supérieur par Menger est une variante de la rémunération suivant la productivité marginale.

Dans la troisième édition des *Eléments*, parue en 1896, Walras adaptera son système d'équations pour introduire la rémunération des facteurs suivant leur productivité marginale[1].

Autre concept appelé à devenir fondamental : le coût. Et son principe-même fait débat : d'une part, Alfred Marshall fait revenir le coût de production dans la théorie de la valeur. Le prix résulte de la rencontre de la demande qui dépend de l'utilité avec l'offre qui dépend du coût de production. La théorie de l'utilité marginale et celle du coût de production sont donc également nécessaires, comme les deux lames d'une paire de ciseaux, pour reprendre sa propre métaphore. Les partisans d'une théorie de la valeur purement subjective lui opposent la doctrine du *coût d'opportunité*. Les plus radicaux sont les Autrichiens ainsi que l'Anglais Philip Wicksteed.

Les partisans de cette doctrine refusent d'envisager le coût comme quelque chose d'objectif. Le coût d'opportunité peut se définir ainsi : la valeur de ce à quoi il faut renoncer pour obtenir le bien considéré. Si pour produire une unité supplémentaire du bien *A*, il faut investir du capital et du travail qui auraient pu servir à produire deux unités supplémentaires du bien *B* auxquelles il faudra renoncer, le coût d'opportunité du bien *A* vaut deux unités du bien *B*. Pendant un quart de siècle, Marshall et Edgeworth débattront avec Wicksteed et les Autrichiens quant à savoir si

[1] Les coefficients techniques anciennement fixes deviennent des inconnues ; des équations sont ajoutées. Nous n'approfondirons pas.

la désutilité du travail et l'abstinence du capitaliste constituent un facteur objectif ultime pour le coût.

3.2.1. Clark et la productivité marginale

Dès 1889, John Bates Clark énonce sa théorie de la répartition, voulant que chaque facteur est rémunéré sur base de sa productivité marginale. Elle fut ensuite énoncée avec force et clarté dans son ouvrage « The Distribution of Wealth : a Theory of Wages, Interest and Profits » paru en 1899.

Dans son exposé, Clark considère qu'il y a deux facteurs de production : le travail et le capital. Mais qu'est-ce que le capital ? Clark insiste sur la distinction entre le CAPITAL et les BIENS CAPITAUX. Les *biens capitaux* sont les outils et les bâtiments, les matières premières ainsi que l'encours de production et le stock de produits finis. La terre est un bien capital parmi les autres et n'est donc plus un facteur de production indépendant comme chez d'autres économistes. Le capital est le fonds financier qui s'incarne dans ces biens capitaux ; il est défini ainsi : « a permanent fund of productive wealth, what is commonly meant by « money » invested in productive goods, the identity of which is forever changing »[1]. Clark déplore l'usage ambigu du terme capital qui souvent désigne aussi bien le capital que les biens capitaux.

Pour Clark, c'est le capital qui est essentiel et non les biens capitaux, c'est à lui que s'applique la loi de la productivité marginale en tant qu'elle détermine l'intérêt, comme elle détermine le salaire du travail. Les biens capitaux restent superficiels, même s'il est évident que sans eux, le capital n'a pas de réalité.

Clark distingue l'analyse statique et l'analyse dynamique. Les lois statiques expliquent les différentes grandeurs comme la valeur, l'intérêt et le salaire dans un cadre statique qui se définit par :

- un état donné de la technologie
- une population active donnée
- un volume de capital donné
- un mode d'organisation donné
- des goûts donnés

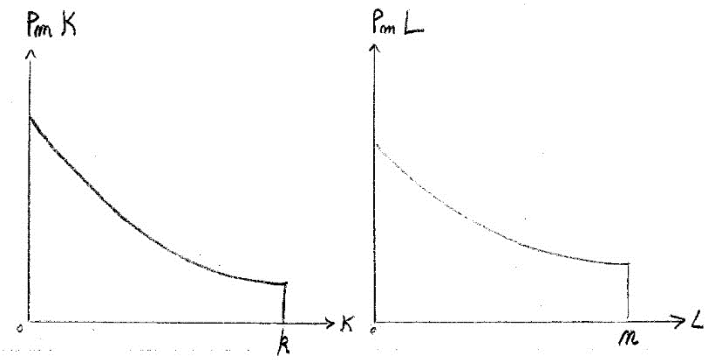
Dans sa quasi-totalité, l'ouvrage traite des lois statiques.

Etudions la productivité marginale dans une économie comportant n unités de travail et k unités de capital. Commençons par celle du travail. Dans un effort d'abstraction, imaginons que nous avons extrait les n unités de travail hors de l'échiquier économique et qu'ensuite, nous les y réintroduisions une à une, de façon à analyser la situation à chaque insertion. La première unité disposera pour elle seule d'un capital de k et aura de ce fait une énorme productivité. L'insertion d'une deuxième unité fait tomber le capital par tête à $k/2$. Ensemble, les deux travailleurs produiront plus que le premier lorsqu'il était seul, mais leur production par tête sera moindre car ils seront moins bien équipés. Et ainsi de suite… Cette baisse progressive du capital par tête ne doit pas être interprétée de façon simpliste. Elle signifie, non pas un nombre moins élevé d'outils identiques, mais de moins bons outils, un moindre degré de mécanisation.

[1] Clark [54] § 11.1

Face à des volumes de main d'œuvre différents, le capital *k* s'incarnera dans des gammes de biens capitaux différentes.

Graphique 3.6 : la productivité marginale



Continuons à ajouter des unités de travail jusqu'à l'unité n qui apportera la contribution la plus faible car le capital par tête sera réduit à k/n. Comme les *n* travailleurs devront recevoir le même salaire, ce salaire ne pourra être supérieur à la productivité de l'unité n. Sinon, celle-ci coûterait plus qu'elle ne rapporte et il serait avantageux pour l'entrepreneur d'engager moins que n unités. Mais le salaire ne sera pas non plus inférieur à la productivité marginale, du fait de la concurrence entre les entrepreneurs. Tant qu'un travailleur rapporte plus qu'il ne coûte, les entrepreneurs ont intérêt à tenter d'obtenir ses services, fût-ce en le débauchant grâce à un salaire plus attrayant.

En fait, il est peu probable que la productivité marginale baisse dès la deuxième unité ; sans doute un effectif d'un travailleur (ou de quelques uns) est-il trop réduit pour profiter pleinement de k unités de capital. Les premières adjonctions de travailleurs feront hausser la productivité marginale, mais l'important est qu'à partir d'un certain niveau, celle-ci baissera nécessairement et définitivement.

L'exercice que nous venons de réaliser avec le travail, recommençons-le maintenant avec le capital. Injectons dans l'économie comportant n travailleurs de façon fixe, les *k* unités de capital (valant par exemple chacune $1.000.000), tour à tour. Nous verrons la productivité marginale du capital baisser. Au fur et à mesure que du capital supplémentaire s'ajoute, il faudra lui trouver des usages de moins en moins productifs puisque les premiers usages vers lesquels on le destine sont évidemment les plus productifs. Le capital quantitativement de plus en plus élevé achètera des outils de moins en moins efficaces. L'intérêt sera donné par la productivité de l'unité k du capital, pour la même raison que le salaire était donné par la productivité de l'unité n de la main d'œuvre. Clark raisonne pour une économie où la concurrence rabote le profit. La répartition se joue exclusivement entre le salaire et l'intérêt, tous deux soumis à un taux uniforme entre les différentes branches. Tant qu'on reste dans le cadre statique, il n'y a pas de profit net en plus de l'intérêt.

La dernière unité de capital, à laquelle nous avions donné l'indice k, n'est pas un outil déterminé. L'augmentation du capital est d'ordre qualitatif autant que quantitatif. « The final increment of the capital of this railroad corporation is, in reality, a difference between two kinds of plants for carrying goods and passengers. One of these is the railroad as it stands, with all its equipments brought up to the highest pitch of perfection that is possible with the present resources. The other is the road built and equipped as it would have been if the resources had been by one degree less (…) the final increment of the capital of this industry is not one that

can be physically taken out of it »[1]. Le remplacement du matériel usagé par des équipements plus performants est une façon très courante d'augmenter le capital.

Peut-on parler de la rémunération des biens capitaux ? Indirectement oui, car ils pourraient être loués. Dans ce cas, leur propriétaire toucherait un loyer. Celui-ci se compose de deux parts : la première est le montant nécessaire pour remplacer l'instrument lorsqu'il sera totalement usé ; la seconde est le gain net. La totalité du gain net de tous les instruments appartenant à un capital correspond à l'intérêt rapporté par ce capital ; il doit y avoir égalité entre eux.

L'optique de Clark combine savamment les points de vue de l'entreprise individuelle et de l'économie prise globalement. Si un entrepreneur A peut occuper un facteur plus productivement que l'entrepreneur B, il acceptera de mieux le rétribuer ; les facteurs, qui sont mobiles, répondront toujours à ces appels. Les facteurs sont attirés là où ils sont les plus productifs. Mais en quittant un secteur B où ils sont moins productifs, ils y font augmenter la productivité marginale tandis qu'ils la font baisser dans le secteur A, qu'ils rejoignent. Le capital et le travail se répartissent entre les secteurs et les entreprises, de façon à égaliser la productivité marginale et les rémunérations dans l'ensemble de l'économie. La productivité marginale du capital dans une entreprise donnée dépend de la quantité qui y est utilisée. Mais en définitive, c'est la quantité de capital disponible dans la société qui en détermine la productivité marginale dans toutes les entreprises. Il en va de même pour le travail.

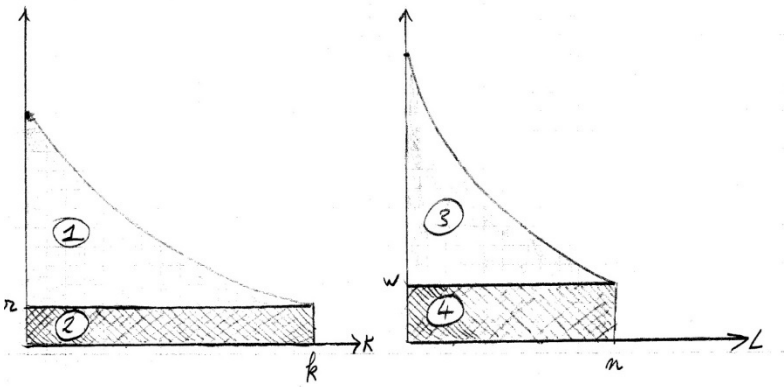
L'analyse dynamique intervient si l'une des cinq grandeurs qui définissent le cadre statique se modifie. Il en résulte un changement dans le niveau NORMAL de la valeur, de l'intérêt et du salaire ; l'analyse dynamique met en lumière ces variations et leur sens. Une fois les nouveaux niveaux normaux déterminés, l'analyse statique reprend le flambeau pour expliquer comment les variables économiques tendent vers leur nouveau niveau normal. Clark prend l'exemple d'un progrès technique permettant de réduire le coût et donc le prix de vente. L'entrepreneur innovateur récoltera d'abord un profit. Mais sous l'effet de la concurrence, celui-ci fondra et bientôt il sera absorbé par le salaire et l'intérêt : toute la société profitera de l'innovation. Le profit n'est jamais que temporaire mais pourtant il est nécessaire ; sans lui, l'innovation manquerait d'un incitant. Dans l'économie réelle, le changement est permanent : un nouveau changement arrive avant que l'équilibre rompu par les précédents ait pu être rétabli. Le salaire, l'intérêt et la valeur courent toujours derrière leurs niveaux d'équilibre qu'ils ne rattrapent jamais.

Revenons à la distribution du produit dans notre économie à k unités de capital et n unités de travail et analysons d'abord la productivité du travail comme facteur variable (courbe de droite). Les travailleurs de 1 à $(n-1)$ ont une productivité supérieure à leur rémunération basée sur la productivité de n. Le revenu qui leur échappe est la zone (3) sur le graphique 3.7. Ne sont-ils dès lors pas « exploités » ? Non dit Clark, la zone (3) est, contrairement aux apparences, imputable au capital et non au travail. « The surplus is (…) not merely a remainder that is left after paying wages, but it is the difference between one product and another. It is the difference between the product of aided labor and that of the labor that is virtually unaided… »[2]. Vers la gauche du graphique, les travailleurs ont une productivité marginale plus élevée, car étant moins nombreux, ils bénéficient d'un capital par tête plus élevé. Ce surplus n'est d'ailleurs rien d'autre que la rémunération du capital. Il y a égalité entre les zones (2) et (3) sur le graphique.

[1] Clark [54] § 17.10
[2] Clark [54] § 13.20

Ce que nous avons dit du travail vaut également pour le capital. La productivité des unités de capital infra-marginales au-delà de l'intérêt « paye » le salaire : il y a aussi équivalence entre les zones (1) et (4) sur le graphique.

Graphique 3.7 : la rémunération des facteurs suivant la productivité marginale



Sur cette base, Clark, qui était un apologiste du capitalisme, estime avoir démontré la justice distributive du système: chacun reçoit l'équivalent de ce qu'il a donné. Il écrit : « what a social class gets is, under natural law, what it contributes to the general output of industry »[1]. Cette prétention a été critiquée par la grande majorité des économistes.

La critique la plus évidente fut formulée par l'économiste Frank Knight, néoclassique et marginaliste comme Clark : « The income doesn't go to factors but to their owners and can in no case have more ethical justification than has the fact of ownership. The ownership of personal or materiel productive capacity is based upon a complex mixture of inheritance, luck and effort, probably in that order of relative importance. What is the ideal distribution from the standpoint of absolute ethics may be disputed, but of the three considerations named certainly none but the effort can have ethical validity »[2].

Deuxième contre-argument : la distribution suivant la productivité marginale augmente la rémunération d'un facteur quand l'autre facteur est abondant et diminue sa rémunération lorsque lui-même est abondant. Le travailleur n'a pourtant aucun mérite à ce que le capital soit abondant, pas plus qu'il ne démérite si la natalité sature le marché du travail. Clark se contredit donc.

Outre ces critiques adressées aux conclusions que Clark tire sur le plan social, on peut mentionner une faiblesse de son modèle économique. Comme l'indique le graphique 3.7, Clark présume que l'addition des rémunérations basées sur la productivité marginale égale exactement le produit. Comme nous le verrons plus loin, il reviendra à Wicksell d'exposer les conditions de validité de cette hypothèse.

La rémunération des facteurs dépend de leur productivité marginale et celle-ci dépend de la quantité des facteurs entrant dans la production. Comment se détermine ce volume ? Dans son ouvrage de 1899, Clark a tendance à considérer que la quantité utilisée des facteurs est tout simplement donnée (d'où nos indices k et n). Cette façon de voir a également été critiquée. A la suite de Marshall, la plupart des économistes néoclassiques considéreront que le prix d'un

[1] Clark [51] p. 313.
[2] Knight [190] p. 598.

facteur se détermine par l'offre et la demande sur son marché ; la théorie de la productivité n'explique que la demande du facteur ; elle doit donc être complétée par une théorie de l'offre du facteur. Nous y reviendrons.

3.2.2. Böhm Bawerk et la théorie autrichienne du capital

Eugen von Böhm Bawerk est le représentant le plus illustre de l'école autrichienne, dont Menger est le fondateur et qui restera active jusque dans l'entre-deux-guerres avec Friedrich von Hayek et Ludwig von Mises. Sur le plan social et politique, elle se caractérise par l'attachement au laissez faire.

Dès 1884, Böhm Bawerk avait publié une étude critiquant toutes les théories antérieures de l'intérêt, des plus obscures aux plus connues comme la théorie marxiste de la plus-value. En 1889, il se mouille lui-même en publiant sa propre théorie dans l'ouvrage « Positive Theory of Capital » que nous analyserons à présent. Le lecteur doit savoir que cet ouvrage énonce une des idées les plus importantes de la science économique : la préférence pour les biens présents sur les biens futurs, comme source de l'intérêt sur le capital[1]. Cette thèse sera développée plus tard, suivant des voies différentes, par Wicksell et Fisher, deux des plus grands économistes néoclassiques (cf. sous-chapitres 3.2.5 et 3.2.7) et elle sert encore de base aux théories contemporaines de l'intérêt.

LA PRODUCTION ET LE CAPITAL

Böhm Bawerk aborde la production de façon inverse à Clark, pour qui le capital et le travail étaient les deux facteurs de production. Selon Böhm Bawerk, toute production se résout en dernier ressort à deux et seulement deux ingrédients : le travail et la terre. Böhm Bawerk donne l'exemple suivant : si d'abord, je fabrique des briques avec de la terre, puis je fais du ciment et ensuite avec ces deux ingrédients, je construis ma maison : « Is there any part of the wall of which it can be said that it was not made by me and the forces of nature ? »[2]. Le but ultime de toute production est la création d'une utilité consommable. On peut obtenir cette utilité, soit par la production DIRECTE qui y mène immédiatement soit par la production INDIRECTE dont le résultat est un bien intermédiaire qu'un ou plusieurs actes de production ultérieurs transformeront en utilité consommable[3]. Le capital n'est rien d'autre que le stock de biens intermédiaires, dont la société dispose à un moment donné.

Vu sous cet angle, ce qu'on appelle le capital n'est qu'un ensemble de biens de consommation qui ne sont pas encore parvenus à maturité, qui sont encore à un stade inachevé dans le circuit de la production. Ceci est valable pour le capital circulant comme pour le capital fixe, même si c'est moins évident pour ce dernier. Le capital est le produit en devenir. Il n'est donc pas un facteur de production, le troisième en compagnie du travail et de la terre (souvent appelés *facteurs primaires*), comme l'ont cru erronément beaucoup d'économistes. Bien comprendre ce point de vue est plus aisé si on se place au niveau de l'économie globale pour avoir une vue qui transcende la division du travail.

Un bien peut être produit suivant toute une diversité de techniques. Selon Böhm Bawerk, la caractéristique majeure d'une technique de production est d'être plus ou moins *détournée*. Elle sera plus détournée si la succession des phases où les travailleurs façonnent des biens intermédiaires est plus longue. Autrement dit, si une part importante du travail et de la terre est

[1] L'idée avait été émise dès 1834 par l'économiste écossais John Rae.
[2] Böhm Bawerk [374] p. 98
[3] Il est évident que la production directe est surtout présente dans les sociétés primitives.

absorbée par des tâches qui ne débouchent pas IMMEDIATEMENT sur le produit consommable. Comme les biens intermédiaires sont la substance du capital, une production plus détournée est aussi une production plus intense en capital.

Une production plus détournée a un avantage : elle est plus productive. Et un désavantage : plus de temps s'écoule entre le moment où les facteurs terre et travail commencent à être appliqués et la fin du processus quand le bien final est consommable. Les pêcheurs qui utilisent des bateaux avec de grands filets obtiendront plus de poisson que ceux qui se contentent juste de petits filets, mais le travail de fabrication des bateaux et filets ne donne à manger que beaucoup plus tard qu'il ne commence. Cette attente est la clé du raisonnement de Böhm Bawerk.

Il y a donc toute une gradation dans les détours de la production qui caractérisent une économie. Böhm Bawerk veut disposer d'un outil qui mesure le niveau capitalistique de l'économie ; c'est à cette fin qu'il introduit sa fameuse *période de production*. La période de production ABSOLUE est la durée qui s'écoule entre la première application de facteurs primaires jusqu'à ce que le bien soit prêt à être consommé. Ce concept souffre du grave défaut qu'elle peut facilement atteindre plusieurs siècles. Pensons au creusement des galeries de la mine d'où provient le métal d'un couteau.

Böhm Bawerk résout cette difficulté en préférant le concept de période MOYENNE de production. Donnons en la formule à l'aide d'un exemple. Pour une production disponible en l'année *t*, il faut appliquer 15 jours de travail en l'année t-10, 20 jours en l'année t-2 et 10 jours en l'année *t*. La durée moyenne d'attente de ces 45 journées de travail égale 4,22 années, soit le résultat du calcul :

$$(10x15 + 2x20 + 0x10)/45.$$

Normalement, les applications de travail très reculées ne représentent qu'une part infime du travail total : le coefficient de pondération les neutralisera quasiment dans le calcul. A noter que si le travail est appliqué de façon uniforme et constante, la période moyenne vaudra exactement la moitié de la période absolue, comme l'avait déjà montré Jevons. A titre de simplification, Böhm Bawerk envisage toujours ce cas dans les exemples chiffrés de son ouvrage.

Böhm Bawerk envisage la productivité du facteur temps de façon très radicale. « It could be confidently maintained that there is no area of production which could not materially increase its product by its present method. It would require no new invention, simply the intercalation of some familiar capitalist intermediate factor, such as an engine here a well-devised transmission there, an ingenious regearing, a blast mechanism, a lever, a governor or the like »[1]. Böhm Bawerk reconnaît qu'il peut y avoir des exceptions[2].

Böhm Bawerk montre que l'épargne et la production indirecte sont intimement liées. Si une tribu primitive consacre une partie de son temps à fabriquer des armes de chasse plutôt que de procéder à la cueillette des fruits sauvages, les fruits non cueillis et donc non consommés pendant la fabrication des armes (qui ne nourrit pas immédiatement) constituent une épargne.

LA PREFERENCE POUR LE PRESENT

Dans l'échange, nous acquérons et vendons des biens, les uns utilisables immédiatement, les autres disponibles pour utilisation seulement à une date ultérieure. Tous font l'objet d'une

[1] Böhm Bawerk [374] p. 85

[2] Beaucoup de commentateurs, principalement I. Fisher, pensent que ce processus a une limite, qu'à partir d'une certaine durée, l'allongement de la production n'augmente plus le produit, voire le diminue.

valorisation selon l'utilité marginale, mais pour les biens futurs, il s'agit de l'utilité marginale actuelle, non celle inconnue, qu'ils auront le moment venu. Cette utilité est certes influencée par notre anticipation des circonstances futures (pénurie ou abondance), mais également, dans notre subjectivité, l'attente affecte l'utilité marginale actuelle. Concernant les valeurs respectives des biens présents et futurs, Böhm Bawerk écrit : « Present goods are as a general rule worth more than future goods of equal quality and quantity. That sentence is the nub and the kernel of the theory of interest I have to present »[1]. Böhm Bawerk attribue trois causes à cette préférence pour les biens présents :

1- L'abondance plus grande de biens dans le futur.

L'agent, qui prévoit une amélioration de sa situation économique, attribue une utilité marginale moindre à ces biens disponibles en plus grande quantité. Böhm Bawerk reconnaît que le cas inverse n'est pas rare, celui de personnes dont la vieillesse rimera avec baisse des revenus. Mais de nombreux biens sont durables : des biens présents peuvent donc servir à la fois au présent et dans le futur, alors que les biens futurs voient leur utilité restreinte au futur. Certes, cet argument ne s'étend pas aux biens périssables.

2- La sous-estimation du futur.

« We feel less concerned about future sensations of joy and sorrow simply because they do lie in the future »[2]. Böhm Bawerk met en cause la connaissance insuffisante des besoins futurs, un manque de volonté et l'incertitude quant à la durée qui nous reste à vivre. Nous ne sommes pas sûrs de profiter du revenu que nous ne consommons pas immédiatement. Il reconnaît que chez certains individus, ceux qui sont sujets à l'avarice, le comportement est exactement l'inverse, mais ils constituent l'exception.

3- La supériorité technique des biens présents.

Comme nous l'avons vu, l'allongement de la période de production accroît la productivité. Böhm Bawerk en conclut que, commencée plus tôt, la production permet d'arriver à un résultat supérieur.

Le discours de Böhm Bawerk sur les trois causes a fait couler beaucoup d'encre. De multiples critiques ont été émises. Voyons ici les commentaires de Wicksell et Fisher, tous deux acquis au principe de la préférence pour le présent, mais réticents vis-à-vis de la justification avancée par Böhm Bawerk.

Fisher accepte les causes un et deux mais rejette fermement la troisième. Le tableau ci-dessous que Fisher reprend de Böhm Bawerk était censé montrer par un exemple que le maximum de valeur qu'un mois de travail (identique) est capable de produire diminue selon qu'il est disponible l'année 1889, 1890, 1891 ou 1892. Ce maximum est donné par le montant souligné dans la dernière colonne de chaque sous-tableau. La deuxième colonne indique la quantité produite par ce travail selon l'année où on en recueille le fruit. La troisième colonne montre la valeur instantanée de ce produit et la quatrième la valeur présente. La cause trois se manifeste par la croissance de la deuxième colonne. La valeur totale s'obtient par la multiplication de la deuxième colonne par la quatrième.

[1] Böhm Bawerk [374] p. 259
[2] Böhm Bawerk [374] p. 268

A MONTH'S LABOR AVAILABLE IN 1888 YIELDS

For the Economic Period	Units of Product	True Marginal Utility of Unit	Marginal Utility Reduced to Perspective	Amount of Value of Entire Product
1888	100	5	5	500
1889	200	4	3.8	760
1890	280	3.3	3	840
1891	350	2.5	2.2	770
1892	400	2.2	2	800
1893	440	2.1	1.8	792
1894	470	2	1.5	705
1895	500	1.5	1	500

A MONTH'S LABOR AVAILABLE IN 1890 YIELDS

For Economic Period	Units	True Marginal Utility	Reduced Marginal Utility	Value
1888	—	5	5	—
1889	—	4	3.8	—
1890	100	3.3	3	300
1891	200	2.5	2.2	440
1892	280	2.2	2	560
1893	350	2.1	1.8	630
1894	400	2	1.5	600
1895	440	1.5	1	440

A MONTH'S LABOR AVAILABLE IN 1889 YIELDS

For Economic Period	Units	True Marginal Utility	Reduced Marginal Utility	Value
1888	—	5	5	—
1889	100	4	3.8	380
1890	200	3.3	3	600
1891	280	2.5	2.2	616
1892	350	2.2	2	700
1893	400	2.1	1.8	720
1894	440	2	1.5	660
1895	470	1.5	1	470

A MONTH'S LABOR AVAILABLE IN 1891 YIELDS

For Economic Period	Units	True Marginal Utility	Reduced Marginal Utility	Value
1888	—	5	5	—
1889	—	4	3.8	—
1890	—	3.3	3	—
1891	100	2.5	2.2	220
1892	200	2.2	2	400
1893	280	2.1	1.8	504
1894	350	2	1.5	525
1895	400	1.5	1	400

D'après Fisher les troisième et quatrièmes colonnes sont assimilables respectivement à la cause un et la cause deux. Si la troisième cause est valide, c'est d'elle que dépend la décroissance des maximums. Or Fisher montre que si la deuxième colonne était constante, si par exemple elle restait à 100 à chaque ligne, la décroissance des maximums se produirait quand même. Par contre, les maximums ne diminueraient pas si les troisième et quatrième colonnes étaient constantes ; ils seraient juste décalés. Le résultat est donc imputable aux causes un et deux. La cause trois est inopérante.

La troisième cause ne convainc pas non plus Wicksell. Celui-ci se montre également sceptique envers la première. Il se méfie d'une cause qui serait inopérante dans une économie stationnaire. De plus, peu de biens peuvent être conservés sur une longue période sans risque et sans frais.

Fisher et Wicksell ne sont donc d'accord ensemble qu'avec la deuxième cause. Mais celle-ci est combattue par Menger et Wieser, les collègues autrichiens de Böhm Bawerk. Ils l'estiment en opposition avec le postulat essentiel de l'économie néoclassique, la rationalité des agents économiques. Néanmoins, comme le remarque Blaug, cet argument ne démontre pas que la deuxième cause est invalide.

L'ORIGINE DE L'INTÉRÊT

« The natural difference in value between present and future goods, the existence of which I have set down in the preceding chapter, is the fountainhead from which all interest takes its origin »[1]. Il en va ainsi dans les deux principales situations donnant lieu à l'intérêt : le prêt à la consommation et le capital investi dans les entreprises ; le phénomène est plus patent dans la première : l'échange entre le bien présent et le bien futur est clairement identifié. Par exemple, €1000 prêtés aujourd'hui contre €1050 remboursés dans un an est un échange équilibré, puisque les euros présents valent plus. L'intérêt est l'agio compris entre ces deux prix.

Mais quel échange entre biens présents et biens futurs produit-il l'intérêt que rapporte le capital investi dans les entreprises ? Le principe est simple : puisque leur utilité n'est pas consommable

[1] Böhm Bawerk [374] p. 290

immédiatement, les biens intermédiaires et les facteurs primaires, la terre et le travail, fonctionnent comme des biens futurs et font donc l'objet d'une décote vis-à-vis des biens consommables immédiatement. Voyons-le avec un exemple très simplifié[1]. Mettons que 10 heures de travail actuel de l'ouvrier agricole produisent 100 kilos de blé consommables dans un an. Si par suite de la préférence pour le présent, 100 kilos de blé dans un an sont évalués à 95 kilos de blé aujourd'hui, le salaire des 10 heures de travail se montera à 95 kilos de blé. Mais dans un an, en vendant 100 kilos de blé, l'entrepreneur empochera à ce moment un surplus valant cinq kilos de blé. C'est cela l'origine de l'intérêt : la différence entre la valeur présente du blé futur qu'est le travail et la valeur du bien présent que ce blé deviendra dans un an : « For his future good gradually ripens into a present good during the course of the production process, and thus grows into possession of the full value of a present good… And that increase in value is originary interest »[2].

Le marché du travail n'est d'ailleurs qu'une manifestation d'un marché plus large et plus abstrait que Böhm Bawerk nomme *marché des biens de subsistance*. La demande y vient des salariés qui doivent recevoir pendant toute la période de production les biens de consommation nécessaires à leur subsistance. L'offre est assurée par les capitalistes et provient du capital qu'ils ont accumulé[3]. Certes, les capitalistes n'investissent pas des biens de consommation dans leur entreprise, mais le capital joue indirectement le rôle *de fonds de subsistance*. La dispersion de la production dans de nombreuses entreprises dont la production n'est pas synchronisée a pour résultat que les travailleurs trouvent en permanence les biens de consommation désirés. Plus grand est le fonds de subsistance, plus longue est la période moyenne de production qu'une société peut se permettre.

Quelles sont les motivations des offreurs et demandeurs sur le marché des biens de subsistance ? Concernant les capitalistes, Böhm Bawerk émet l'hypothèse, assez étonnante compte tenu de ce qui précède, qu'ils mettent sur pied d'égalité les biens présents (subsistance) qu'ils offrent et le bien futur (travail) qu'ils demandent. S'il n'en tenait qu'à eux, il n'y aurait pas d'agio dans l'échange. Par contre, les salariés, s'ils devaient travailler sans la subsistance offerte par les capitalistes, ne pourraient se permettre qu'une période de production très courte, donc peu productive (économie primitive). Il est plus avantageux pour eux d'accepter les biens de subsistance des capitalistes avec un agio que de consommer le fruit d'une production propre très restreinte.

Contrairement à la thèse marxiste, il n'y a donc pas d'exploitation du travail : le travail est acheté à sa valeur en tant que bien futur. Le produit du travail sera vendu à sa valeur en tant que bien présent. Aucune trace de spoliation n'est décelable dans ces échanges. L'intérêt est légitime. Böhm Bawerk ajoute : « The difference in value between present and future goods is a basic phenomenon that is independent of all human arrangements »[4]. Même dans une communauté socialiste, l'intérêt continuerait à exister sous une forme ou sous une autre, car l'Etat ne pourrait attribuer aux biens futurs une valeur aussi élevée qu'aux biens présents ; cela conduirait à une période de production excessivement longue car sans frein naturel, et les

[1] Cet exemple est simplificateur, car il ne considère que le capital circulant et un seul facteur primaire, le travail. Mais Böhm Bawerk consacre une section de l'ouvrage à la genèse de l'intérêt pour les biens durables, tant des biens de consommation que ceux qui servent comme capital fixe. La théorie repose sur la distinction entre la valeur des biens durables et celle des services qu'ils rendent. Ces services sont étalés sur plusieurs années et plus ils sont reculés, moins ils sont valorisés. La théorie explique le lien entre ces valeurs et le taux d'intérêt, mais la place nous manque pour l'exposer complètement.

[2] Böhm Bawerk [374] p. 301.

[3] Dans tout le raisonnement, Böhm Bawerk assimile le capitaliste à l'entrepreneur.

[4] Böhm Bawerk [374] p. 342

travailleurs en subiraient des privations excessives. Böhm Bawerk revient régulièrement avec cette idée, très discutée, que sans la préférence pour le présent et plus spécifiquement les deux premières causes, la société privilégierait une période de production tendant vers l'infini, vu la croissance de la productivité qui en résulterait.

LA DETERMINATION DU TAUX D'INTERET

Le modèle bawerkien de la détermination du taux d'intérêt est en fait une une théorie de la répartition. Dans le modèle simplifié à deux biens (le travail et un bien de consommation), la détermination du taux d'intérêt et celle du salaire ne font qu'une, car sur le marché des biens de subsistance, l'offre de travail est une demande de capital (biens présents) et l'offre de capital est une demande de travail (biens futurs).

Menger et Clark expliquaient toute rétribution des agents économiques par la règle de la productivité marginale. Böhm Bawerk applique cette règle aux facteurs primaires mais non au capital.

Comme nous l'avons vu, l'allongement de la période de production entraîne nécessairement l'accroissement de la productivité, mais d'après Böhm Bawerk, cette hausse est elle-même en décélération. La productivité marginale de l'accroissement de la période de production est donc décroissante, comme la productivité marginale du capital chez Clark. Böhm Bawerk travaille avec un exemple chiffré dont le tableau ci-dessous reproduit les données[1].

WAGE RATE $2,500

Production Period In Years	*Annual Product Per Worker*	*Annual Gain Per Worker*	*Number Of Workers Employed*	*Total Annual Gain Per $50,000 Of Capital*
1	$1750	—$ 750	40	Loss
2	2250	— 250	20	Loss
3	2650	150	13.33	$2,000.00
4	2900	400	10	4,000.00
5	3100	600	8	4,800.00
6	3250	750	6.66	**5,000.00**
7	3350	850	5.71	4,857.14
8	3425	925	5	4,625.00
9	3475	975	4.44	4,333.33
10	3500	1000	4	4,000.00

Lorsque la période de production passe de un an à deux ans, le produit par travailleur augmente de 1750 à 2250, soit un écart de 500. De deux à trois ans, l'écart n'est plus que de 400 et le lecteur peut vérifier que la baisse continue. Böhm Bawerk considère cette baisse comme une évidence, qu'il ne se donne donc pas la peine d'argumenter.

Dans l'optique de Böhm Bawerk, les données exogènes sont :
- l'offre de capital (le fonds de subsistance)
- l'offre de travail (le nombre de travailleurs)
- l'évolution de la productivité (cf. tableau ci-dessus).

Il faut déterminer le salaire et la période de production et, sur base de ceux-ci, le taux d'intérêt. Pour raison didactique, Böhm Bawerk pose d'abord que le salaire est connu et fixé à 2.500 par an. La période de production optimale, est celle qui maximise le rendement du capital (fixé à 50.000) en cinquième colonne. Dans notre exemple, elle est de 6 ans. Chaque travailleur rapporte alors 750 en plus que les 2500 qu'il coûte annuellement. L'effectif que le capital peut

[1] Böhm Bawerk [374] p. 358. La première colonne indique la période de production ABSOLUE.

payer dans ces conditions est donné par la fraction 50.000 / (0,5 x 2.500 x 6), dont le dénominateur égale le salaire total payé pendant 6 ans, divisé par 2, compte tenu de ce qui fut dit plus haut sur le rapport entre la période de production absolue et la période de production moyenne dans le cas d'une application de travail constante. Cela donne 6,66 travailleurs rapportant chacun 750. Le rapport annuel est de 5.000, ce qui correspond à un taux d'intérêt de 10%.

Si le salaire avait été supérieur à 2.500, la période de production optimale eût été plus longue avec un effectif inférieur et inversement s'il avait été inférieur. Dans l'économie réelle, c'est le nombre de travailleurs qui est fixé ; il revient ainsi au salaire et à la période de production de s'adapter (de concert) pour réaliser l'équilibre. Le salaire d'équilibre sera celui qui, à la fois :

- correspond à la période de production optimale (comme ci-dessus)
- éponge exactement l'offre de capital et celle de travail, suivant le principe qu'aucun travailleur et aucune unité de capital ne peuvent rester inemployés. Si le salaire est plus élevé que ce niveau d'équilibre, les travailleurs inoccupés feront pression pour qu'il baisse. Dans le cas contraire, le capital inoccupé fera pression pour qu'il hausse.

Parmi les multiples combinaisons entre le salaire et la période de production, une seule satisfait ces deux conditions.

Böhm Bawerk montre, comme Jevons avant lui, que l'intérêt égale la productivité relative du dernier allongement de la période de production rendu possible par le fonds de subsistance. Pour Clark, l'intérêt était la productivité marginale du capital. Pour Böhm Bawerk, dans cette proposition, il faut remplacer CAPITAL par ATTENTE.

Quels facteurs influencent-ils le taux d'intérêt ? « The interest rate in a given economy will rise in inverse ratio to the subsistence fund, in direct ratio to the working population which that fund must support and in direct ratio to the degree of productivity that marks continue prolongation of the production period »[1].

La théorie de Böhm Bawerk est une construction impressionnante mais pas sans faille :

1- La productivité de Böhm Bawerk connaît une croissance vers une asymptote horizontale, lorsque la période de production tend vers l'infini. Selon Fisher, après quelques années de croissance, la productivité devrait atteindre un maximum puis se réduire.

2- L'explication sur la genèse de l'intérêt dans la production et la détermination de son taux semble pouvoir se passer complètement des trois causes de la préférence pour le présent[2]. La productivité de l'allongement de la période de production joue un rôle essentiel, mais comme l'a montré Fisher, la préférence pour le présent est indépendante de ce facteur. Beaucoup de commentateurs, dont Fisher et Wicksell, s'accordent à considérer que l'explication de l'intérêt par Böhm Bawerk se ramène en définitive à la productivité du capital, une explication à laquelle, paradoxalement, il ne cesse pas d'opposer la sienne.

   Certes, les travailleurs préfèrent les biens de subsistance offerts par les capitalistes à ceux, moins abondants, qu'ils pourraient produire eux-mêmes sans capital. Ce fait n'est manifestement pas explicable par la deuxième cause ; il a des similitudes avec la première cause. Pourtant l'avantage de travailler avec du capital est valable même dans une économie

[1] Böhm Bawerk [374] p. 365

[2] Wicksell estime que l'exposé de Böhm Bawerk se serait avantageusement passé de ce chapitre. Selon lui, un discours sur l'origine de l'intérêt, indépendant de la détermination du taux d'intérêt, n'a pas de sens.

stationnaire, alors que la première cause un implique nécessairement une expansion de la consommation.

3- Comme le fait remarquer Wicksell, le modèle de Böhm Bawerk ignore l'intérêt composé[1]. Sans cette simplification, la période de production ne serait pas indépendante du taux d'intérêt et ne pourrait donc pas servir à le déterminer.

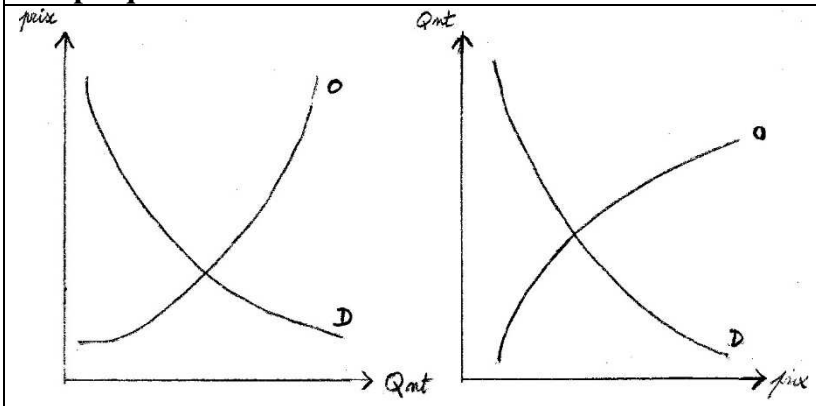
3.2.3. Marshall : l'offre et la demande

Alfred Marshall, qui était établi à l'université de Cambridge, a dominé la science économique en Angleterre jusqu'à sa mort. En 1890 parut son ouvrage « Principles of Economics ».

LES COURBES DE L'OFFRE ET DE LA DEMANDE

La figure 3.8-A est un standard de la science économique, connu de toute personne ayant été en contact même superficiellement avec cette discipline. On y voit une courbe indiquant que l'offre d'un produit est plus abondante lorsque le prix est plus élevé et une autre courbe montrant que la demande est plus abondante lorsque le prix est plus bas. Le prix du marché est celui qui égalise l'offre et la demande, ce qui se traduit graphiquement par l'intersection des deux courbes. Marshall est le père de ce graphique.

Graphique 3.8 : les courbes d'offre et de demande



La figure de gauche indique comment les courbes sont dessinées depuis Marshall ; le graphique de droite, comment il faudrait logiquement les dessiner.

Marshall n'est pas le premier à avoir dessiné les courbes d'offre et de demande ; Cournot et Walras le firent avant lui. Mais Walras les dessina comme sur la figure 3.8-B. Les deux graphiques se ressemblent, mais l'abscisse et l'ordonnée sont inversées. Est-ce important ?

Chez Walras, la causalité allait nettement du prix vers la quantité offerte ou demandée. L'inversion graphique des coordonnées, signifie-t-elle que Marshall inverse le sens de la causalité ? Ce n'est pas certain. Il se montre assez ambigu sur la question. Toujours est-il que la version graphique de Marshall s'est imposée, ce qui est paradoxal et même illogique puisqu'en même temps la causalité walrassienne est unanimement acceptée. La variable dépendante se trouve normalement sur l'ordonnée et la variable indépendante sur l'abscisse.

Ce graphique est représentatif de la méthode de l'*équilibre partiel*, à laquelle le nom de Marshall est attaché. Cette appellation se justifie par le fait que le graphique 3.8 (*A* autant que

[1] Cf. la théorie de capital de Jevons.

B) isole du reste de l'économie, le marché du bien dont l'offre et la demande sont analysées. On oppose souvent l'équilibre PARTIEL marshallien et l'équilibre GENERAL walrassien.

Walras, plus scrupuleux que Marshall, montre que d'un point de vue rigoureusement scientifique, ce graphique n'a que valeur d'approximation. Appelons *X* le produit faisant l'objet du graphique 3.8, avec à l'équilibre le prix p_X et la quantité q_X. Une variation de l'offre ou de la demande de *X* affectera directement p_X, mais modifiera également l'équilibre d'autres marchés ; en retour, ces variations entraîneront le déplacement des courbes d'offre et de demande de *X*. Le graphique 3.8 ne peut embrasser toute la complexité de la problématique.

L'UTILITE ET LA DEMANDE

Marshall rappelle la règle à suivre pour que l'agent maximise l'utilité qu'il tire de son revenu : « marginal utility of a shillings' worth of goods on each line of expenditure shall be the same. »[1]. Si l'utilité du dernier shilling dépensé pour les boissons dépasse celle du dernier shilling dépensé pour les vêtements, l'agent gagnerait à dépenser moins en vêtements (ce qui réduirait peu son utilité) et plus en boissons (ce qui l'augmenterait plus). Cette règle exprime avec d'autres mots l'équation (3.5) de Walras.

La quantité demandée diminue quand le prix augmente, toutes autres choses restant égales. La courbe de demande a donc une pente négative ; cela paraît un truisme et Cournot l'avait conçu ainsi. Au contraire, Marshall, comme Walras, fait dépendre la forme de la courbe de demande de celle de l'utilité marginale. La demande du marché n'est que la somme des demandes individuelles ; la demande d'un bien par un individu diminue lorsque le prix de ce bien augmente, tout simplement parce que l'utilité marginale du bien est décroissante.

Son raisonnement est plus simple et moins rigoureux que celui de Walras. Marshall donne l'exemple d'un amateur de thé qui adapte sa consommation au prix unitaire. Supposons que sa courbe de demande soit ainsi :

Prix en shillings	20	14	10	6	4	3	2
Quantité en livres	1	2	3	4	5	6	7

Marshall assimile l'utilité particulière de chaque unité consommée au prix qu'il veut bien offrir pour l'obtenir. Par exemple, la première unité a une utilité de 20, la deuxième de 14…

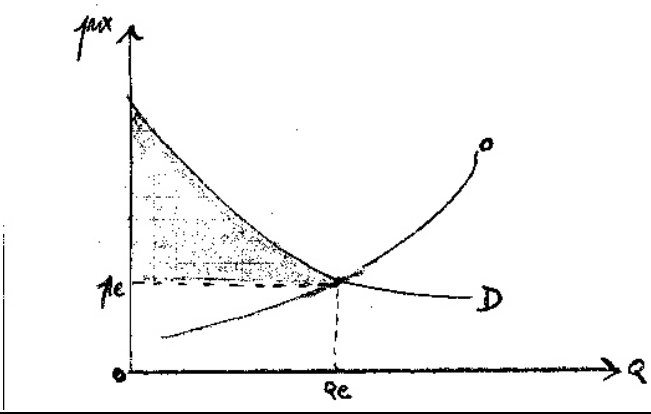
Marshall fut bien obligé de reconnaître une faiblesse de ce raisonnement : l'utilité marginale du revenu décroît comme celle des biens. Le dernier shilling d'un riche qui en possède beaucoup aura moins d'utilité que celui d'un pauvre. Or justement, lorsque le prix d'une livre de thé passe de 2 à 10, notre amateur de thé s'appauvrit, si le revenu NOMINAL reste constant. Chacun des 10 shillings aura donc individuellement plus de valeur que chaque shilling du groupe de deux. La troisième livre de thé semble avoir cinq fois plus d'utilité que la septième. En réalité, on ne peut pas tirer cette conclusion, car les shillings dans les deux cas n'ont pas la même utilité marginale. Marshall s'en tient toutefois à son raisonnement, estimant un peu présomptueusement que dans la grande majorité des cas, les différences d'utilité marginale de la monnaie sont négligeables.

Marshall pose la question de savoir « how far the price which is actually paid for a thing represents the benefit that arises from its possession ». Dans l'exemple ci-dessus, imaginons que le prix du marché s'établisse à 3s. Notre consommateur achètera 6 livres de thé pour

[1] Marshall [253] § 3.5.7

lesquelles il déboursera 18s. Mais l'utilité totale qu'il en retirera se monte à 20+14+10+6+4+3 = 57s. Le consommateur bénéficie donc gratuitement d'un excédent d'utilité valant 39s. Cet excédent d'utilité est appelé le *surplus du consommateur*. Sur le graphique 3.9, il correspond à la surface hachurée. Le concept de surplus du consommateur tel qu'il est présenté ici est évidemment entaché de la faiblesse mentionnée ci-dessus quant à l'utilité marginale de la monnaie dépensée.

Graphique 3.9 : le surplus du consommateur



L'un des apports les plus importants de Marshall est l'*élasticité de la demande*, une notion devenue essentielle dans la science économique[1]. « The elasticity (or responsiveness) of demand in a market is great or small according as the amount demanded increases much or little for a given fall in price and diminishes much or little for a given rise in price »[2]. La formule de l'élasticité (notée ε) est :

$$\varepsilon = (-dQ/Q) / (dP/P) \qquad (3.25)$$[3]

Le numérateur est la variation de la quantité demandée et le dénominateur est la variation du prix, à la marge. Mais attention, il s'agit dans les deux cas de la variation RELATIVE (dQ/Q, dP/P) et non de la variation absolue (dQ, dP). Par exemple, si une hausse du prix de 2% entraîne une baisse de la quantité de 4%, l'élasticité sera de 2.

L'élasticité peut être calculée en chaque point de la courbe et variera souvent (mais pas nécessairement) d'un point à l'autre d'une même courbe. Les courbes B, C et D du graphique 3.10 montrent des courbes dont l'élasticité est constante et vaut respectivement l'unité, l'infini et zéro[4]. Marshall attire l'attention sur la figure 3.10-B : les courbes de demande sont des hyperboles équilatères dont la fonction est Q.P = C où *C* est une constante. Autrement dit, pour une courbe pareille, la dépense totale (Q.P) est constante, quel que soit le prix.

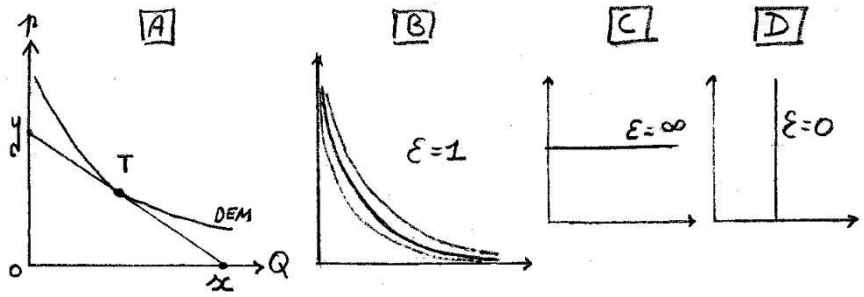
[1] Marshall ne s'est intéressé qu'à la demande mais le concept d'élasticité est aussi couramment appliqué à l'offre.

[2] Marshall [253] § 3.4.2. Il semble se rallier ici à la causalité walrassienne mais ce n'est pas toujours aussi évident.

[3] Le signe de dQ est inversé pour neutraliser l'opposition de signes entre dQ et dP.

[4] Mentionnons deux courbes particulières : la droite horizontale, dite demande *parfaitement élastiques* où ε = ∞ en tous points (figure C) et la droite verticale, dite demande *parfaitement inélastiques* où ε = o en tous points (figure D). La verticalité de la courbe parfaitement inélastique montre bien l'absence de réaction de la quantité.

Graphique 3.10 : l'élasticité de la demande



La figure A représente l'élasticité au point T, à l'aide de la tangente à la courbe de demande en ce point ; elle vaut le rapport Tx/Ty. Les figures B,C et D représentent des cas particuliers.

Marshall examine ensuite les facteurs qui influencent l'élasticité de la demande :

- La part du bien dans le budget du consommateur : concernant le sel, par exemple, « it is doubtful that any fall of price would induce a considerable increase in the consumption »[1]
- Dans la classe peu aisée, des produits comme la viande peuvent avoir une élasticité assez élevée, mais dans les classes nanties, ce type de produit aura une faible élasticité.
- L'existence de substituts : si le prix du bœuf augmente, sa consommation se réduira d'autant plus fort que les consommateurs pourront se rabattre sur des produits comparables comme le mouton par exemple.

LA PRODUCTION ET L'EQUILIBRE DU MARCHE

Marshall est le premier économiste à avoir étudié en détail la question des rendements d'échelle, c'est-à-dire l'influence de l'échelle de la production sur le coût. Les facteurs qui réduisent le coût de la production à plus grande échelle sont appelés *économies d'échelle* ; ceux qui accroissent le coût *déséconomies d'échelle*. Si les économies d'échelle l'emportent sur les déséconomies, on parle de *rendements croissants* et dans le cas contraire, on a affaire à des *rendements décroissants* ; lorsqu'elles se compensent, les rendements sont constants.

Concernant les rendements d'échelle, Marshall distingue les *économies INTERNES* que la firme recueille grâce à sa grande taille et les *économies EXTERNES* qui sont des avantages dont elle profite, qui résultent de la production à grande échelle dans la branche ou dans la région où la firme opère. Les économies externes sont souvent liées à la localisation surtout quand une région devient un pôle de développement de telle ou telle industrie : la main d'œuvre qualifiée y abonde, les moyens de communication sont denses, des revues spécialisées diffusent la connaissance technique et commerciale…

Les avantages de la grande taille des entreprises sont d'ordre technique, commercial ou financier. Les machines les plus performantes ne sont généralement rentables qu'au-delà d'un seuil de production ; la division du travail peut être optimisée avec un effectif de main d'œuvre plus important… Mais la petite taille a également des avantages : proximité du contrôle sur le travail, évitement des problèmes de communication entre départements. D'après Marshall, certaines branches sont propices aux rendements croissants et à la grande taille, d'autres aux rendements décroissants et à la petite taille. Les rendements décroissants se rencontreront surtout dans les secteurs où le facteur « nature » joue un rôle important (agriculture, mines). Les rendements croissants seront principalement présents dans les secteurs où un bien homogène peut être produit en très grande quantité ou dans ceux où la publicité joue un rôle important pour étendre les possibilités de commercialisation.

[1] Marshall [253] § 3.4.6

D'une façon générale, les avantages des grandes firmes semblent suffisamment importants pour qu'on s'attende à une monopolisation tendancielle de l'économie. Or Marshall constate que cette tendance ne se réalise pas ; il l'explique de façon un peu légère : les entreprises connaîtraient un cycle de vie : créées par un entrepreneur énergique, elles passent d'abord par une phase de pleine vitalité. Mais les successeurs ne sont pas toujours à la hauteur du fondateur et la décadence devient inévitable après une ou deux générations. Si elles deviennent des sociétés anonymes, elles sont menacées par la bureaucratie et la sclérose. Au moment où elles pourraient décoller vers le monopole, le déclin commence et une nouvelle firme plus jeune s'impose sur le marché.

Pour analyser la relation entre la taille de l'entreprise et son coût de production, Marshall crée le concept de *firme représentative* qui lui sert à neutraliser toutes les circonstances spécifiques, notamment le fait que l'entreprise soit dans telle ou telle phase de son cycle de vie. Il s'agit d'une firme ayant déjà connu une certaine expansion mais gardant un potentiel de croissance, normalement bien gérée, ayant un accès normal aux économies internes et externes. C'est d'elle qu'il sera question lorsque nous approfondirons la relation entre le coût et l'offre.

Revenons aux deux lames de notre paire de ciseaux, l'offre et la demande comme déterminants de la valeur. Selon Marshall, la règle est la suivante : plus courte est la période considérée, plus l'influence de la demande prévaut ; plus longue elle est, plus l'offre sera déterminante[1]. Dans cette optique, Marshall définit trois périodes :

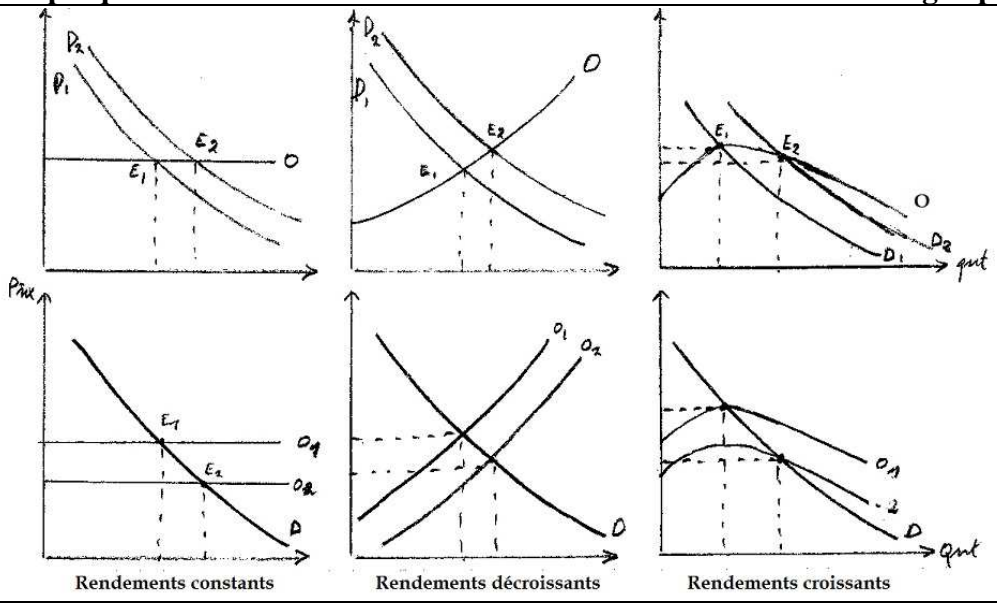
1- la <u>période de marché</u>, tellement courte que l'offre n'a pas le temps de s'ajuster à la demande. L'offre est fixe et le prix varie considérablement en fonction des variations de la demande au jour le jour.
2- le <u>court terme</u> : la firme représentative ne peut pas varier de taille mais, malgré les rigidités qui bloquent l'accroissement de certains facteurs de production, elle peut accroître sa production dans une certaine mesure grâce à d'autres facteurs dont la quantité est plus flexible ; cette production accrue se fera nécessairement dans des conditions non optimales quant à la proportion entre les facteurs.
3- le <u>long terme</u> : les firmes ont le temps d'ajuster la quantité de tous les facteurs, y compris ceux qui s'ajustent lentement.

Que l'on considère le court ou le long terme, le **prix d'offre** est celui qui est juste suffisant pour inciter l'industrie à produire la quantité demandée. Le coût à prendre en considération est le coût marginal, c'est-à-dire le coût de production de ces unités qui sont à la marge de ne pas être produites du tout et qui ne le seraient pas si le prix attendu était inférieur.

A court terme, une hausse de la production implique toujours une hausse du prix d'offre puisque l'accroissement de la production ne peut se faire qu'à coût croissant : soit parce que les conditions techniques ne sont pas optimales, soit parce qu'il faut accroître la rémunération accordée à des facteurs rares pour les attirer dans la production. Par contre à long terme, le prix d'offre intègre naturellement les rendements d'échelle, ce qui rend le tableau plus complexe. Divers cas sont concevables : le prix d'offre constant, croissant, décroissant ou même alternativement décroissant et croissant. Une augmentation de la quantité produite aura donc un effet sur le prix d'offre de longue période, qui est variable d'une industrie à l'autre.

[1] Cette idée rappelle celle de Smith selon laquelle le prix du marché oscille autour de la valeur déterminée par le coût de production, en fonction des aléas de l'offre et la demande.

Graphique 3.11 : effets d'une hausse de l'offre ou la demande en longue période



Marshall montre les effets d'une hausse de l'offre ou de la demande à long terme à l'aide du graphique 3.11. Les trois figures du haut se rapportent à une hausse de la demande, qui passe de D_1 à D_2. Quels que soient les rendements d'échelle, la quantité produite et échangée croît. Quant au prix, il ne varie pas si les rendements d'échelle sont constants, il augmente si les rendements sont décroissants et il diminue si les rendements sont croissants. Les trois figures du bas montrent les effets d'une hausse de l'offre, qui passe de O_1 à O_2. Quel que soit le type de rendements d'échelle, le prix diminue et la quantité échangée augmente.

La courbe d'offre en cas de rendements croissants pourrait intriguer le lecteur, peu habitué à voir une courbe d'offre décroissante ; en tout cas, elle a étonné Marshall lui-même. L'aspect le plus inquiétant du cas envisagé, est que la firme semble ne pas connaître de frein au désir de toujours augmenter le volume de sa production jusqu'à l'infini. La firme n'a pas de point d'équilibre. Pour s'extraire de cette difficulté, Marshall ressort sa théorie de la décadence inévitable des firmes ayant atteint une certaine taille. Rappelons qu'il s'agit ici de la courbe d'offre à long terme ; la courbe d'offre à court terme est toujours croissante comme l'est le coût lorsqu'il n'est pas possible d'optimiser les conditions de la production.

LES FACTEURS DE PRODUCTION

Marshall compte quatre facteurs de production : les trois traditionnels : la terre, le travail et le capital, auxquels il ajoute l'*organisation*, c'est-à-dire la fonction d'entrepreneur. Malheureusement, Marshall ne parvient pas à montrer en quoi l'organisation est plus qu'une variété de travail spécifique, celui de l'entrepreneur[1].

En quelle quantité les différents facteurs seront-ils utilisés et à quel taux seront-ils rémunérés ? L'entrepreneur règle en permanence la proportion des facteurs pour obtenir la combinaison qui minimise son coût ; c'est ce que Marshall appelle le *principe de substitution*. Dans cet exercice, tous les facteurs sont sujets à une loi des rendements décroissants qui correspond au fait qu'appliqué en excès, tout facteur voit son rendement décroître : « Excessive applications of any means to the attainment of any end are indeed sure to yield diminishing returns in every

[1] Cette conception sera critiquée par Wicksell qui estime que l'organisation manque de la précision quantitative pour être considérée comme un facteur.

branch of business ; and, one may say, in all the affairs of life »[1]. Marshall met en garde contre la confusion entre cette loi des rendements décroissants d'un facteur utilisé en mauvaise proportion avec les autres facteurs et les rendements d'échelle décroissants. La ressemblance n'est que de façade. La première loi, qui n'est autre que celle de la productivité marginale décroissante, joue sur le court terme, alors que la seconde joue sur le long terme.

L'entreprise optimise sa rentabilité lorsque chaque facteur est utilisé dans une proportion telle que le rapport entre son produit marginal et son prix est le même que celui de tous les autres facteurs. A l'équilibre, il y a indifférence dans le choix des facteurs ; le dernier shilling dépensé doit rapporter autant, quel que soit le facteur dans lequel il est investi. C'est une autre manière de présenter la solution déjà connue ; à l'équilibre, les deux doivent donc être égaux.

Chaque facteur continue à être demandé tant que sa productivité marginale dépasse son prix. La productivité marginale d'un facteur est le prix maximum que l'entrepreneur payera pour l'acquérir. On peut donc considérer que la courbe de productivité marginale du facteur dans une entreprise donnée est en même temps la courbe de demande de ce facteur par cette entreprise. En additionnant les demandes des différentes entreprises, on obtient la demande du marché, qui sera confrontée avec l'offre pour fixer le prix du facteur.

L'offre des facteurs de production dépend de leur abondance, mais également de la volonté des détenteurs de les investir dans la production. Marshall reprend le point de vue jevonnien basé sur l'abstinence[2] pour le capital et la désutilité pour le travail. Les courbes d'offre des facteurs ont une pente croissante parce qu'un salaire et un intérêt plus élevés incitent à plus de travail et plus d'épargne. Comme Jevons, Marshall reconnaît que la rémunération plus élevée peut également avoir l'effet inverse, mais ce cas lui semble plus rare.

Dans deux passages de son ouvrage, Marshall s'écarte un peu de la conception orthodoxe de la productivité marginale. Il l'associe à ce qu'il appelle le produit net marginal qu'on pourrait définir ainsi : supposons que la proportion entre les facteurs soit optimale en ce sens qu'elle permet la maximisation du profit de l'entreprise. Si celle-ci augmente d'une unité le volume de la main d'œuvre, le produit net de cet incrément est le supplément de profit qui adviendra si la quantité des autres facteurs est adaptée de façon à maximiser le profit avec le nouveau volume de la main d'œuvre. En réalité, Marshall est ambigu quant à son adhésion à ce principe. Blaug explique son irruption par la volonté de « réfuter la critique de **Hobson** selon laquelle la quantité utilisée d'un facteur ne peut pas varier sans que les quantités de tous les facteurs soient modifiées »[3]. Notre chapitre 3.4. est consacré à cet économiste hétérodoxe, avec qui Marshall et son école étaient en débat pendant des années.

3.2.4. Wicksteed et la fonction de production

Prêtre de l'Eglise unitarienne, théologien, spécialiste de Dante, intellectuel engagé dans les débats de société, l'économiste anglais Philip Wicksteed avait des sympathies pour le mouvement socialiste. Il convainquit beaucoup de Fabiens de se détourner de la théorie de la valeur-travail et les convertit au marginalisme, notamment Georges Bernard Shaw, avec qui il se liera d'amitié.

---

[1] Marshall [253] § 5.8.12

[2] Au terme « abstinence », Marshall préfère l'« attente », mais cette substitution terminologique laisse le fond inchangé : c'est le fait de différer une consommation qui est considéré comme un sacrifice.

[3] Blaug [35] p. 483.

Wicksteed se distingue dans l'explication des motifs qui guident les agents économiques. La conception dominante est celle de l'homo economicus, guidé par le seul motif de son intérêt individuel conçu dans une optique restrictivement hédoniste. Wicksteed estime au contraire qu'il n'y a pas une science des comportements économiques distincte de l'explication des comportements humains en général. « And when we are considering whether we will live in the country or the town, we may find on examination, that we are carefully equating increments and decrements of such apparently heterogeneous indulgences as those associated with fresh eggs and friendship »[1]. Toutes les satisfactions, égoïstes comme altruistes, sont prises dans le cadre d'un vaste choix unique qui égalise les satisfactions marginales.

LA FONCTION DE PRODUCTION

En 1894, Wicksteed publie un ouvrage intitulé « The Coordination of the Laws of Distribution ». Cet ouvrage expose les règles présidant à la distribution du produit entre les facteurs de production, en recourant aux mathématiques plus que de coutume à cette époque. C'est dans cet ouvrage qu'apparaît pour la première fois la fameuse *fonction de production*, l'un des concepts les plus importants de l'économie néoclassique.

$$\mathrm{P} = \mathrm{F(A,B,C...K...N)} \qquad (3.26)$$

La quantité produite (P) est une fonction *F* des quantités utilisées de chacun des *n* facteurs de production. Soient les variables *A, B…K…N* pour représenter ces quantités[2]. L'idée sous-jacente est d'établir entre la quantité produite d'une part et celle des inputs d'autres parts une relation mathématique pouvant être mise sous forme d'équation.

F est une fonction à plusieurs variables. Il est évident que ses dérivées partielles donnent la variation de *P* engendrée par une variation marginale de la quantité du facteur considéré ; elles correspondent aux productivités marginales des facteurs. Par exemple, pour le facteur *K*, on a :

$$\partial \mathrm{P}/\partial \mathrm{K} = \mathrm{Pm\ K} \qquad (3.27)$$

La théorie néoclassique de la répartition prétend que chaque facteur est rémunéré à sa productivité marginale. Pour le facteur *K*, la rémunération totale de l'ensemble des unités vaudra :

$$(\partial \mathrm{P}/\partial \mathrm{K}).\mathrm{K} = \text{rémunération totale de K} \qquad (3.28)$$

La validité de cette théorie nécessite que la somme des rémunérations ainsi calculées s'égalise avec la quantité produite à distribuer. Ce problème est connu sous l'appellation *l'épuisement du produit*. L'équation de l'épuisement du produit est :

$$P = A.\frac{\partial P}{\partial A} + B.\frac{\partial P}{\partial B} + \ \ldots + N.\frac{\partial P}{\partial N} \qquad (3.29)$$

Wicksteed apporte une longue démonstration mathématique de cette égalité, mais une hypothèse importante lui est indispensable : les rendements d'échelle doivent être constants. Si 3 agriculteurs sur 2 hectares produisent 100, 30 agriculteurs sur 20 hectares doivent produire 1000. Ceci revient à postuler que la fonction *F* est homogène de degré un ce qui, mathématiquement, s'exprime ainsi :

$$\text{si } \mathrm{P} = \mathrm{F(A,B,C)}, \text{ alors :} \qquad \mathrm{m.P} = \mathrm{F(m.A,\ m.B,\ m.C)} \qquad (3.30)$$

---

[1] Wicksteed [388] p. 5.

[2] Pour la facilité, j'utilise les mêmes lettres pour « étiqueter » les facteurs et pour exprimer la quantité qui en est utilisée.

où m est une constante quelconque

Wicksteed analyse cette condition : est-il réaliste de la supposer ? La question se pose à deux niveaux, car c'est évidemment le produit marginal en valeur qui rémunère les facteurs, soit le produit marginal physique multiplié par le prix de vente du produit.

1. Wicksteed ne doute pas que le produit physique est une fonction homogène de degré un des inputs, ce qui paraît douteux au vu de l'analyse marshallienne des rendements d'échelle.
2. Mais lorsque la quantité physique P augmente, le prix[1] n'est-il pas appelé à diminuer ? En concurrence parfaite, la firme peut normalement augmenter sa production sans baisser le prix, mais ce n'est pas le cas du monopoleur. Finalement, Wicksteed conclut : « the proposition is approximately true for small increments ».

Si l'offre d'un facteur diminue, sa productivité marginale augmentera et donc sa rémunération également. Cette relation ne peut être généralisée à l'ensemble des facteurs simultanément. En cas de réduction simultanée de tous les facteurs, ils ne seront pas mieux rémunérés. Si un facteur est en pénurie, sa rémunération ne hausse que parce que la productivité marginale de tous les autres facteurs diminue.

On ne peut parler d'augmentation équiproportionnelle de tous les inputs dans la fonction de production que si chaque facteur est homogène et peut être mesuré dans sa propre unité physique. Wicksteed propose donc d'abandonner la traditionnelle distinction entre trois facteurs, la terre, le capital et le travail : « We must regard every kind and quality of labour that can be distinguished from other kinds and qualities as a separate factor. Still more important is it to insist that instead of speaking of so many £ worth of capital, we shall speak of so many ploughs, so many tons of manure, and so many horses… »[2]. Le nombre de facteurs varie d'une entreprise à l'autre et peut être très élevé.

En fait, les facteurs de production de Wicksteed correspondent aux biens d'ordre supérieur de Menger. Tout comme son prédécesseur, Wicksteed insiste sur deux d'entre eux, l'attente ou avance des capitaux et la fonction d'entrepreneur.

L'entrepreneur est un facteur de production et Wicksteed semble penser que l'on peut quantifier son utilisation et lui appliquer la théorie de la distribution valable pour les autres facteurs. Le rôle de l'entrepreneur est d'engager les autres facteurs dans la proportion idéale et de négocier leur rémunération. Même si sa rémunération est un résidu, elle tendra vers sa productivité marginale. Marshall comptait également l'entrepreneur parmi les facteurs de production, mais sa théorie était très ambiguë et il n'est pas certain qu'il considérait que la loi de la productivité marginale s'applique à ce facteur. La majorité des économistes actuels n'ont pas suivi Wicksteed et ne comptent pas l'entrepreneur parmi les facteurs de production, car sa quantité n'est pas fractionnable en unités homogènes. Dans cette optique, le profit est ce qui reste quand les autres facteurs ont été rémunérés suivant leur productivité marginale. Peut-être Wicksteed et Marshall ont-ils eu le tort de ne pas suffisamment distinguer le TRAVAIL de l'entrepreneur, qui est une forme de travail parmi d'autres, et la FONCTION d'entrepreneur, plus impalpable, qui consiste à faire exister son entreprise.

Le facteur de production ***avances*** met à la disposition de l'entrepreneur, une épargne constituée préalablement. Wicksteed dissèque le phénomène de l'épargne, dans un sens proche de la

[1] Wicksteed ne fait pas intervenir les prix ; chaque facteur reçoit un morceau du produit P. Lorsque Wicksteed introduit l'élasticité de la demande du produit pour distinguer le monopole de la concurrence, son exposé manque de clarté.

[2] Wicksteed [389] p. 33

théorie autrichienne. Les entrepreneurs ne sont pas les seuls demandeurs d'avances. Les consommateurs sont également demandeurs sur le *marché des avances*, notamment pour financer les achats de biens durables. Sur le marché des avances se rencontrent ceux qui veulent avancer le flux des dépenses par rapport au flux des revenus et ceux qui veulent le retarder. Même si tous les agents ont une préférence entre le présent et le futur allant dans le même sens, des variations dans l'intensité de cette préférence suffisent à justifier l'avantage pour chacun, de prêter ou emprunter, moyennant un intérêt qui satisfait les deux parties.

Revenons à notre fonction de production. Wicksteed reconnaît que l'application du calcul différentiel suppose que la production est une fonction CONTINUE par rapport à tous les facteurs, ce qui ne saurait être le cas si certains ne sont fractionnables qu'en unités de grande taille, dont le nombre ne varie que par à-coups. De tels facteurs existent (l'entrepreneur, certaines machines…). Wicksteed estime qu'on ne peut tirer argument de difficultés de ce type, inhérent à la science économique, pour faire de ces facteurs, des cas particuliers dans la théorie de la distribution.

Quelle est la place du capital dans la conception wicksteetienne ? Si le capital est « the sum of resources with which a man starts in business (expressed in gold) … in this sense, capital is not one of the factors of production but all of them »[1]. Wicksteed met en garde contre l'utilisation du terme « capital » pour désigner l'agrégation de tous les facteurs qui ne sont pas du travail ou de la terre, dont la monnaie serait l'unité naturelle comme l'acre est l'unité de la terre et l'heure celle du travail. D'ailleurs, le capital d'une entreprise comporte l'encours de production et le stock de produits finis qui sont du travail et des matières cristallisés ; sa nature diffère donc de celle de ces facteurs de production.

Wicksteed montre qu'une économie n'atteint un équilibre de la production que si cette double condition est satisfaite :

1- le dernier shilling que les entreprises investissent dans chacun des facteurs doit avoir la même rentabilité. C'est le principe de substitution de Marshall.
2- Le rapport entre les productivités marginales des facteurs doit être le même dans les différents secteurs industriels. Sinon, certains échanges de facteurs entre les secteurs pourraient se révéler rentables.

L'OFFRE, LA DEMANDE ET LE COUT D'OPPORTUNITE

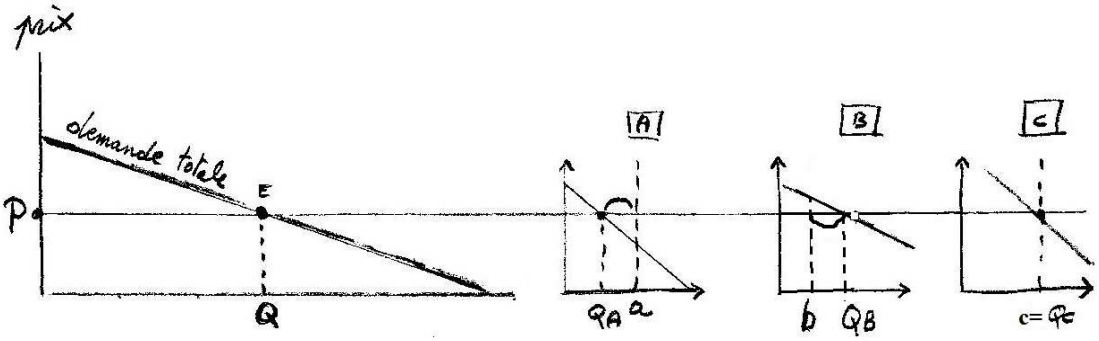
En 1910, Wicksteed publia « The common sense of political economy » et en 1913 un article intitulé « The Scope and Method of Political Economy ». Ces pages attaquent la conception suivant laquelle le coût de production pourrait servir d'explication à la valeur d'échange des biens. Mais Wicksteed va plus loin : il attaque la conception marshallienne de la détermination du prix par l'offre et la demande et défend une conception déjà présente implicitement chez Walras. Il ne craint pas d'affirmer : « What about the supply curve that usually figures as a determinant of price, co-ordinate with the demand curve ? I say it boldly and baldy: there is no such thing (…) what is usually called the supply curve is in reality the demand curve of those who possess the commodity »[2].

L'offre ne doit pas être rejetée en tant que concept, mais elle n'agit pas sur pied d'égalité avec la demande, qui est l'élément fondamental, dont l'offre n'est qu'un dérivé. Voyons ce modèle de plus près.

[1] Wicksteed [389] p. 40.
[2] Wicksteed [388] p. 13.

Les agents sur le marché disposent d'une allocation initiale des différents biens. Pour chaque bien, tous les agents présents sur le marché ont une courbe de demande individuelle, qui indique leur besoin total du bien, indépendamment de la quantité qu'ils possèdent déjà. Cette demande-besoin est donc un concept différent de la demande-acquisition à laquelle nous sommes plus habitués. En additionnant les demandes individuelles d'un bien, on construit la courbe de demande collective. Connaissant la quantité disponible du bien (Q), cette courbe de demande collective suffit à déterminer le prix du bien sans qu'intervienne une courbe d'offre. A ce prix du marché, certains agents demandent plus du bien qu'ils n'en possèdent et seront acheteurs ; d'autres en demandent moins qu'ils n'en possèdent et seront vendeurs (offreurs). On n'est donc pas acheteur ou vendeur a priori mais seulement en fonction du prix du marché.

Graphique 3.12 : l'offre et la demande suivant Wicksteed



Pour une quantité disponible Q, le prix du marché s'établit à P. A ce prix, l'agent A demande du bien une quantité QA inférieure à celle qu'il possède (a) et il sera donc offreur. L'agent B est dans la situation inverse. L'agent C n'achètera ni ne vendra car la quantité dont il dispose est celle qui lui convient à ce prix

Cette conception, logique dans une économie d'échange pur, semble fort éloignée de la réalité industrielle moderne. Mais Wicksteed l'étend à l'économie industrielle. Voici comment. Les facteurs sont demandés par les entreprises en fonction de leur productivité. Chaque facteur se répartira entre les secteurs de façon à ce que sa productivité marginale soit identique dans chacun d'eux, du fait de la concurrence que les secteurs se livrent pour les obtenir. En fonction de la quantité des facteurs ainsi dirigés vers les différents secteurs se détermine la quantité disponible des différents biens (la quantité Q sur le graphique).

Dans cette optique, l'offre d'un facteur, comme celle des produits, n'est qu'une demande sous-jacente inversée. Rappelons-nous l'équation (3.14) de Walras. C'est en égalisant l'utilité marginale du temps libre avec celle des autres biens, que le ménage détermine le nombre d'heures de travail qu'il offre. L'offre de travail n'est que l'inverse de la demande de temps libre. Le travail n'est pas un coût en soi mais seulement le coût d'opportunité de renoncer à du temps libre.

S'opposent donc deux conceptions de l'offre de travail : la conception utilitariste de la peine, présente chez Jevons, Marshall et Edgeworth, branchée sur le concept de COUT DE PRODUCTION, et la conception walrassienne, partagée par les Autrichiens[1] et reprise par Wicksteed, qui ne laisse place qu'au COUT D'OPPORTUNITE. Dans sa critique de l'ouvrage de Walras, Edgeworth lui adresse ce contre-argument ; l'équilibre général n'aboutit pas nécessairement à une égalisation de la désutilité du travail dans les différents emplois. Il y a donc un résidu de peine dans différentes industries qui s'y manifeste de façon irréductible comme un coût de production.

[1] Menger consacre un paragraphe des *Grundsätze* à contester la pénibilité du travail et un deuxième à contester la théorie de l'abstinence.

La même opposition se retrouve dans l'explication de l'offre de capital : d'un côté, on a *l'abstinence*, une peine, donc un coût en soi, qu'on retrouve chez Jevons, Clark et Marshall, bien que celui-ci n'appréciait pas ce terme. En face, on a l'*attente*, une notion introduite par les Autrichiens et reprise par Wicksteed. Ici, l'offre de capital, c'est le choix de consommer ce revenu demain plutôt qu'aujourd'hui. « In fine, every man who is not living absolutely from hand to mouth will make some attempt so to distribute his resources over time as to apply them where they will give the best psychic return »[1]. Il le fera de façon à égaliser l'utilité marginale de la dernière livre sterling dépensée aux différentes périodes.

Tant pour le travail que pour le capital, la conception du coût d'opportunité s'est imposée.

3.2.5. Wicksell : le capital en long et en large

Knut Wicksell est un homme de synthèse : il a assimilé du Walras, du Böhm Bawerk, du Wicksteed et du Marshall et même du Ricardo. Toutefois, il est avant tout l'héritier de la conception autrichienne du capital. Il a développé les idées de Böhm Bawerk sous une forme nettement plus mathématique et les a incorporées dans un équilibre général à la walrassienne. Nous allons résumer brièvement ses « Lectures on Political Economy » parus en 1901. Plus lucide et scrupuleux que Böhm Bawerk, il évitera ses erreurs, mais il devra bien reconnaître que sa rigueur conduit à des résultats parfois décevants.

THEORIE DE LA VALEUR ET GAIN DE L'ECHANGE

La théorie de la valeur basée sur l'utilité marginale nous est maintenant familière. L'exposé qu'en donne Wicksell est brillant mais n'innove pas. Nous nous contenterons donc d'épingler quelques remarques éclairantes.

La plupart des marchandises sont disponibles en toute une gamme de qualités, de la moins bonne à la meilleure. Wicksell donne l'exemple d'un acheteur qui a le choix entre trois chevaux dont les prix respectifs sont de 500, 550 et 575 shillings. S'il préfère le deuxième, c'est qu'il estime la différence de qualité entre les deux premiers à plus de 50s et celle entre le deuxième et le troisième à moins de 25 s. L'utilité marginale peut, dans ce cas, être assimilée à la qualité de la marchandise et l'acheteur choisira le niveau de qualité des différentes marchandises de telle façon à égaliser le rapport entre l'utilité marginale et le prix. Bref, on choisit le niveau de qualité d'une marchandise suivant la même règle que pour fixer la quantité qu'on en consomme.

Les prix constatés dans le commerce de détail peuvent sembler anarchiques au point de ne répondre à aucune loi de la valeur. Pourtant, confirme Wicksell, la loi de la valeur joue bel et bien ; comment s'expliquent alors les différences de prix, parfois importantes, pour une même marchandise dans des magasins concurrents ?

1. La valeur des articles n'est qu'une part du coût des détaillants, grands et petits. Le coût de distribution est très important. D'une part, il n'existe pas de règle absolue quant à la répartition de ce coût entre les différents articles distribués. En outre, le niveau des coûts de distribution peut varier d'un établissement à l'autre ; les clients accepteront de payer un prix comportant un coût de distribution supérieur s'ils bénéficient de meilleurs conseils en contrepartie.

[1] Wicksteed [389] § 1.7.50

2. Les consommateurs souhaitent une certaine stabilité dans les prix. Les commerçants ne les modifient pas chaque fois qu'un changement temporaire le justifierait mais seulement lorsque le changement est assez important ou durable.
3. Un magasin peut être un monopole dans son quartier parce que les clients sont réticents à se déplacer. Dans ce cas, c'est la loi de la formation de prix en monopole qui joue.

Wicksell ne mentionne pas la cause qu'avait identifiée Mill : les consommateurs ne font pas toujours l'effort pour acheter au meilleur rapport qualité/ prix.

Wicksell aborde ensuite à la problématique du *gain de l'échange*. Il reproche à Walras ses affirmations relatives à la maximisation de l'utilité sociale par la libre concurrence. Elles vont au-delà de cce que le modèle peut effectivement prouver et servent de légitimation du système social existant à des forces politiques intéressées. Il assène plusieurs arguments, dont le principal est celui-ci :

Les prétentions de Walras quant au maximum d'utilité sociale n'ont de validité que si les utilités des individus composant la société peuvent être additionnées[1]. Ceci implique que tous les membres de la société ont a priori une égale capacité à jouir des biens consommés. Tout en n'y croyant pas vraiment, Wicksell joue le jeu de cette hypothèse. Vient alors un phénomène social que Walras ignore. Le pauvre conserve une utilité marginale élevée des différents produits, vu la restriction qu'il a dû apporter à sa consommation, alors que l'utilité marginale du riche sera très faible pour la plupart des produits. Partant d'un équilibre walrassien, on pourrait donc imaginer des échanges qui augmentent l'utilité sociale parce que les pauvres y trouvent un avantage supérieur à la perte des riches. Mais cet état serait inaccessible aux seules forces du marché.

Ce résultat fait ensuite l'objet d'une généralisation mathématique. Wicksell démontre que « in normal cases, there can always be found a system of uniform prices at which exchanges will produce a larger sum of utility than at competitive prices »[2]. L'Etat ou un accord entre partenaires sociaux pourrait imposer ces prix. Mais ce sont toujours les prix concurrentiels qui permettent de maximiser la production. Le meilleur moyen de maximiser l'utilité consiste donc à laisser agir la libre concurrence dans un environnement social égalitaire plutôt que de détourner l'équilibre concurrentiel.

LA PRODUCTION SANS LE CAPITAL

Avec Böhm Bawerk et contrairement à d'autres auteurs, Wicksell considère que la nature du capital est différente de celle des deux facteurs primaires. Comme l'analyse du capital se révèle très complexe, il commence par examiner une économie simplifiée où n'interviennent que les facteurs primaires : la terre et le travail. Voyons-la brièvement.

Wicksell ressort le problème de l'épuisement du produit. Si un entrepreneur loue la terre et le travail, qu'il rémunère suivant leur productivité marginale, son profit correspondant au produit résiduel, tendra-t-il vers zéro ? Wicksteed, le premier, avait découvert la condition de l'épuisement, en l'occurrence les rendements d'échelle constants. Wicksell estime que ce cas est plutôt exceptionnel. Mais un autre cas répond aux conditions mathématiques de l'épuisement du produit : lorsque les rendements d'échelle sont d'abord croissants puis

[1] Les équations de l'équilibre général n'additionnaient ou ne comparaient jamais des utilités de personnes différentes. Walras perd cette prudence au moment de tirer des conclusions.
[2] Wicksell [385] v1 p. 80. L'expression *prix concurrentiels* (« competitive prices ») est devenue courante ; elle désigne les prix correspondant à l'équilibre général en concurrence parfaite.

décroissants, le passage de l'un à l'autre comporte un point infinitésimal où les rendements d'échelle sont constants. Un cas que, de façon surprenante, Marshall n'a pas approfondi dans son analyse des rendements d'échelle. « As a rule, the best returns are obtained at some particular scale of operations for the firm in question; if this is exceeded, the advantages of centralization are outweighed by the increased costs which are encountered when larger areas must be exploited for the provision of raw or auxiliary materials or else for the marketing of the product. This scale of production is, under the given circumstances, the "optimum" towards which the firm must always, economically speaking, gravitate… »[1].

Que se passe-t-il lorsque la condition de l'épuisement du produit n'est pas satisfaite ?

- Soit les rendements d'échelle sont **croissants**. Alors, pour toute entreprise qui n'est pas la plus grande sur son marché, le coût total des facteurs dépasse le produit et l'entrepreneur supporte une perte. On arriverait rapidement à la monopolisation de l'économie.
- Soit les rendements d'échelle sont **décroissants** et le coût des facteurs n'épuise pas le produit, laissant un profit net à l'entrepreneur. Ce profit attirerait toujours de nouveaux entrepreneurs et les entreprises tendraient vers une taille lilliputienne.

Appelons k le rapport entre le produit et le coût total des facteurs. On a $k = 1$, lorsque le produit est juste épuisé par la rémunération des facteurs. Que se passe-t-il si k est différent de un ? Vu la substituabilité entre les facteurs, le rapport entre la productivité marginale du facteur et sa rémunération serait le même pour la terre et le travail, mais ce rapport ne vaudrait plus un (égalité de la rémunération et de la productivité marginale). Notons par F la fonction de production, par a la quantité de travail et par b la quantité de terre, par l le salaire et r la rente. On aurait :

$$\frac{\partial F/\partial a}{l} = \frac{\partial F/\partial b}{r} = k \qquad (3.31)$$

LA PRODUCTION AVEC LE CAPITAL

Nous arrivons enfin au cœur de l'économie wicksellienne. Lui-même juge cette problématique très complexe.

La définition du capital retenue par Wicksell est celle-ci : « produced means of production »[2]. Wicksell critique l'approche de Walras qui met le capital sur le même pied que le travail et la terre. Les mille et un types de capitaux sont mesurés en unité physiques et pour chacun d'eux on détermine la quantité et le prix d'équilibre, comme pour tout produit ou service. Une telle approche ne peut produire que des résultats tautologiques. Wicksell met en avant une caractéristique importante du capital, qu'il ne partage pas avec la terre et le travail : il offre un autre étalon de mesure : le fait que chacun de ses composants est une valeur et qu'on peut donc traiter la valeur totale du capital dans la société.

Si on considère le capital en tant que somme de valeur d'échange, la similitude avec la terre et le travail est rompue. La terre et le travail sont mesurés en hectares ou heures, c'est-à-dire en unités immanentes à leur nature ; la rente ou le salaire qui les rémunèrent sont au contraire exprimés en valeur et sont donc d'un autre ordre que le facteur qu'ils rémunèrent. Le capital

[1] Wicksell [385] v1 p. 129.
[2] Wicksell [385] v1 p. 145.

montre les caractéristiques inverses : les capitaux sont mesurés dans une unité extérieure à leur nature (l'unité de valeur) ; l'intérêt qui les rémunère s'exprime dans cette même unité de valeur.

Les conséquences de cette dissymétrie sont importantes. Concernant la terre et le travail, la loi qui explique la rémunération par la productivité marginale s'applique aussi bien au niveau de l'économie globale qu'à celui de l'entreprise individuelle. En est–il de même du taux d'intérêt ? Au niveau de l'entreprise, les rémunérations des facteurs sont imposées par le marché ; on peut donc conclure à l'égalité de l'intérêt avec la productivité marginale du capital. Par contre, au niveau de l'économie globale, les variations du capital disponible affectent les rémunérations de tous les facteurs, d'une façon qui engendre des interactions entre elles. Ceci rend inapplicable au capital la théorie qui détermine la rémunération par la productivité marginale. Selon Wicksell, la clé de la compréhension est à rechercher du côté de la théorie de l'attente de Böhm Bawerk.

Avant d'entrer dans le vif du sujet, précisons les hypothèses simplificatrices du modèle.
1- Wicksell raisonne dans le cadre d'une économie STATIONNAIRE.
2- Le capital fixe, jugé trop complexe, est éliminé. Le capital consiste donc exclusivement en CAPITAL CIRCULANT.

La statique comparative

Pour analyser le lien entre le taux d'intérêt et l'intensité capitalistique de la production, Wicksell inaugure une méthode d'analyse, la *statique comparative*, qui jouira d'un franc succès par la suite. La méthode d'analyse apparemment la plus naturelle consiste à imaginer une économie dont le taux passe de r à $r+\Delta r$ au temps t ; Comment les variables endogènes s'ajusteront-elles à partir du temps t+1 ? L'alternative, la statique comparative, consiste à comparer deux économies aux variables exogènes parfaitement identiques, à l'exception du taux d'intérêt : la première conserve un taux d'intérêt de r sur toute la période de 0 à T (où $T > t$). La seconde garde un taux constant de $r+\Delta r$ durant cette même période. En quel rapport les variables endogènes des deux économies différeront-elles ? Les deux économies se reproduisent à l'identique de période en période. Cela permet de ne comparer que des situations d'équilibre.

Pour mesurer le capital en unités physiques, Wicksell recourt à l'astuce suivante : « All capital goods, however different they may appear, can always ultimately be resolved into labour and land »[1]. Le capital est donc une quantité de travail capitalisée plus une quantité de terre capitalisée. Prenons l'exploitation d'un boulanger : son travail courant est celui par lequel il transforme la farine en pain ; sa terre courante est assez peu importante et correspond à la surface au sol occupée. Son capital circulant comporte ses stocks de produits finis, semi-finis ou de matières premières, principalement la farine. De toute évidence, cette farine peut être ramenée à du travail passé dont le produit n'a pas été consommé mais capitalisé ; celui du fermier et celui du meunier, ainsi qu'à la terre, très importante dans le travail agricole, dont le résultat a également été capitalisé.

Pour simplifier, divisons l'échelle du temps en deux périodes annuelles : l'année courante où le boulanger traite la farine et l'année (-1) où le paysan et le meunier l'ont produite. Soient a et b, respectivement les quantités de travail et de terre utilisés l'année courante, et a_1 et b_1 les

[1] Wicksell [385] v1 p. 149

quantité de travail et de terre de l'année (-1) qui ont été capitalisées. On a alors la fonction de production :

$$P = F(a, b, a_1\ b_1) \qquad (3.32)$$

Les dérivées partielles de *a* et *b* sont notées *l* et *r* ; ce sont évidemment le salaire et la rente courants. Les dérivées partielles de a_1 et b_1 sont l_1 et r_1 ; ce sont des variables plus abstraites, correspondant aux productivités marginales du travail et de la terre capitalisés. D'après Wicksell, sauf dans le cas d'une pléthore de capital peu probable, la terre et le travail capitalisés sont plus productifs que la terre et le travail courants. On a donc $l_1 > l$ et $r_1 > r$. Si l'on reprend notre exemple du pain, ceci ne signifie pas que la productivité marginale du fermier dans la ferme est supérieure à celle du boulanger dans la boulangerie mais que la productivité marginale du travail du fermier capitalisé dans la boulangerie est supérieure à celle du boulanger. Cette comparaison peut sembler un peu absurde mais c'est notre exemple qui n'est pas idéal ; la productivité supérieure du travail capitalisé serait plus évidente si on comparait la productivité marginale du boulanger avec celle du travail capitalisé du fabricant de ses outils. Mais cet exemple serait moins adéquat car il a trait au capital fixe, qui, par hypothèse, est exclu du modèle.

A l'équilibre le rapport l_1/l doit être égal au rapport r_1/r ; si $l_1/l > r_1/r$, il serait avantageux de capitaliser a_1, ce qui ferait baisser l_1, jusqu'à ce que l'égalité soit atteinte. L'intérêt provient justement du surplus de productivité des facteurs capitalisés par rapport aux facteurs courants. A l'équilibre, on a :

$$(l_1 - l)/l = (r_1 - r)/r = i \qquad \text{où i est le taux d'intérêt} \quad (3.33)$$

Compliquons et considérons maintenant une échelle comportant plus de deux périodes annuelles. Imaginons par exemple que les services de la terre agricole et du fermier datent de la période (-2), ceux du meunier de la période (-1) et ceux du boulanger de la période courante. Tout ce qu'on vient de dire du rapport entre l_1,r_1, *l* et *r* est valable pour le rapport entre l_2, r_2, l_1 et r_1. A l'équilibre, on a :

$$(l_1 - l)/l = (l_2 - l_1)/l_1 = i \qquad (3.34)\ [1]$$
$$(r_1 - r)/r = (r_2 - r_1)/r_1 = i$$

Le capital total dont dispose la société va se répartir entre les facteurs capitalisés de un an et ceux de deux ans selon telle ou telle proportions, ce qui influencera les productivités marginales respectives. A l'équilibre, la proportion se fixera de façon à ce que les équations (3.34) soient respectées. Ce processus suppose une certaine substituabilité entre les facteurs capitalisés d'âges différents comme il en existe entre les facteurs courants et les facteurs capitalisés ou entre des facteurs différents de même « âge ». On choisira par exemple une méthode de production, qui en imposant un peu plus de travail ou de terre au fermier, en épargnera au meunier ou vice versa. A l'équilibre, le taux d'intérêt annuel est donc indépendant de la durée de l'investissement ; une telle conclusion semble contraire à l'expérience. Wicksell la justifie en ceci que les écarts de taux selon le terme d'un investissement, communs dans la réalité économique, s'expliquent par les différences de risque, un élément extérieur à la problématique présente.

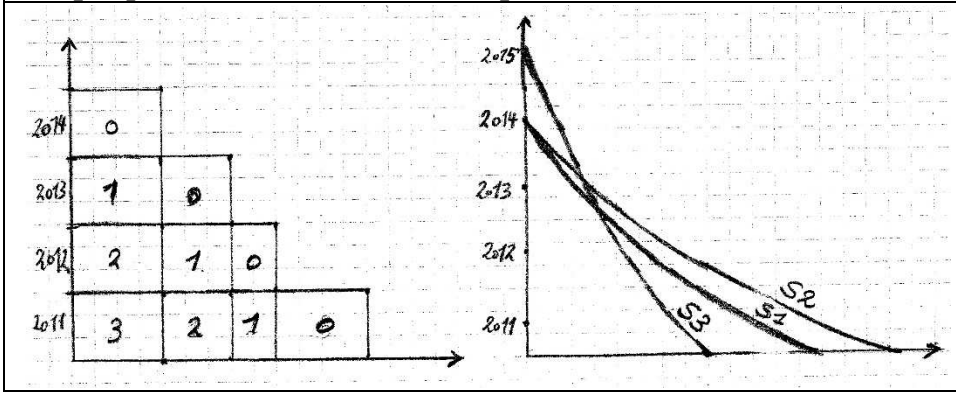
Les capitaux diffèrent par leur structure d'âge et pour Wicksell, il s'agit là d'une caractéristique essentielle d'un capital. Pour en rendre compte, Wicksell dessine le graphique 3.13[1] qui expose

---

[1] $l_2/l = (1+i)^2 \quad \Leftrightarrow \quad l_2 = l(1+i)^2$
$r_2/r = (1+i)^2 \quad \Leftrightarrow \quad r_2 = r(1+i)^2$

la double dimension du capital : le capital a une largeur (en abscisse) et une hauteur[2] (en ordonnée). La LARGEUR représente la quantité de facteurs primaires incorporés et la HAUTEUR la durée de cette incorporation avant la disponibilité pour la consommation. Prenons un exemple où la période de production absolue est de quatre ans et où chaque année une certaine quantité de facteurs primaire vient s'ajouter au produit en cours. Les rectangles sur une même ligne ont en commun qu'ils concourent à un produit qui sera consommable la même année (qui étalonne l'ordonnée). Les rectangles d'une même colonne sont d'une même phase (de gauche à droite, première, deuxième … dernière année de la période de production). Le chiffre dans le rectangle indique l'âge actuel (en 2011)[3]. Par exemple, le rectangle de gauche de la deuxième ligne représente le travail de 2010 (qui a donc un an en 2011) en vue du produit à consommer en 2013. La largeur du rectangle représente la quantité de facteurs primaires concernée.

Graphique 3.13 : la structure du capital selon Wicksell



Sur la figure de droite, le temps et la quantité de facteurs incorporés sont des variables continues. La structure de départ est S1. Si le capital disponible s'accroît et que cet accroissement s'opère par un élargissement, tous les rectangles se dilatent en largeur et on arrive à la structure S2. Si le capital augmente en hauteur (approfondissement), alors que son volume total reste constant, la largeur doit être réduite en compensation et on obtient la structure S3.

Wicksell examine ensuite les conséquences d'une augmentation du capital disponible sur la rémunération des facteurs courants. On doit s'attendre à un élargissement. Mais Wicksell démontre de la manière suivante que le plus souvent, il y aura également un approfondissement. La loi de la productivité marginale décroissante fera baisser l_1 et l_2 par rapport à l et de même pour la terre. Supposons que la baisse de l_2 et celle de l_1 par rapport à l sont égales et représentons ce taux par $1/(1-\varepsilon)$. Alors l_2 et r_2 ont moins diminué que l_1 et r_1 et l'équilibre de l'équation (3.34) n'est plus respecté. Wicksell démontre en effet que se vérifie l'inégalité :

$$l.(1+i)^2(1-\varepsilon) > l.[(1+i)(1-\varepsilon)]^2 \qquad (3.35)[4]$$

En conséquence, l'utilisation des facteurs capitalisés de deux ans d'âge devient plus profitable et le capital se déplacera de a_1 et b_1 vers a_2 et b_2 ; peut-être même que a_3 et b_3, jusque là non

[1] Ce graphique a quelque connexion avec le graphique 3.3 de Jevons.

[2] Cette distinction restera importante dans la théorie du capital postérieure à Wicksell. Mais on préférera ultérieurement le terme « profondeur » à « hauteur ». « Capital-deepening » signifie l'allongement des détours de la production.

[3] L'âge zéro représente les facteurs courants (*l* et *r* dans les formules ci-dessus). Ces rectangles ne font donc pas vraiment partie du capital actuel en 2011.

[4] Le membre de gauche indique ce que l_2 est devenu. Le membre de droite ce qu'il serait devenu si la proportion entre l_2 et l_1 avait été conservée. Tout ceci vaut également pour la terre capitalisée.

rentables le deviendront et attireront également une partie du capital. Ainsi jusqu'à ce que l'équilibre soit rétabli. Le capital augmente en hauteur, autrement dit la période de production s'allonge.

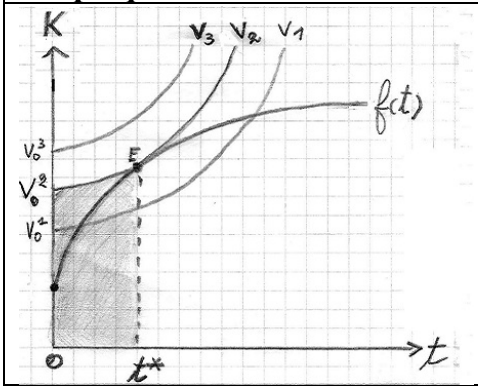
Quel est l'effet de cet accroissement du capital sur la rémunération des facteurs primaires ? Il y a une double influence contradictoire :
- D'une part, l'ELARGISSEMENT du capital augmentera la demande des facteurs primaires courants et donc leur rémunération.
- D'autre part, l'APPROFONDISSEMENT est défavorable au salaire et à la rente ; si le capital doit « entretenir » les facteurs primaires pendant plus longtemps dans le capital, il devra se contenter d'une absorption annuelle moindre de ces facteurs.

Selon Wicksell, le premier effet doit toujours l'emporter.

Wicksell illustre ensuite sa théorie par un exemple. C'est le fameux **modèle du vin**, qu'on peut éventuellement lire comme une allégorie de ce que nous avons vu jusqu'ici. Les capitalistes achètent le jus de raisin aux viticulteurs et le laissent fermenter en tonneau. La valeur de vente du vin augmente avec le temps de fermentation. Les viticulteurs représentent le travail et la terre et le prix qui leur est payé pour le jus représente le salaire et la rente. Le temps de fermentation est le temps pendant lequel ce travail et cette terre sont capitalisés, autrement dit la période de production. Son allongement augmente la productivité représentée ici par la qualité du vin. Le problème consiste à déterminer le temps de fermentation qui maximise le profit des capitalistes (correspondant au taux d'intérêt en concurrence parfaite).

Cet exemple se distingue de celui du graphique 3.13 et est plus simple en ceci que les facteurs primaires sont appliqués en une fois la première année de la période de production, puis se contentent de « mûrir »[1]. Cela n'empêche pas que l'intervention des facteurs de production soit continue, mais les facteurs primaires appliqués pendant des années différentes concourent à la production des millésimes différents.

Graphique 3.14 : le modèle du vin



[1] Nous avons donc rencontré deux types d'application des facteurs primaires : en flux et ponctuelle. La livraison du produit pour la consommation peut également être en flux ou pnctuelle. En définitive, quatre situations sont possibles :
- entrée ponctuelle – production ponctuelle
- flux d'entrée – production ponctuelle
- entrée ponctuelle – flux de production
- flux d'entrée – flux de production.

Le modèle du vin appartient à la première catégorie (la plus simple).

Le graphique 3.14 illustre cet exemple où le temps intervient comme une variable continue. Bien qu'il soit une inconnue du système, Wicksell, pour des raisons didactiques, pose d'abord que le taux d'intérêt *i* est connu. Il marque la courbure des courbes de valeur acquise $V_1,V_2,V_3\ldots$ La courbe f(t) indique le prix de marché du vin, tel qu'il évolue par rapport au temps de fermentation *t*. Son point de départ est le prix de marché du jus de raisin. La période de production optimale t* pour un taux d'intérêt donné *i*, est celle qui maximise la valeur actuelle (actualisée au taux *i*) du prix de vente du vin. C'est le point où la courbe f(t) est tangente à une courbe de valeur acquise.

Le capital circulant (le seul par hypothèse) est constitué par l'ensemble des fûts d'âges différents stockés dans les caves des capitalistes. Le capital consiste exclusivement en un stock de vin attendant la maturité. La valeur du capital (*K*) est donc la somme des valeurs acquises par les jus de fruits des différents âges (en nombre infini) entre zéro et t*. Elle est donnée par la surface hachurée sur la figure de gauche.

Après cet exemple pédagogique, Wicksell revient aux fondamentaux. La variable exogène essentielle est la valeur du capital. Il démontre mathématiquement ainsi qu'une augmentation du capital (K) entraîne :

- la hausse du salaire, de la période de production et de la production
- la baisse du taux d'intérêt, bien que la valeur totale de l'intérêt augmente (mais moins que celle du capital)[1].
- Pour ce qui est des parts respectives du travail et du capital dans le revenu national, a priori, l'évolution peut être favorable à l'un ou à l'autre selon que la productivité moyenne du travail augmente moins ou plus que sa productivité marginale.

Wicksell calcule aussi le produit marginal du capital, qui n'est autre que la différentielle df(t)/dK. Il démontre qu'il est inférieur au taux d'intérêt, ce qui confirme la non-équivalence entre ces deux grandeurs. Par contre, il ressort que l'intérêt est égal à la productivité marginale de l'attente df(t)/dt, comme chez Jevons et Böhm Bawerk.

Dans son ouvrage « Value, Capital and Rent », Wicksell avait attiré l'attention sur un autre aspect de cette problématique : « This increase of the national capital is accompanied by an increase in wages which partially swallows it up, with the result that the lengthening of production actually achieved always falls short of the lengthening of production when the rate of wage remains unchanged »[2]. Une partie du nouveau capital finance la hausse des salaires induite et n'est donc pas opérationnelle pour accroître l'intensité en capital de la production. Le capital physique et le capital en valeur n'évoluent pas proportionnellement. On a appelé ce phénomène l'*effet Wicksell*.

RAFFINEMENTS

1- Application des facteurs primaires en continu

Wicksell n'a pas éludé le cas, plus complexe mais également plus réaliste, où les facteurs primaires sont appliqués en continu pendant toute la période de production, comme chez Böhm Bawerk. Il construit la fonction de production :

$$P = f(a,b,t,t') \qquad (3.36)$$

[1] Il existe toutefois un seuil au-delà duquel le capital devient pléthorique et le montant absolu de l'intérêt diminue.

[2] Wicksell [386] p. 137

La valeur produite est fonction de la quantité de travail *a*, de la quantité de terre *b*, de la période de production moyenne du travail *t* et de la période moyenne de la terre *t'*. Cherchant les valeurs de *a,b,t* et *t'* qui maximisent le taux d'intérêt, il se heurte à une difficulté majeure. Le problème ne peut recevoir une solution exacte que si *t* et *t*' sont indépendantes du taux d'intérêt. Il est facile de voir que ce n'est pas le cas : si on investit €100 de travail par an au cours d'une période de production de trois ans, l'âge moyen des capitaux ne sera pas de 1 an et demi comme le pensait Böhm Bawerk ; il sera légèrement supérieur à un an et demi, car au terme des trois ans, l'investissements de la première année aura une valeur de $100(1+i)^2$, celui de la deuxième année vaudra $100(1+i)$ et celui de la troisième 100; la période moyenne dépend du taux d'intérêt !

Wicksell montre que sans la simplification faite inconsciemment par Böhm Bawerk, le problème devient insoluble. Au mieux, le modèle est une bonne approximation de la réalité si la période de production est courte et le taux d'intérêt peu élevé. Mais, Wicksell doit l'avouer, sa tentative de démontrer que l'intérêt est la productivité marginale de l'attente n'est pas un franc succès, lorsque les facteurs primaires sont capitalisés en continu.

2- Analyse de longue période

Après l'analyse stationnaire, vient l'analyse dynamique. Puisqu'à chaque période, la valeur du capital est un des facteurs déterminant les caractéristiques de l'équilibre, il est intéressant d'analyser l'évolution de l'économie lorsque cette donnée connaît une croissance continuelle. Wicksell pense qu'au niveau social, l'accumulation du capital ne devrait pas s'arrêter tant qu'un taux d'intérêt positif subsistera et jouera le rôle d'incitant.

Etonnamment, Wicksell généralise les résultats de la statique comparative à l'analyse dynamique et historique. L'accumulation du capital entraîne dans son sillage la baisse du taux d'intérêt. Après Smith, Ricardo, Marx, Walras et Jevons, il prophétise la baisse tendancielle du taux d'intérêt, allant même jusqu'à son annulation.

Face à l'absence de confirmation empirique, Wicksell recense diverses causes possibles qui contrecarrent et retardent cette évolution naturelle, notamment :

- L'incertitude qui freine l'incitation à investir, encore que le système financier moderne tende à la diminuer
- La sous-évaluation subjective du futur, le deuxième motif de la préférence pour le présent selon Böhm Bawerk
- La destruction du capital par les guerres qui reviennent régulièrement

Après ces considérations, Wicksell envisage quand-même que « the most important reason why the rate of interest has not fallen is probably that our modern societies differ in a high degree from the stationary type »[1]. Il évoque deux facteurs qui écartent l'économie réelle de l'économie stationnaire : le progrès technique et la croissance démographique.

3- La prise en compte du capital fixe

En 1923, **Akerman**, un ancien étudiant de Wicksell, publie un ouvrage décrivant l'équilibre d'une économie où le capital fixe coopère avec le travail, une approche courageuse d'un problème qui a fait fuir une majorité d'économistes, y compris Wicksell lui-même en 1901. Aiguillonné par cet ouvrage, Wicksell, à plus de soixante-dix ans, se remet à l'ouvrage et publie cette même année un article présentant son propre modèle avec capital fixe.

[1] Wicksell [385] v1 p. 213

Le nouveau modèle est réellement plus complexe que le précédent. Nous sommes dans une économie bi-sectorielle produisant un bien de production et un bien de consommation. Il y a maintenant deux fonctions de production, à partir desquelles sont construites tout un système d'équations. Wicksell démontre mathématiquement que dans cette nouvelle économie, ses conclusions précédentes restent valables, avec la nuance que maintenant l'intérêt peut être aussi bien plus élevé que plus bas que la productivité marginale du capital.

3.2.6. Pareto : l'équilibre et l'optimum

Aujourd'hui on se rappelle principalement le Pareto économiste, mais de son temps, Vilfredo Pareto était surtout connu comme sociologue et polémiste politique. Le fond de sa pensée est le scepticisme vis-à-vis de tous les théories, les mouvements et les partis qui veulent libérer les classes dominées. Inévitablement, une nouvelle aristocratie prendra le relais de la précédente. Cet élitisme fait de lui un héraut de la pensée contre-révolutionnaire. Mussolini en fut l'admirateur.

C'est sa pensée économique qui nous intéresse. Pareto succéda à Walras en 1893 à la chaire d'économie politique de la faculté de droit de l'Université de Lausanne. Entre les deux, la filiation est évidente, même si les différences ne manquent pas. On désigne le binôme Walras-Pareto sous l'appellation « Ecole de Lausanne ».

Au début de son « manuel d'économie politique » (1906), Pareto écrit : « L'objet principal de notre étude est l'équilibre économique. Nous verrons bientôt que celui-ci résulte de l'opposition qui existe entre les goûts des hommes et les obstacles à les satisfaire. Notre étude comprend donc trois parties bien distinctes : 1° l'étude des goûts ; 2° l'étude des obstacles 3° l'étude de la façon dont se combinent ces deux éléments pour arriver à l'équilibre »[1].

A quels obstacles, Pareto fait-il allusion ? En échange pur, la quantité des biens à répartir est fixée à un niveau inférieur aux besoins, ce qui fait obstacle à la satisfaction de ceux-ci. Lorsqu'il y a production, la limitation des ingrédients nécessaires à celle-ci (y compris le travail) y fait obstacle semblablement.

Pareto définit l'équilibre ainsi : « l'état qui se maintiendrait indéfiniment s'il n'y avait aucun changement dans les conditions dans lesquelles on l'observe »[2]. L'équilibre auquel il s'intéresse est un équilibre général, ce qui établit clairement la filiation avec Walras.

La théorie de la production de Pareto sera largement améliorée par ses disciples (cf. chapitre 4.2). Le présent exposé s'en tiendra donc à l'échange pur. Signalons toutefois une avancée importante par rapport à Walras : au même titre que le consommateur, le producteur est maintenant reconnu comme un agent cherchant à optimiser sa situation (en maximisant non pas son utilité mais son profit) ; chez Walras, la production suivait la demande par une espèce d'automatisme.

L'essence d'une situation concurrentielle se trouve dans le comportement des agents en *preneurs de prix* (« price takers ») [3]. La caractéristique du monopole est au contraire la volonté

[1] Pareto [273] p. 150
[2] Pareto [273] p. 154
[3] Les économistes parlent de la *fonction paramétrique* des prix pour désigner cette situation où les agents considèrent que la seule décision qu'ils ont à prendre est la détermination des quantités achetées et vendues et que le système de prix existant est une information suffisante à cette fin.

de manipuler les prix. Pareto étudie ces deux comportements mais notre exposé se limitera à la concurrence. A propos des agents qui acceptent les prix au lieu de vouloir les manipuler, Pareto note ce paradoxe : ce sont justement leurs décisions d'achat et de vente qui déterminent ces prix : l'influence joue donc dans les deux sens mais elle est inconsciente et diluée dans un grand nombre de décisions individuelles.

LES GOUTS : LA CONSOMMATION

Pareto estime que le terme *utilité* rend mal compte du concept qui joue le rôle que nous connaissons dans l'économie néoclassique. Sa connotation est trop objective. « La morphine n'est pas utile, au sens ordinaire du mot, puisqu'elle est nuisible au morphinomane ; elle lui est au contraire utile économiquement, puisqu'elle satisfait un de ses besoins, alors même qu'il est malsain »[1]. Pareto lui préfère un terme de son cru : l'« *ophélimité* ».

Comme Edgeworth avant lui, Pareto tient compte de ce que les ophélimités procurées par les consommations de plusieurs biens ne sont pas indépendantes entre elles ; cette interdépendance peut prendre deux formes opposées : la *complémentarité* et la *substituabilité*. « Pour avoir de la lumière, il faut une lampe et aussi du pétrole »[2]. Le pétrole et la lampe sont des biens *complémentaires*.

La *substituabilité* consiste en ce qu'un bien peut être utilisé à la place d'un autre avec un résultat plus ou moins semblable. Un cas particulier est celui où plusieurs biens répondent à un même besoin, mais en procurant des niveaux de plaisir différents. « Celui qui est très pauvre mange beaucoup de maïs, peu de pain et très rarement de la viande. Ses ressources augmentant, il mangera plus de pain et moins de maïs ; si sa situation s'améliore encore, il mangera du pain et de la viande… »[3]. Pareto appelle *biens inférieurs*, les biens dont la consommation diminue quand le pouvoir d'achat augmente. Comme le montre l'exemple ci-dessus, un même bien peut être supérieur pour un niveau de revenu (le pain quand il remplace le maïs) et inférieur pour un revenu plus élevé (le pain, lorsqu'il est lui-même remplacé par la viande).

Le choix du consommateur consistera à déterminer la combinaison des biens à consommer qui maximise son ophélimité totale, tout en respectant son budget. « Le pauvre se demandera s'il vaut mieux pour lui acheter un peu de saucisson ou un peu de vin ; le riche s'il préfère acheter une automobile ou un bijou ; mais tous, plus ou moins, résolvent des problèmes de ce genre »[4].

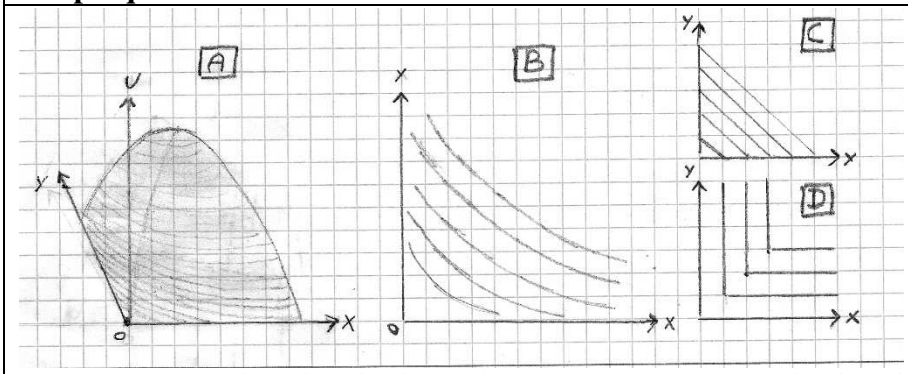
Pour analyser ce choix, Pareto ressort les *courbes d'indifférence* inventées un quart de siècle plus tôt par Edgeworth, mais il les modifie pour leur donner les propriétés bien connues des étudiants en économie. Les variables en abscisse et en ordonnée ne sont plus les quantités ECHANGEES des biens *X* et *Y*, comme chez Edgeworth, mais les quantités DETENUES par l'agent. Les points formant les courbes, comme elles apparaissent sur le graphique 3.15, représentent des allocations de biens différentes, dont l'agent se voit doté, que ce soit à l'origine ou à la suite d'un échange.

Pareto compare les courbes d'indifférence avec les courbes de niveau indiquant l'altitude sur une carte topographique. Une courbe d'indifférence rassemble des combinaisons des biens, qui ont une altitude commune sur la « *colline du plaisir* ». Mais l'altitude sur la colline du plaisir,

[1] Pareto [273] p. 157
[2] Pareto [273] p. 253
[3] Pareto [273] p. 258
[4] Pareto [273] p. 161

autrement dit l'ophélimité, est-elle mesurable ? Pareto ne le croit pas. En fait, les courbes d'indifférence servent plus à hiérarchiser les altitudes différentes qu'à les mesurer[1]. « Un homme peut savoir que le troisième verre de vin lui procure moins de plaisir que le second ; mais il ne peut en aucune façon savoir quelle quantité de vin il doit boire après le second verre pour avoir un plaisir égal à celui que lui a procuré ce second verre de vin »[2]. Dans le même ordre d'idées, Pareto considère que l'ophélimité est trop subjective pour qu'on puisse comparer celles de deux individus différents ou les additionner entre elles.[3].

Graphique 3.15 : les courbes d'indifférences de Pareto



La figure A montre la colline du plaisir ; la figure B ne conserve que les deux dimensions de la base, sur lequel apparaissent les courbes d'indifférence dont le niveau d'ophélimité augmente quand on se déplace vers le nord-est.

On constate que les courbes d'indifférence sont convexes par rapport à l'origine. Cela s'explique ainsi : « pour compenser les manques d'une petite quantité, toujours la même, d'une marchandise donnée, il faut d'autant moins d'une autre qu'on en possède davantage de la première »[4]. Il faut mentionner deux cas particuliers : la complémentarité parfaite se traduit par des courbes d'indifférence en angle droit (figure D), car l'augmentation de la quantité d'un seul des deux biens n'augmente absolument pas l'ophélimité totale. La substituabilité totale se traduit par des lignes d'indifférence droites (figure C).

Soit un indice *I* étalonnant les courbes d'indifférence, qui remplace la mesure de l'utilité. Dans un champ d'indifférence, les courbes peuvent être définies de deux façons : par une fonction f_I ou par une fonction θ_I:

$$f_I(x,y,z...I) = 0 \Leftrightarrow \quad I = \theta_I(x,y,z...) \qquad (3.37)$$[5]

Sur un graphique, on ne peut représenter les courbes d'indifférence que pour deux biens *x* et *y*, mais algébriquement, l'équation (3.37) permet de traiter un nombre quelconque de biens.

Pareto considère l'altitude de la colline du plaisir comme non significative, mais sa forme a clairement les propriétés suivantes :

1. Elle est globalement montante et en tout cas jamais décroissante. Le plaisir ne peut diminuer quand la consommation augmente.
2. Sa pente faiblit. Comme les autres économistes néoclassiques, Pareto constate la décroissance de l'utilité marginale, qu'il appelle *ophélimité élémentaire*.

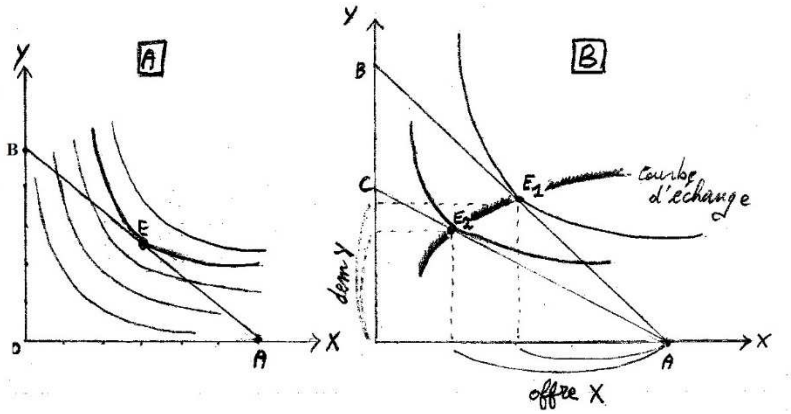
[1] Il s'agit d'une différence avec les cartes topographiques, où l'altitude peut être mentionnée
[2] Pareto [273] p. 264
[3] Le très utilitariste Edgeworth croyait au contraire à la possibilité aussi bien qu'à la nécessité de comparer et additionner les utilités individuelles.
[4] Pareto [273] p. 271
[5] L'indice *I* accolé à la fonction *f* ou θ indique que la forme de la fonction est propre à la valeur de *I*.

3. La pente faiblit même de plus en plus. « Non seulement le second verre de vin procure moins de plaisir que le premier, et le troisième moins que le second, mais la différence entre le plaisir que procure le troisième et celui que procure le second est moindre que la différence entre le plaisir que procure le premier et celui que procure le second »[1].

L'ophélimité élémentaire d'un bien décroît lorsque sa consommation augmente. Mais elle dépend également de la quantité consommée des autres biens. Elle diminue lorsque s'élève la quantité consommée de biens substituables mais elle augments en cas de hausse de la consommation de biens complémentaires.

Nous pouvons enfin aborder **l'échange**. L'agent va tenter d'adapter le panier de biens dont il dispose pour optimiser son ophélimité. Cela se traduit par des déplacements dans le champ d'indifférence, que Pareto appelle des *sentiers*. Nous voyons sur la figure 3.16-A un agent qui possède la quantité *OA* du bien *X* (que nous pouvons imaginer comme étant la monnaie) et qui ne possède pas de *Y* mais voudrait en obtenir une certaine quantité. Il s'engage donc sur le sentier partant de *A* vers *B*. L'inclinaison de cette droite, donnée par le ratio *OB/OA*, indique le rapport de prix entre X et Y ; s'il dépense la totalité de sa provision *OA* de *X*, l'agent peut obtenir la quantité *OB* de *Y*. Le segment de droite AB exprime donc la limite qu'est le budget ; au nord-est de cette droite, les combinaisons de *X* et *Y* sont inaccessibles. En *E,* ce sentier est tangent à une courbe d'indifférence, qui est bien sûr la plus élevée qu'il peut atteindre. E est donc son meilleur choix.

Graphique 3.16 : l'optimum du consommateur : deux biens



Sur la figure 3.16-B, nous faisons varier le rapport de prix, qui passe de *AB* à *AC*, ce qui représente un renchérissement de *Y* par rapport à *X*. L'équilibre passe alors de E_1 à E_2. La consommation de *Y* régresse. Pareto appelle *courbe d'échange* le regroupement des points d'équilibre aux différents rapports de prix. Comme l'indique Pareto, il s'agit d'une véritable courbe d'offre et de demande ; elle indique l'offre de *X* et la demande de *Y* de cet agent aux différents rapports de prix.

Pareto en profite pour rappeler l'interdépendance walrassienne de l'ensemble des biens : si *X* et *Y* sont deux biens extraits d'une multitude pour faire l'objet d'une observation fine, la demande de *Y* dépend non seulement du rapport de prix de *Y* en *X* mais aussi des prix des autres biens qui n'apparaissent pas sur le graphique.

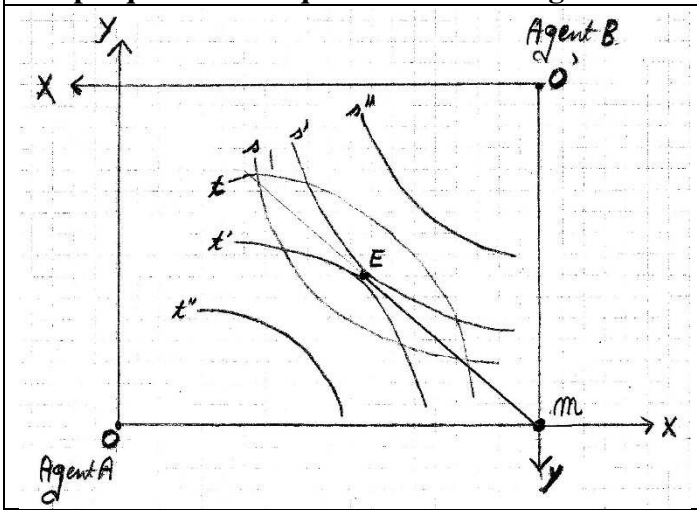
[1] Pareto [273] pp. 267-268. Ceci correspond à une courbe d'utilité marginale concave.

L'EQUILIBRE GENERAL

Nous restreignons toujours l'analyse à l'échange pur. Commençons par étudier l'équilibre de l'échange entre DEUX agents, à l'aide du graphique 3.17. Pareto y emboîte leurs champs d'indifférence en tête-bêche. L'agent A ne possède que le bien *X* en quantité *Om* et l'agent *B* ne possède que du bien *Y* en quantité *O'm*. Les courbes d'indifférence de l'agent *A* et de l'agent *B* sont respectivement *s,s',s"* et *t,t',t "*. La répartition initiale est donnée par le point *m*. L'échange que les agents vont effectuer est celui qui permet d'arriver à une nouvelle répartition des biens *X* et *Y* qui optimise leur situation à tous deux en suivant la règle de la tangence formulée ci-dessus. Pareto énonce le théorème suivant : « s'il existe un point où un sentier parcouru par les agents qui contractent, est tangent aux courbes d'indifférence de ces individus, c'est là un point d 'équilibre »[1]. C'est le cas du point *E* sur le graphique.

Comme E est un point d'optimum pour les agents *A* et *B*, leurs courbes d'échange s'y coupent. Ceci assure l'équilibre entre l'offre et la demande tant pour le bien *X* que pour le bien *Y*. En effet, la courbe d'échange de *A* est une courbe d'offre de *X* et une courbe de demande de *Y*, alors que la courbe d'échange de *B* est une courbe d'offre de *Y* et une courbe de demande de *X*.

Graphique 3.17 : l'équilibre de l'échange



Concernant la manière d'arriver à l'équilibre, autrement dit concernant la technique d'échange utilisée, la conception de Pareto est plus proche de celle d'Edgeworth que de celle de Walras. Ce sont les agents qui tâtonnent pour trouver le « bon » sentier et non le commissaire-priseur qui tâtonne pour trouver le bon prix.

Passons maintenant au stade suivant : l'équilibre général avec une multitude d'agents et de marchandises. Ici, Pareto reprend la méthode walrassienne consistant à mettre les conditions d'équilibre en équations. Le système d'équations permet de déterminer l'ensemble des prix et des quantités échangées si il y a l'égalité entre le nombre des équations et celui des inconnues. Supposons qu'il y ait *n* marchandises (dont la monnaie) et *m* individus.

Il y a trois groupes d'équations :

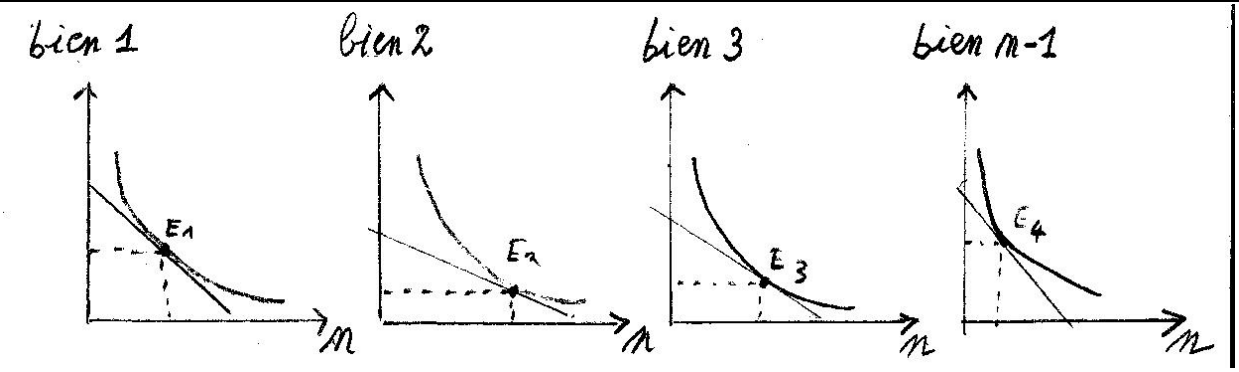
1) Celles qui disent que chaque individu égalise entre elles les *ophélimités pondérées*[2] des différents biens. Ces équations sont au nombre de (n–1).m.

[1] Pareto [273] p. 189

[2] Tel est le terme utilisé par Pareto pour désigner le rapport entre l'ophélimité marginale et le prix

Le graphique 3.18 illustre ce groupe d'équations pour un individu particulier. Si chacun des (n-1) biens autres que la monnaie a son ophélimité pondérée égalisée avec celle de la monnaie, toutes les ophélimités pondérées sont égales. Les points de tangence, correspondant à l'égalisation des ophélimités pondérées, déterminent les quantités demandées (en ordonnée) et les quantités de monnaie[1] consacrées à ces achats (en abscisse).

Graphique 3.18 : L'optimum du consommateur : de nombreux biens



Le graphique montre les champs d'indifférence de l'individu pour les n-1 produits avec la monnaie (bien n).

2) Celles (au nombre de *m*) qui égalisent pour chaque individu le total de ses dépenses et le total de ses recettes.

3) Celles, au nombre de (n-1), qui égalisent pour chaque produit le total des achats et le total des ventes dont il fait l'objet[2]. Comme l'avait déjà montré Walras, si tous les marchés sauf un sont en équilibre, le dernier doit l'être également.

Ces équations sont en nombre égal aux inconnues, que sont les (n-1) prix (autres que la monnaie) et les m.n quantités de chaque produit échangées par chaque individu.

Ayant ainsi déterminé les prix par un système d'équations simultanées, Pareto en conclut que l'équilibre général EST la théorie de la valeur. Toute autre théorie de la valeur est fausse parce qu'incomplète ; la valeur dépend d'une multitude de causes et c'est une erreur de mettre l'une d'elles en exergue. Pareto reproche son ambiguïté à Walras, qui d'une part, le premier avait déterminé les prix à partir d'un système d'équations simultanées, mais qui de l'autre côté prétendait voir en l'utilité marginale la cause de la valeur.

Clôturons l'exposé sur Pareto avec un concept qui a particulièrement contribué à sa réputation et par lequel il continue à influencer la science économique contemporaine. Une répartition donnée du produit, faisant suite au processus d'échange, assure aux membres d'une collectivité le *maximum d'ophélimité*, si « tout petit déplacement à partir de cette position a nécessairement pour effet s'augmenter l'ophélimité dont jouissent certains individus et de diminuer celle dont jouissent d'autres… »[3]. Si on s'écarte d'une telle répartition, il y a nécessairement des gagnants et des perdants. Le vocabulaire de Pareto est évidemment tombé en désuétude ; aujourd'hui, on dit d'une telle répartition qu'elle est *Pareto-optimale*.

Pareto énonce et démontre mathématiquement que lorsque l'équilibre général concurrentiel est atteint, les membres de la collectivité considérée jouissent du maximum d'ophélimité au sens

[1] Contrairement à Walras, Pareto ne distingue pas entre la monnaie et le numéraire, ce qui est une régression
[2] On aurait aussi pu dire (m-1) conditions du type 2 et *n* conditions du type 3. Cela reviendrait au même.
[3] Pareto [273] p. 354.

défini ci-avant. Walras aboutissait déjà à la même conclusion, mais le critère de Pareto pour évaluer le caractère optimal d'une répartition est plus rigoureux que la construction walrassienne[1].

Le maximum d'ophélimité rappelle immanquablement la *courbe de contrat* d'Edgeworth. Une courbe de contrat réduite autant que possible puisque les agents parétiens prenaient les prix comme donnés. Comme nous le verrons au chapitre 4.2, la répartition d'équilibre est dépendante de la répartition initiale des biens entre les agents. Il y a donc potentiellement de nombreuses répartitions *Pareto-optimales*.

3.2.7. Fisher, l'épargne et l'investissement

Les apports d'Irving Fisher sont multiples. Dès 1892, il publiait « Mathematical Investigations in the Theory of Value and Price », consacré à la théorie de la valeur et des prix. Comme nous le verrons plus loin, il apporta également beaucoup à la macroéconomie. Mais c'est à sa théorie du capital et de l'intérêt que nous nous attacherons ici. Cette théorie fut d'abord présentée dans des ouvrages parus en 1906 et 1907 et ensuite synthétisée très didactiquement dans sa « Theory of Interest » parue en 1930, livre que nous allons résumer ci-après.

Les mécanismes très complexes des économies développées cachent parfois la simplicité des principes fondamentaux de l'économie. C'est le cas de tout ce qui touche aux investissements. Un revenu peut être soit consommé soit investi. L'opposition entre ces deux actions est moindre qu'il n'y paraît. Il n'y a de revenu que là où il y a de l'utilité (« enjoyment »). Cette vérité vaut pour l'investissement autant que pour la consommation. La différence entre les deux se ramène à l'intervalle de temps qui nous sépare de la satisfaction du besoin. Si j'achète des pommes, la satisfaction est quasi immédiate, si j'investis dans un verger, elle est reportée au moment où des pommes seront effectivement mangées.

Fisher combat ce qu'il considère comme une méprise répandue, à savoir que le taux d'intérêt serait au capital ce que le salaire est au travail. Selon cette conception, un capital rapporte un revenu annuel résultant de la multiplication de sa valeur par le taux d'intérêt. Fisher prend le problème par l'autre bout : les biens capitaux rapportent à leur propriétaire un flux de revenu, -selon le cas- fixe ou variable d'année en année. La valeur du capital n'est justement que l'actualisation de la somme de ces revenus futurs au taux d'intérêt. Fisher inverse donc la causalité entre la valeur du capital et son revenu.

LA PREFERENCE POUR LE PRESENT

Fisher fait reposer sa théorie sur deux piliers : la *préférence pour le présent* et *l'opportunité d'investir*. Etudions-les successivement.

Pour son premier pilier, Fisher s'est évidemment inspiré de Böhm Bawerk, dont le nom figure dans la dédicace de l'ouvrage. La valeur attribuée à un bien présent est supérieure à celle qu'on accorde au même bien livrable plus tard. Fisher fait lui-même référence à la sous-estimation du futur, le deuxième motif invoqué par Böhm Bawerk pour justifier cette préférence. Fisher utilise indistinctement les expressions « time preference » ou « impatience ». Pour la justifier, il articule sa démonstration en deux arguments consécutifs : d'abord, les biens de production : on

[1] Rappelons-nous la critique de Wicksell contre Walras : comme la théorie parétienne est plus prudente, elle échappe à cette critique. Mais Wicksell reproche à la théorie de Pareto de n'être qu'une tautologie ; selon lui, on peut résumer sa théorie par cette évidence que les agents continuent à échanger tant qu'ils y ont avantage.

préfère en disposer immédiatement plutôt que, par exemple, dans un an, car sinon, nous ne percevrions leur revenu qu'un an plus tard. Mais pourquoi préférer ce revenu plus tôt, si ce n'est parce la satisfaction de nos besoins, rendue possible par ce revenu, est-elle aussi préférée plus tôt que plus tard ? « The manufacturer prefers present weaving than future weaving because the earlier the weaving takes place, the sooner will he be able to sell the cloth and realize his enjoyment income (…) All preference, therefore, for present goods resolves itself, in the late analysis, into a preference for early enjoyment income over deferred enjoyment income »[1].

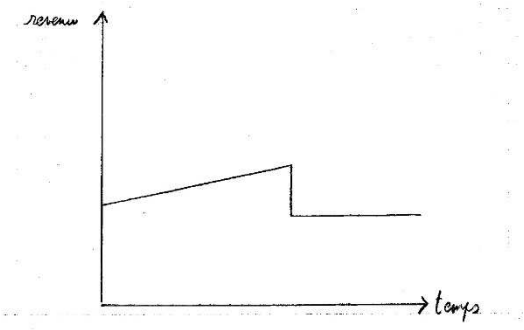
Pratiquement, comment s'incarne cette impatience ? Elle est visible et mesurable. Fisher parle du *taux d'impatience*. Si pour disposer de €100 aujourd'hui, je suis d'accord d'abandonner €110 dans un an, mon taux d'impatience est de 10%. Si une contrepartie de 95 dans un an me suffit, le taux d'impatience est de –5%. Une impatience négative est concevable dans des cas particuliers mais la préférence pour les biens présents implique que son signe normal est positif.

Le taux d'impatience varie d'un individu à l'autre ainsi que pour un même individu suivant les circonstances. De quels facteurs dépend-il ? De la personnalité de l'individu et des caractéristiques du flux de revenus attendus.

L'individu plus prudent aura un taux d'impatience plus bas car il voudra toujours « conserver une poire pour la soif ». La capacité de résister aux multiples tentations de la société de consommation joue également. Mais interviennent également des facteurs personnels non caractériels. Souhaiter aider ses descendants et leur laisser un héritage fait baisser le taux d'impatience.

Mais les caractéristiques du revenu futur de l'individu sont les plus importantes pour la détermination de son taux d'impatience : l'ampleur du revenu, la forme du flux de revenu (« income shape ») et le risque qui lui est lié. L'ampleur du revenu joue en ceci qu'un revenu trop faible force l'individu à une impatience totale. La priorité doit être assurée à la survie physique, qui peut nécessiter la consommation immédiate du revenu total. Le fait d'être constamment dans le besoin n'incite pas non plus à voir à long terme. Par contre, un individu, qui possède une fortune tellement élevée qu'il serait vain de la consommer immédiatement, pourrait avoir un taux d'impatience négatif.

Graphique 3.19 : la forme du revenu



Comme exemple, voici la forme qu'aurait le flux de revenu d'un individu assuré de recevoir un salaire dépendant de l'ancienneté et ensuite une pension de retraite.

La *forme du revenu* est sans doute le critère le plus influent. Un individu peut s'attendre à un revenu qui sera constant, croissant ou décroissant ou qui connaîtra une alternance de phases

[1] Fisher [93] pp. 102-103

différentes. Le fait de s'attendre à un revenu futur supérieur au revenu présent pousse l'impatience à la hausse et le cas inverse calme l'impatience.

Le dernier facteur est le risque. Un revenu plus risqué doit inciter à la prudence.

L'OPPORTUNITE D'INVESTIR

Il s'offre normalement au détenteur d'un capital le choix entre plusieurs opportunités pour son investissement. Par exemple, le propriétaire d'un terrain peut l'affecter soit à l'agriculture, soit à la construction de logements, soit à l'exploitation forestière. Choisir entre ces opportunités d'investissement, c'est choisir entre des flux de revenus qui diffèrent tant par leur ampleur que par leur forme. Par exemple, les revenus des cinq premières années de l'exploitation forestière pourraient être –5, 0, 5,10, 10 et ceux de l'exploitation agricole 4, 3, 4, 3, 4. Le meilleur choix est celui qui maximise la valeur actuelle du flux de revenus, compte tenu du taux d'intérêt en vigueur. Lorsque ce taux est élevé, les opportunités rapportant à court terme (l'exploitation agricole dans notre exemple) sont avantagées ; lorsqu'il est bas, l'avantage peut revenir à celles dont les revenus sont plus différés.

Imaginons maintenant qu'il y ait une infinité d'opportunités différentes. D'après Fisher, cela correspond à la réalité. Les activités dans lesquelles peut se tenir l'investissement sont nombreuses, mais en outre, l'intensité de l'investissement dans chacune d'elles inclut toute une gradation, dont les nombreux niveaux (investir 100, 105, 110… dollars ou milliers de dollars) impliquent une prise de décision.

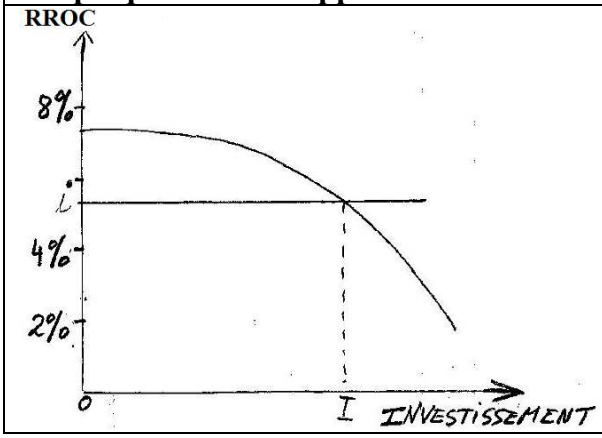
Pour établir la règle de l'investissement optimal, Fisher crée un concept ad hoc : le *taux de rendement au-dessus du coût* (RROC) [1]. A chaque niveau de la gradation dans l'investissement, le coût initial devient plus élevé et, normalement, un rendement total s'élève également. Le RROC est le taux d'escompte qui égalise les valeurs actuelles des flux de revenus nets de deux niveaux successifs dans la gradation de l'investissement. Comme les activités sont réalisées dans l'ordre décroissant de l'avantage rapporté, à chaque avancée dans la gradation, le RROC baisse : « each successive choice compared with its predecessor follows the law of decreasing return » [2].

Cette décroissance du RROC avec le montant investi le rapproche progressivement du taux d'intérêt. L'investisseur subirait une perte en réalisant une opportunité dont le RROC est inférieur au taux d'intérêt. La règle d'optimisation de l'investissement consiste donc à égaliser le RROC marginal (celui de la dernière opportunité retenue) avec le taux d'intérêt.

[1] Rate of Return Over Cost

[2] Fisher [93] p. 198. Le concept de RROC tel que défini ici pourrait sembler inutilement compliqué. N'eût-il pas suffi de prendre en considération le *taux interne de rentabilité* de chaque investissement calculable sans comparaison avec un investissement alternatif ? Le lecteur trouvera l'explication au sous-chapitre 9.6.2.

Graphique 3.20 : les opportunités d'investissement et le RROC



L'EQUILIBRE INDIVIDUEL

L'équilibre individuel associe l'optimisation aux niveaux de l'impatience et de l'investissement. Concernant l'investissement, nous savons déjà qu'il faut appliquer toutes les opportunités d'investissement jusqu'à l'égalisation du RROC avec le taux d'intérêt. Du point de vue de l'impatience, l'individu dispose également de moyens d'améliorer sa situation : l'emprunt et le prêt.

Logiquement, l'agent a avantage à emprunter si le taux d'impatience dépasse le taux d'intérêt ; il gagnerait à prêter dans le cas inverse. Prenons l'exemple d'un taux d'intérêt de 5%. L'individu dont le taux d'impatience est de 3% estime que €100 aujourd'hui valent €103 dans un an, alors que sur le marché, il trouvera des emprunteurs potentiels prêts à lui en rembourser 105; l'agent dont le taux d'impatience est de 10% estime que €100 aujourd'hui valent €110 dans un an, alors qu'il peut se les procurer moyennant un remboursement de €105 sur le marché des prêts. Emprunter et prêter reviennent à modifier la forme du revenu ; l'emprunteur augmente son revenu présent au détriment de son revenu futur et réduit donc son taux d'impatience ; le prêteur accroît son taux d'impatience. Ces prêts auront pour effet d'égaliser les taux d'impatience de tous les agents avec le taux d'intérêt. Lorsque cette égalité sera atteinte, les agents ne pourront plus améliorer leur situation.

L'optimum individuel est donc la combinaison adéquate de ces actions : investir et prêter ou emprunter, de façon à égaliser le RROC et le taux d'impatience avec le taux d'intérêt. Fisher adopte le point de vue de l'entrepreneur individuel : si le RROC de son affaire est inférieur au taux d'intérêt, il doit privilégier les placements externes à son entreprise ; si son RROC est supérieur, il doit investir jusqu'à égalisation, quitte à emprunter s'il ne possède pas les fonds.

Fisher reconnaît que dans la réalité, les agents auront des réticences et même des difficultés à emprunter aussi souvent que la règle optimale le leur commande, et ce pour deux raisons principales :

1. les emprunteurs qui n'ont pas de garanties à offrir ne trouvent pas de prêteur
2. les empunteurs craignent les énormes désagréments que leur causerait, le cas échéant, la situation d'insolvabilité.

L'EQUILIBRE DU MARCHE

La règle veut que les agents prêtent, empruntent et investissent de façon à égaliser leur taux d'impatience et leur RROC avec le taux d'intérêt comme si celui-ci était donné. Mais en même temps, insensiblement, toutes ces actions influencent le taux d'intérêt. L'équilibre du système n'implique aucun raisonnement circulaire ; ce type d'interdépendance peut être traité par un

système d'équations linéaires indépendantes, à la walrassienne. Comme son prédécesseur, Fisher met le problème en équations et montre qu'il comporte une solution parce que le nombre des équations et celui des inconnues se valent. Le taux d'intérêt résulte de la conjonction de forces subjectives (l'impatience) et de forces objectives (le rendement des investissements). Le jeu de l'offre et la demande fait émerger le taux d'équilibre.

Les inconnues sont :

- les soldes des opérations sur le marché financier par les *n* agents chacune des *m* années. Nombre = *m.n*
- les taux d'intérêt de chaque année par rapport à la précédente. Nombre = *m-1*.
- les taux d'impatience de chaque individu d'une année à la suivante. Nombre = *n.(m-1)*
- les revenus annuels des investissements de chaque individu. Nombre = *n.m*
- les RROC de chaque agent pour chaque année sauf la dernière. Nombre = *n.(m-1)* ; ces inconnues peuvent être dérivées des précédentes et ne sont pas reprises dans le décompte.

Les cinq groupes d'équations disent que :

- les taux d'impatience (entre deux périodes consécutives) sont fonctions des revenus attendus. Nombre = n.(m-1)
- entre deux périodes consécutives, les taux d'impatience s'égalisent avec le taux d'intérêt. Nombre = n.(m-1)
- chaque année, il y a égalité entre l'argent que certains reçoivent sur le marché des prêts et l'argent que d'autres payent sur ce marché. Nombre = *m*
- pour chaque agent, il y a égalité entre les sommes empruntées et la valeur actualisée des remboursements. Nombre = *n,* dont on en retire une car ces équations ne sont pas indépendantes de celles du groupe précédent
- chaque agent a une série de revenus annuels, fonction de son choix d'investissement. Nombre = *n*
- entre deux périodes consécutives, les RROC s'égalisent avec les taux d'intérêt. Nombre = n.(m-1).

L'équilibre conçu par Fisher implique qu'il y ait un taux d'intérêt particulier à chaque année entre cette année et la suivante. Certes, tous ces taux ne se rencontrent pas dans l'économie réelle. Par exemple, il n'existe pas en 2003 un taux pour les prêts de 2005 à 2006. Fisher maintient que les taux rencontrés dans la vie sont des aménagements à partir des taux annuels tels qu'ils apparaissent dans son système d'équations. Par exemple, le taux de 2003 à 3 ans n'est que la moyenne des taux annuels implicites 2003-2004, 2004-2005 et 2005-2006.

L'économie réelle offre une énorme variété de placements (actions, obligations, comptes de dépôt…) et les niveaux de risque afférents à ces placements sont également variables. Cette multiplicité ne contredit pas la théorie. Le taux d'intérêt théorique tel qu'il sort du système d'équations se décline simplement en une multitude de taux pour tenir compte de la variété des niveaux de risque. Un placement plus risqué doit rapporter un taux d'intérêt supérieur.

Concluons. Comme clin d'œil à Marshall, Fisher évoque la paire de ciseaux pour illustrer que ni l'impatience, ni le rendement des investissements ne suffisent à bâtir une théorie de l'intérêt ; les deux aspects, le subjectif et l'objectif, sont complémentaires. Fisher semble toutefois privilégier le principe d'opportunité. Il donne ce cas d'école : sur une île déserte, une seule activité serait possible, dont le rendement est physiquement déterminé : le stockage de produits non périssables. Dans ce cas, seul un taux d'intérêt nul pourrait prévaloir, quels que soient les degrés d'impatience des individus. Si l'activité unique concernait le stockage de produits périssables, le seul taux d'intérêt possible serait négatif et correspondrait au taux de détérioration. Il en va de même pour les activités rentables : « Man can obtain from the forest or

the farm more by waiting than by premature cutting of trees or by exhausting the soil. In other words, Nature's productivity has a strong tendency to keep up the rate of interest »[1].

Fisher reconnaît lui-même qu'il est impossible de démontrer que le signe normal du taux d'intérêt est positif, sur base du seul principe d'impatience en l'absence du principe d'opportunité. Vu les opportunités d'investissement, un taux d'intérêt nul ou négatif mènerait à une situation où il y aurait beaucoup de candidats emprunteurs par rapport aux candidats prêteurs, ce qui pousserait le taux à la hausse. On retrouve un peu le même paradoxe qu'avec Böhm Bawerk ; d'abord, ces auteurs insistent sur l'impatience ou la sous-estimation du futur. Ensuite, ils construisent un modèle de la détermination du taux d'intérêt où ce facteur ne joue qu'un rôle (très ?) accessoire dans le fait qu'il est positif.

TAUX D'INTERET MONETAIRE ET TAUX REEL

Dans les sections précédentes, « the rate of interest was described as the percentage premium on present goods over future goods of the same kind ». Mais habituellement, c'est dans un bien particulier qu'est exprimé le taux d'intérêt : la monnaie. Si la valeur de la monnaie ne garde pas un rapport stable avec celle des autres biens, il s'opère un dérapage entre le *taux d'intérêt monétaire* et celui qui ressort des sections précédentes, que nous appellerons le *taux d'intérêt réel*. En fonction du bien ou du panier de biens choisis comme référence, il y a potentiellement une multitude de *taux d'intérêt réels*. Pour déterminer celui-ci, Fisher préconise de rapporter le *taux d'intérêt monétaire* au coût de la vie.

C'est le *taux d'intérêt réel* qui intéresse les prêteurs et les emprunteurs. S'ils anticipent une dépréciation de la monnaie (hausse du coût de la vie), ils adapteront le taux monétaire en conséquence. Normalement, tout pourcent de hausse du coût de la vie prévu par les agents économiques doit produire un pourcent de hausse du *taux d'intérêt monétaire* et tout pourcent de baisse du coût de la vie prévu par les agents économiques doit produire un pourcent de baisse du *taux d'intérêt monétaire*. Il y a toutefois une exception : le *taux d'intérêt monétaire* ne peut devenir négatif : la monnaie étant un bien qui peut être conservé indéfiniment quasiment sans coût, il serait alors préférable de thésauriser la monnaie plutôt que de la prêter.

Fisher a développé une analyse statistique très sophistiquée pour déterminer si la réalité conforte la théorie ci-dessus. La conclusion est décevante. Les hausses de prix déclenchent certes des hausses du taux d'intérêt nominal, mais ces hausses s'étalent sur de nombreuses années avant que la compensation soit effective. L'écart-type du *taux d'intérêt réel* est sensiblement supérieur à celui du taux monétaire alors qu'on s'attendrait à l'inverse. Il attribue cette situation à un « universal lack of foresight ».

3.2.8. Différentes conceptions du capital

Tous les économistes du titre III représentent l'école néoclassique, mais ils nous offrent une palette variée de conceptions du capital.

CAPITAL PHYSIQUE OU CAPITAL VALEUR ?

Nous avons vu comment **Wicksell** met en avant la double nature du capital, à la fois collection physique hétérogène et fonds financier, ainsi que ses efforts pour venir à bout de cette difficulté.

[1] Fisher [93] p. 193

Les autres économistes sont moins conscients du caractère problématique de cette dualité. En témoigne la divergence entre les chemins qu'ils empruntent. Clark, Böhm Bawerk et Marshall considèrent le capital comme un fonds financier ; Clark distingue explicitement le capital des biens-capitaux, ceux-ci ne jouant qu'un rôle accessoire. Walras, Menger et Wicksteed prennent la direction inverse, axée sur les biens capitaux ; ceux-ci se voient attribuer individuellement une productivité marginale déterminant leur rémunération.

Ce que **Walras** appelle « un capital » est un *bien capital* au sens de Clark. Le véritable facteur de production n'est pas tant le capital que les services qu'il rend. Leur rémunération est une espèce de loyer fictif incluant l'intérêt et l'amortissement.

Ce loyer fictif rémunère aussi les nombreux biens-capitaux qui constituent des biens d'ordre supérieur chez Menger ou des inputs de la fonction de production chez Wicksteed. L'attente constituant un facteur à part chez ces deux auteurs, le loyer fictif des biens capitaux ne devrait comporter que l'amortissement, sans l'intérêt. L'attente et la charrue sont toutes deux présentes dans la fonction de production de Wicksteed. L'intérêt égale la productivité marginale de la première ; la productivité marginale de la seconde est mal définie. La présentation de Walras semble plus cohérente.

Walras, **Menger** et **Wicksteed** incluent dans le revenu à partager entre les facteurs, la totalité des recettes des entreprises ; une partie de ce revenu n'est pas un revenu net mais un retour des avances. Clark n'envisage que la *valeur ajoutée* par l'entreprise ; celle-ci se partage entièrement entre salaires et profit. Le fermier perçoit son profit après avoir rétribué ses ouvriers ; le prix de la charrue quant à lui se répartit entre le profit de l'industriel qui l'a vendue et les salaires de ses ouvriers. Le métal de la charrue se répartit entre les profits du maître des forges et les salaires des ouvriers sidérurgistes. Chaque dépense des entreprises correspondant à l'acquisition de produits peut être ramenée au duo profit & salaire. Le revenu national tout entier peut être décomposé en ces deux rémunérations. La conception du capital en tant que fonds financier, plus en phase avec la comptabilité nationale, me paraît la plus sensée.

Chez **Walras**, pour obtenir la valeur globale du capital social, il suffit de multiplier la quantité de chaque capital par le prix correspondant et de sommer. En se préoccupant des deux natures du capital, contrairement à ses collègues, Wicksell permet à leurs grandeurs d'évoluer disproportionnellement. Un accroissement du capital fait hausser la demande de travail[1], ce qui entraînera la modification du salaire et du taux d'intérêt et donc du rapport entre le volume physique et la valeur du capital.

CAPITAL TEMPS OU CAPITAL STOCK ?

Clark ne dut pas apprécier une théorie qui nie que le capital est un facteur de production et que l'intérêt est sa rémunération. En 1895, il prend l'initiative d'une controverse qui l'oppose à **Böhm Bawerk** par articles de presse.

Clark reconnaît que la création de capital implique une attente. Si je démarre une exploitation forestière, il me faudra effectivement attendre les fruits de l'investissement et plus les arbres plantés sont d'une variété lente à croître, plus j'attendrai. Mais une théorie de l'intérêt doit rester valable dans le cadre de l'analyse statique, dont nous savons qu'elle servait de perspective à Clark. Une fois que l'attente initiale arrive à terme, chaque année, je coupe une rangée d'arbres et je la replante. Peu importe alors que la durée de croissance d'un arbre soit de

[1] En supposant la constance des méthodes et techniques de production.

vingt ou de quarante ans. Dans le cadre de l'analyse stationnaire, la notion de période de production ne correspond à rien. Il n'y a pas d'attente : à chaque moment, simultanément nous mangeons des fruits et entretenons des arbres fruitiers. Comment Böhm Bawerk expliquerait-il l'intérêt quand l'attente ne joue pas ? C'est la conception *synchrone* de la production qui s'oppose à celle de l'attente.

Le capital, rappelle Clark, en même temps qu'il crée un revenu qui s'appelle l'intérêt, se renouvelle lui-même automatiquement. Le prix de vente des produits couvre leur prix de revient, dont fait partie l'amortissement des équipements. Le renouvellement d'un matériel arrivant en fin de vie se fait donc naturellement, sans épargne, financé par les seules transactions commerciales de l'entreprise. On chercherait en vain la nécessité d'une épargne dans une économie stationnaire, ce qui disqualifie totalement l'attente de Böhm Bawerk.

Clark a dans son viseur notamment la *période de production*. Selon lui, elle tendrait couramment vers l'infini. Il écrit : « If we are to trace the origins of any goods to the earliest labour that has contributed to making them, we shall follow the series of instruments backward to a point at which no capital existed »[1]. Et il voit en Adam et Eve des producteurs d'une part infime de tous les biens capitaux existants. Clark oublie que pour Böhm Bawerk, ce n'est pas la période ABSOLUE qui compte, mais la période MOYENNE.

Clark conteste également que le taux d'intérêt soit fonction de la période de production. Prenons un capital *A* se montant à 1000 et ayant une période de production de deux ans et un capital *B*, également de 1000, ayant une période de production de quatre ans. S'ils ont la même productivité marginale, ils rapporteront nécessairement un intérêt identique et leurs périodes de production divergentes seront sans effet ; celle-ci n'est donc pas le déterminant du taux d'intérêt. Clark oublie que Böhm Bawerk se plaçait au niveau macroéconomique. La population active et le stock de capital étant donnés, la période de production et le taux d'intérêt sont tous deux des variables endogènes. Dans cette optique, comparer des capitaux identiques avec des périodes de production différentes n'a pas de sens.

Autre point de friction : l'abstinence du capitaliste. Clark, qui reprend l'idée de Senior et Mill, en fait l'origine du capital. Böhm Bawerk le conteste et précise que ce qu'il appelle ATTENTE n'a rien à voir avec le concept d'ABSTINENCE. Rappelons-nous qu'il prête la préférence pour les biens présents sur les biens futurs non pas aux capitalistes mais aux salariés. D'ailleurs, l'abstinence des capitalistes est une fausse piste : « saving can involve sacrifice and meritorious action, but it doesn't necessarily do so. The man who has only a small income will concededly not be able to lay aside something of the little he has without appreciable deprivation and the display of vigorous self-control. But the man with an annual income of one million, who is content to live on only a half million while he lays aside the other half million, need perform no heroic act of renunciation. The formation of capital requires only that the saving be a factual reality and any moral admixture of sacrifice or merit is a matter of complete indifference so far as the result is concerned »[2]. Clark maintient qu'il y a abstinence : « By extending the lines of railroad that he owns, the multi-millionaire denies himself the enlarging of his palatial residences »[3].

[1] Clark [53] p. 270.
[2] Böhm Bawerk [374] p. 115. Menger avait déjà exprimé le même avis.
[3] Clark [52] p.98.

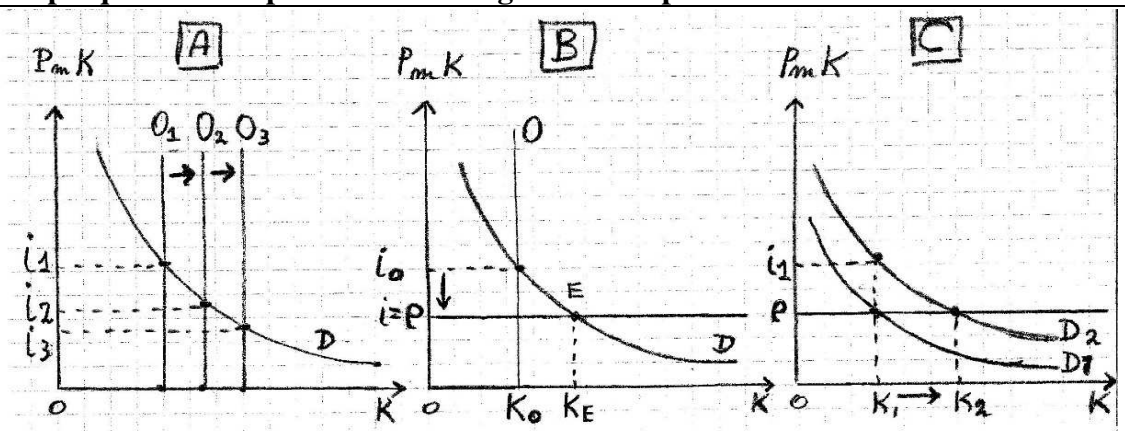
Cette controverse rebondira dans les années trente, avec Frank Knight reprenant le flambeau de Clark et Hayek celui de Böhm Bawerk. Nous rendrons compte de cet épisode dans le chapitre 9.2 « Faut-il croire en la théorie autrichienne ? ».

CAPITAL OU INVESTISSEMENT ?

Selon Clark, la productivité marginale du capital est à l'intérêt ce que la productivité marginale du travail est au salaire. Mais comment cette théorie opère-t-elle pour transformer la productivité marginale en taux d'intérêt ? Clark fut peu explicite à ce sujet. On trouve une réponse à cette question dans l'article « A Mathematical Theory of Saving » de Frank **Ramsey**[1] (1928).

La courbe de productivité marginale correspond à la demande de capital par les entreprises. Mais quid de l'offre ? Ramsey l'assimile au stock de capital existant. L'offre de capital étant donnée à court terme, elle apparaît sur la figure 3.21-A comme une droite verticale. A long terme, le stock de capital varie du fait de l'épargne investie, ce que le graphique montre par le déplacement latéral de l'offre verticale. A tout instant *t*, le taux d'intérêt se détermine par l'intersection de l'offre et de la demande de capital.

Graphique 3.21 : la productivité marginale du capital et le taux d'intérêt



En abscisse, K représente la quantité de capital ; en ordonnée figurent la productivité marginale, le taux d'escompte du futur et l'intérêt.

Le taux d'intérêt tel que déterminé ci-dessus n'est pas nécessairement le taux d'équilibre à long terme. Celui-ci est atteint lorsqu'il n'y a plus d'incitation à l'investissement (ou au désinvestissement). La donnée fondamentale est ce que Ramsey appelle le *taux d'escompte du futur* qu'on peut rapprocher du taux d'impatience fisherien. Supposons, comme sur la figure 3.21-B, que le capital de la société vaille K_0, avec un taux d'intérêt de i_0. A ce taux, l'incitation à épargner est élevée, parce que le taux d'intérêt dépasse le taux d'escompte du futur, noté ρ. L'investissement de cette épargne pousse progressivement le capital à la hausse avec un taux d'intérêt descendant le long de la demande. Ce processus durera jusqu'à ce que le taux d'intérêt atteigne *i* qui est son niveau d'équilibre où il égale le *taux d'escompte du futur*. Ramsey pense que le taux d'intérêt peut rester durablement au dessus de son niveau d'équilibre.

[1] Frank Plumpton Ramsey (1903- 1930), brillant intellectuel de l'université de Cambridge, fut philosophe, mathématicien et économiste. Mort à 26 ans, il eut le temps d'introduire le philosophe linguiste autrichien Wittgenstein en Angleterre et de le traduire dans la langue de Shakespeare. Deux de ses contributions économiques marquèrent cette science : un ouvrage sur le lien entre les probabilités et l'incertitude (1926) et le présent article de 1928, sur lesquels nous reviendrons ultérieurement.

Dans l'analyse statique, l'équilibre est un point d'inertie. Evidemment, les changements incessants de l'économie réelle affectent cet équilibre. Par exemple, le progrès technique fait hausser la productivité marginale, ce qu'illustre le passage de la courbe D_1 à la courbe D_2 sur la figure C. Dans ce cas, l'intérêt augmente jusqu'au niveau i_1 qui est au dessus de son niveau d'équilibre à long terme. L'accumulation du capital est de nouveau stimulée, faisant revenir le taux d'intérêt au niveau ρ, mais cette fois avec un capital total plus élevé (K_2).

Cette théorie peut déconcerter : l'économie néoclassique ne nous a pas habitués aux courbes d'offre inélastiques. Pourquoi en arrive-t-on là ? Ramsey écrit : « The difficulty is that the rate of interest functions as a demand price for a whole quantity of capital but as a supply price, not for a quantity of capital, but for a rate of saving »[1]. En le comparant avec leur taux d'escompte du futur, les ménages utilisent le taux d'intérêt pour déterminer leur épargne annuelle qui est un *flux*. De l'autre côté, les entreprises ajustent leur capital total, qui est un *stock*, pour égaliser la productivité marginale avec le taux d'intérêt. Il y a là un hiatus.

La théorie de **Fisher** évite cette difficulté, car elle raisonne en termes de flux, tant du côté de l'épargnant que de celui du producteur. Le taux d'intérêt gouverne ces flux en s'égalisant à la fois avec le taux d'impatience et avec le RROC.

Entre le capital et l'investissement, on a la relation suivante :

$$I_t = \Delta K = K_{t+1} - K_t \qquad (3.38)$$

L'investissement de l'année *t* égale le capital en l'année t+1 moins le capital en l'année *t*. Il s'agit évidemment de l'investissement NET, qui prend en compte la détérioration du matériel.

Chez Ramsey, les entreprises déterminent K_{t+1} par leur politique d'optimisation ; I_t en découle puisque K_t est connu. Chez Fisher, l'optimum conduit les firmes à déterminer I_t ; K_{t+1} se déduit de K_t et I_t. Il semble donc que les entreprises ne puissent avoir en même temps une politique d'investissement optimal et une politique optimisant leur capital. Le caractère énigmatique de cette proposition stimulera les économistes. Le débat qui en résultera est analysé au chapitre 9.6.

[1] Ramsey [293] p. 556.

3.3- LES NEOCLASSIQUES FACE AUX CLASSIQUES

Les promoteurs de la pensée néoclassique ont initié une nouvelle théorie de la valeur et de la répartition. Il n'est pas étonnant qu'ils aient jugé la théorie qui précédait la leur et que l'appréciation fut plutôt négative.

Critiquant la théorie ricardienne selon laquelle la valeur est déterminée par le coût de production, **Jevons** écrit : « Repeated reflection and inquiry have led me to the somewhat novel opinion that value depends entirely upon utility ». Il se réfère évidemment à l'utilité MARGINALE, mais celle-ci varie avec la quantité consommée ; le coût de production aurait donc quand même un rôle dans la détermination de la valeur s'il influençait ladite quantité. Sans doute est-ce le sens du fameux aphorisme de Jevons :

> Cost of supply determines supply
> Supply determines final degree of utility[1]
> Final degree of utility determines value

Marshall consacre un chapitre de ses « Principles » à la théorie de la valeur de Ricardo. Finalement, il ménage respectueusement le vieux maître et adresse ses flèches à la critique jevonienne. Notamment, il tourne en dérision le couplet ci-dessus ; si on retire les éléments intermédiaires, il reste que le coût de production détermine la valeur. On se rappelle la thèse de Marshall : le prix de marché vient de l'égalisation de l'offre avec la demande ; la demande est réglée par l'utilité et l'offre par le coût. La vision de Jevons est donc aussi « one-sided and fragmentary » que celle de Ricardo et même « much more misleading »[2].

L'attaque la plus virulente mais aussi la plus étayée est le fait de **Walras** qui consacre trois leçons respectivement à la théorie de la valeur, à la rente foncière et à la rémunération des facteurs de production.

On se souvient de la distinction opérée Par Ricardo et Mill entre les biens singuliers dont le prix dépend de l'offre et la demande sans autre qualification et les biens reproductibles dont le prix tend vers le coût de production. Walras adresse un double critique :

1- Aucune production ne peut être multipliée, selon les termes utilisés par Walras « à l'infini » ou « sans limite ». Il emprunte ce « sans limite » à une citation de Mill.

   Son insistance est excessive, car la multiplication à l'infini est une condition excessive. L'essentiel que la circulation des facteurs entre les industries ne soit pas entravée.

2- Les classiques ne dévoilent pas les éléments exogènes qui devraient déterminer ce coût qui attire le prix à lui. Selon Walras, ce coût qui précède le prix n'existe tout simplement pas ; au contraire, la rémunération des facteurs s'adapte aux prix des produits : si la demande élevée d'un produit élève son prix, les entreprises concernées demanderont massivement des facteurs qui participent à cette production et ces facteurs verront leur prix augmenter.

   Ici, très judicieusement, Walras se rend compte qu'il faut distinguer deux cas : certains facteurs sont spécifiques à une industrie et d'autres sont généralistes et susceptibles d'être transférés d'une industrie à l'autre. Les premiers voient leur prix déterminé définitivement par le prix du produit, alors que cette influence n'est que temporaire pour les seconds. L'argument de Walras manque donc partiellement son but : la théorie ricardienne de la

[1] *Final degree of utility* est le terme utilisé par Jevons pour l'*utilité marginale*.
[2] Marshall [253] § I.14

valeur est une théorie d'équilibre à long terme. La dépendance à court terme des facteurs généralistes vis-à-vis du prix des produits ne la contredit donc pas.

Walras s'attaque ensuite à la théorie de la rente, envers laquelle il émet une triple critique.

1. Prenant l'exemple de Ricardo des investissements successifs de 1000 livres à la fois sur de nouvelles terres et sur des terres déjà cultivées, Walras met le problème en équations et fait du capital une variable continue et non plus discrète. Il constate que dans ces conditions, la terre marginale apporte également une rente, sauf cas particulier.

   Kurz et salvadori, défenseurs de Ricardo, démontrent que la rente marginale non nulle ne découle pas de la variable CONTINUE, mais qu'elle est liée à la rente intensive[1].

2. Ricardo fait intervenir l'investissement de capitaux successifs d'une même VALEUR alors que la productivité marginale décroissante concerne les éléments PHYSIQUES de ce capital ; il y a donc une hypothèse implicite que le prix de ces éléments de capital reste constant, ce qui réduit la portée de cette théorie. Si leur prix connaissait une baisse sur le long terme, des doses identiques de capital en valeur pourraient maintenir leur productivité.

   Selon Kurz et Salvadori, la logique de l'argumentation ricardienne implique des doses successives en unités physiques identiques. Ricardo aurait commis une erreur d'expression.

3. Ricardo confère à la rente un statut spécial a priori. Il en explique la détermination, l'intérêt et le salaire étant donnés ; mais pourquoi pas l'intérêt sur base de la rente et du salaire ou le salaire sur base de l'intérêt et de la rente ?

Cette dernière question nous amène à la critique walrassienne contre la théorie classique de la rémunération des facteurs. Il reproche à cette école de n'avoir pas de théorie d'ensemble à cet égard. Il s'attaque à la *théorie du fonds des salaires*, énoncée par Mill mais à laquelle Ricardo semble préférer l'idée que le salaire est rivé au minimum vital. Concernant le profit, Walras reproche aux classiques de n'avoir pas distingué l'intérêt rémunérant le capitaliste et le profit rémunérant l'entrepreneur. Même lorsque les deux fonctions reposent sur une même tête, la distinction reste utile. Comment Ricardo déterminait-il le profit ? On a l'équation P = I + S où *P* est la valeur produite, *I* est l'intérêt et *S* est le salaire ; la rente ne fait pas partie du coût. Cette équation détermine *P* à partir de *I* et *S*. *S* est expliqué. *I* est le surplus laissé par *P* au-delà de *S*. Cette équation unique doit déterminer deux inconnues, *I* et *P*. Il y a donc indétermination. On retrouve ce même argument chez Jevons : « But such a doctrine is radically fallacious : it involves the attempt to determine two unknown quantities from one equation »[2].

Wicksell s'intéresse également à la théorie ricardienne. Son jugement est nuancé. Il analyse la question à deux niveaux :

1- La cohérence de la théorie ricardienne. Il écrit : "On the other hand, the way in which Ricardo develops his argument— totally free from the fantastic ideas and dialectic leaps of many of the later schools—is a model of strictly logical reasoning about a subject which seems, at first glance, to admit of so little precision"[3]. Comme l'écrivent Kurz et Salvadori, Walras n'avait pas entièrement compris la théorie ricardienne qu'il a jugée à l'aune de ses propres principes car il n'avait pas conçu qu'une autre théorie puisse avoir des bases aussi différentes de la sienne. En réalité, le modèle ricardien, comme tout modèle joue avec des

[1] Kurz [203] p. 384.
[2] Jevons [165] § 8.5.
[3] Wicksell [386] p. 35.

variables dépendantes et des variables indépendantes. Ses variables indépendantes permettent bien de déterminer ses variables dépendantes, que sont le prix, le profit et la rente. Parmi les variables indépendantes, il y en a une qui avait échappé à Walras : le volume et la composition de la production sociale. Lorsque Walras met la théorie ricardienne en équations, il l'ignore alors qu'il ajoute une équation très néoclassique mais étrangère à la pensée de Ricardo selon laquelle le total du capital investi est donné. Contrairement à Walras, Wicksell avait parfaitement compris le modèle ricardien.

2- Il reproche à la théorie ricardienne son côté formaliste. Elle est cohérente mais dépend de beaucoup de simplifications qui la rendent irréaliste. Comme Walras, Wicksell estime que Ricardo n'a pas réussi à expliquer le salaire et l'intérêt d'une façon indépendante du prix. Selon lui, la determination est réciproque : "…costs of production and exchange values cannot stand in the simple relation of cause and effect which Ricardo supposes. As we shall see later, they are mutually conditioned like the various elements in a single economic system in equilibrium"[1]. Lorsqu'intervient une modification dans la distribution de la demande entre les biens, et par là des quantités produites, des rentes peuvent s'accroître ou diminuer dans telle ou telle industries. La demande du produit influence alors la rémunération des facteurs.

De notre analyse, il ressort que le ricardisme ne se limite pas à la détermination du prix par le coût de production. Il s'agit d'un modèle d'équilibre global. La prééminence du coût de production peut s'affirmer dans le cadre d'autres alternatives au modèle néoclassique. Nous en étudierons une au chapitre 5.5, alors que les chapitres 5.2 à 5.4 seront consacrés à une résurgence du ricardisme dans la seconde moitié du XXe siècle.

Wicksteed critique lui aussi la conception classique de la valeur, mais il vise à côté de la cible. Son leitmotiv est que le caractère irrévocable de la production terminée fait de son coût effectif une donnée obsolète. Ce qui a été produit à un coût supérieur au prix du marché sera vendu à perte, ce qui fut produit à un coût inférieur rapportera un profit supplémentaire ; l'établissement du prix de marché n'a que faire des productions non rentables ou sur-rentables.

Pour les économistes favorables à la thèse ricardienne, le coût déterminant n'est pas le coût historiquement réalisé mais celui auquel le bien peut voir sa production entamée aujourd'hui et rapporter le profit moyen dans des conditions normales.

Du temps doit s'écouler entre l'émergence de surprofits dans une industrie et le transfert de ressources pour y accroître l'offre et y réduire le prix et la marge. L'optique classique est celle d'un équilibre à long terme. Qu'elle n'explique pas le court terme est une limitation assumée. Il est injuste de lui opposer des arguments centrés sur le court terme.

Wicksteed répète avec insistance que le coût de production, plutôt que de déterminer le prix, joue un autre rôle : il départage les productions qui seront effectuées et celles qui ne le seront pas. Cette idée, loin de contredire la théorie classique, avait déjà été affirmée par Ricardo et Mill. Wicksteed ajoute toutefois cette remarque pertinente : l'infériorité du prix par rapport au coût de production sert souvent de stimulant à l'accroissement de la productivité.

[1] Wicksell [385] t.I p. 26

3.4- UNE VOIX DISCORDANTE

HOBSON CONTRE LA THEORIE MARGINALISTE DE LA REMUNERATION DES FACTEURS

Hobson est un des rares économistes qui ne s'inscrit ni dans la tradition classique, ni dans la tradition néoclassique, bien qu'il fut influencé par les deux. Au début du XXe siècle, il fit figure de résistant à la vague néoclassique. Nous avons vu plus haut qu'il pourrait avoir influé sur la pensée marshallienne.

En 1909, il publie « The Industrial System », un livre dont le thème central est le partage du revenu entre les quatre groupes de facteurs que sont, selon lui : le travail, le capital, la terre et l'entreprise qu'il appelle « ability ». Il part de ce postulat : le volume de la production est fonction de la motivation des facteurs qui, elle-même, dépend de la rémunération. Sur cette base, Hobson divise le revenu en trois fonds :

1- le coût du maintien de la production à l'état stationnaire. Pour chacun des quatre facteurs, il y a une rémunération minimale en dessous de laquelle l'état du facteur se dégrade, ce qui menace le maintien de l'activité à son niveau actuel. Hobson la situe un rien au dessus du minimum vital.
2- Le surplus productif : c'est le surplus de rémunération nécessaire pour stimuler les facteurs à augmenter la production.
3- Le surplus improductif, qu'il appelle « unearned income ». C'est un supplément de revenu en sus des deux précédents qu'un facteur peut obtenir du fait de sa rareté relative. On pourrait le réduire ou l'annuler sans mettre en danger la production.

La rémunération de chacun des quatre facteurs est traversée par cette division en trois fonds. Il y a donc un intérêt gagné et un intérêt non gagné, un profit gagné et un profit non gagné, un salaire gagné et un salaire non gagné…[1] Le même principe vaut pour tous les facteurs, y compris le profit qui rémunère l'entreprise, mais il y a des nuances entre eux. Ainsi, la rente foncière est presque entièrement du *revenu non gagné* car aucun effort n'est nécessaire pour offrir la terre ; les dépenses pour prévenir la dégradation de la terre sont seules comptées dans le coût de maintenance.

Le coût de maintenance du capital n'est pas l'intérêt mais l'amortissement, car l'épargne n'est nécessaire que pour accroître le capital. L'intérêt est donc entièrement inclus dans les fonds deux et trois. Une grande part de l'épargne, celle des plus riches, ne représente pas un effort, mais cette constatation n'invalide pas la nécessité de l'intérêt comme stimulant : « So long as the easy automatic saving of the Duke of Westminster needs to be supplemented by the hard-felt saving of John Smith, of Oldham (NDLR : Monsieur tout le monde), the former must receive the same price per unit of saving as the second »[2]. Le fait qu'il existe d'autres motivations à épargner que l'intérêt n'en dément pas la nécessité non plus. Selon Hobson, qui était socialiste, la problématique de l'équité sociale « applies not to the justice or necessity of paying interest for saving, but to the process by which much of the accumulation which enables saving to be made has been achieved »[3].

[1] Hobson oppose sa théorie des revenus non gagnés à la théorie marxiste qui considère comme provenant de l'exploitation, la totalité des revenus autres que le salaire.
[2] Hobson [157] p. 71
[3] Hobson [157] p. 71

L'impression, correcte, que la croissance économique a besoin d'une grande masse d'épargne ne doit pas conduire à l'illusion, fausse, que l'économie peut investir à long terme n'importe quel taux d'épargne[1]. Deux causes limitent le taux d'épargne assimilable :

- il faudra bien consommer la masse de biens produits à l'aide des investissements
- on ne peut prévoir les besoins et les techniques à très long terme

Il y a un taux d'épargne optimal. Le dépassement répété du seuil d'épargne efficace est source de gaspillage. Hobson est un des promoteurs de la théorie de la sous-consommation (ou de la sur-épargne) pour expliquer les crises périodiques.

Revenons à la répartition du revenu entre les facteurs. La répartition optimale, celle qui permet de maximiser le produit, prévaut rarement. Alors que le partage du fonds de maintenance n'est pas source d'opposition, seul le rapport de forces règle la répartition du surplus. L'atout essentiel d'un facteur est sa rareté, naturelle ou artificielle. A ce jeu, le gagnant est habituellement le facteur entreprise. Il est en position de force dans la négociation avec les autres facteurs ; en outre, les entrepreneurs restreignent la concurrence entre eux pour que le prix de vente de leur produit incorpore un profit non gagné confortable ; l'inclusion de revenus non gagnés dans le prix varie beaucoup d'un produit à l'autre. Selon Hobson, c'est une évidence que le profit réel des différentes firmes échappe à toute corrélation avec le niveau qui serait nécessaire pour motiver leur entrepreneur.

A l'inverse de l'entreprise, le travail, souffre de son abondance, qui le contraint régulièrement à accepter un salaire proche du minimum vital. Néanmoins, certains salaires peuvent s'élever de façon à comprendre un surplus improductif, notamment lorsqu'il y a pénurie de travailleurs qualifiés dans telle ou telle spécialité.

La répartition du revenu dépend des raretés relatives des facteurs. Aucune autre loi économique, et certainement pas de celles qui font intervenir la productivité, ne détermine les rémunérations. C'est dans cette optique qu'Hobson s'attaque à la théorie de la productivité marginale. Il écrit: « So intimate is the interdependence of the factors upon one another, and of the several parts of each factor upon the other parts of its own factor and upon all the parts of the other factors, that no separate productivity can rightly be attributed to any factor, still less to any part of a factor »[2]. Entre les facteurs, entre leurs différentes unités, il y a une « organic co-operation » qui fait du produit un tout indécomposable. L'insistance de Hobson sur le caractère ORGANIQUE des combinaisons de facteurs aurait fait pencher Marshall vers le *produit net*.

Contre le marginalisme, Hobson émet aussi cette autre critique : « It is far more accurate to say that the price per unit causes the margin to be where it is, than to attribute any causative power to the margin, as margin, in relation to the price per unit »[3]. C'est le prix du facteur qui détermine la marge de son utilisation et non la productivité à la marge qui détermine son prix. Mais, comme nous l'avons vu, cette affirmation n'est pas incompatible avec la conception marshallienne où la productivité marginale ne détermine que la DEMANDE du facteur.

De la théorie ci-avant découle l'avis que porte Hobson sur le syndicalisme. « Trade unionism will rank as an organised attempt to divert rents, excessive interests and profit and other

[1] On ne peut généraliser à l'économie globale le cas de l'individu ou de la nation qui peuvent épargner sans limite pourvu qu'ils trouvent des partenaires qui demandent du capital.
[2] Hobson [157] p. 106
[3] Hobson [157] p. 103

'unearned' incomes into wages »[1]. Le plus généralement, des marges sont disponibles ; pour le mouvement ouvrier, tout l'art consiste à être assez fort pour les cueillir.

Hobson apprécie l'intervention de l'Etat en faveur d'une distribution de la richesse plus équitable et plus efficace. Ici aussi, la politique doit viser à détourner les revenus non gagnés. Trois moyens sont envisageables.

1. augmenter les salaires dans les secteurs à hauts profits. Hobson y est néanmoins réticent si le résultat revient à convertir des profits non gagnés en salaires non gagnés.
2. baisser les prix des produits de façon à transférer les surplus à l'ensemble des consommateurs.
3. utiliser le surplus pour financer des services d'intérêt général.

Hobson semble préférer la troisième voie ; deux façons de centraliser le surplus s'offrent à l'Etat :

- être lui-même opérateur industriel et engranger les profits rapportés par ses entreprises. Hobson la préconise pour les industries à monopole naturel.
- appliquer une imposition progressive des revenus, partant de l'hypothèse que les revenus élevés comportent une part non gagnée plus importante. Hobson semble assez séduit par cette forme d'intervention étatique.

Hobson s'intéresse ensuite à un sujet cher à Marshall : l'échelle de la production. Il introduit une idée fort intéressante : la taille de la production et la taille de l'entreprise doivent être envisagées séparément. Il existe pour la plupart des unités de production une taille optimale au-delà de laquelle les gains de la division du travail ne compensent plus les déséconomies de coordination. Si cette taille n'est pas très élevée, il peut être avantageux pour une firme de posséder plusieurs unités de production. Lorsque les économies d'échelle sont plus sensibles en matière commerciale et financière que dans la production, la concentration du capital sera normalement plus poussée que la concentration de la production. Comme l'échelle de la production, la taille de l'entreprise peut également connaître un seuil dont le franchissement serait désavantageux. Cette limite est toutefois contournable par les entraves à de la concurrence. « The growth of uneconomically large businesses will normally occur in an industry where such growth may itself become an instrument for repressing competition »[2]. Ce dernier argument est au moins partiellement contradictoire : la firme peut difficilement s'appuyer sur une de ses faiblesses pour asseoir son pouvoir. Selon Hobson, les dirigeants d'entreprises sont motivés par le profit TOTAL et préféreraient gagner 9% sur un capital de 500.000 que 12% sur un capital de 300.000. Cet argument est discutable, compte tenu de la possibilité qu'ont les capitalistes de diversifier leurs placements.

[1] Hobson [157] p. 214
[2] Hobson [157] p. 197

TITRE IV- DEVELOPPEMENT DE L'ECONOMIE NEOCLASSIQUE

4.1- LA FIRME ET LE MARCHE : COUTS, PRIX ET PROFIT

Durant tout le premier tiers du XXe siècle, la théorie néoclassique de la productivité et des coûts va se raffiner. Parmi les économistes ayant le plus contribué à ce mouvement, citons Francis Edgeworth (1911, 1913), Frank Knight (1921), Cecil Pigou (1913, 1920), Jacob Viner (1932) et Roy Harrod (1930, 1931). L'année 1933 fut importante, car parurent deux ouvrages novateurs de Joan. Robinson et Edward Chamberlin.

Marshall est la principale source d'inspiration des auteurs analysés dans ce chapitre.

4.1.1. Productivité et coût

ANALYSE A COURT TERME

Voici comment Viner définit le *court terme* : « The "short run" is taken to be a period which is long enough to permit of any desired change of output technologically possible without altering the scale of plant, but which is not long enough to permit of any adjustment of scale of plant »[1]. Il prend comme hypothèse qu'à court terme, on peut classer les facteurs en deux catégories : les *facteurs FIXES* (c'-à-d dont le volume ne peut varier pendant la période considérée) et les *facteurs VARIABLES*. Viner précise : « Scale of plant will be used as synonymous with the size of the group of factors which are fixed in amount in the short run »[2].

On considère généralement le travail comme l'exemple du facteur variable et le capital comme l'exemple du facteur variable et le capital (équipements durables) comme l'exemple du facteur fixe. Pour ces deux facteurs, j'utiliserai régulièrement les abréviations *L* et *K*.

Pour analyser la **productivité** des facteurs, il est commode de distinguer trois concepts qui sont complémentaires.

1. la *productivité TOTALE* qui correspond au total de la production
2. la *productivité MOYENNE* est la productivité totale divisée par le nombre d'unités du facteur variable : c'est par exemple la production par homme / heure
3. la *productivité MARGINALE*, supplément de produit dû à la dernière unité du facteur variable

J'utiliserai régulièrement les abréviations, respectivement : PT, PM et Pm.

Un exemple numérique fera bien comprendre le lien entre ces trois grandeurs. L'exemple ci-dessous est donné par Edgeworth.

[1] Viner [372] p. 26
[2] Viner [372] p. 26

nombre de jours travail	récolte totale en boisseaux	incrément des doses successives	boisseaux par jour de travail
…	…	…	…
13	220	…	16,02
14	244	24	17,43
15	270	26	18
16	294	24	18.38
17	317	23	18.65
18	339	22	18.83
19	360	21	18.95
20	380	20	19
21	396	16	18.86
…	…	…	…

La première colonne est la variable indépendante. Le contenu de la deuxième (PT) est constaté empiriquement ou, ici, donné ; la troisième (Pm) et la quatrième (PM) sont déduites des deux premières.

La partie gauche du graphique 4.1 explique comment *PT*, *PM* et *Pm* évoluent en fonction de la quantité du facteur variable *L*. Cette évolution, caractérisée par la *loi des rendements variables*, peut être divisée en trois secteurs séparés par les lignes *b-b'* et *c-c'*.

Le premier secteur (I) est appelé *rendements croissants*[7]. La productivité marginale y est supérieure à la productivité moyenne : les doses additionnelles de *L* font donc augmenter la productivité moyenne et l'on constate que la courbe *PM* est effectivement croissante. A noter que le secteur (I) peut lui-même être divisé en deux parties. Avant le point *a*, la courbe *PT* est convexe et Pm est croissante. Après ce point, *PT* devient concave et *Pm* décroissante. Mais cette division est moins importante que la première.

Après b et b', commence le secteur des *rendements décroissants* (II). Comme *Pm* y est inférieure à *PM*, les doses supplémentaires de *L* réduisent la productivité moyenne, dont la courbe est devenue décroissante. A partir du point *c*, l'accroissement de *L* réduit *PT*. Il est donc en excès. C'est le secteur III, qu'on pourrait qualifier de rendements négatifs.

Les productivités moyenne et marginale apparaissent implicitement sur le graphique 4.1-A représentant la productivité totale ; la productivité marginale en un point y est donnée par la pente de la tangente à la courbe *PT* en ce points ; la productivité moyenne en un point est donnée par la pente du segment de droite reliant le point considéré de *PT* à l'origine.

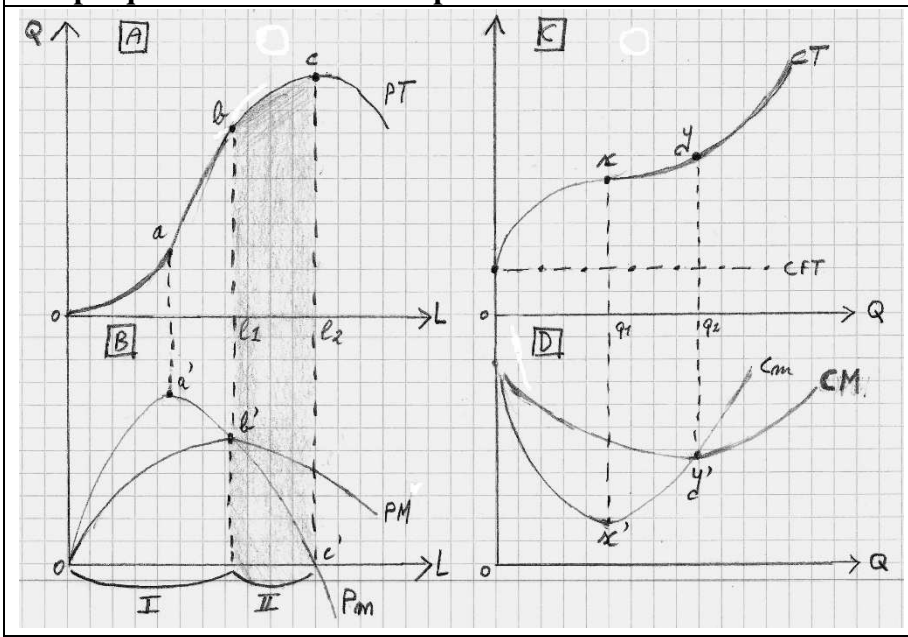
De tous les segments de droite partant de l'origine et touchant la courbe *PT*, c'est *o-b* qui a la pente la plus élevée. La productivité moyenne est donc maximale en *b*. Cette conclusion concorde avec le fait que la courbe *PM* atteint son sommet pour une quantité l_1 du facteur variable. En outre, en ce point, *o-b* est tangent à la courbe et sa pente donne la productivité marginale : les courbes *PM* et *Pm* se coupent donc en *b*'.

A la base, PT, PM et Pm sont exprimées en unités physiques. On sait que la rémunération du facteur s'égalise avec sa productivité marginale. Dans ce cas, Pm est exprimée en valeur. L'expression en valeur requiert la multiplication de Pm par le prix de vente du produit. En concurrence parfaite, cela revient à la multiplier par une constante, puisque le volume de production de la firme individuelle n'influence pas le prix. Dans ce cas, la courbe a exactement la même forme en valeur ou en unités physiques.

[7] On trouve le terme RENDEMENT utilisé comme synonyme pour PRODUCTIVITE. Il faut toutefois veiller à ne pas confondre ces rendements à court terme avec les *rendements d'échelle*

Ne perdons pas de vue ce que le graphique ne montre pas, c'est-à-dire qu'une quantité fixe d'autres facteurs se combine avec le ou les facteurs variables. En faisant varier la quantité du facteur variable *L*, on fait également varier la proportion entre *K* et *L*. Tant l'excès de *K* que l'excès de *L* sont nuisibles à la production. Comme l'ont montré, Knight et Sraffa, seul le secteur II sur le graphique est économiquement significatif. Dans le secteur I, le facteur fixe est en excès et la production serait augmentée si une part de celui-ci était délaissée. Dans le secteur III, c'est le facteur variable *L* qui est en excès, à tel point que la production serait accrue si *L* était réduite.

Graphique 4.1: les courbes de productivité et de coût



Le **coût** est un concept lié à celui de productivité, entretenant avec elle une relation inverse. Pour construire les courbes de coût, la quantité produite est placée en abscisse[1].

Le coût résulte de la multiplication de la quantité des facteurs utilisée par leur prix. Lorsque nous sommes en concurrence parfaite, la firme est trop petite pour influencer le prix des facteurs. Ces prix sont donc invariables par rapport à la quantité utilisée. Les variations éventuelles du coût ne peuvent alors provenir que de considérations technologiques.

Comme pour la productivité, nous distinguons trois concepts :
- le *coût total* (CT) est le coût généré par la production des Q unités
- le *coût moyen* (CM) est le coût par unité produite (= CT/Q)
- le *coût marginal* (Cm) est le supplément de coût total généré par l'accroissement de la production d'une unité.

Cette classification en croise une autre. Le *coût fixe* est le coût des facteurs fixes et le *coût variable* est le coût des facteurs variables. Le coût total (CT) est l'addition du coût fixe total (CFT) et du coût variable total (CVT). Le coût moyen (CM) est l'addition du coût fixe moyen (CFM) et du coût variable moyen (CVM). Par contre, le coût marginal ne peut évidemment comporter que des coûts variables.

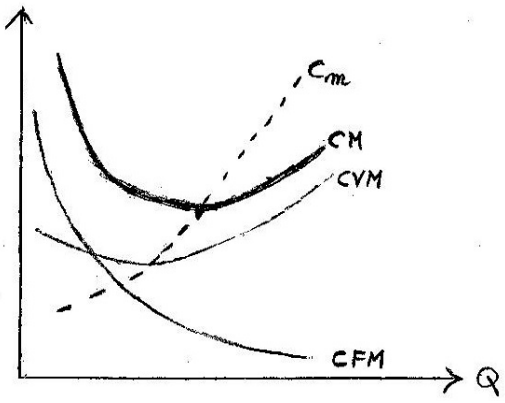
[1] Elle correspond à PT, mais il est d'usage d'utiliser la lettre *Q*.

Le coût fixe n'est fixe qu'au niveau du coût total (courbe CFT horizontale sur la figure 4.1-C). Lorsqu'on divise cette constante par la variable indépendante Q, on obtient l'hyperbole équilatère CFM sur la figure 4.1-D. Le CFM n'est donc pas constant.

Viner attribuait la forme en U de la courbe CM à l'addition de l'hyperbole CFM avec la courbe croissante convexe CVM. Mais beaucoup d'auteurs reconnaissent cette forme en U à la courbe CVM également, car elle correspond à la loi des rendements variables : d'abord les rendements croissants (ce qui revient à des coûts décroissants), ensuite des rendements décroissants (ce qui revient à des coûts croissants).

Le coût marginal coupe la courbe CM en son minimum. En effet, si Cm >CM, CM doit être croissant, alors que si Cm< CM, CM doit être décroissant. Il doit en être de même de CVM si elle a un minimum.

Graphique 4.2: le coût moyen et ses composants

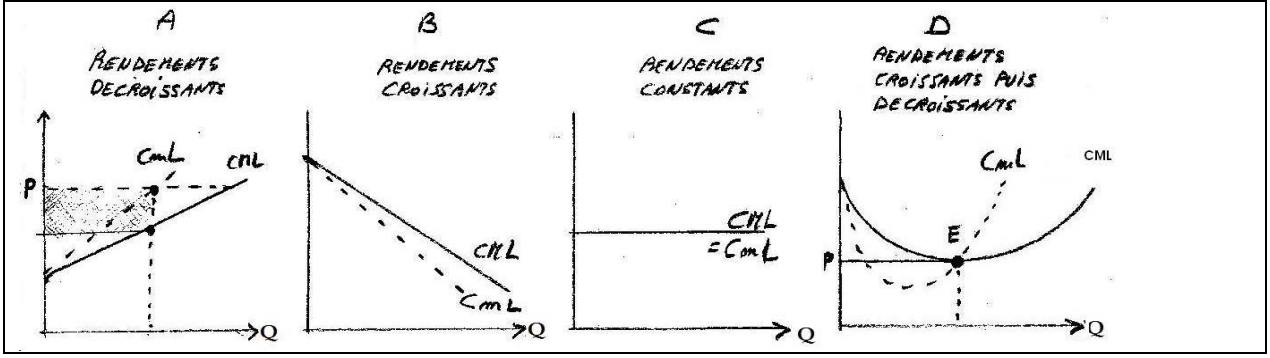


ANALYSE A LONG TERME

Dans le long terme, les quantités de tous les facteurs peuvent varier. Il n'y a donc plus de coût fixe.

Le graphique 4.3 représente les courbes de coût moyen à long terme (CML) et de coût marginal à long terme (CmL). Quelle est la forme des courbes de coût à long terme ? Elle exprime le type des rendements d'échelle. Le graphique 4.3 montre les principaux cas pouvant se présenter.

Graphique 4.3 : le coût moyen à long terme et les rendements d'échelle



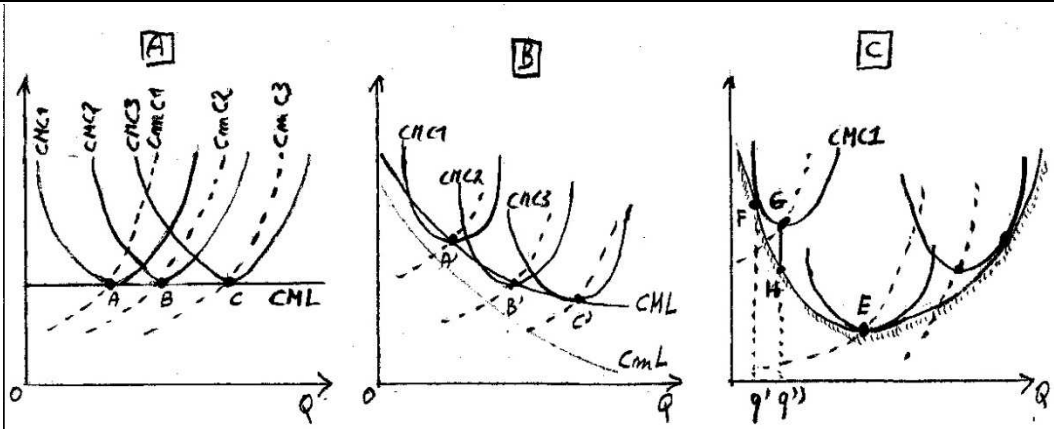
A la suite de Wicksell (cf. supra), beaucoup d'économistes croient en la prépondérance de la figure *D*. Chamberlin l'argumente ainsi : « It is important to avoid identifying the emergence of

the problems of coordination arising from increasing complexity with the minimum point on the cost curve. Where they first become important will depend upon the product in question and the techniques and circumstances under which its production is undertaken at different times and places, but in general they will begin to appear for quite small outputs (…) In the early stages they are submerged by the overwhelming gains from further specialization and more efficient techniques already discussed. But since these latter tend to exhaust themselves, with larger aggregates, whereas complexity steadily increases, it appears certain that the diseconomies must sooner or later outweigh the economies and beyond this point predominate. The forces making for economies and for diseconomies are in balance at the minimum of the *AC* curve »[1].

Une relation doit évidemment exister entre les courbes de coût à court terme (graphiques 4.1 à 4.2) et les courbes de coût à long terme (graphique 4.3). Une courbe à court terme suppose une taille donnée de la firme. Il en existe donc une pour chaque taille possible, C'est Viner, qui, le premier, eut l'idée de la *courbe enveloppe* : en reliant un point bien choisi de chacune des courbes de coût moyen à court terme (*CMC*), on obtient la courbe de coût moyen à long terme (*CML*).

Viner pensait qu'il fallait relier les points minimums des courbes CMC. C'est ce que montrent les figures 4.4-A et B, respectivement pour les rendements constants et les rendements croissants. Le graphique 4.4-B comporte une anomalie : pour certaines valeurs de *Q*, la courbe CMC est en dessous de la courbe CML, ce qui n'est pas normal puisque la courbe CML permet en chacun de ses points d'ajuster au mieux la proportion entre les facteurs ; elle ne peut jamais être au-dessus des courbes CMC. En fait, le seul point de CML qui correspond au minimum d'une courbe CMC est son minimum à elle (*E*).

Graphique 4.4 : la courbe de coût moyen à long terme en tant que courbe enveloppe



La figure B montre l'erreur de Viner. La figure C montre la correction apportée par Chamberlin

Le graphique 4.4-C est de Chamberlin. Supposons que l'entreprise a la taille T1. Pour cette taille, en produisant *q*', elle est sur la courbe CML ; en augmentant la production au niveau *q*'', elle la quitte, alors qu'elle atteint le coût minimum pour sa taille. En fait pour la quantité *q*'', le point *H* est le moins coûteux. Correspondant à une taille plus élevée que la taille T1, il n'appartient pas à la courbe CMC1 mais à une courbe CMC2 non dessinée. Un point qui n'est pas le minimum de CMC2 peut être situé plus bas que le minimum de CMC1. Ce paradoxe s'explique par les économies d'échelle apportées par la taille T2, qui dominent l'avantage de l'optimisation à court terme. La courbe enveloppe a ceci de traître qu'une des variables essentielles, la taille de l'entreprise, n'y apparaît pas explicitement.

[1] Chamberlin [47] p. 247-8. Sa courbe « AC » correspond à notre courbe « CML ».

4.1.2. Equilibre du producteur et du marché en concurrence parfaite

Concurrence parfaite est le terme utilisé par les économistes pour désigner cette situation déjà évoquée au chapitre précédent qui se caractérise par l'incapacité pour tout vendeur ou tout acheteur individuels d'influencer le prix par son offre ou sa demande.

Pour qu'il en soit ainsi, un certain nombre de conditions fonctionnelles doivent être satisfaites. Voyons comment quelques économistes importants présentent cette problématique.

J.Robinson[1] compte deux grandes conditions :
- un grand nombre d'offreurs
- la perfection du marché qui implique que face à l'offre, tous les demandeurs affichent une réaction identique. Plusieurs facteurs rendent cette condition peu réaliste :
    - le coût du transport, variable suivant les localisations respectives de l'acheteur et du vendeur
    - la perception subjective de la qualité du produit
    - l'aspect personnel de la relation entre l'acheteur et le vendeur
    - la perception des prix élevés comme un gage de qualité
    - la publicité

Chamberlin distingue quatre conditions :
- Le nombre élevé d'offreurs
- L'homogénéité du produit entre les producteurs
- Le libre accès de nouveaux producteurs au marché
- L'absence d'entente entre les producteurs

Chamberlin considère que ces quatre conditions instituent ce qu'il appelle la *concurrence pure*, mais que la concurrence parfaite implique en plus la parfaite information des intervenants sur le marché, et la mobilité parfaite des facteurs entre les branches.

Stigler établit la liste des conditions comme suit :
- la firme la plus grande n'a qu'une part de marché réduite
- « These many firms (…) are assumed to act independently »
- « Complete knowledge of offers to buy and sell by the participants in the market »
- La divisibilité du produit
- "Resources are mobile among uses; and their owners are informed as to yields in these various uses"[2]

ANALYSE A COURT TERME

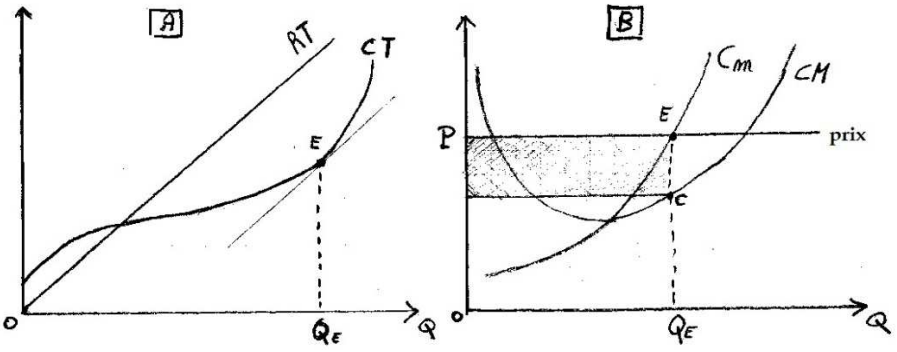
Abordons maintenant l'**optimum** de la firme : quelle quantité doit-elle produire pour maximiser son profit[3] ? La présentation graphique est la plus aisée. Edgeworth a préféré recourir au graphique du coût total (4.3-A) et Viner à celui du coût unitaire (4.3-B) mais les deux

[1] J'accole systématiquement l'initiale du prénom de Joan Robinson à son nom, car il sera aussi question de son mari Austin Robinson.
[2] Stigler [348] pp. 6-7
[3] Edgeworth rappelle fort à propos que les points des courbes de coût sont déjà le fruit d'une première optimisation. Pour chaque quantité produite, plusieurs techniques sont possibles, d'efficacités variables. Les courbes des graphiques 4.1 et 4.2 donnent le coût de la technique la plus efficace pour chaque quantité produite.

concordent. Sur le graphique 4.5-A, la *recette totale* de la firme *(RT*), son chiffre d'affaires, est donné par la droite montante à partir de l'origine. Le prix de vente unitaire est donné par la pente de cette droite, qui n'est droite que parce que le prix est indépendant de la quantité produite, ce qui est une condition de la concurrence parfaite. Pour que le profit soit maximisé, il faut que la courbe de coût total *CT* soit en dessous de la courbe *RT* et que l'écart vertical entre les deux soit maximum La première condition implique que la courbe soit convexe au point d'écart maximum, sinon il y aurait maximisation de la perte plutôt que du profit. L'optimum du producteur se trouve donc nécessairement dans le secteur des rendements décroissants. La deuxième condition est rencontrée lorsque la tangente à la courbe *CT* est parallèle à la courbe *RT*. Autrement dit lorsque le prix est égal au coût marginal puisque la pente de la courbe *CT* (et donc de sa tangente) indique le coût marginal.

La figure B expose la même règle d'optimisation au niveau des recettes et coûts unitaires. La firme a intérêt à produire plus tant que le prix dépasse le coût marginal, c'-à-d jusqu'au point Q_E où les deux grandeurs sont égales. Pour ce niveau de production, la marge bénéficiaire unitaire vaut le segment CE, soit le supplément du prix par rapport au coût moyen. Le profit total vaut le profit unitaire multiplié par Q_E, et est donné par la surface hachurée.

Graphique 4.5: la maximisation du profit à court terme



Le prix apparaît comme une droite horizontale, car il est insensible à la quantité produite de la firme. Cette horizontalité exprime également qu'au prix du marché, la firme peut vendre une quantité qui ne connaît aucune limite extérieure. C'est ce qu'explique Viner : « If this particular producer is an insignificant factor in his industry, i.e., if atomistic competition prevails, he may reasonably assume that no change in his output, and especially no change consistent with the maintenance of the scale of plant at its original level, will have any appreciable effect on the price of his product. Under these conditions, the partial demand curve for his product may be taken as a horizontal line whose ordinate from the base is equal to the prevailing price »[1].

Puisque la condition de l'optimum est l'égalité du coût marginal avec le prix, la courbe *Cm* sera la courbe d'offre de courte période de la firme. Pour chaque prix, c'est l'abscisse de la courbe *Cm* qui donne la quantité à offrir qui maximise le profit. Plus précisément, seule la portion croissante de *Cm* formera la courbe d'offre, puisque seuls les rendements décroissants sont significatifs.

[1] Viner [372] p. 28. La DEMANDE DU MARCHE reste bien sûr déclinante. On a affaire à deux concepts de demande liés mais distincts. C'est précisément l'intersection des courbes de demande et d'offre du marché, illustrée par le graphique 3.8 de Marshall, qui détermine l'altitude de la droite de prix sur la figure 4.5.

ANALYSE A LONG TERME

Voyons comment se détermine l'équilibre de la firme à long terme. Le problème est de déterminer l'échelle de la production, autrement dit la taille de la firme, qui minimisent le coût.

Pigou fait le raisonnement suivant :

- si le prix est supérieur au coût marginal, la firme gagnerait à augmenter sa production ; si le prix est inférieur au coût marginal, elle gagnerait à réduire sa production. A l'équilibre, le prix doit donc égaler le coût marginal.
- Si le prix est supérieur au coût moyen, la firme obtiendra un profit, qui attirera de nouveaux arrivants ; la concurrence fera baisser le prix jusqu'à ce qu'il ne soit plus supérieur au coût moyen[1]. Si le prix est inférieur au coût moyen, la firme encourra une perte. La baisse de l'offre le ramènera au niveau du coût moyen. A l'équilibre, le prix doit donc égaler le coût moyen.

En conclusion, l'équilibre de longue période implique que le prix s'égalise à la fois avec le coût moyen et le coût marginal, qui sont donc égaux entre eux.

Pigou inspecte ensuite les quatre cas du graphique 4.3 pour voir si cette condition d'équilibre peut être remplie.

- Dans les cas *A* et *B*, elle ne peut l'être.
- Dans le cas *C*, si elle est remplie en un point, elle l'est en tout point. L'équilibre est donc indéterminé.
- Dans le cas *D*, le point *E* est un point d'équilibre unique et stable. Le prix s'égalise avec le coût moyen au point où celui-ci atteint son minimum. Wicksell avait montré qu'en ce point, les rendements étant constants, la rémunération des facteurs à leur productivité marginale épuise exactement le produit (cf. supra). Il n'y a ni profit ni perte.

Il y a unanimité parmi les économistes pour considérer que les rendements d'échelle croissants sont incompatibles avec la concurrence. Viner écrit : « The familiar proposition that net internal economies of large scale production and long run stable equilibrium are inconsistent under competitive equilibrium is clearly illustrated by this chart (NDLR : 4.4-B). (…) Provided that no change in its output will affect market price, it will pay the concern to enlarge its plant whatever the price may be, and whatever its existing scale of plant may be. If thereby, it grows so large that its operations exert a significant influence on price, we pass out of the realm of atomistic competition and approach that of partial monopoly »[2].

Sraffa montre que l'hypothèse opposée, les rendements décroissants, n'est pas plus tenable. Le processus est exactement inverse. Il écrit : « it cannot display increasing costs for all of its length : because in such a case, competition would tend to make every firm infinitely small and the number of firms infinitely large »[3]. Si l'on regarde la figure B, on constate que l'égalité entre le coût moyen et le coût marginal n'est possible que pour une production nulle.

Passons maintenant de la firme individuelle au **marché** dont elle fait partie. En principe, la courbe d'offre du marché s'obtient par l'addition horizontale des offres des producteurs individuels : pour une même valeur de l'ordonnée (coûts et prix), on additionne les quantités (abscisse) des divers producteurs. Les variations de l'offre du marché peuvent emprunter deux

[1] Ceci implique l'absence de barrières a l'entrée sur le marché.
[2] Viner [372] p. 37
[3] Sraffa [344] p. 28

voies : une variation de la production des firmes individuelles ou la variation du nombre de firmes.

Voyons comment se comporte l'offre du marché pour les différents types de rendements d'échelle. Pour simplifier le problème, supposons que toutes les firmes ont des courbes de coût identiques.

- Lorsque les rendements d'échelle sont croissants ou constants. Viner estime que l'offre du marché est indéterminée
- Lorsque ces rendements sont constants, Pigou considère que le coût moyen rectiligne horizontal peut servir de courbe d'offre de l'industrie.
- Lorsque la courbe CML des entreprises est en U, Pigou estime que leur courbe d'offre individuelle à long terme tient en un seul point : le minimum de CML. Puisque chaque entreprise est rivée à une quantité produite optimale, l'industrie dans son ensemble ne peut adapter son offre à une variation de la demande que par un changement du nombre d'entreprises et non par des variations de la quantité produite des entreprises existantes. Comme toutes les entreprises ont le même coût, l'offre à long terme de l'industrie est une droite horizontale, comme dans le cas des rendements constants.

De toute évidence, l'hypothèse du coût identique dans toutes les firmes n'est pas confirmée par la réalité. Toutefois, explique Viner, les écarts de coût sont révélateurs d'un ajustement à long terme non abouti. Voici pourquoi l'équilibre de long terme postule le nivellement des coûts: « All long run differences in efficiency as between concerns will be assumed, however, to be compensated for by differential rates of compensation to the factors responsible for such differences, and these differential rates will be treated as part of the ordinary long run money costs of production of the different concerns »[1]. C'est un phénomène bien connu. Les commerces jouissant d'une situation géographique particulièrement favorable ne donnent pas lieu à un taux de profit supérieur mais paient des loyers supérieurs. Si une entreprise fait des surprofits parce qu'elle occupe un facteur particulièrement productif, cette situation n'est tenable qu'à court terme, car les autres entreprises feront monter les enchères pour débaucher ce facteur, dont le prix augmentera au dessus du minimum nécessaire pour le convaincre de s'investir[2]. A ce sujet, Chamberlin, remarque que l'égalité des minimums des courbes CML n'implique pas l'identité des courbes. Les producteurs plus efficaces obtiendront leur coût moyen minimum pour un niveau de production plus élevé que leurs concurrents moins performants.

Evidemment, ces ajustements prennent du temps. L'équilibre à court terme postule l'égalité entre les coûts marginaux des producteurs, puisque chacun égalise le sien avec le prix. Par contre, « under short run equilibrium, the average costs, including the fixed costs, of any particular producer need bear no necessary relationship to price, except that the average direct costs must not exceed price »[3].

[1] Viner [372] p. 25

[2] Ce facteur est nécessairement rare, car s'il ne l'était pas, les firmes en acquerraient une plus grande quantité et produiraient plus, le prix du produit diminuerait et le profit confisqué par ce facteur s'évanouirait. La part de sa rémunération en sus du minimum nécessaire est généralement appelée *rente* lorsque sa rareté est due à une cause permanente (exemple : la situation géographique). Elle est appelée *quasi-rente* lorsque cette rareté est due à une cause temporaire (exemple : pénurie d'une certaine spécialité de main d'œuvre).

[3] Viner [372] pp. 43-44

4.1.3. Le producteur et le marché hors concurrence parfaite

LE MONOPOLE

Il existe deux acceptions du terme *monopole*. Soit il définit de façon restrictive la situation où un seul offreur alimente un marché. Soit, on le considère comme un pouvoir, le pouvoir d'influencer le prix du marché, afin d'en tirer le maximum de profit. Dans ce cas, un cartel de producteurs sera considéré comme un monopole. Comme le remarque Pigou, le monopole est le plus souvent (mais pas nécessairement) protégé par des barrières à l'entrée de nouveaux producteurs. L'absence de ces barrières rendrait les accords sur les prix trop précaires pour en valoir la peine.

Le pouvoir du monopole n'est pas absolu ; la demande des consommateurs exerce également une influence. C'est ce que montre Menger : « if the monopolist wishes to sell a particular quantity of the monopolized good, he cannot fix the price at will. And if he fixes the price, he cannot determine the quantity that will be sold at the price he sets »[1]. Suivant les circonstances, la vente d'une quantité élevée à bas prix rapportera plus, autant ou moins que la vente d'une quantité moindre à prix plus élevé.

La plupart des économistes ont étudié la maximisation du profit du monopoleur ; les diverses explications sont concordantes. Wicksell rappelle que Cournot avait déjà apporté un traitement algébrique satisfaisant à la question.

La demande individuelle de la firme, en situation de concurrence parfaite est le mieux représentée par une droite horizontale. Celle du monopoleur correspond évidemment à la demande du marché dans sa globalité et est donc décroissante : pas de hausse de la quantité sans baisse de prix.

« it will be the advantage of the monopolist to fix his prices so high that he will obtain the maximum net profit. Every rise in price causes, we may assume, a falling off in demand. But so long as the falling is less than proportionate to the increased profit per unit of the commodity resulting from higher prices, the total net profit will increase. But when the decrease in sales is more than proportionate to the increased profit per unit, any further increase in price will be disadvantageous. The ideal monopoly price is thus to be found precisely at the meeting point of both theses tendencies »[2]. Wicksell donne l'exemple numérique suivant [3]:

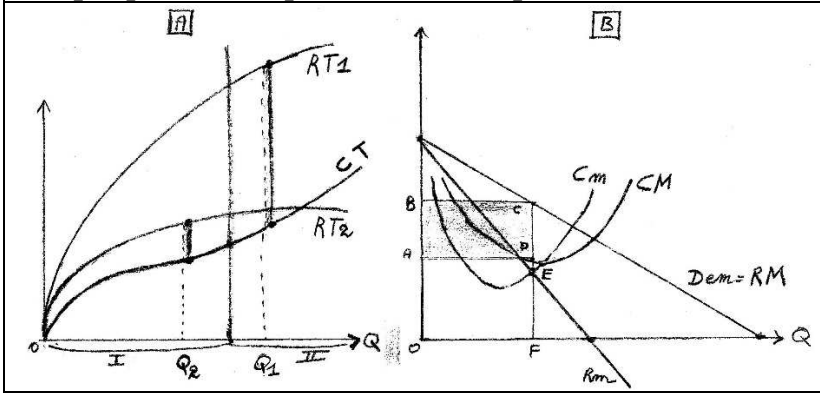
[1] Menger [256] p. 210
[2] Wicksell [385], t.I p. 90
[3] Une erreur d'impression s'est glissée dans la dernière colonne pour les prix 20, 18, 16 et 14. La dernière colonne doit être le produit des deux précédentes.

Selling Price. £	*Costs.* £	*Profit per Unit.* £	*Turnover.* *Pieces.*	*Total Net Profit.* £
22	2	20	—	—
20	2	18	200	—
18	2	16	400	3,600
16	2	14	600	6,400
14	2	12	800	8,400
13	2	11	900	9,900
12	2	10	1,000	10,000
11	2	9	1,100	9,900
10	2	8	1,200	9,600
8	2	6	1,400	8,400
6	2	4	1,600	6,400
4	2	2	1,800	3,600
2	2	—	2,000	—

Le prix qui maximise le profit est évidemment £12. Wicksell montre aussi que le niveau des coûts fixes n'influence pas le prix optimum, contrairement à celui des coûts variables. Si par exemple, on devait déduire de chaque ligne de la dernière colonne un coût de 1000, cela ne changerait pas le classement des profits totaux. Par contre, si le coût unitaire (2e colonne) passait de £2 à £4, le prix optimum passerait à £13 et la quantité vendue correspondante se réduirait de 1000 à 900.

Ce qui est vrai des coûts l'est également des taxes. Cournot l'avait déjà remarqué, tout comme Marshall, qui écrit : « on the other hand, a tax proportional to the amount produced gives an inducement to the monopolist to lessen his output and raise his price »[1]. Par contre, remarque Marshall, une taxe en pourcentage du profit de monopole n'affecterait pas l'optimum.

Graphique 4.6 : l'équilibre du monopoleur



La figure 4.6-A est la version pour monopole de la courbe 4.5-A, proposée par Edgeworth. La recette totale devient une courbe concave, dont la pente décroissante révèle la décroissance du prix lorsque la quantité augmente. Edgeworth mentionne cette différence avec la concurrence parfaite : l'équilibre peut cette fois tomber dans le secteur des rendements croissants. Ce sera le cas si l'élasticité de la demande est très faible, ce qui se marquera par une forte concavité de RT (comme RT2 sur le graphique). L'analyse de l'équilibre de la firme en monopole et même en concurrence parfaite semblera plus claire à la prochaine section, car tout ce savoir sera prodigieusement formalisé, structuré par J.Robinson.

Comment Pigou envisage-t-il la courbe d'offre du monopoleur ? Le prix d'offre ne peut évidemment être inférieur ni au coût moyen (sinon, on est en perte), ni au coût marginal (sinon la dernière unité ne serait pas offerte). Par contre, en monopole, le prix d'offre peut être supérieur au coût moyen, car dans l'analyse précédente, c'était précisément la concurrence

[1] Marshall [253] § 5.4.13

entre producteurs qui ramenait le prix d'offre au coût moyen. Il écrit: « in a one-firm industry the supply price of any given quantity of output is equal to average cost or to marginal cost, according as the one or the other of theses is the greater »[1].

Beaucoup d'auteurs blâment le monopole, car, par rapport à la concurrence, il pousse l'équilibre vers la gauche du graphique 4.6. Les consommateurs se partagent une production moindre qu'ils payent plus cher. Marshall tempère : peut-être la courbe d'offre concurrentielle serait-elle surélevée par rapport à celle du monopoleur. En effet, le prix d'offre du monopoleur bénéficie de ses avantages en matière de coût. Il y a toutes les économies d'échelle, à quoi s'ajoutent les économies en marketing, car le monopoleur ne doit pas s'évertuer à convaincre les consommateurs de lui être fidèles. Pigou relève un autre avantage du monopole, dans le cas des industries « where temporary low prices lead to the development of new demands »[2]. En vendant à perte ou à profit réduit pendant un temps, le monopoleur peut stimuler la demande et rendre possible des économies d'échelle, qui grossiront ses bénéfices futurs. Une firme en compétition avec d'autres ne se le permettrait pas, parce que seule une trop faible part du gain futur lui échoirait.

Clôturons notre analyse du monopole par l'examen d'une forme particulière, le *monopole discriminatoire*, où le monopoleur demande aux différents consommateurs ou groupes de consommateurs, des prix différenciés. Résumons brièvement le chapitre que Pigou lui consacre. Le cas peu réaliste où chaque unité du produit serait mise aux enchères et vendue au plus offrant n'est guère intéressant. Examinons le cas où le marché est segmenté en quelques sous-marchés où règnent des prix différents. Imaginons que le bien soit vendu sur le marché *A* (plus cher) et le marché *B* (moins cher). L'intérêt du monopoleur serait d'assurer une segmentation telle que « the lowest demand price in sub-market A exceeded the highest demand price in sub-market B ». Pigou estime qu'en pratique, « the monopolist's freedom of action is limited by the need already referred to, of acting on general rules. This consideration makes it necessary that he shall choose for his sub-markets, groups that are distinguishable from one another by some readily recognisable mark ». Autrement dit, « he will not distinguish and bring together entirely new groups but will make use of distinctions already given in nature »[3].

Pour que la discrimination ait des chances de succès, deux conditions sont nécessaires:

1. Les biens ne peuvent pas être facilement transférables d'un marché à l'autre. L'existence de droits de douane ou la cherté du transport sont des facteurs favorables. C'est d'ailleurs dans le commerce international que la discrimination est la plus courante. Le monopoleur pourra aussi inscrire l'interdiction de revente parmi les clauses de son contrat pour sécuriser le cloisonnement des marchés.
2. « No unit of the demand, proper to one market, can be transferred to another market »[4]. Pigou donne l'exemple d'un médecin ou d'un avocat qui appliquerait un tarif différent selon que le client soit riche ou pauvre. Les riches ne voudront pas devenir pauvres simplement pour payer moins cher leur consultation chez le médecin. Ceci montre toutefois la nécessité de segmenter le marché sur base de critères simples et reconnaissables. Sinon, le risque de fraude deviendrait excessif.

[1] Pigou [283] p. 802
[2] Pigou [283] p. 271
[3] Pigou [283] p. 281.
[4] Pigou [283] p. 275.

Pigou examine l'effet de la discrimination sur le volume des ventes en comparant le monopole discriminatoire avec le monopole simple. Reprenons notre exemple du segment *A* (plus cher) et du segment *B* (moins cher). Si le monopole simple avait lui-aussi alimenté les deux secteurs, il est impossible de déterminer l'effet de la discrimination a priori. Si le monopole simple n'avait alimenté que le segment *A*, il est très probable que le monopole discriminatoire élargirait l'offre et il est certain qu'il ne la réduirait pas. Les chances d'élargissement sont maximales si la demande en *B* est forte et élastique et si les rendements d'échelle sont croissants. Ces conclusions sont d'autant plus vraies si le monopole simple n'avait alimenté ni l'un ni l'autre des secteurs *A* et *B*. Pigou compare également le monopole discriminatoire à la concurrence parfaite. « Under conditions of constant and of increasing supply price it is obviously impossible for discriminating monopoly of any degree to make output greater than it would be under simple competition ». Par contre, si le prix d'offre est décroissant, le monopole discriminatoire est susceptible d'élargir l'offre.

ROBINSON ET LA CONCURRENCE IMPARFAITE

La plupart des ouvrages économiques examinés ci-avant se focalisaient sur la concurrence parfaite et consacraient un chapitre au monopole, présenté comme l'exception à la normalité. Concernant les situations entre ces deux extrêmes, l'économie politique était peu diserte. L'année 1933 voit l'arrivée de deux ouvrages extrêmement importants qui vont casser ce moule simpliste : « The Economics of Imperfect Competition » de Joan Robinson et « The Theory of Monopolistic Competition » de Chamberlin.

La concurrence parfaite impliquait la réalisation de deux conditions fondamentales :
- un nombre élevé de vendeurs
- un produit homogène

La non-satisfaction de la première condition produit l'*oligopole*, celle de la seconde donne la *concurrence monopolistique*. Ces deux formes de marché peuvent être regroupées sous l'appellation *concurrence imparfaite*. Les deux ouvrages novateurs diffèrent largement entre eux : celui de Chamberlin est occupé de façon prépondérante par la concurrence monopolistique et consacre quelques pages à l'oligopole. Joan Robinson, au contraire envisage globalement les différentes structures de marché. Elle veut procurer « to the analytical economist a box of tools »[1]. Dans son optique, les structures de marché de la concurrence parfaite au monopole en passant par l'oligopole diffèrent essentiellement par l'élasticité de la demande qui s'adresse à chaque entreprise. Infinie dans le cas de la concurrence parfaite, elle se réduit quand diminue le nombre d'offreurs, ce qui revient à dire que la courbe de demande devient alors plus pentue. Chaque structure de marché peut donc être étudiée comme un cas particulier d'une règle générale qui les embrasse tous.

L'impressionnant ouvrage de J.Robinson formalise des théories qui existaient précédemment à l'état latent, mais apporte également des innovations. Il définit nombre de concepts éclairant la firme et le marché. Il établit les relations entre eux par une algèbre simple et en offre de multiples représentations graphiques.

Nous avions défini le coût moyen (CM) comme le quotient du coût total (CT) sur la quantité produite et le coût marginal (Cm) comme le supplément de coût total dû à un accroissement marginal de la production. La même distinction peut être opérée quant à la recette. La recette moyenne (RM) est la recette totale (RT) divisée par la quantité produite et la *recette marginale* (Rm) est le supplément de recette dû à l'accroissement marginal de la production (en supposant

[1] Robinson [301] p. 1

que la production soit vendue). Comme on le verra, la *recette marginale* est l'innovation décisive pour la détermination de l'équilibre.

Nous connaissons déjà la demande du consommateur, la demande totale sur un marché, la demande de la firme à l'égard des facteurs de production. Arrive un nouveau concept de demande : celle qui s'adresse à la firme individuelle pour son produit. Ce concept est assez abstrait en ce sens que, si la concurrence règne, les consommateurs ne se focalisent pas a priori sur une firme précise. La transposition de la demande vers la firme implique donc une opération intellectuelle de conversion ; c'est ainsi qu'on obtient la demande horizontale dans le cas de la concurrence parfaite. La courbe de cette demande n'est autre que la courbe RM puisque la recette moyenne, c'est le prix.

En concurrence imparfaite, comme en monopole, la demande adressée à la firme est décroissante : pas de hausse des ventes sans diminution du prix. La courbe Rm est donc sous la courbe RM. La recette marginale peut même être négative : ce sera le cas si l'effet positif de la quantité vendue plus élevée ne compense pas l'effet négatif de la baisse du prix, autrement dit si l'élasticité de la demande est inférieure à l'unité.

Lorsqu'il est question de demande plus ou moins élastique, on se réfère implicitement à un point donné sur l'abscisse car en dehors de courbes particulières (la demande horizontale en concurrence parfaite, par exemple), l'élasticité varie le long de la fonction. Qu'elle soit convexe, concave ou linéaire, la demande a une élasticité infinie à sa jonction avec l'axe des ordonnées et une élasticité nulle à sa jonction avec l'axe des abscisses. Entre les deux, elle passe par toutes les valeurs, notamment l'unité. Pour la quantité où l'élasticité de la demande est unitaire, la recette marginale est nulle[1]. Lorsque la courbe RM est linéaire, la courbe Rm l'est également et son éloignement horizontal de l'axe des ordonnées vaut précisément la moitié de l'éloignement de la recette moyenne.

La firme maximise son profit en poussant sa production tant que Rm est supérieur à Cm et en l'arrêtant exactement en ce point. C'est ce que montre la figure 4.6-B. Le point E, lieu d'intersection des courbes Rm et Cm détermine la quantité à produire (OF). L'équation de l'équilibre à court terme est donc :

$$\mathrm{Rm} = \mathrm{Cm} \qquad (4.1)$$

Le prix est donné par l'altitude de la courbe RM pour cette quantité. La recette totale égale le prix multiplié par la quantité, soit le rectangle OBCF. Le coût total égale le coût moyen (FD) multiplié par la quantité, soit le rectangle OADF. Par soustraction, le profit est le rectangle hachuré ABCD. Ce schéma explique tant le profit monopolistique que le profit oligopolistique. Sa mécanique est même valable pour la concurrence parfaite. Simplement dans ce cas, vu l'horizontalité de RM (droite de prix), celle-ci se confond avec Rm, comme on peut le voir sur le graphique 4.5-B. Dans la section relative à la concurrence parfaite, la règle consistait à égaliser le prix avec Cm. Puisque Rm équivaut au prix, l'égalité Rm = Cm vaut également pour la concurrence parfaite.

Robinson démontre que la relation entre RM et Rm est donnée par l'équation (4.2) faisant intervenir l'élasticité de la demande ε.

[1] A l'équilibre, la recette marginale est normalement positive puisqu'elle égale le coût marginal. On peut en conclure que l'équilibre se trouve habituellement dans la zone où l'élasticité de la demande est encore supérieure à l'unité.

$$RM = Rm.\frac{\varepsilon}{\varepsilon - 1} \qquad (4.2)$$

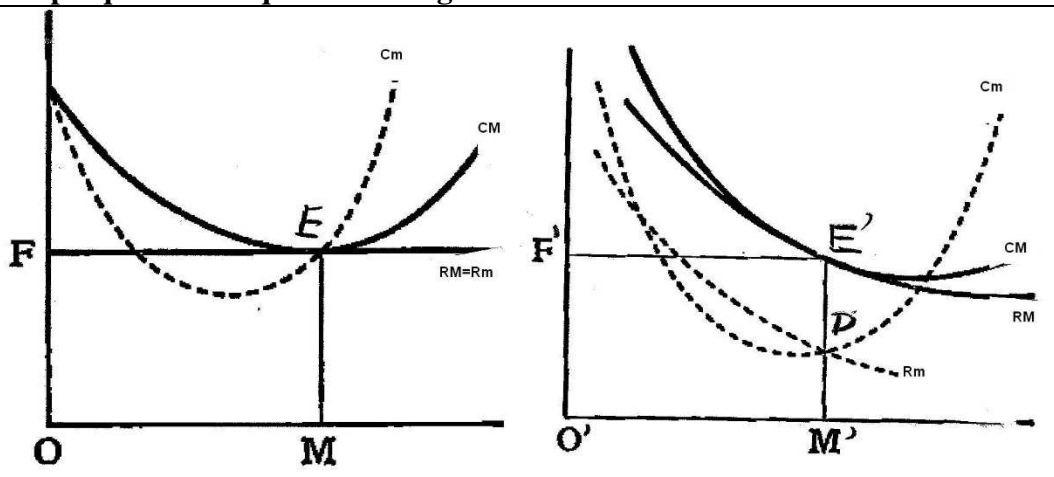
$$P^* = Cm.\frac{\varepsilon}{\varepsilon - 1} \qquad (4.3)$$

Vu l'équation (4.1), à l'équilibre, la relation entre le prix et le coût marginal est donnée par l'équation (4.3). Cette équation est importante, car elle établit la formule de la marge bénéficiaire en concurrence imparfaite[1]. On peut conclure de (4.3) qu'une modification de la demande qui n'affecte pas son élasticité laissera le prix inchangé.

Rappelons que le profit maximisé par la formule (4.1) est en fait un surprofit, car le coût du capital (profit normal) est inclus dans le coût. Pour autant que l'accès au marché soit suffisamment libre[2], l'existence de ce surprofit attirera de nouveaux concurrents, tout comme une perte pousserait des firmes présentes vers la sortie. L'équilibre donné par (4.1) est donc de courte période. Robinson appelle « full equilibrium » celui où il n'y a pas d'incitation au changement du nombre de firmes[3]. Il est forcément sans surprofit. Le coût moyen doit donc y être égal à la recette moyenne en plus de l'égalité entre Rm et Cm. Cette situation est exposée à la figure 4.7 pour la concurrence parfaite (figure de gauche) et pour la concurrence imparfaite (figure de droite).

L'arrivée de nouvelles firmes réduit la demande individuelle des firmes en place. Sur la figure de droite, RM se déplace vers le sud-ouest et ce mouvement s'arrête lorsque RM cesse d'être supérieur à CM, c'est-à-dire lorsque les deux courbes sont tangentes. La tangence entre RM et CM implique l'intersection entre Rm et Cm pour la même quantité.

Graphique 4.7 : l'équilibre à long terme



Le principe est le même en concurrence parfaite et en concurrence imparfaite. Il y a toutefois deux différences :

- A droite, comme la demande est déclinante, la tangence a lieu à gauche du minimum du coût moyen. L'équilibre est donc moins optimal. Comme nous l'avons déjà vu au sous-chapitre 4.1.2, l'équilibre concurrentiel de longue période est au minimum de CM.

[1] Plus la concurrence est imparfaite, moins la demande est élastique et plus la formule (4.3) produira une marge bénéficiaire élevée. Dans un article publié en 1935, **Lerner** préconisera la formule (P-Cm)/P pour mesurer le degré de pouvoir monopolistique d'une firme.

[2] J.Robinson semble sous-estimer l'importance de cette condition, que l'oligopole risque de malmener.

[3] C'est l'équilibre de longue période déjà abordé au sous-chapitre 4.1.2.

- En concurrence parfaite, le point E est à la croisée des quatre courbes (qui ne sont que trois). En concurrence imparfaite, le point E' est le lieu de tangence entre RM et CM. Mais l'intersection entre Cm et Rm a lieu au point D situé en dessous (pour la même quantité).

Partant du *full equilibrium* en concurrence imparfaite (E' sur le graphique 4.7), un accroissement de la demande du marché générera un profit qui attirera de nouvelles firmes. La demande de la firme individuelle reculera après avoir augmenté. Si la variation de la demande n'en affecte pas l'élasticité, l'augmentation de la production globale proviendra exclusivement de l'augmentation du nombre de firmes ; chaque firme reviendra exactement à l'équilibre initial. Si la demande de chaque firme est devenue moins élastique, le nombre de firmes augmentera d'autant plus, car l'équilibre de chacune recule vers la gauche ; le prix s'élèvera. L'inverse se passera si la demande devient plus élastique. A terme, la hausse de la demande du marché pourra donc aboutir à la hausse, la baisse ou la constance du prix.

L'ouvrage accorde beaucoup d'attention à la demande des facteurs de production[1] et plus particulièrement le travail, tant par la firme individuelle que par l'industrie. Il s'intéresse notamment au lien entre la demande du facteur et l'offre du produit. Il analyse également les caractéristiques concurrentielles propres au marché du facteur. Pour ce faire, Robinson établit une symétrie avec le marché des produits. La firme fait face à une OFFRE pour ses facteurs de production au même titre qu'à une DEMANDE pour son produit. Avec une taille de concurrence parfaite, la firme ne peut pas plus influencer le prix sur le marché du facteur que sur celui du produit ; « sa » courbe d'offre du facteur est donc horizontale. Mais une firme dominante peut être un monopole à l'achat ; pour cette situation, Robinson crée le terme *monopsone*. Le monopsoneur fait face à une courbe d'offre croissante du facteur : il doit payer plus pour acheter plus, puisqu'il représente une part importante du marché voire sa totalité.

Des prédécesseurs de Robinson, nous savons qu'en concurrence parfaite, le prix de demande d'un facteur vaut sa productivité marginale physique multipliée par le prix du produit et que le salaire d'équilibre est la valeur de cette courbe où elle rencontre l'offre du facteur. Cette règle simple ne suffit pas à l'ambition de Robinson. Elle fait appel aux concepts suivants :

- Le *produit marginal net*, contrairement au *produit marginal*, suppose qu'on adapte la quantité des autres facteurs de façon à la conserver optimale. Il égale la différence entre la variation de la recette totale[2] et la variation du coût total engendrées. Ce concept était déjà présent -mais mal défini- chez Marshall et Pigou.
- Le *coût moyen* d'un facteur est sa rémunération unitaire (par exemple, le salaire horaire) ; le *coût marginal* du facteur est le supplément de coût induit par l'embauche d'une unité supplémentaire du facteur. Ces deux variables[3] traduisent l'offre du facteur à l'intérieur de l'entreprise ; elles coïncident lorsque cette offre est parfaitement élastique.

La règle bien connue pour maximiser le profit consiste à égaliser la recette marginale avec le coût marginal du produit. Robinson amène une nouvelle règle qui conduit au même résultat : l'égalisation du produit marginal net des facteurs avec leur coût marginal.

A la fin de l'ouvrage, Robinson met elle-même à profit sa *box of tools* pour échafauder une théorie de l'exploitation du travail. Comme on le sait, selon la théorie néoclassique, le salaire

[1] Le concept d'*élasticité de substitution* entre les facteurs est introduit à cette occasion (cf. sous-chapitre 4.2.2).

[2] Celle-ci dépend de la variation de la quantité et, lorsque la concurrence est imparfaite, de la variation du prix.

[3] Il ne faut pas confondre les coûts moyen et marginal du facteur et les coûts moyen et marginal de la quantité produite.

normal concurrentiel est celui qui s'égalise avec la productivité marginale du travail en valeur, c'est-à-dire le produit marginal physique multiplié par le prix du produit. Pigou avait considéré comme exploitation le paiement d'un salaire inférieur (cf. chapitre 9.1). Robinson reprend cette même définition, mais apporte une explication neuve à la mécanique de l'exploitation, basée sur l'imperfection des marchés. Deux causes d'exploitation sont donc possibles :

1- L'exploitation due aux imperfections sur le marché des produits, Dans ce cas, la recette marginale est inférieure au prix et donc le *produit marginal net* est inférieur à la *valeur du produit marginal physique*. Le salaire payé, égalisé avec le produit marginal net, sera donc inférieur à sa valeur normale.

2- L'exploitation due aux imperfections du marché du facteur travail. Plusieurs cas sont envisageables, dont voici le plus simple : le salaire payé par l'entreprise (monopsone) suit une courbe qui croît avec l'effectif embauché. Le *coût marginal du travail* est supérieur à son coût moyen, donc au salaire, qui de ce fait sera inférieur à la valeur du produit marginal physique.

Comme on le verra au chapitre 5.3, J. Robinson tournera plus tard le dos à l'orthodoxie néoclassique. Elle prendra même ses distances par rapport à son ouvrage de 1933. A ce propos, Flatau écrit: "Looking back on her life's work, Robinson was very critical of the static neoclassical framework adopted in (Economics of Imperfect Competition). However, she remained attached to her results concerning exploitation"[1].

CHAMBERLIN ET L'OLIGOPOLE

Selon Chamberlin, les caractéristiques du marché oligopolistique sont :
1- peu de vendeurs font face à beaucoup d'acheteurs
2- le produit commercialisé est parfaitement standardisé
3- les vendeurs sont indépendants (ce qui exclut les filiales d'un même groupe) et ne concluent pas d'accords entre eux, ni ouvertement, ni tacitement.

Alors qu'il analysera la concurrence monopolistique avec force détails, Chamberlin est nettement plus discret à propos de l'oligopole. Il s'attache principalement à critiquer les théories antérieures que nous avons évoquées au sous-chapitre 2.4.1. Il leur reproche la naïveté attribuée aux vendeurs qui augmentent leur offre ou diminuent leur prix pour étendre leur part de marché, en pensant que les concurrents ne réagiront pas. Cette candeur revêt deux formes.

La première version est : *je me sens libre d'accroître mon offre, car je ne pense pas que mes concurrents adapteront la leur*. Soit un duopole où la firme *A* va modifier la quantité qu'elle offre, pour l'optimiser en fonction de l'offre de son concurrent *B*, en considérant naïvement que celui-ci ne modifiera pas la QUANTITE qu'il offre. Ensuite, *B* adaptera son offre à la nouvelle donne, en considérant que *A* ne modifiera pas la sienne. Et ainsi de suite… C'est le cas de **Cournot**. Rappelons sa conclusion : le prix se fixera à un niveau qui s'écartera d'autant plus de celui du monopole que le nombre d'offreurs est élevé (pour tendre vers son niveau de concurrence pure).

La deuxième version est : *je me sens libre de diminuer mon prix, car je crois que mes concurrents garderont le leur constant*. En fait, chacun à son tour réduira son prix de façon à le rendre légèrement inférieur à celui du concurrent. Ce petit jeu durera jusqu'à ce que le profit ait

[1] Flatau [94] p.8

entièrement disparu. C'est la conception de **Bertrand**. Le prix de l'oligopole s'établirait au niveau de celui de la concurrence pure.

Chamberlin estime que ni l'une ni l'autre de ces hypothèses de comportement ne sont fondées. Il vaut mieux admettre que les vendeurs sont conscients de leur dépendance mutuelle. Lorsqu'un vendeur adapte son prix ou la quantité qu'il offre, cette action a un effet direct sur son profit, mais également un effet indirect de par la réaction du concurrent qu'elle induit. Avant de commettre une action, le vendeur raisonnable en appréciera l'impact sur son profit, tant par les effets indirects que par les effets directs. En vertu de ce principe, quel comportement adoptera-t-il ? Selon Chamberlin, il s'abstiendra de déclencher une guerre des prix ou des quantités. Il fixera son prix à un niveau correspondant à celui du monopole et se contentera, sans batailler, d'une offre valant Q/n où Q est la quantité d'équilibre du monopole (cf. graphique 4.6) et n le nombre de vendeurs. Pour arriver à ce résultat, aucun accord de prix entre les offreurs concurrents n'est nécessaire. La simple recherche par chacun de son profit maximum suffit.

Le profit oligopolistique de Chamberlin ne semble pas attirer de nouvelles firmes, contrairement à celui de Robinson, qui s'évanouissait à long terme.

Chamberlin reconnaît toutefois que deux facteurs peuvent pousser un vendeur en oligopole à espérer qu'une réduction de son prix augmente sa part de marché et soit plus profitable que le partage du profit monopolistique.

1- Le délai de réaction des concurrents. S'il est suffisant, les surprofits temporaires peuvent justifier les actions offensives.
2- L'incertitude de chaque vendeur quant à la réaction des autres à ses initiatives. Chamberlin note plusieurs causes à cette incertitude dont celle-ci : lorsque le nombre de vendeurs est dans la zone critique, l'ampleur de la réduction de la demande subie par les autres vendeurs est presque imprévisible. Or leur réaction en dépend. Il y a donc ici une certaine indétermination de l'équilibre oligopolistique.

LA CONCURRENCE MONOPOLISTIQUE

La différenciation du produit

Chamberlin définit ainsi la différenciation : « A general class of product is differentiated if any significant basis exists for distinguishing the goods (or services) of one seller from those of another. Such a basis may be real or fancied so long as it is of any importance whatever to buyers, and leads to a preference for one variety of the product over another »[1]. Dans le commerce de détail, la différenciation porte non seulement sur le produit lui-même mais également sur le service du distributeur : il peut être plus ou moins compétent, plus ou moins courtois… Même si l'on imagine deux vendeurs totalement identiques offrant exactement les mêmes produits, leur localisation géographique différente suffirait à créer des préférences dans l'esprit des consommateurs (qualité de l'environnement ou, tout simplement, proximité). En tenant compte de ces facteurs, « it is evident that virtually all products are differentiated, at least slightly and that over a wide range of economic activity, differentiation is of considerable importance »[2]. Pour se rendre compte de l'importance de la différenciation dans l'économie, il suffit de penser qu'en son absence, c'est le hasard et non la préférence qui assemble les paires acheteur-vendeur sur le marché.

[1] Chamberlin [47] p.56
[2] Chamberlin [47] p.57

En quelques mots, voici le fond de la pensée chamberlinienne : l'économie politique traditionnelle a trop opposé le monopole et la concurrence sans se rendre compte que dans l'économie réelle, ils sont souvent associés, en fait chaque fois que le produit est différencié. Dans ce cas, chaque offreur peut être considéré comme un monopoleur pour son propre produit mais il est néanmoins en concurrence avec les producteurs des autres produits qui sont des substituts proches du sien. Alors qu'en concurrence parfaite, les marchés des producteurs rivaux sont fusionnés en un seul, en concurrence monopolistique, on a affaire à un réseau de marchés distincts mais liés. La concurrence monopolistique est parfaitement compatible avec un nombre élevé de vendeurs et c'est cette situation qui sert de contexte à l'analyse qui suit.

Le coût des ventes

Les firmes déploient une partie de leur activité pour s'attacher des clients, un fait négligé par l'économie politique traditionnelle, qui ne reconnaît que l'activité de production. La demande est réputée préexistante et ne coûtant rien à l'entreprise. Il est vrai qu'en concurrence pure, le marketing est sans objet puisque la courbe de demande est horizontale : vendre n'est donc jamais un problème. En monopole, son inutilité est encore plus patente puisque le fournisseur est incontournable. Les coûts de vente sont donc un phénomène inhérent à la concurrence monopolistique.

Chamberlin distingue les *coûts de production* et les *coûts de vente* comme suit : les premiers servent à satisfaire la demande telle qu'elle est ; les seconds tentent de déplacer la courbe de demande de la firme vers la droite et éventuellement d'altérer sa forme (élasticité). Les coûts de vente incluent principalement les dépenses publicitaires et l'entretien d'un réseau de distribution (rémunération des vendeurs etc.), mais pas le transport ni le stockage. Ils visent à influencer le consommateur mais également les détaillants. Certains types de dépenses ont un caractère mixte (production & vente).

Comment les dépenses de marketing modifient-elles la demande ? « The explanation may be related to the two factors of (a) imperfect knowledge, and (b) the possibility of altering people's wants … »[1]. L'information qui met l'accent sur les prix rendra la courbe de demande plus élastique ; l'information peut également mettre le produit en avant, afin de le sortir de l'anonymat ou de le rendre plus désirable. « If the information is truthful, wants are more effectively satisfied; if not, they are less effectively satisfied »[2].

Les coûts des ventes sont-ils sujets aux rendements croissants, constants ou décroissants ? D'une part, l'efficacité de la répétition des slogans et la cherté de l'accès aux grands médias créent des économies d'échelle. Mais à partir d'un certain seuil, il coûtera de plus en plus cher de convaincre le consommateur d'acheter plus d'unités du produit[3]. En conclusion, le coût des ventes par unité du produit a une forme en « U », au même titre que le coût de production, mais les deux « U » d'un même produit peuvent être décalés et dissemblables.

L'équilibre de la firme

Quelle politique, la firme doit elle mettre en œuvre pour maximiser son profit ? Quelles corrections faut-il apporter à l'analyse que nous avons consacrée à la concurrence parfaite et au monopole pour appréhender la concurrence monopolistique ? Le problème du producteur est ici beaucoup plus complexe, car il a non plus une mais trois variables sur lesquelles il peut agir. Ce sont :

[1] Chamberlin [47] p.118
[2] Chamberlin [47] p.119
[3] En supposant que le prix reste constant, quelle que soit la quantité offerte

1- Le **prix**. Puisque le vendeur est considéré comme monopoleur pour son produit, il fait face à une demande de forme décroissante et doit déterminer le couple prix-quantité qui maximise son profit.
2- Les **caractéristiques qualitatives** du produit qui rendent celui-ci plus ou moins attrayant pour les consommateurs, mais l'attrait (la qualité) a un coût.
3- Les **dépenses de marketing** : celles-ci augmentent simultanément la demande et le coût des biens.

Ignorant la concurrence monopolistique, la théorie économique traditionnelle ne prenait en compte que la concurrence sur les prix. Mais « more and more is price competition evaded by turning the buyer's attention towards a trade mark or by competing on the basis of quality of service »[1].

La méthode de Chamberlin consiste à analyser successivement la position optimale de la firme pour chacun des trois ajustements pris isolément, c'est-à-dire comme si les deux autres variables restaient fixes. Chamberlin attire l'attention sur le fait que la position optimale combinée des trois ajustements pourrait très bien n'être optimale pour aucun d'eux en particulier.

L'analyse de ces trois types d'ajustements avec tous ses cas particuliers est assez fastidieuse et ne sera pas exposée. Contentons-nous d'une conclusion peu surprenante qui peut en être tirée : économiquement, la concurrence monopolistique tend à être moins efficace que la concurrence parfaite ou pure. La concurrence sur les prix place l'équilibre de la firme sur un point de la courbe CM à gauche de son minimum. Par rapport à la concurrence parfaite, le prix est supérieur et la quantité produite moindre. Le prix plus élevé a pour cause unique le coût de production. Il n'y a pas de surprofit comme en monopole ou en oligopole. Toute autre chose restant égale, les dépenses de marketing élèvent également le prix ; par contre, l'effet sur la quantité vendue ira dans un sens ou dans l'autre, en fonction de la forme des courbes de demande et de coût.

La concurrence monopolistique et la théorie de la rétribution des facteurs

Pour les raisons déjà exposées dans la section consacrée à J. Robinson, le théorème égalisant la rémunération des facteurs avec la productivité marginale en valeur s'avère inadéquat en régime de concurrence monopolistique.

Dans tous les cas où la courbe de demande est décroissante, la dernière unité du facteur rapportera un produit dont le prix est en baisse ; le produit-revenu marginal sera donc inférieur à la valeur du produit marginal.

Donc, en concurrence monopolistique et dans toutes les structures de marché où la demande n'est pas horizontale, le produit total est insuffisant pour payer les facteurs à leur productivité marginale. La théorie qui explique la rémunération des facteurs par leur productivité marginale en valeur est donc valide uniquement pour la concurrence pure.

C'est d'autant plus paradoxal que dans certaines conditions, la demande décroissante est précisément la circonstance qui permet à l'entreprise d'obtenir un profit. Selon Chamberlin, si ce profit devait être partagé entre les facteurs, il ne suffirait pas pour payer chacun à sa

[1] Chamberlin [47] p.73

productivité marginale. Pour cette raison, il se refuse à user du terme « exploitation », comme J. Robinson.

En outre, la concurrence monopolistique cause un deuxième dommage à la théorie de la productivité marginale, plus grave selon Chamberlin. Peut-on dire que les facteurs engagés dans les activités de marketing sont payés suivant la valeur de leur produit marginal ? Pour que la réponse soit positive, il faudrait pouvoir déterminer celle-ci. Si la demande additionnelle du produit promu vient en positif, la demande d'autres produits supprimée par déplacement doit venir en négatif. Dans l'hypothèse d'un revenu total inchangé, les deux se compensent et ce produit marginal est nul. Conclusion : « To hold that factors employed in selling activities are paid in accord with the value of their marginal products would be a manifest absurdity »[1].

[1] Chamberlin [47] p.187

4.2- LE RENOUVEAU PARETIEN DE L'ENTRE-DEUX GUERRES

A la fin des années vingt, l'économie néoclassique, toujours fort marquée par Marshall, tourne en rond dans l'analyse de l'équilibre partiel qui la maintient dans une espèce d'approximation qui bloque son développement mathématique.

Ce sont les parétiens qui lui apporteront un deuxième souffle. Dans les années trente et quarante, de nombreux économistes inspirés par Pareto vont échafauder ensemble un arsenal microéconomique cohérent. Les principaux noms sont John Richard Hicks, Abba Lerner, Maurice Allais, Oskar Lange, Paul Samuelson, Arthur Bowley, Tibor Scitovsky, Lionel Robbins et Harold Hotelling.

Pareto était loin d'avoir épuisé la veine d'idées économiques qu'il avait découverte. Marchant sur ses pas et ceux de Walras, les parétiens raisonnent dans le cadre d'un équilibre général. L'utilisation des mathématiques fait un bond en avant.

4.2.1. Le comportement du consommateur individuel

C'est un article de Hicks, « A Reconsideration of the Theory of Value »[1] paru en 1933, qui lance le renouveau parétien. Hicks écrit : « Of all Pareto's contributions there is probably none that exceed in importance his demonstration of the immeasurability of utility (…) Pareto definitely abandoned this and replaced the concept of utility by this of a scale of preferences. »[2].

LE TAUX MARGINAL DE SUBSTITUTION

S'adressant à ses collègues, Hicks leur propose d'adopter de nouveaux outils et il sera largement entendu. « If total utility is not quantitatively definable, neither is marginal utility. But the theory of value does not need any precise definition of marginal utility. What it does need is only this : that when an individual's system of wants is given, and he possesses any given set of goods, X, Y, Z…, we should know his *marginal rate of substitution* between two goods. The marginal rate of substitution of any good X for any good Y is defined as the quantity of good Y which would just compensate (the individual) for the loss of a marginal loss of X »[3].

Ecrivons maintenant la formule du taux marginal de substitution, noté dorénavant TMS(X/Y), indiquant la quantité de *Y* que pèse en utilité l'unité marginale de *X* :

$$\mathrm{TMS}\ (\mathrm{X}/\mathrm{Y}) = \mathrm{dY}/\mathrm{dX}\ [4] \qquad (4.4)$$

---

[1] L'article, coécrit, comporte deux parties : un exposé des principes par Hicks, suivi d'un développement mathématique par Allen.

[2] Hicks [149] p. 33

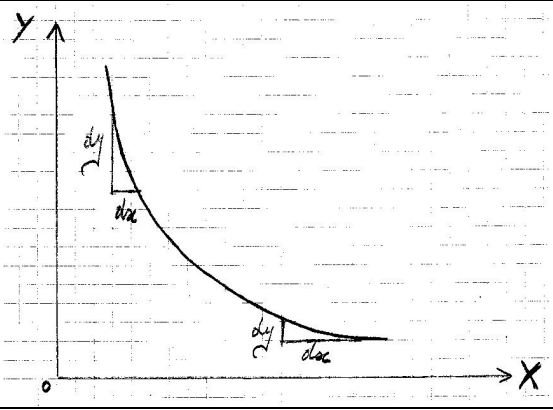
[3] Hicks [149] p. 36. En réalité, l'article dit : "The marginal rate of substitution of any good *Y* for any good *X* is defined as the quantity of good *Y* which would just compensate (the individual) for the loss of a marginal loss of *X*". Peu de temps après avoir écrit cet article, Hicks changea d'avis sur la terminologie et intervertit TMS(X/Y) et TMS(Y/X). Ma citation ci-dessus a été corrigée pour anticiper cette correction, qui s'est imposée dans la microéconomie moderne.

[4] dY et dX sont les variations des quantités consommées de X et Y. La lettre « d » signifie qu'on a affaire à des différentielles et qu'il s'agit donc de variations à la marge. Comme dY et dX sont de signes contraires,

Hicks commente: « It will be evident to the reader that this marginal rate of substitution is nothing else than what we have been in the habit of calling the ratio of the marginal utility of X to that of Y ».

Le principe de l'utilité marginale décroissante est remisé au placard. Lui succède le principe du *taux marginal de substitution décroissant* quand *X* augmente.

Graphique 4.8 : le taux marginal de substitution



Le taux marginal de substitution, TMS(X/Y), correspond à la pente de la courbe d'indifférence et est donc forcément différent en chaque point ; en fait, il décroît lorsqu'on se déplace vers la droite du graphique car la quantité de *Y* nécessaire pour compenser la baisse de *X* devient de plus en plus faible au fur et à mesure qu'on possède plus de *X* par rapport à *Y*. Cette baisse du TMS reflète donc la convexité des courbes d'indifférence. Ceci s'explique par notre préférence pour une consommation diversifiée plutôt que de consommer beaucoup d'un bien et (presque) rien des autres. L'explication du TMS décroissant est donc plus subtile que le simple fait de l'utilité marginale décroissante des biens.

Que devient la règle d'optimisation du consommateur déjà exprimée dans l'équation (3.5) de Walras et le graphique 3.16-A de Pareto ? Hicks n'a plus qu'à traduire la méthode de Pareto dans la nouvelle terminologie. Au point de tangence entre la courbe d'indifférence et la droite de budget, la pente de la droite et celle de la courbe sont nécessairement égales. On a donc l'égalité entre le TMS et le rapport de prix. La nouvelle équation de l'équilibre devient donc :

$$\text{TMS}\ (\text{X}/\text{Y}) = \text{PX}/\text{PY} \qquad (4.5)$$

« This is the form in which we now have to write the law of proportionality between marginal utilities and prices »[1].

Une courbe d'indifférence plus incurvée signifie un TMS décroissant plus nettement, ce qu'on peut interpréter comme une difficulté plus grande de substituer *Y* à *X*. L*'élasticité de substitution* est l'outil permettant de mesurer la substituabilité entre deux biens[2]. Hicks la définit par ce ratio :

le TMS est normalement négatif ; pour éviter cette conséquence, certains auteurs ajoutent le signe *moins* devant le membre de droite de (4.1).

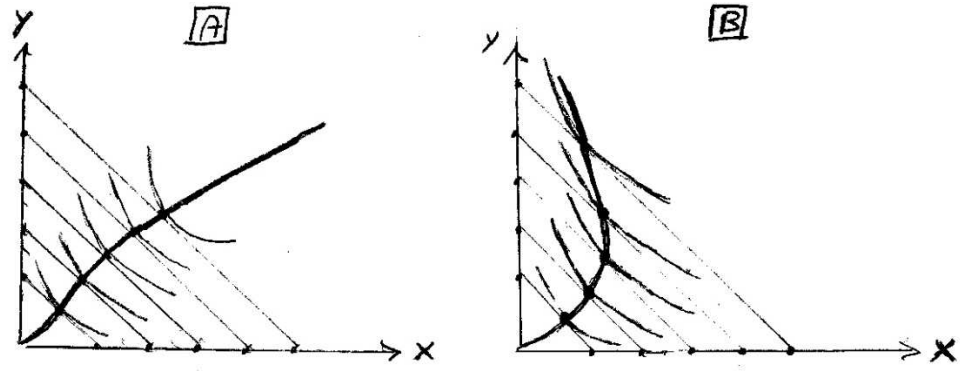
[1] Hicks [149] p. 37

[2] Le cas extrême est celui de biens totalement complémentaires. L'élasticité de substitution sera nulle et les courbes d'indifférences seront des angles droits.

$$\frac{\text{Relative increase in the proportion possessed of two commodities (Y/X)}}{\text{Relative increase in the marginal rate of Y for X}}$$ [1]

Hicks examine ensuite l'**influence du revenu** (budget) sur la consommation. Pour ce faire, il trace sur le graphique du champ d'indifférence d'un individu, des droites de budgets correspondant à une gradation du revenu, différant par l'éloignement de l'origine mais de pente égale (niveau de prix constant). Chaque segment sera tangent à une et une seule courbe d'indifférence. Si l'on relie tous ces points de tangence entre eux, on obtient une *courbe de dépense* qui indique comment évolue la répartition de la consommation entre *X* et *Y* lorsque le revenu augmente. Lorsque la courbe est orientée vers la droite, la consommation des deux biens augmente avec le revenu. Lorsque la courbe tourne vers la gauche à partir d'un certain revenu, cela signifie que la consommation de *X* diminue avec le revenu et que *X* est donc un bien inférieur.

Graphique 4.9 : la courbe de dépense



Dans la figure A, les biens X et Y sont normaux. Dans la figure B, X est un bien inférieur

On peut mesurer l'influence du revenu sur la demande de X par son *élasticité-revenu*, que Hicks mesure par le ratio :

$$\frac{\text{Relative increase in demand for X}}{\text{Relative increase in income}}$$ [2]

La formule suppose évidemment que les prix des biens soient constants. Si l'élasticité-revenu est élevée, la demande de ce bien est très sensible aux variations du revenu (ce pourrait être le cas des biens de luxe). Si l'élasticité-revenu égale 2, une hausse du revenu de 10% entraîne une hausse de la demande de 20%. Dans le cas d'un bien inférieur, l'élasticité est négative.

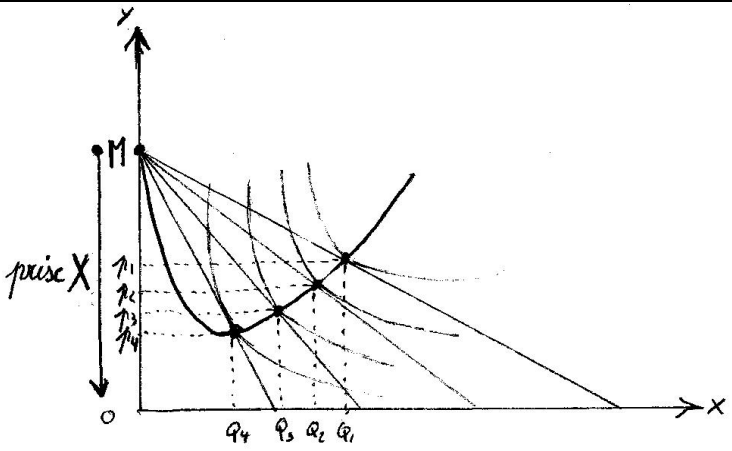
LA COURBE DE DEMANDE

Hicks montre ensuite comment déduire la courbe de demande du bien *X* par un individu à partir de son champ d'indifférence. Pour ce faire, il utilise la même technique que Pareto au graphique 3.16-B (courbe d'échange). Partant d'une quantité fixe *M* du bien *Y*, qui est considérée comme le budget de l'agent, il tire des segments de droite de moins en moins inclinés, qui correspondent à des prix de *X* décroissants. Les points de tangence de ces segments avec les courbe d'indifférence indiquent en abscisse la quantité consommée de *X* aux différents prix. On peut facilement déduire la courbe marshallienne bien connue (graphique 3.8) de la

[1] Hicks [149] p. 40
[2] Hicks [149] p. 45.

présente courbe. Il suffit de construire une nouvelle ordonnée valant l'écart entre *M* et l'ordonnée de la courbe hicksienne ; l'abscisse est parfaitement identique.

Graphique 4.10 : la courbe de demande de Hicks



Une rapide inspection du graphique permet à Hicks de constater que la forme de la courbe de demande est influencée par celle de la *courbe de dépense* et par l'incurvation des courbes d'indifférence. Cette constatation rejoint les conclusions mathématiques d'Allen, selon lesquelles l'élasticité de la demande du bien *X* par rapport à son prix (formule 3.25 de Marshall) peut être décomposée suivant la formule :

$$\varepsilon = k_x.\varepsilon r + (1-k_x).\varepsilon s \qquad (4.6)$$

où ε est l'élasticité marshallienne de la demande
k_x est la part du bien X dans la consommation totale
εr est l'élasticité revenu telle que définie ci-avant
εs est l'élasticité de substitution tel que définie ci-avant

« The increase in demand for a commodity *X*, which results from a fall in its price, depends partly upon the income elasticity of demand for *X*, and partly upon the elasticity of substitution between *Y* and *X*. We can in fact look upon the increase in demand as consisting of two parts, one of which is due to the increase in real income which a fall in the price of *X* entails, the other to the opportunity of substituting *X* for other goods which results from the fall in the relative price of *X* »[1]. Ces deux facteurs sont appelés *effet de revenu* et *effet de substitution*[2]. Logiquement, l'effet de revenu gagne en importance lorsque la part de *X* dans la consommation totale augmente. Il est évident que cette décomposition en deux effets joue également sur la demande du bien *X* lorsque c'est le prix d'un autre bien qui en est la cause.

L'effet de substitution est toujours positif. L'effet de revenu l'est généralement mais il est négatif pour les biens inférieurs. Pour un bien inférieur représentant une part importante de la consommation, cet effet revenu négatif pourrait l'emporter sur l'effet de substitution, d'où il résulterait que la consommation de ce bien diminue quand son prix baisse. La courbe de demande croissante est connue sous le nom de *paradoxe de Giffen*.

[1] Hicks [149] p. 47
[2] En 1915, l'économiste-mathématicien russe, Eugen Slutsky, disciple de Pareto, était déjà parvenu à ce résultat, qui était passé inaperçu jusqu'à ce que Hicks le redécouvre.

Friedman et la demande marshallienne
En 1953, Milton Friedman, économiste important dont il sera question plus loin, publie son article « The Marshallian Demand Curve ». Il y met en doute le bien-fondé du concept d'effet revenu. De deux choses, l'une : soit la variation du prix de *X* est liée à des phénomènes de marché et révèle juste une variation des prix relatifs, auquel cas l'effet de revenu est incongru. Soit la variation du prix de *X* est due à une modification des conditions de sa production. Toutes autres choses restant égales, la hausse de la productivité augmente le revenu total. Mais alors on a affaire à deux faits distincts : d'une part, la variation du prix et de l'autre, celle de la productivité et du revenu ; il est préférable des les étudier séparément. Friedman, disciple de Marshall, était sceptique à l'encontre de l'école de Lausanne. Il est paradoxal que ce soit l'adepte de l'équilibre partiel qui corrige un parétien dans un sens orienté vers l'équilibre général.

Notons que cet effet dual concerne également l'offre des facteurs ; c'est lui qui explique le paradoxe constaté par divers économistes, dont Jevons et Knight, qu'une courbe d'offre de travail décroissante à partir d'un certain point est loin d'être exceptionnelle. En gagnant plus par heure, l'agent peut se permettre de travailler moins d'heure. L'individu a un champ d'indifférence entre la consommation matérielle et le loisir ; l'effet de substitution et l'effet revenu jouent concomitamment et la forme des courbes d'indifférence propre à tel individu peut rendre l'effet revenu prépondérant.

4.2.2. Le comportement du producteur

LA SUBSTITUTION ENTRE LES FACTEURS ET LA SUBSTITUTION ENTRE LES PRODUITS

Le problème du producteur révèle une certaine similarité avec celui du consommateur. Celui-ci devait maximiser son utilité en optimisant le choix entre des produits consommés. Le producteur doit maximiser son profit en optimisant ses choix. Il a deux décisions à optimiser :
- quels biens produire et en quelles quantités ;
- quels facteurs de production utiliser et en quelles quantités.

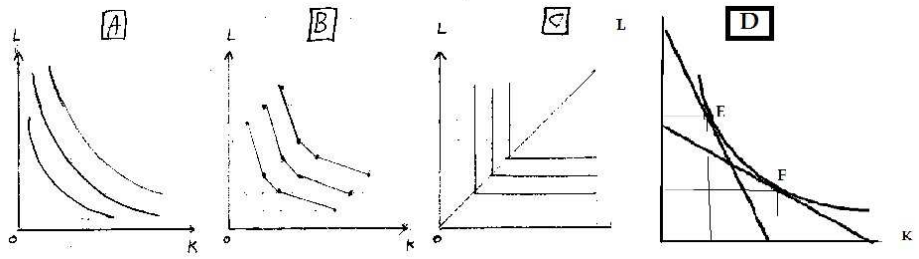
Examinons d'abord le choix relatif aux facteurs de production et supposons que la firme n'offre qu'un seul produit. Les outils analytiques mis en œuvre sont calqués sur ceux de la consommation. On a ainsi conçu une courbe d'indifférence du producteur, qui fut baptisée *isoquante*[1] : une isoquante relie les combinaisons de deux facteurs de production qui rendent une production totale égale[2], ce que montre le graphique 4.11 où les deux facteurs sont désignés par les lettres *K* et *L*. Dans le cas de la *colline du plaisir*, la hauteur avait été « gommée » parce que trop subjective pour être observable ; par contre, la hauteur de la *colline de production* est objective. On démontre que lorsque la fonction de production (et donc de la colline de production) est concave, les isoquantes sont nécessairement convexes par rapport à l'origine.

Les économistes parétiens apprécient tout particulièrement les isoquantes continues et différentiables, comme sur la figure 4.11-A. Pareille courbe suppose une variabilité totale des coefficients de production. N'importe quelle quantité d'un facteur est combinable avec n'importe quelle quantité de l'autre facteur, ce qui est une hypothèse forte. Le graphique 4.11-B

[1] Cette courbe apparut pour la première fois en 1924 chez Bowley. En français, certains l'appellent « isoquant » au masculin, d'autres « isoquante » au féminin.
[2] Il y a en fait potentiellement plus de techniques de production différentes que de points sur le champ d'isoquantes, car des techniques différentes peuvent nécessiter une même combinaison des inputs. Par hypothèse, en ce cas, le point de l'isoquante correspond à la plus efficace de ces techniques.

représente des isoquantes discontinues. Le graphique 4.11-C représente un cas particulier d'isoquante discontinue : l'*isoquante en forme de L*, relative au cas des *coefficients de production fixes* : l'accroissement d'un facteur ne fait pas monter d'isoquante quand l'autre facteur n'augmente pas dans une proportion déterminée.

Graphique 4.11 : les isoquantes



La figure D montre le déplacement de l'équilibre lorsque le rapport de prix se modifie.

Semblablement au taux marginal de substitution entre les biens, il y a un *taux marginal de substitution technique* (TMST) entre les facteurs ; c'est le rapport entre la réduction infinitésimale de l'un d'eux (*K*) et la hausse de l'autre (*L*) nécessaire pour maintenir la production constante. Quel lien, ce taux a-t-il avec la théorie de la productivité marginale maintenant bien connue ? Très logiquement, le taux auquel il faut substituer les facteurs dépend de leurs productivités marginales respectives. Si on réduit un facteur qui a une productivité marginale élevée, il faudra d'autant plus de l'autre facteur pour compenser cette réduction. Lerner écrit : « If *K*'s *mp* (NDLR : *marginal product*) is twice *L*'s, a unit of *K* is substitutable for two units of *L* and its *M* (NDLR: le TMST) for *L* is 2 and its *M* for any other factor is twice *L*'s *M* »[1]. On a donc toujours l'égalité :

$$\text{TMST (K/L)} = \text{PmK/PmL} \qquad (4.7)$$

Dans le cas particulier des *coefficients fixes*, la productivité marginale des facteurs est indéterminée et le concept de TMST n'a pas de sens. Lerner le prouve ainsi : si on réduit la quantité de n'importe quel facteur de 1%, le produit sera réduit de 1% comme si on avait réduit la quantité de tous les facteurs de 1% (en coefficients fixes, les rendements d'échelle sont toujours constants) : si on doit rémunérer les facteurs suivant cette productivité marginale, chacun pourrait revendiquer la totalité du produit, ce qui est absurde. Pour contourner le problème, il préconise de recourir au *produit marginal net* de Marshall.

L'analogie entre le consommateur et le producteur continue quand il faut déterminer l'équilibre. Il est défini par la tangence entre la droite de budget du producteur et une isoquante. La droite du budget est ici appelée *isocoût* parce que les points, qui la composent, correspondent à un coût de production total identique (mais réparti différemment entre les facteurs). Le déplacement de cette droite vers le nord-est correspond à un accroissement du coût ou du budget de production. La pente de cette courbe indique le rapport entre les prix des facteurs de production (qui- ne l'oublions pas- est l'inverse du rapport entre les quantités).

La tangence entre l'isocoût et l'isoquante est exprimée par l'équation :

$$\text{TMST (K/L)} = \text{PK/PL} \qquad (4.8)$$

Lorsque TMST (K/L) > PK/PL, cela signifie que le facteur *K* est utilisé en proportion insuffisante par rapport à *L*, que la productivité marginale de *K* est supérieure à son prix (et celle de L inférieure) et qu'il est souhaitable d'augmenter la quantité de *K* jusqu'à ce que

[1] Lerner [223] p. 138. J'ai changé les lettres désignant les facteurs pour rester cohérent.

l'égalité (4.8) soit réalisée et que la productivité marginale des facteurs égale à leur prix (donné par le marché). A l'équilibre, investir un euro de plus dans le facteur *K* ou dans le facteur *L* aura le même effet sur la production. Tant qu'il est plus rentable d'acheter un facteur que l'autre, le producteur a intérêt à modifier la proportion entre les facteurs dans ce sens.

Venons en maintenant au cas d'une entreprise multiproduits. Elle devra déterminer les quantités entrantes des facteurs et les quantités sortantes des produits de façon à réaliser simultanément une triple égalité que Hicks décrit ainsi[1] :

1. Le rapport de prix de deux facteurs quelconques doit être égal au taux marginal de substitution technique entre eux
2. Le rapport de prix de deux produits quelconques doit être égal au taux marginal de substitution entre les deux produits
3. Le rapport de prix d'un facteur et d'un produit quelconque doit être égal au taux marginal de transformation entre le facteur et le produit (c'est-à-dire, au produit marginal du facteur en terme du produit dont il s'agit).

La première égalité correspond à l'équation (4.8). Le concept de TMS intervenant dans le deuxième principe correspond à l'accroissement de la production du produit *Y* nécessaire pour maintenir constant le profit lorsqu'on réduit infinitésimalement la quantité du produit *X*, toutes autres choses restant égales[2]. Puisqu'à l'équilibre, le producteur et le consommateur égalisent tous deux le TMS(X/Y) avec le rapport de prix entre les deux produits, les TMS des producteurs s'égalisent avec ceux des consommateurs pour toute paire de produits.

Hicks explique qu'avec un seul produit et un seul facteur, la condition d'équilibre serait simplement l'égalité entre le prix et le coût marginal. Les trois conditions précitées aboutissent d'ailleurs à cette égalité, mais sont ensemble plus riches de signification. C'est également l'avis exposé par Lerner dans « The Economics of Control » (1946). Il y étudie l'optimum du bien-être dans une économie collectiviste et dans une économie capitaliste en mettant l'accent sur les ressemblances plus que sur les différences. A l'intention des gestionnaires des entreprises collectivisées, Lerner a établi une règle simple. "… (where *p* stands for the price of the product and *vmf* stands for the value of the marginal amount of factor needed to produce one more unit of the product), the correct formulation of the Rule is as follows: *When p is greater than vmf, apply more of the factor, when p is less than than vmf, apply less of the factor; and when p = vmf, continue to produce at the same rate because the optimum position has been reached"*[3]. La recherche du profit pousse les entreprises privées à appliquer cette règle spontanément. Selon Lerner, sa règle a un double avantage sur l'égalisation du coût marginal avec le prix : d'une part, elle n'implique pas la concurrence parfaite ; d'autre part, elle indique non seulement de combien faire varier la production, mais aussi comment.

Hicks aborde ensuite la problématique de la *complémentarité* des facteurs. Soit une entreprise qui produit le bien unique *X* à partir de trois facteurs *K*, *L* et *M*. Si le prix de *K* augmente, l'effet de substitution fera augmenter les demandes de *L* et de *M*. Mais si *M* est complémentaire par rapport à *K*, sa demande diminue.

Hicks relève un paradoxe : comme le montre la figure 4.11-D, lorsqu'il n'y a que deux facteurs, si on maintient la production constante, ils sont nécessairement des substituts. Par contre, si on étudie la variation de la production, ils sont nécessairement complémentaires, ce que Hicks

[1] Il faut ajouter les conditions de stabilité : TMS croissant entre les produits, TMST décroissant entre les facteurs.

[2] Contrairement au TMS acheteur, le TMS vendeur est croissant.

[3] Lerner [223] p. 128

démontre. S'il y a plus de deux facteurs, cette tendance à la complémentarité entre eux existe également, bien que des facteurs peuvent être substituables deux à deux au sein de cet ensemble complémentaire. Il faut donc se méfier de la figure 4.11-D qui met en exergue un aspect de la relation et -selon Hicks- pas le principal.

Revenons au cas d'un produit *X* et deux facteurs ; quel sera l'effet d'une diminution du prix de *K*, outre le fait que sa demande augmentera ? Hicks envisage trois possibilités :
- L'offre de *X* augmente et la demande de *L* augmente
- L'offre de *X* augmente et la demande de *L* diminue, car la substitution domine la complémentarité
- L'offre de *X* diminue et la demande de *L* également.

Ces cas sont énoncés dans l'ordre décroissant de probabilité. Le troisième cas, que Hicks appelle *régressivité entre X et K*, est très particulier : *K* et *L* sont des substituts, *K* convient mieux pour la production à petite échelle, contrairement à *L* et les ressources de l'entrepreneur pour accroître l'échelle de la production sent réduites.

L'ELASTICITE DE SUBSTITUTION

L'*élasticité de substitution* fut inventée indépendamment par Hicks dans l'ouvrage « The Theory of Wages » (1932) et par J.Robinson dans l'ouvrage « The Economics of Imperfect Competition » (1933). Elle mesure la substituabilité entre les facteurs de production. Il ne s'agit pas d'un outil à proprement parler parétien mais il présente une indéniable consonance parétienne comme en témoigne l'élasticité de substitution entre les produits, au niveau du consommateur, présentée au sous-chapitre 4.2.1. Le parallélisme est évident[1].

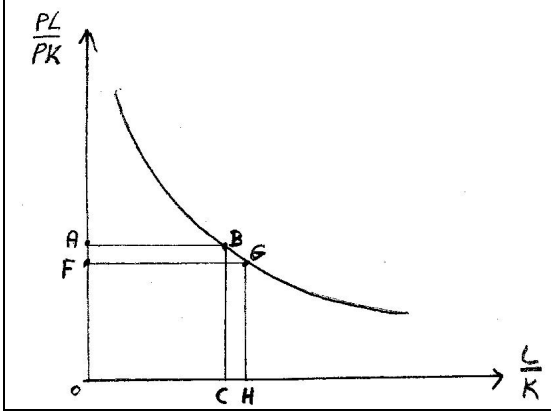
Le concept apparaît dans le chapitre de l'ouvrage de **Hicks** intitulé « Distribution and Economic Progress », qui s'intéresse aux effets du progrès sur la répartition du revenu entre les facteurs de production. Ce chapitre, comme le reste de l'ouvrage, provoqua un large débat en 1933 auquel prirent part nombre d'économistes. Notamment, la généralisation de la théorie de la productivité marginale à l'échelle de l'économie globale suscita des objections. Comme suite, Hicks publia en 1936 une version révisée de ce chapitre, moins ambitieuse et plus simple, que nous examinons à présent.

Hicks travaille d'abord avec des hypothèses simplificatrices, puis il examine ce qu'il reste de sa théorie lorsqu'on les abandonne. Les hypothèses sont :
- il n'y a que deux facteurs de production (que je désignerai par *K* et *L*).
- la concurrence parfaite règne sur le marché des produits et les marchés des deux facteurs.
- Les rendements d'échelle sont constants
- on se situe au niveau de la firme ; il n'y a qu'un seul produit ; les facteurs de production sont rémunérés par le partage de son produit physique.

La concurrence parfaite et les rendements d'échelle constants font coïncider la rémunération avec la productivité marginale ; la productivité marginale des facteurs ne dépend que du ratio de leurs quantités et non de la quantité absolue qui est utilisée. On peut donc dessiner une courbe affichant la relation entre le rapport de prix des facteurs et le rapport entre leurs quantités. Hicks reprend le graphique 4.12 de Lerner, l'un des économistes ayant participé au débat.

[1] Il serait logique de parler d'« élasticité de substitution technique », mais l'épithète « technique » est généralement ignoré.

Graphique 4.12 : l'élasticité de substitution



L'élasticité de substitution (σ) n'est autre que l'élasticité de cette courbe, c'est-à-dire le rapport entre la variation relative de la variable en abscisse sur la variation relative de la variable en ordonnée. On a donc :

$$\sigma = \frac{\frac{\Delta(L/K)}{(L/K)}}{\frac{\Delta(PL/PK)}{(PL/PK)}} \qquad (4.9)$$

« That the elasticity of substitution of *K* for *L* is the same as the elasticity of substitution of *L* for *K* may be sufficiently indicated by remarking that a relative increase in K is the same thing as a relative decrease of *L* »[1]. Si on avait K/L en abscisse et PK/PL en ordonnée, la courbe du graphique 4.12 serait identique et l'élasticité de substitution également.

Chaque point de la courbe inscrit un rectangle entre les axes. La surface de ce rectangle, valant PL. L / PK. K, représente le rapport entre les parts de *L* et de *K* dans le revenu à partager. La variation de cette surface est fonction de l'élasticité de la courbe. Si σ >1, la surface du rectangle augmente vers la droite du graphique : l'accroissement de l'intensité en travail L/K élève alors la part relative du travail dans le revenu. Lorsque σ = 1, cette part reste constante ; lorsque σ < 1, la part relative du travail diminue[2].

Hicks constate que l'abandon des hypothèses simplificatrices est très difficile à gérer. Qu'on pense seulement que la courbe du graphique 4.12 n'est plus définie sans les hypothèses des rendements d'échelle constants et de la concurrence parfaite. Que la symétrie n'existe plus s'il y a plus de deux facteurs. Nous laisserons de côté ces difficultés.

Lorsqu'un facteur voit son offre augmenter alors que celle des autres reste constante, sa part RELATIVE dans le revenu national s'accroîtra si σ >1. Sa part ABSOLUE augmentera si l'élasticité de sa demande est supérieure à un. La part ABSOLUE des autres facteurs pris comme un ensemble s'accroîtra toujours mais pas nécessairement celle de chacun d'eux.

Robinson fait intervenir l'élasticité de substitution dans un chapitre intitulé « The Demand Curve for Labour of an Industry ». Elle en définit le concept : « The degree to which substitution of factors is possible can best be measured by considering the change in the ratio of the factors which occurs when the relative prices alter (…) It appears appropriate to call the proportionate change in the ratio of the amounts of the factors employed divided by the proportionate change in the ratio of their prices to which it is due, the *elasticity of substitution*,

[1] Hicks [155] p. 290

[2] Il faut rappeler que, hors le cas d'une hyperbole équilatère, σ varie d'un point à l'autre. Hicks semble simplifier en considérant σ constant le long de la courbe.

by analogy with elasticity of demand or of supply. The elasticity of substitution is determined by the technical conditions of production »[1]. Robinson limite son analyse à la concurrence parfaite et pose que les rendements d'échelle sont constants.

La question essentielle est l'impact d'une modification du salaire sur la quantité totale employée des facteurs de production et notamment du capital. Supposons une baisse du salaire. Le capital employé aura tendance à diminuer du fait de la substitution entre les facteurs dans une mesure d'autant plus grande que σ est élevé. Parallèlement, il tendra à augmenter parce que la baisse du prix du produit stimulera la production, dans une mesure d'autant plus grande que la demande du produit est plus élastique. Notons cette dernière élasticité par ε. Le rapport entre ces deux élasticités sera déterminant pour l'évolution de la quantité de capital. Si $\varepsilon > \sigma$, la quantité de capital augmentera, si $\varepsilon = \sigma$, elle ne variera pas, si $\varepsilon < \sigma$, elle diminuera.

Si une réduction du salaire entraîne la hausse du capital employé dans une industrie, ce qui implique que $\varepsilon > \sigma$, on peut en conclure que l'élasticité de la demande de travail (ε') est a fortiori supérieure à σ, vu la hausse relative de la quantité de travail employée.

Selon Marshall, l'élasticité de la demande d'un facteur est d'autant plus grande que :

1- L'élasticité de la demande du produit est élevée.
2- Les possibilités de substitution entre les facteurs sont importantes.
3- Le facteur représente une part importante du coût total du produit.
4- L'offre des autres facteurs est plus élastique.

Selon Robinson, les deux derniers facteurs doivent être nuancés. Elle démontre que la troisième proposition n'est correcte que si $\varepsilon > \sigma$. Lorsque $\varepsilon = \sigma$, ε' est indépendante de la part du travail dans le coût total et lorsque $\varepsilon < \sigma$, ε' varie en sens inverse de cette proportion. Il faut apporter une restriction un peu semblable à la quatrième proposition. Si $\varepsilon = \sigma$, une variation du salaire n'affectera pas la quantité de capital utilisée et le prix du capital restera inchangé, quelle que soit l'élasticité de son offre. En fait, l'élasticité ε' n'est affecté par l'élasticité de l'offre de capital que s'il y a divergence entre ε et σ et le sera d'autant plus que l'écart est important.

*

L'élasticité de substitution se marque dans la forme des isoquantes. Une incurvation plus forte signifie une faible substituabilité. L'élasticité de substitution σ varie entre zéro et l'infini. Si $\sigma = 0$, les biens sont parfaitement complémentaires et les isoquantes sont en angle droit (figure 4.11-C). Si $\sigma = \infty$, l'isoquante sera une droite et la substituabilité sera totale, ce qui est un cas très particulier : si deux facteurs sont à ce point substituables, on peut les considérer comme un seul facteur.

L'ECHELLE DE LA PRODUCTION

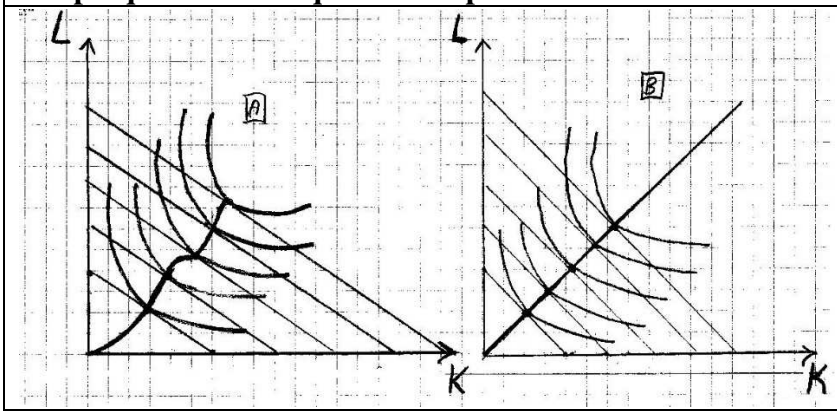
Des isocoûts parallèles représentent des échelles de production différentes, les prix des facteurs étant donnés. Pour chaque isocoût peut être déterminé un optimum. La courbe reliant ces optimums s'appelle le *chemin d'expansion* de la firme (graphique 4.13). L'avancement sur le chemin d'expansion en s'éloignant de l'origine correspond à l'accroissement de la taille de la firme[2]. Cette définition de la taille de l'entreprise ne correspond pas exactement à celle de Viner (quantité du facteur fixe : cf. supra).

[1] Robinson [301] p. 256

[2] Comme l'a fait remarquer Chamberlin, vu les indivisibilités, il est faux de croire que le chemin d'expansion doit nécessairement partir de l'origine.

Sur la figure de gauche, le chemin d'expansion évolue en zigzag car la proportion idéale entre les facteurs change lorsque la taille augmente. Dans l'exemple de droite, le chemin est rectiligne car cette proportion est constante.

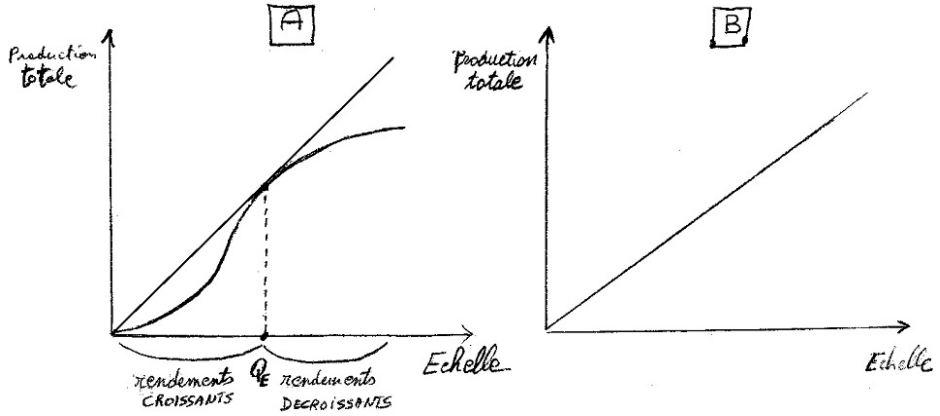
Graphique 4.13 : L'équilibre du producteur et le chemin d'expansion



Y a-t-il une taille (ou échelle de production) optimale en ce sens qu'elle minimise le coût de production moyen ? Autrement dit, sur le chemin d'expansion, y a-t-il un optimum plus optimal que les autres ? Manifestement, le graphique 4.13 (A ou B) donne la proportion optimale entre les facteurs pour chaque taille, mais rien n'y indique la taille idéale. Pour déterminer celle-ci, il faut faire appel à la troisième dimension de la colline de production, la hauteur, que nous avions éliminée. C'est ce que fait le graphique 4.14, qui montre une coupe verticale de la colline de production tout au long du chemin d'expansion[1] ; autrement dit la quantité produite en fonction de la taille. La forme de la courbe dépend donc du type de rendements d'échelle auquel nous avons affaire.

Dans le cas des rendements constants, le chemin d'expansion est nécessairement rectiligne. L'inverse n'est pas vrai : un chemin d'expansion rectiligne (proportions constantes) est compatible avec tous les types de rendements d'échelle.

Graphique 4.14 : Coupe verticale de la colline de production



[1] Cette courbe rappelle la courbe de la production totale du graphique 4.1. Il y a toutefois une différence importante entre elles. Le long du chemin d'expansion, les deux facteurs de production sont variables, alors que le graphique 4.1 comportait un facteur fixe et un facteur variable : elle était une coupe de la colline de production perpendiculaire à l'axe du facteur fixe.

Nous retrouvons la conclusion contenue dans les sous-chapitres 3.2.5 et 4.1.1. Il n'y a de taille optimale que si les rendements d'échelle sont croissants puis décroissants ; elle se situe alors à la limite entre ces deux zones (point Q_E sur la figure A).

LES INDIVISIBILITES

Nous nous référerons à la théorie de **Lerner** en la matière. Commençons par définir de quoi il s'agit : « … when we consider that factors are often available only in large units like waterways, that products are often produced in naturally large units like ocean liners or skyscrapers, and that methods of production are often of a minimum size even if the factors and the products are perfectly divisible, like an assembly plant for automobiles or a continuous strip-steel rolling mill »[1].

Les considérations qui suivent montrent que, manifestement, Lerner n'aime pas les indivisibilités.

Dans la suite de l'exposé, nous supposerons un facteur *K* sujet à une *indivisibilité* importante, combiné avec le facteur *L* parfaitement *divisible*. Pour accroître sa production dans une proportion modérée, l'entreprise doit accroître la quantité de *L* en gardant constante celle de *K*, ce qui implique une variation de la proportion entre les facteurs qui n'aurait pas été choisie si tous les facteurs étaient divisibles. Il s'avère ainsi que les rendements d'échelle ne peuvent pas être constants en présence d'indivisibilités.

La productivité marginale du facteur indivisible, qui ne se manifeste que lors de « sauts » plus ou moins grands, peut être soit positive, soit nulle, soit négative. Envisageons le cas où elle est négative : le facteur divisible sera alors nécessairement dans une phase de rendements croissants (secteur I sur le graphique 4.1-B). Si les rendements restent croissants au-delà du point où le produit marginal en valeur de *L* égale son prix, le facteur coûtera plus à l'entreprise qu'il ne lui rapporte puisqu'en rendements croissants, la productivité marginale réglant la rémunération est supérieure à la productivité moyenne réglant la recette. Si règne la concurrence parfaite, la perte est inévitable. Par contre, le système collectiviste pourrait atteindre la répartition optimale des ressources ; il aurait avantage à pousser la production jusqu'à l'égalisation de la productivité marginale en valeur de *L* avec son prix, malgré la non-couverture de la rémunération de *L*.

« The indivisibility becomes significant when it is large enough to destroy perfect competition. To overcome the indivisibility, the output of the firm might have to be so large that only one firm or perhaps a small number of firms could satisfy the demand… »[2]. Même dans le cas où de nombreuses firmes peuvent se payer le facteur indivisible et où sa productivité marginale est positive, celles qui pourront acquérir une quantité supérieure du facteur divisible auront un coût moyen plus bas car elles pourront atteindre une proportion entre les facteurs où l'excès du facteur indivisible est moindre, voire totalement résorbé. Si tous les facteurs étaient divisibles, les rendements d'échelle constants permettraient à toutes les firmes d'atteindre la proportion optimale entre les facteurs et avec la même aisance[3]. On peut donc dire que les indivisibilités, à la fois, avantagent la grande taille (ce qui nuit à la concurrence) et qu'elles entravent l'accession à la répartition optimale des facteurs.

---

[1] Lerner [223] p. 174.
[2] Lerner [223] p. 176.
[3] D'après Lerner, en rendements constants, l'équilibre à long terme qui implique le profit nul, n'incite pas les firmes à l'accroissement de l'échelle de la production.

Chamberlin conteste cette conception, largement répandue, qui loue la *divisibilité*. Dans la recherche de l'efficacité, il y a deux variables : la proportion entre les facteurs et l'échelle. Chez beaucoup d'économistes, la première est devenue hégémonique. Pourtant, les avantages ou inconvénients de la grande taille sont nombreux et de tous ordres. « In summary, it appears that indivisibilities play no part whatever in explaining economies of scale. Where all factors are perfectly divisible, efficiency remains nevertheless a function of size »[1]. Il déplore cette dichotomie entre l'optimisation de la taille et celle de la proportion. Dans bien des cas, il est impossible de faire varier l'une sans modifier l'autre.

4.2.3. L'optimum

Rappelons-nous que l'*optimum de Pareto* correspond à une situation où il est impossible d'accroître le bien-être d'un agent sans réduire celui d'un autre. Quelles sont les caractéristiques qui distinguent une telle économie ? Quatre conditions sont à la fois nécessaires et suffisantes (simultanément). Il doit y avoir :

- répartition optimale de la consommation par les ménages
- offre optimale des facteurs par les ménages
- répartition optimale des facteurs entre les firmes
- adéquation optimale des produits offerts par les firmes aux besoins des consommateurs.

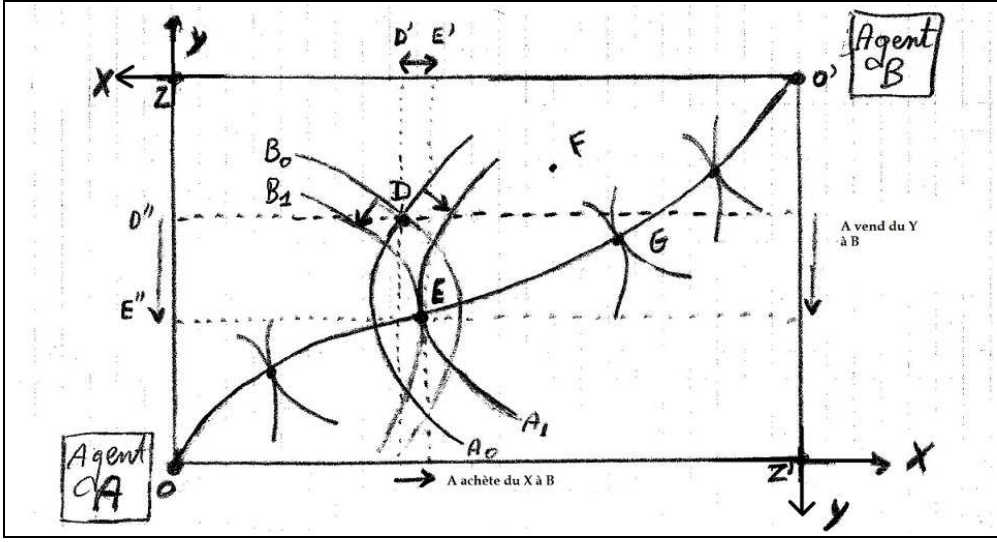
Certains économistes démontrent ces conditions par l'algèbre. D'autres ont élaboré une présentation graphique simplifiée, impliquant deux consommateurs (*A* et *B*) consommant deux biens (*X* et *Y)* produits chacun par une firme, où l'offre des facteurs (*K* et *L*) est considérée comme une donnée exogène. Examinons cette présentation devenue très courante.

L'OPTIMUM DE LA CONSOMMATION

Pour l'analyser, dessinons le graphique imaginé par Arthur Bowley, qu'on appelle la *Boite d'Edgeworth* et qui s'inspire à la fois du graphique 3.4 d'Edgeworth et du graphique 3.17 de Pareto. Emboîtons les champs d'indifférence des individus *A* et *B* en tête bêche ; nous obtenons une boite dont la largeur et la hauteur représentent respectivement les quantités totales de *X* et *Y* disponibles pour la consommation. Chaque point à l'intérieur de la boite représente donc une *allocation*, une répartition de *X* et *Y* entre *A* et *B*. Un échange entre eux est donc marqué par le passage d'un point à un autre.

[1] Chamberlin [47] p.244.

Graphique 4.15 : la boite d'Edgeworth-Bowley



Posons que la répartition de départ soit *D*, c'est-à-dire que *A* possède Z-D' de *X* et *B* en possède une quantité D'-O'. *A* possède O-D" de *Y* et *B* en possède Z-D". *A* est sur la courbe d'indifférence A_0 et B sur la courbe B_0. S'ils passent en *E*, ils ont tous les deux augmenté leur utilité car les courbes A_1 et B_1 sont supérieures aux courbes initiales. Par contre, en *E*, il est impossible que tous deux passent à une courbe supérieure, puisque leurs courbes sont tangentes ; *E* est donc une répartition optimale au sens de Pareto. Si on relie tous les points qui répondent à cette caractéristique, on obtient la *courbe de contrat* O-O'[1].

La multiplicité des optimums parétiens (tous les points de la courbe de contrat) s'explique par le fait que des allocations initiales différentes aboutissent à des allocations d'équilibre différentes. Si, au lieu de *D*, la répartition initiale avait été *F*, la répartition d'équilibre aurait été *G*. L'optimum parétien est un optimum par rapport à une répartition initiale et non un optimum absolu. Lorsqu'on progresse de O à O', l'agent *A* passe à des courbes d'indifférence supérieures et l'agent *B* à des courbes inférieures. Dans l'optique parétienne, on ne peut pas hiérarchiser les allocations sur la courbe de contrat ; même l'allocation *F* et l'allocation *E* ne pourraient l'être hiérarchisées. Par contre, *E* est indiscutablement supérieur à *D*.

Puisque les courbes d'indifférence des consommateurs sont tangentes, nous pouvons énoncer la première condition de l'optimum :

$$\mathrm{TMS_A(X/Y) = TMS_B(X/Y)} \qquad (4.10)$$

Si on la généralise, elle devient : tous les consommateurs doivent avoir le même taux marginal de substitution entre les produits pris deux à deux.

La courbe de contrat ci-dessus relie des optimums limités à la seule consommation. Nous avons supposé données les quantités disponibles des biens *X* et *Y* ; autrement dit, nous avons ignoré la production, que nous introduisons maintenant.

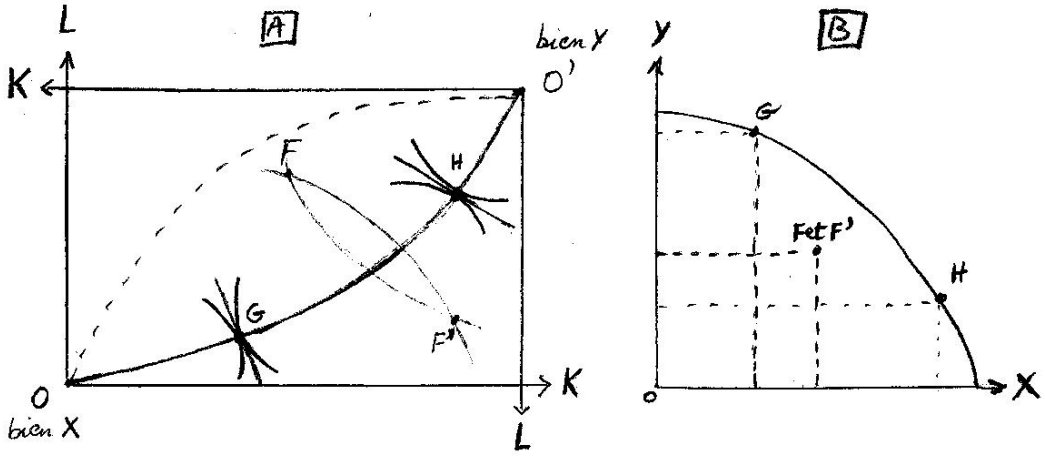
---

[1] La présente *courbe de contrat* a une signification totalement différente de celle d'Edgeworth au graphique 3.4, malgré la similitude apparente. Dans le graphique d'Edgeworth, chaque point du champ représente non pas une allocation mais un échange. Edgeworth montrait qu'on ne pouvait pas prévoir lequel des échanges sur sa courbe de contrat émergerait en tant qu'équilibre. Dans l'optique parétienne, ce problème d'indétermination de l'échange est ignoré car les prix sont de toute façon considérés comme donnés par le marché.

L'OPTIMUM DE LA PRODUCTION

L'optimum consiste à maximiser la production totale à partir de ressources données en facteurs *K* et *L*. Appliquons la même procédure que pour l'optimum à la consommation ; construisons une *boite d'Edgeworth à la production* avec les champs d'isoquants des industries produisant les biens *X* et *Y*. Les allocations optimales des facteurs, celles à partir desquelles un transfert de facteurs ne peut augmenter la production d'une industrie qu'au détriment de celle de l'autre, se caractérisent par la tangence entre les isoquants. Elles forment ensemble la *courbe de contrat à la production*, exposée au graphique 4.16.

Graphique 4.16 : La boite d'Edgeworth à la production



La forme de la *courbe de contrat à la production* reflète le rapport entre l'intensité K/L dans l'industries *X* et l'intensité K/L en *Y*. La courbe en trait continu sur la figure A correspond au cas où la production du bien *X* est plus intense en facteur *K* et celle de *Y* est plus intense en facteur *L*. Lorsque l'on passe du point G au point H, la production de *X* est augmentée au détriment de celle de *Y*. Le plein emploi des facteurs étant postulé, les facteurs *K* et *L* ainsi libérés par la production de *Y* sont reconvertis dans l'industrie *X*, mais ils ne le sont pas dans la proportion correspondant à la technologie existante ; comme le marché connaît un excédent de facteur *L*; son prix baisse par rapport à celui de *K* et les deux industries adaptent leur consommation des facteurs en conséquence. La pente de la tangente plus faible en H qu'en G exprime cette variation du rapport de prix. La courbe de contrat pointillée est celle qui aurait prévalu si la production de *X* avait été plus intense en *L* et celle de *Y* plus intense en *K*.

La deuxième condition de l'optimum de Pareto est donc :

$$\mathrm{TMST_X(K/L)} = \mathrm{TMST_Y(K/L)} \qquad (4.11)$$

Si on la généralise, elle devient : toutes les firmes doivent avoir le même taux marginal de substitution technique entre les facteurs pris deux à deux. Si cette condition n'est pas remplie, la production de certains biens pourrait être augmentée (sans réduction d'autres productions) en transférant des facteurs vers les entreprises où leur produit marginal est relativement plus élevé[1].

[1] Wicksteed avait déjà exposé cette condition d'équilibre (cf. supra).

L'ADEQUATION DE LA PRODUCTION A LA CONSOMMATION

Examinons le graphique 4.16-B. La surface entre l'abscisse et l'ordonnée correspond à l'ensemble des combinaisons des productions de *X* et de *Y*. La courbe qui y apparaît est appelée *frontière des possibilités de production* (FPP) ou *courbe de transformation*. Elle relie les combinaisons optimales, celles de la courbe de contrat à la production. En deçà se retrouvent les combinaisons infra-optimales, c'-à-d les autres points de la boite d'Edgeworth à la production ; les points au dessus de la courbe représentent les combinaisons inaccessibles. La courbe décline évidemment ; sa concavité signifie que l'augmentation unitaire de la production d'une industrie nécessite le sacrifice d'une quantité croissante de l'autre production ; elle s'explique par la différence d'intensité dans l'usage des facteurs entre les deux industries. Le *taux marginal de transformation* (TMT) est précisément le rapport à la marge entre l'accroissement de la quantité produite d'un bien et la réduction de la production de l'autre. Il correspond à la pente de la courbe de transformation.

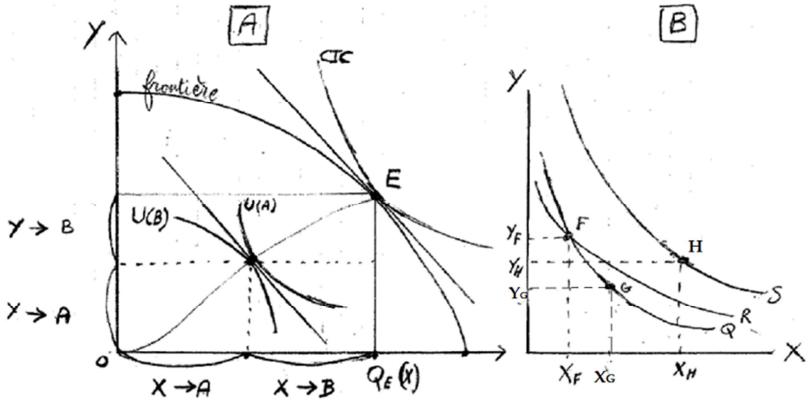
La troisième condition de l'optimum est :

$$\mathrm{TMS(X/Y)} = \mathrm{TMT(X/Y)} \qquad (4.12)$$

Les quantités produites des différents biens doivent être déterminées de telle façon que le *taux marginal de transformation* entre les produits deux à deux soit égal à leur *taux marginal de substitution*. La satisfaction sera maximisée si, à la marge, les éléments qui conduisent les entreprises à produire telles quantités des différents biens trouvent entre eux le même équilibre que celui qui guide les choix des consommateurs.

Visualisons cette condition sur le graphique 4.17. Chaque point de la courbe de transformation définit un rectangle entre les axes. Chacun de ces rectangles est une boite d'Edgeworth à la consommation, puisque la longueur et la largeur définissent les quantités disponibles des deux produits. Dans chacune de ces boites, il y a un point candidat à l'optimum où la pente des courbes d'indifférence est la même que celle de la courbe de transformation (tangentes parallèles). Quel point de la courbe de transformation, autrement dit quelle combinaison de *X* et *Y* choisir ? Comment déterminer la combinaison qui offre le maximum de satisfaction à *A* et *B* ensemble tout en respectant la condition de non-additivité des satisfactions individuelles ? Si on pouvait concevoir des courbes d'indifférence qui indiquent les niveaux de satisfaction commun de *A* et *B*, l'optimum se situerait à la tangence entre une telle courbe et la courbe de transformation.

Graphique 4.17 : L'adéquation de la production à la consommation



De telles courbes existent ; elles s'appellent *courbes d'indifférence collectives* (CIC) et ont pour père Tibor **Scitovsky**. Il les a conçues pour analyser les effets de l'instauration de tarifs

douaniers ; on sait que dans une même communauté nationale, ceux-ci font des gagnants et des perdants. Il s'interrogeait comment savoir si la communauté y gagne ou y perd globalement ?

Selon Scitovsky, la décision politique doit prendre en compte l'efficience et la justice distributive, mais l'économie politique doit les analyser séparément. Une allocation *H* est évidemment plus efficiente que l'allocation *F* si le passage de *F* à *H* améliore directement la situation de chacun. Mais elle le serait également avec en *H* des gagnants et des perdants directs si une redistribution appropriée des biens *X* et *Y* entre les intervenants pouvait les rendre tous gagnants, que cette redistribution ait lieu ou non. Ce qui revient à dire ceci: ceux qui bénéficient du passage à *H* continueraient à en bénéficier, même après avoir indemnisé ceux à qui *F* est désavantageux : "It will be convenient to express this by saying that we make sure whether the people who would benefit by the change could profitably bribe those harmed into accepting it. Secondly, we must make sure that the people who are against the change would be incapable of bribing those in favor to vote against it, without thereby losing more than they would if the change were carried"[1].

Sur la figure 4.17-B, il y a indifférence entre les états *F* et *G* si le lot $(X_G - X_F)$ permet de compenser EXACTEMENT les pertes d'utilité des membres perdants. En reliant toutes les allocations qui ont cette relation avec l'allocation *F*, on obtient la CIC *Q*. En fait, une infinité de CIC passent par chaque point, qui se coupent forcément en ce point ; elles correspondent à des répartitions différentes de la satisfaction initiale entre les membres de la communauté. Mais, « once the distribution of welfare is given, the shape of the community indifference curve is uniquely determined »[2]. Lorsque sur la figure 4.17-B, en se dirigeant vers le nord-est, on passe sur une nouvelle CIC qui ne coupe pas la précédente, l'efficience s'est accrue.

Les CIC de Scitovsky ne résultent donc pas de l'agrégation de courbes d'indifférence des individus constituant la collectivité, par une espèce d'additivité. Samuelson a d'ailleurs démontré qu'une telle agrégation était impossible.

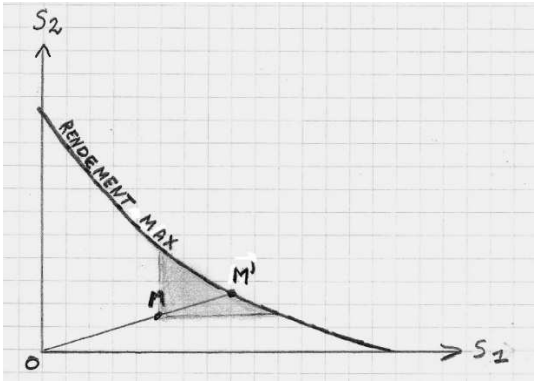
Le point E de la figure 4.17-A donne l'impression que le modèle dégage un optimum unique. Mais des distributions initiales différentes engendrent des CIC différentes. Il y a donc un optimum pour chaque distribution initiale.

*

Revenons à la distinction entre l'efficience et la justice avec Maurice **Allais** qui dessine le graphique 4.18. Différents états économiques sont localisés entre les axes en fonction de la satisfaction obtenue respectivement par l'agent un (abscisse) et l'agent deux (ordonnée)[3]. Sous la plume d'Allais, l'optimum de Pareto se nomme *maximisation du rendement social*. Il appelle *surface de rendement maximum* la courbe correspondant aux états pareto-optimaux, déclinante puisqu'on ne peut y accroître la satisfaction de l'un qu'en diminuant celle de l'autre. La justice de la distribution sur cette courbe peut être débattue, mais l'efficience y est maximisée.

[1] Scitovsky [330] pp. 60-61.
[2] Scitovsky [330] p. 63.
[3] On est dans l'hypothèse simplificatrice d'une économie à deux agents.

Graphique 4.18: rendement social d'un état économique



Le long des demi-droites partant de l'origine, la répartition de la satisfaction est constante.

Envisageons l'état économique correspondant au point *M* : tous les états de la zone hachurée lui sont pareto-supérieurs, puisque la satisfaction augmente pour au moins un agent sans sacrifice de l'autre. Pour déterminer le rendement de l'état *M*, il faut effectuer une comparaison de satisfaction entre ce point et un des points de la surface de rendement maximum bordant l'aire grisée. Allais prend comme référence de comparaison, l'état *M'* qui se caractérise par une répartition identique de la satisfaction entre les agents. Allais calcule ainsi le rendement d'un état économique : un rendement de n% signifie que l'accession au rendement maximum implique l'accroissement de la satisfaction de chaque individu d'un pourcentage égal à (100–n)/100.

Allais écrit : « l'existence d'un rendement social inférieur constitue donc au point de vue économique l'équivalent d'une fuite de vapeur due à la mauvaise étanchéité d'un joint dans une usine. Il y a là une perte sèche, qui est introduite par l'état même du système et à laquelle il peut être remédié sans dommage pour personne »[1].

EQUILIBRE CONCURRENTIEL ET OPTIMUM

A l'aide des nouveaux outils, les parétiens ont redémontré l'existence d'un équilibre général en concurrence parfaite. Ils ont également prouvé la coïncidence entre cet équilibre et l'optimum de Pareto. L'une des premières démonstrations complètes est celle d'Allais (1943)[2]. Allais énonce ainsi le premier théorème fondamental du bien-être :
« *pour que le rendement social soit maximum, il est nécessaire et suffisant que l'état économique considéré puisse être regardé comme l'état d'équilibre d'une économie de concurrence, c'est-à-dire qu'il existe, explicitement ou implicitement, un système de prix concurrentiel* »[3].

Allais montre que dans une économie comportant à la fois des secteurs monopolisés et des secteurs concurrentiels, il est également possible de maximiser le rendement social ; la condition en est que les secteurs monopolisés fixent leur prix au niveau de leur coût marginal et que la quantité qu'ils offrent minimise leur coût moyen. Mais nous avons vu au sous-chapitre 4.1.1 que ces règles ne correspondent pas au comportement d'un monopoleur privé maximisant son profit. Allais préconise donc la nationalisation des monopoles.

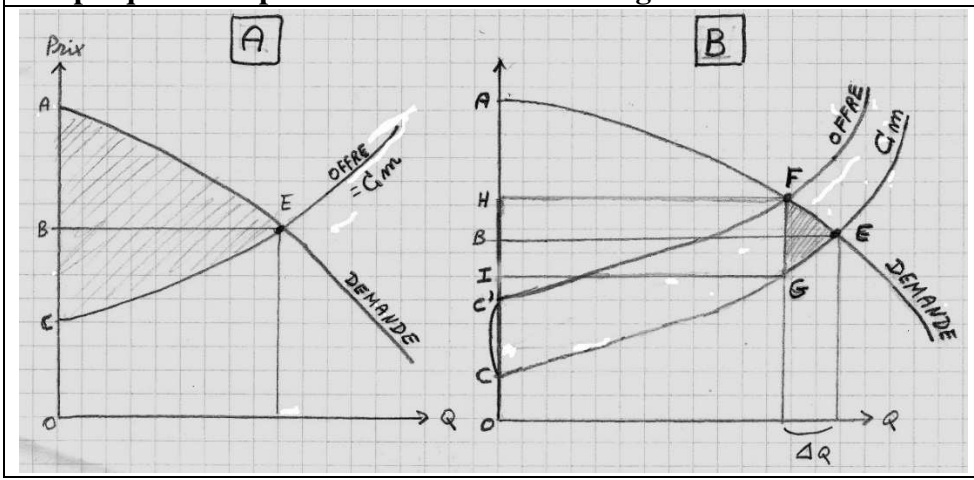
---

[1] Allais [8] p. 510
[2] A cette occasion, Allais remarque que la démonstration de Pareto comportait quelques lacunes, aussi bien mathématiques que de logique économique.
[3] Allais [8] p. 619

Ceci nous amène à évoquer la thèse de la ***tarification au coût marginal*** défendue par **Hotelling**, dans son article « The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates » (1938). Dans cet article, il démontre que le bien-être est maximisé lorsque, pour tous les biens, le prix payé par les consommateurs équivaut au coût marginal.

La figure 4.19-A montre l'équilibre sur le marché d'un bien particulier, avec la courbe d'offre croissante (correspondant au coût marginal) qui coupe la courbe de demande descendante. La surface ABE correspond au *surplus du consommateur* ; la surface BCE correspond au *surplus du producteur*[1]. La surface ACE donne donc le bien être total.

Graphique 4.19 : prix s'écartant du coût marginal



Supposons maintenant qu'une accise forfaitaire frappe ce produit. Comme le montre le graphique 4.19-B, la courbe d'offre se trouve surélevée par rapport à celle du coût marginal d'un écart CC'. Elle coupe la demande en F, ce qui entraîne une baisse de la quantité échangée. Le prix payé par le consommateur vaut OH et le prix perçu par le vendeur égale OI. Le bien être social comporte trois parties :

- le surplus du consommateur AHF
- le surplus du producteur CIG
- le revenu fiscal HIGF

Par rapport au bien-être antérieur à la taxation, il y a donc une perte équivalente à la surface GEF.

En fait, la démonstration de Hotelling est algébrique et mieux en phase avec l'équilibre général parétien que notre exposé basé sur le graphique 4.19. Mais lui-même trouve que ce graphique illustre didactiquement le problème ; nous nous en contenterons.

Il y a au moins quatre situations où le prix payé par le consommateur s'écarte du coût marginal :

- les taxes sur les produits (et de façon inverse les subventions)
- les coûts fixes importants[2].
- Les rendements croissants

[1] Tout comme le consommateur aurait payé plus cher pour les unités précédentes (cf. Marshall : graphique 3.9), le producteur se serait contenté d'un prix inférieur, correspondant au coût marginal.

[2] Samuelson met en garde contre l'idée courante que le prix égal au coût marginal est nécessairement insuffisant pour couvrir le coût unitaire complet du fait que le coût marginal ne comporte que des coûts variables. Au point d'équilibre, le coût marginal peut être supérieur au coût moyen variable. Par contre, il est exact que si les coûts fixes sont élevés, le coût marginal risque d'être inférieur au coût complet.

- les monopoles (cf. graphique 4.6-B).

Concernant les taxes et les subventions affectant les produits, Hotelling conclut à leur nocivité. Le bien-être serait accru (au sens de Pareto) si elles étaient remplacées par la fiscalité sur les revenus et les successions (pour une recette fiscale totale donnée). On constate toutefois sur le graphique 4.19-B que la perte de bien-être GFE est moindre dans le cas où les courbes d'offre et de demande ont une pente élevée. En conséquence, Hotelling admet que des taxes indirectes frappent les biens dont l'offre est inélastique, comme la terre, lorsque le financement de l'Etat par les impôts directs s'avère trop difficile. L'économiste norvégien Frisch montrera que ces recommandations de Hotelling ne sont pas pleinement correctes. Des taxes et subventions, parfaitement proportionnelles, qui n'affecteraient pas les prix relatifs, conserveraient l'optimum[1].

Les entreprises avec des coûts fixes et celles travaillant en rendements croissants encourraient une perte si elles vendaient leur produit au coût marginal. Les monopoles sacrifieraient une part de leur surprofit. Pour néanmoins les inciter à vendre au coût marginal, Hotelling préconise que la perte ou le sacrifice encourus soient compensés par des subventions financées par des impôts directs, si possible au bénéfice de toutes les entreprises et en tout cas pour les entreprises d'utilité publique (gaz, électricité, eau, chemins de fer etc.) Partant de là, Hotelling conteste l'idée répandue selon laquelle des projets publics ne doivent être réalisés que si les recettes escomptées couvrent les coûts fixes. Si l'amortissement de la construction d'un pont était couvert par des subsides plutôt que par les tarifs imposés aux utilisateurs, cela permettrait d'accroître le bien-être grâce à l'utilisation plus intense de l'infrastructure. Tout un débat s'ensuivit. Des économistes plus réticents à l'intervention publique, comme Coase et Vickrey, objectent que si l'entreprise de service public est dégagée de la nécessité de couvrir ses coûts fixes, elle ne sera pas incitée à les minimiser.

OPTIMUM ET REPARTITION DU BIEN-ETRE

Revenons à nouveau à la question de l'efficience et de la juste répartition. Citons encore Allais, qui insiste comme Scitovsky et d'autres sur la séparation nécessaire entre ces deux analyses. « Ainsi, la recherche de la maximisation du rendement social est essentiellement un problème de technique économique dont la solution est absolument indépendante de la conception politique et sociologique, que l'on peut avoir sur la répartition des services consommables »[2].

En fait, les économistes parétiens ont démontré ce qui est parfois présenté comme le deuxième théorème fondamental du bien-être :
Toute allocation pareto-optimale peut être atteinte en laissant l'équilibre se dégager de la concurrence parfaite à partir d'une allocation initiale ad hoc.

Ce théorème est très important par ses conséquences : si un point de la courbe de contrat est jugé plus équitable que les autres, il « suffira » au pouvoir politique de redistribuer préalablement les richesses de façon à ce que la concurrence parte d'un point d'où elle mène à l'optimum recherché. Cette manière de rechercher l'équité, via l'impôt direct progressif ou la sécurité sociale, est économiquement plus efficace que celle visant à altérer artificiellement le prix de certains biens de consommation par des subventions. Comme l'écrit Allais, « la méthode la plus avantageuse consiste à modifier directement les modalités de répartition des

[1] Les accises sont souvent forfaitaires. La TVA est proportionnelle, mais avec des distorsions, vu la multiplicité des taux et les exonérations.
[2] Allais [8] p. 643

services consommables sans toucher en aucune façon aux conditions de concurrence du secteur différencié »[1].

Sur base du *deuxième théorème fondamental du bien-être*, plusieurs auteurs parétiens vont mener une réflexion visant à déterminer l'état économique optimal, à la fois sur les plans économique et éthique ; ceci implique le recours à un critère permettant de hiérarchiser les situations pareto-optimales[2].

Parmi les pistes explorées, la plus importante est la *fonction de bien-être de Bergson-Samuelson*. L'article de **Bergson** « A Reformulation of Certain Aspects of Welfare Economics » (1938) est important parce qu'y apparaît pour la première fois la fonction qui exprime « the welfare of the community during any given period of time ». Elle a comme variables indépendantes les consommations par chaque individu de chacun des biens, la quantité de travail fournie par chaque individu pour chaque produit (sources de désutilité), les quantités globales des autres facteurs de production consommées. En jonglant avec les dérivées partielles de cette fonction, Bergson reformule des principes du bien-être émis par Pareto, Barone, Marshall et Lerner. Certaines conclusions relatives à la justice distributive se dégagent.

Dans son article « Social Indifference Curves » (1956), **Samuelson** part d'un constat : l'unité posant les choix économiques est généralement la famille plutôt que l'individu. La famille sélectionne ses dépenses en fonction de la satisfaction des besoins de ses membres. « The family acts as if it were maximizing their joint welfare function »[3]. Samuelson généralise à la société la fonction de bien-être qui est un fait familial.

Dans la formule (4.13), u_1 et u_2 … sont les fonctions d'utilité ordinales des membres de la famille ou de la société, dont les paramètres sont les quantités des biens *X*, *Y*… qu'ils consomment. *U* est la fonction de bien-être de la famille ou de la société et « *f* is an ordinal indicator function that grows when any u_i alone increases ».

$$U = f(u_1, u_2,\ldots) \text{ où :} \qquad (4.13)$$
$$u_1 = u_1\,(X_1,Y_1,\ldots),\ u_2 = u_2\,(X_2,Y_2,\ldots),\ \ldots$$

La société doit maximiser le bien-être moyennant la contrainte de rester sur la courbe de contrat. Le graphique 4.20 expose le problème pour une économie simplifiée à deux individus. La BSSWF étant ordinale, elle repose sur le champ d'indifférence social. Les *contours d'indifférence sociaux* lient les combinaisons entre l'utilité de l'agent un et celle de l'agent deux qui sont jugées également satisfaisantes en vertu du critère éthique sous-jacent à la BSSWF. La fonction de bien-être sociale traduit un jugement éthique qui relève de la sphère politique : Samuelson commente : « it is not a scientific task of the economist to deduce the form of this function »[4].

---

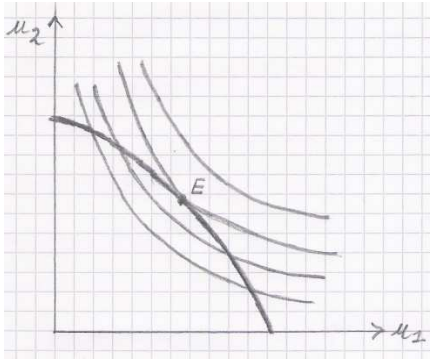
[1] Allais [8] p. 643. Dans la terminologie d'Allais, le secteur différencié est le secteur concurrentiel.

[2] Les parétiens ne sont pas toujours unanimes. Samuelson reproche par exemple à Lange et d'autres de considérer la mesure cardinale de l'utilité comme superflue dans l'analyse des comportements des consommateurs mais nécessaire pour que l'économie du bien-être puisse émettre des jugements normatifs. Selon Samuelson, un classement ordinal des états économique doit suffire pour ces deux objectifs.

[3] Samuelson [317] p. 10.

[4] Samuelson [315] p. 387. La forme des contours d'indifférence sociale traduit ce jugement de valeur. Toutes autres choses restant égales, une plus forte incurvation signifie que les répartitions très inégalitaires sont peu appréciées.

Graphique 4.20 : maximisation de la fonction de bien-être social



La contrainte du maximum d'utilité possible est représentée par la courbe dessinée en gras, qui n'est autre que la *surface de rendement maximum* d'Allais (cf. graphique 4.18) et que Samuelson appelle la *frontière de l'utilité*[1]. Elle est dérivée de la *courbe de contrat à la consommation*, comme la *frontière de la production* l'est de la *courbe de contrat à la production* (cf. graphique 4.16). L'optimum social est évidemment le point de tangence entre un *contour d'indifférence sociale* et *la frontière des possibilités d'utilité*, soit le point *E*.

L'égalité $R = R_1+R_2+\ldots+R_n$ indique le partage du revenu entre les *n* membres. A l'optimum social, on a nécessairement :

$$\partial U_1/\partial R_1 = \partial U_2/\partial R_2 = \ldots = \partial U_n/\partial R_n \qquad (4.14)$$

L'*utilité marginale sociale*[2] (« marginal social significance ») des dollars dépensés par les différents membres de la famille ou de la société doit être identique. Pour la réussite de la redistribution des revenus en vue de cet optimum, Samuelson envisage des conditions : elle n'est jamais définitive, car les variations des prix, des goûts… modifient l'équilibre. Elle doit être réalisée en ayant l'équilibre final à l'esprit, ce qui n'est pas aisé si le processus concurrentiel qui s'appuie sur les allocations redistribuées prend du temps pour produire l'équilibre final. Il est également souhaitable que la redistribution porte directement sur du pouvoir d'achat et non sur les moyens d'acquérir du revenu.

La concurrence, mène-t-elle à l'optimum social, compte tenu de la redistribution ? Qu'il y ait multiplicité des équilibres généraux concurrentiels n'est pas à exclure ; ils ne peuvent pas tous être optimaux. Précisément, Samuelson montre qu'un et un seul d'entre eux est optimal. Malheureusement, rien ne garantit que l'équilibre concurrentiel optimal soit stable[3], ce qui signifie que l'économie ne se dirige pas spontanément vers cette position.

Comme le rappelle Samuelson, il ne faut pas confondre ses CONTOURS D'INDIFFERENCE SOCIAUX avec les COURBES D'INDIFFERENCE COLLECTIVES (CIC) de Scitovsky. Ces dernières n'impliquent aucun jugement éthique.

En 1951 paraît un ouvrage important, « Social Choice and Individual Values », de Kenneth **Arrow**. Une opinion répandue[4] veut que la *fonction de bien-être social de Bergson-Samuelson* ait été enterrée par le *théorème de l'impossibilité*, énoncé dans cet ouvrage.

---

[1] Sur le graphique 4.20, je l'ai dessinée convexe, mais Samuelson précise qu'elle peut avoir n'importe quelle forme, même irrégulière, pourvu qu'elle décline en chacun de ses points.

[2] A ne pas confondre avec l'utilité marginale du revenu, mieux connue, qui est $\partial u_1/\partial R_1$, $\partial u_2/\partial R_2$ …

[3] Pour la stabilité de l'équilibre : cf. chapitres 6.2 et 6.3.

[4] Opinion erronée à mon avis.

Au début de l'ouvrage, Arrow rappelle le *paradoxe du vote* bien connu : entre trois options alternatives, l'individu X les classe ainsi : A >B >C ; pour l'individu Y, le classement est B >C >A ; pour l'individu Z : C >A >B. Un vote à la majorité ne peut dégager une synthèse.

« By a *social welfare function* will be meant a process or rule which, for each set of individual orderings $R_1,\ldots,R_n$ for alternative social states (one ordering for each individual), states a corresponding social ordering of alternative social states, *R* »[1]. L'attention est donc principalement portée sur l'AGREGATION des préférences individuelles. Parmi toutes les règles possibles, Arrow en exclut certaines de son champ d'analyse :

- Celles qui relèvent de la dictature et ne respectent pas la souveraineté du citoyen
- Celles qui contreviennent aux axiomes des choix rationnels qu'il présente (divers critères comme la transitivité…).

Un exemple d'une règle acceptable est le vote à la majorité. Arrow démontre mathématiquement que dans des conditions normales (notamment, les individus ne sont pas tous identiques), il est impossible d'élaborer une telle règle d'agrégation lorsque le choix comporte plus de deux options.

Bergson et Samuelson réagirent. De la contribution d'Arrow, Samuelson dira : "It was a very important finding in political science as it showed that the failure of specific voting functions is not due to any lack of cleverness, but is a reflection of general impossibility. However, it had nothing to do with ethics and welfare economics. Arrow use of 'welfare social function' for his 'voting function' was unfortunate"[2].

Le théorème d'Arrow a suscité une abondante littérature et fera l'objet de critiques, notamment par Amartya Sen. A mon sens, Arrow a eu tort de se focaliser sur l'agrégation au niveau du pouvoir politique :

- D'une part, la fonction de bien-être sociale ne nécessite pas nécessairement une telle agrégation. Dans les « Foundations », Samuelson écrivait ironiquement : « Without inquiring into its origins, we take as a starting point for our discussion a function of all the economic magnitudes of a system which is supposed to characterize some ethical belief-that of a benevolent despot, or a complete egotist, or 'all men of good will', a misanthrope, the state, race, or group mind, God etc »[3] .
- D'autre part, l'agrégation arrowienne est en porte-à-faux avec le fonctionnement des démocraties parlementaires modernes où le compromis joue un rôle essentiel et recueille généralement une adhésion, parce qu'il existe une culture du compromis.

La principale lacune de ce concept me paraît être son niveau d'abstraction élevé.

La *fonction de bien-être social* a subi d'autres attaques par la suite, visant à prouver son impossibilité ou prétendant qu'elle implique la mesure CARDINALE de l'utilité. Les adversaires campèrent sur leur position et le débat cessa sans véritable conclusion.

[1] Arrow [12] p. 23. Le terme « ordering » exclut toute mesure cardinale de l'utilité.
[2] Suzumura [351] p. 12
[3] Samuelson [321] p. 221

LES BIENS PUBLICS

Il nous faut distinguer deux catégories de biens :

- les biens PRIVES : ce sont les seuls que nous avons étudiés jusqu'à présent. Lorsque des unités de ce type de bien sont consommées par un individu, elles sont rendues indisponibles pour la consommation d'autres individus. La nourriture en est un exemple.

- Les biens de consommation collective, encore appelés biens PUBLICS, qui –selon la définition de Samuelson- sont ceux « which all enjoy in common in the sense that each individual's consumption of such a good lead to no subtraction from any other individual's consumption of that good »[1]. Les principaux exemples sont la sécurité publique, les égouts, la voierie, la radio ou la TV hertziennes…

Des premiers, on peut dire que la consommation sociale est la somme des consommations individuelles. Pour les seconds, la consommation sociale égale la consommation personnelle de chaque individu. Chaque automobiliste « consomme » (potentiellement, du moins) la totalité du réseau routier.

Il ne faut pas confondre cette distinction avec celle qu'on peut opérer suivant que l'offreur soit un service public ou une entreprise privée. En effet, la réalité sociale abonde d'exemples de biens privés offerts par le secteur public et de biens publics offerts par des entreprises privées (généralement sous concession).

Samuelson consacre aux biens publics une série d'articles publiés en 1954, 1955 et 1958. Il y constate que ce sujet avait auparavant été négligé par les économistes, à l'exception de Wicksell, Lindahl, Musgrave et quelques autres.

En quoi faut-il modifier la caractérisation de l'optimum du bien-être pour y intégrer le cas des biens publics ? Samuelson expose un modèle d'optimum en trois groupes d'équations :

1- Pour chaque bien PRIVE, le TMT entre ce bien et un bien privé de référence est égal au TMS de chaque individu entre ces deux mêmes biens. C'est la condition parétienne traditionnelle.
2- Pour chaque bien PUBLIC, le TMT entre ce bien et ledit bien de référence est égal à la somme des TMS de tous les membres de la société entre ces deux mêmes biens.
3- L'augmentation du bien-être social est équivalente, quel que soit l'individu dont l'indice d'utilité est accru marginalement (maximisation de la BSSWF).

L'élément neuf de ce système est le deuxième groupe d'équations. La différence avec le premier groupe, le fait de sommer les TMS individuels, se justifie ainsi : si je réduis la production de pain d'une unité, un des membres de la société devra subir cette restriction. Si je réduis le réseau routier d'un kilomètre, tous les individus voient leur possibilité de circuler restreinte d'autant.

Samuelson titre une section de son article : « Impossibility of decentralized spontaneous solution ». Le système d'équations ci-dessus a une solution, autrement dit l'optimum existe. Mais le marché ne comporte pas de mécanisme qui y mène : « Although the optimum is definable, rational persons will not, if left to themselves, be led by an invisible hand to the bliss point »[2]. Imaginons une société qui voudrait financer l'offre de biens publics selon une modalité où l'utilisateur paie en proportion de la satisfaction qu'il en retire. Le problème, déjà identifié par Wicksell, tient à ce que les consommateurs de biens publics ne sont pas motivés à

[1] Samuelson [315] p. 387
[2] Samuelson [318] p. 334

révéler leur véritable désir de les consommer. Chacun sera tenté, égoïstement, de dissimuler son besoin réel, en espérant profiter de la contribution d'autrui[1]; si tous se comportent ainsi, aucun moyen ne sera consacré à la production de biens publics et les besoins correspondants ne seront pas satisfaits. Alors que sur les marchés de biens privés, « each person is motivated to do the signalling of his tastes needed to define and reach the attainable bliss point ».[2].

Dans son article ultérieur, Samuelson répond à une critique largement adressée à son modèle : la conception des biens publics y est trop puriste, trop théorique. Dans la réalité, la plupart des services offerts par les gouvernements s'écartent du cadre des biens publics défini dans ce modèle : à un degré variable, il y a presque toujours une part d'exclusivité dans le bénéfice d'un bien public par un individu, réduisant, fût-ce très légèrement, l'avantage disponible pour les autres. Par exemple, les pompiers ou la police ne peuvent pas intervenir partout à la fois. Sur le fond, Samuelson admet la validité de ces critiques, mais son propos était de présenter un modèle simple intégrant les deux types extrêmes de biens. Les biens privés n'ont pas toujours les propriétés de rendements d'échelle constants, de divisibilité, d'absence d'effets de voisinage etc voulues par la théorie orthodoxe. De la même façon, nombre de services offerts par les gouvernements ont leurs caractéristiques de bien public mélangées avec celles des biens privés. Samuelson observe à ce propos que les restrictions imposées à la consommation d'un bien public n'en font pas un véritable bien privé. Il donne l'exemple de la télévision à péage cryptée. Les diffuseurs ne règleront certainement pas le prix sur le coût marginal qui est nul dans ce cas, alors qu'il s'agit là d'une règle pour atteindre l'optimum de Pareto.

4.2.4. Nouvelles mathématiques de l'optimisation

En 1947, Samuelson publie « Foundations of Economic Analysis », un livre qui influencera profondément la science économique pour les décennies à venir. Cet ouvrage se distingue, non par l'énoncé de nouvelles théories, mais par l'élaboration d'un outil d'analyse mathématique qui doit permettre aux économistes de travailler plus rigoureusement.

Samuelson écrit : "The general method involved may be very simply stated. *In cases where the equilibrium values of our variables can be regarded as the solutions of an extremum (maximum or minimum) problem, it is often possible regardless of the number of variables involved to determine unambiguously the qualitative behavior of our solution values in respect to changes of parameters*"[3]. Dans un modèle de ce type, il y a la variable à maximiser ou à minimiser, les autres variables endogènes et les variables exogènes que Samuelson appelle paramètres. L'auteur présente une méthode formelle permettant d'établir le signe de la modification de toutes les variables lorsqu'un paramètre varie.

Nous avons vu que déjà Cournot définissait l'équilibre à partir de la maximisation du revenu par l'annulation de la dérivée de la fonction à maximiser. L'optimisation est plus complexe lorsque la maximisation ou la minimisation s'opèrent moyennant le respect d'une ou plusieurs contraintes. La méthode utilisée est alors celle des *multiplicateurs de Lagrange*. Certes, Samuelson n'est pas le premier à y recourir. Des économistes mathématiciens comme Lange et Hotelling l'avaient déjà fait avant lui. Mais cette méthode est omniprésente dans l'ouvrage et par la suite dans la science économique. Plus précisément, l'optimisation occupe la première partie du livre ; la seconde est consacrée à la stabilité de l'équilibre, un sujet que nous aborderons au titre VI.

---

[1] Les Anglo-Saxons appellent cette stratégie le « free rider ».
[2] Samuelson [315] p. 389
[3] Samuelson [321] p. 21

Supposons une fonction *f* à *n* variables $f(x_1,\ldots,x_n) = 0$ qu'il y a lieu de maximiser, en respectant *m* contraintes données chacune par une fonction $g_i(x_1,\ldots,x_n) = 0$, pour $i = 1\ldots m$. A l'aide de ces m+1 fonctions, on en construit une du type :

$$L = f - \lambda_1.g_1 - \ldots - \lambda_m.g_m \qquad (4.15)$$

Cette fonction a n+m variables : $x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_m$. Pour trouver le maximum recherché, on va annuler ses dérivées partielles par rapport aux n+m variables. On obtient un système de n+m équations à n+m inconnues :

$$\partial L/\partial x_1 = 0 ; \ldots \partial L/\partial x_n = 0 ; \partial L/\partial \lambda_1 = 0 ; \ldots \partial L/\partial \lambda_m = 0 \qquad (4.16)$$

Une technique permet de déterminer la valeur des n+m variables, l'essentiel étant d'obtenir les valeurs de $x_1\ldots x_n$ qui confèrent un extremum à la fonction *f*. Et également de s'assurer que cet extremum est un maximum et non un minimum. Cette dernière exigence est appelée *condition de second ordre*, par opposition à la *condition de premier ordre* expliquée ci-dessus. La condition de second ordre concerne le signe des dérivées partielles secondes de *f* ainsi que les relations entre elles pour déterminer si la fonction est concave (pour un maximum) ou convexe (pour un minimum).

Samuelson remarque que si on change un paramètre, plus il y a de contraintes et plus elles sont sévères, moins ample sera la variation de la variable contrôlée pour assurer l'équilibre. C'est ainsi que l'élasticité d'une demande à long terme sera normalement supérieure à l'élasticité de cette demande à court terme.

Samuelson revisite les théories de l'équilibre du consommateur et de l'équilibre du producteur, en mettant en application les méthodes qu'il a développées. Voyons rapidement les résultats pour la théorie du producteur. Samuelson la divise en deux parties : la première, basée sur la minimisation du coût est purement technologique ; la seconde, fondée sur la maximisation du profit est à la fois commerciale et technologique.

Commençons par l'optimisation technologique. A ce stade, la quantité produite par l'entreprise est imposée ($x = x^\circ$) et ce sont les quantités des inputs (v_i) qui sont les variables à ajuster pour minimiser le coût total *C*. Les prix de ces inputs (w_i) constituent les paramètres du système. Nous devons minimiser la fonction (4.17) moyennant la contrainte (4.18) où φ est la fonction de production.

$$C = \Sigma_1^n w_i v_i \qquad (4.17)$$
$$\text{moyennant } \varphi(v_1\ldots v_n) = x^\circ \qquad (4.18)$$
$$L = \Sigma_1^n w_i v_i - \lambda[\varphi(v_1\ldots v_n) - x^\circ] \qquad (4.19)$$

L'équation (4.19) est la fonction lagrangienne. Après en avoir annulé les dérivées partielles, Samuelson obtient :

$$1/\lambda = \varphi_1/w_1 = \ldots = \varphi_n/w_n \qquad (4.20)$$

Où φ_i est la notation pour la dérivée partielle de la fonction de production φ par rapport à v_i, qui n'est autre que la productivité marginale du facteur concerné. L'équation (4.20) nous ramène donc à la règle familière qu'il faut acquérir de chaque facteur la quantité qui permet d'égaliser le rapport entre la productivité marginale et le prix. Le facteur de proportionnalité est λ et Samuelson montre que dans le cas présent λ correspond au coût marginal.

Après l'optimisation technologique, Samuelson aborde la maximisation du profit, pour laquelle doit évidemment être relâchée l'hypothèse d'une production donnée. Il faut maximiser le profit π dont l'équation est :

$$\pi = \pi(x,w_1...w_n) = R(x) - A - V(x,w_1...w_n) \qquad (4.21)$$

où *R* est la recette totale, fonction de la quantité produite *x*, où *A* est le coût fixe total et *V* est le coût variable total. Samuelson montre que la maximisation de π implique :

$$\partial R/\partial x = \partial V/\partial x \qquad (4.22)$$
$$\partial^2 R/\partial x^2 < \partial V^2/\partial x^2$$

"This is the familiar theorem that at the optimal output the marginal revenue curve must intersect the marginal cost curve from above"[1]. De 4.20 et 4.22, on peut déduire que lorsque le profit est maximisé, le prix d'un facteur de production est égalisé avec la valeur de son produit marginal, un théorème bien connu. Attention de ne pas confondre cette condition avec celle qui fut déduite de la seule équation (4.20), qui lui ressemble, et qui se vérifie à tous les niveaux de production, pour autant que le coût soit minimisé.

[1] Samuelson [321] p. 77

4.3- LA MACROECONOMIE AVANT KEYNES

La révolution marginaliste fut essentiellement microéconomique : elle se préoccupa plus de l'équilibre des marchés et du comportement des acteurs économiques que du niveau des prix et du chômage. En fait, les néoclassiques ne se démarquèrent pas fondamentalement de leurs prédécesseurs classiques sur ces questions. Notamment, la loi de Say reste un postulat. Jevons écrit : « Early writers on Economics were always in fear of a supposed glut arising from the powers of production surpassing the needs of consumers, so that industry would be stopped, employment fail, and all but the rich would be starved by the superfluity of commodities. The doctrine is evidently absurd and self-contradictory (…) Over-production is not possible in all branches of industry at once, but it is possible in some compared with others »[1].

4.3.1. Irving Fisher et l' « équation de l'échange »

Ainsi s'appelle la théorie selon laquelle la valeur de la monnaie varie en proportion inverse de sa quantité, toutes autres choses restant égales. Puisque les prix des autres biens sont exprimés en monnaie, la valeur de la monnaie n'est rien d'autre que l'inverse du niveau général des prix. Comme nous avons pu nous en rendre compte, cette théorie est l'une des plus anciennes de la science économique. Elle avait déjà été exposée par le mercantiliste Locke en 1692. Cantillon, Hume, Ricardo et Mill en furent des hérauts.

Toutefois, c'est dans l'ouvrage d'Irving Fisher, « The Purchasing Power of Money », paru en 1911, que cette théorie sera exposée et argumentée le plus systématiquement. L'idée géniale de Fisher fut de la résumer dans une équation simple, à partir de laquelle il est possible de visualiser les rapports entre les différentes variables. Cette équation, devenue l'une des plus connues de la science économique, Fisher la baptise *équation de l'échange*.
La version de base de cette équation est :

$$M.V = p_1.Q_1 + p_2.Q_2 + \ldots \quad + p_n.Q_n \qquad (4.23)$$

Le membre de droite somme toutes les transactions individuelles de la période considérée, chacune valant la multiplication d'une quantité par un prix unitaire.

A gauche, on a :
- *M* est la masse monétaire en circulation (moyenne pendant la période considérée)
- *V* est la vitesse de circulation de la monnaie

Le membre de gauche s'attache à la monnaie qui circule en sens inverse des biens. Vu la définition de *M* et celle de *V*, il est la somme des paiements monétaires effectués. L'équation des échanges affirme cette évidence qu'est l'égalité entre la monnaie payée et la valeur des biens achetés.

Fisher préfère l'équation de l'échange sous cette forme simplifiée, « where *P* is a weighted average of all the *p*'s and *T* the sum of all the *Q*'s »[2] :

$$M.V = P.T \qquad (4.24)$$

- *P* signifie le niveau général des prix
- *T* signifie le volume REEL des transactions sur les marchés.

[1] Jevons [165] § 5.52
[2] Fisher [92] p. 26

Mais la théorie quantitative de la monnaie va bien au-delà de cette égalité que beaucoup d'économistes considèrent comme tautologique. Elle établit des relations causales entre les différentes variables. En fait, *M*, *V* et *T* sont déterminés par des facteurs exogènes à l'équation, que nous allons analyser. Par contre, *P* est le seul élément passif de l'équation. Il vaut simplement MV/T. Comme en plus, *V* et *T* sont considérés comme stables à court terme, la théorie quantitative se limite à un effet de *M* sur *P*.

Fisher attribue aux variables de l'équation les causes exogènes que voici :

- **T** dépend de tous les facteurs réels qui facilitent la production et le transport des marchandises : ressources naturelles disponibles, division du travail, état de la technique, accumulation du capital, confiance dans les affaires…
- **V** dépend des habitudes des agents en matière de paiement : leur tendance à la thésaurisation, le recours au crédit, l'inclination à payer par chèque… *V* dépend également du système de paiement de la communauté : si les salariés sont payés au mois, ils devront maintenir des encaisses plus élevées que s'ils sont payés à la semaine et *V* sera donc plus faible. La régularité des recettes et dépenses joue également ; plus celle-ci est élevée, moins les agents devront maintenir une encaisse de précaution.
- **M** dépend du système monétaire. Fisher analyse en détail la monnaie métallique mais s'intéresse étonnamment peu aux billets de banque. Interviennent d'abord les transferts de métal précieux entre la sphère monétaire et la sphère non monétaire, (monnayage dans un sens ou fonte dans l'autre sens). Il y a ensuite la production d'or qui est stimulée lorsque le prix de marché, résultant de la demande totale, est supérieur au coût marginal de cette production. La monnaie scripturale, quant à elle, conserverait une proportion constante avec la monnaie fiduciaire, déterminée par l'inclination du public à détenir ses avoirs liquides en monnaie sous telle ou telle forme.

Fisher examine ensuite les conséquences qu'auraient les variations de M, V et T.

- Une variation de *M* retomberait entièrement sur *P* ; *V* et *T* ne seraient pas affectés car les facteurs dont ils dépendent (cf. ci-dessus) sont imperméables à l'influence de la masse monétaire. Le mécanisme de la hausse de *P* serait celui-ci : « everybody will want to exchange this relatively useless extra money for goods and the desire so to do must surely drive up the price of goods »[1].
- Une variation de V, de même, affectera *P* et sera sans influence sur *M* et *T*.
- La variation de *T* entraînera une variation inverse de *P*, mais pourra également avoir d'autres effets. Si la hausse de *T* correspond à l'enrichissement de la communauté, *V* pourra s'accroître, car le rapport de l'encaisse sur le revenu varie en sens inverse du niveau de richesse. L'accroissement de *T* plus important dans un pays que dans les autres devrait également modifier la répartition de l'or dans le monde en sa faveur.

Par contre, l'effet d'une variation exogène de *P* est une question qui ne se pose pas. Fisher s'insurge contre l'idée qu'une augmentation de *P* pourrait résulter directement de l'action des trusts ou des syndicats ouvriers. Une augmentation de *P* ne pourrait provenir que de la baisse de *T* éventuellement causée par ces actions. Pour sa démonstration, Fisher consent à imaginer un doublement exogène de *P*. Quel en serait l'effet sur M, V et T ? Il n'y en aurait aucun, ce qui démontre l'impossibilité d'un tel phénomène. Les facteurs déterminant ces variables sont insensibles à *P*. Ce passage n'est pas très convaincant. Par exemple, selon Fisher, la monnaie scripturale ne varierait pas vu son rapport rigide avec la monnaie fiduciaire. Il paraît douteux que les banques resteraient passives et n'adapteraient pas leur offre de monnaie si la monnaie fiduciaire était figée. Toujours selon Fisher, *V* serait invariable car « (it has) already been

[1] Fisher [92] p. 154

adjusted to suit individual convenience. To double (it) might not be a physical possibility and would certainly be of great inconvenience »[1]. Nécessité, ne ferait-elle pas loi ?

Faisant un pas en arrière, Fisher considère toutefois que l'insensibilité de *T* par rapport à *M*, ne vaut qu'à long terme. Une variation de *M* aura des effets TRANSITOIRES sur *T*. Voyons comment Fisher présente l'enchaînement des faits, dans le cas d'un accroissement de *M*. Cette hausse se répercute sur *P*. Le taux d'intérêt réel baisse en deçà de son niveau d'équilibre puisque Fisher postule que les agents tardent à ajuster les taux d'intérêt nominaux quand les prix varient (cf. supra). Il en résulte une débauche de crédits, qui financent des investissements au-delà de ce qui est raisonnable : *T* augmente donc. Le retour de manivelle vient à son tour. Précipité par une demande de crédit sous pression, l'intérêt réel augmente et freine l'économie, suscitant des faillites parmi les entrepreneurs qui empruntent maintenant à un taux excessif : en un deuxième temps, *T* se contracte. Fisher estime que ce cycle peut durer dix ans. Un délai assez long pour des effets transitoires. En fait, n'est-ce pas le cycle conjoncturel que Fisher introduit dans la théorie quantitative ? De cette conception émerge la conclusion paradoxale, que les effets de *M* sur *T* pourraient être immenses et durables, soit l'exact contraire de ce que voulait prouver Fisher[2].

Le dernier chapitre s'intitule « The Problem of Making Purchasing Power more stable ». Pour minimiser les fluctuations de *P*, le contrôle de *M* est essentiel. Fisher envisage nombre de systèmes et de politiques monétaires susceptibles d'aller dans ce sens.

Fisher est réputé pour être un des promoteurs-précurseurs de l'indice mesurant l'évolution des prix et cette matière occupe une bonne part de son ouvrage. L'équation de l'échange sous sa forme (4.24) exige d'ailleurs un tel indice, non seulement pour *P* mais également pour *T*. Ce dernier ne peut résulter de l'addition de quantités produites exprimées dans des unités différentes. Le calcul de *P* et celui de *T* sont intimement liés.

4.3.2. Wicksell et le taux d'intérêt naturel

En 1906, Wicksell publie le tome II des « Lectures on Political Economy », consacré à la monnaie.

Il commence par examiner les trois fonctions essentielles de la monnaie :

1. intermédiaire d'échange
2. réserve de valeur
3. mesure de la valeur

Servir comme intermédiaire d'échange est la seule fonction inhérente de la monnaie. Les deux autres en découlent. La thésaurisation de monnaie n'est pratiquée que parce qu'elle est en même temps intermédiaire d'échange.

« Money is not a commodity like other commodities ». Pour les autres biens, il y a toujours une proportionnalité entre leur utilité marginale et leur prix. « But money itself has no marginal utility (…) It has perhaps an indirect marginal utility, equivalent to the goods which we could obtain in exchange for it, but this depends in turn on the exchange value or purchasing power of the money itself and it thus does not itself regulate the latter. Similarly, 'supply' and 'demand',

[1] Fisher [92] p. 171

[2] En fait, Fisher se montre ambigu et contradictoire ; emporté par sa description du cycle du crédit, son propos y dépasse ce que lui-même prétend y exprimer.

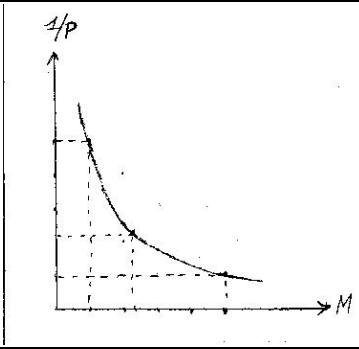
expressions so conveniently applied to almost everything under the sun, become obscure and in reality meaningless when applied to money »[1].

La vitesse de circulation est le temps moyen que la monnaie reste dans la poche de l'agent entre deux transactions. Plus ce temps est long, plus grand sera le besoin de monnaie. A l'opposé de Fisher, Wicksell considère que *V*[2] s'adapte à la pénurie ou l'abondance de monnaie, dans certaines limites.

En avance sur son temps, Wicksell prône le divorce entre l'or et la monnaie, et notamment la cessation du libre monnayage de l'or. Il veut promouvoir la monnaie scripturale, une tendance qui était déjà à l'œuvre à son époque. Le système idéal, selon lui, est celui de la pure monnaie de crédit, qui n'a finalement triomphé que près d'un siècle après qu'il a écrit ces lignes. Vu son coût, la production d'or à des fins monétaires est un gaspillage de ressources.

Wicksell donne son appréciation sur la *théorie quantitative*, la seule théorie monétaire digne de ce nom selon lui, Il apprécie la distinction fondamentale qu'elle établit entre la monnaie et les autres biens. Selon la théorie quantitative, la valeur de la monnaie ne varie pas seulement en SENS inverse de sa quantité mais en PROPORTION inverse. Lorsque la quantité offerte d'un bien quelconque augmente, son prix tend à baisser, mais la proportion de cette baisse est variable selon l'élasticité de la demande de ce bien. Par contre si la quantité offerte de monnaie augmente de 10%, toutes autres choses restant égales, la valeur de la monnaie sera réduite d'exactement 10%, quelle que soit la forme de la courbe de demande de monnaie.

Graphique 4.21: la valeur de la monnaie en fonction de sa quantité



La courbe d'équilibre monétaire du graphique 4.21 est une hyperbole équilatère. Chacun de ses points délimite un rectangle dont la hauteur représente la valeur de la monnaie (l'inverse du niveau général des prix : 1/P) et dont la base exprime la masse monétaire (M). La caractéristique d'une hyperbole équilatère est que tous ces rectangles formés avec les axes ont la même surface, autrement dit que le produit M/P est constant, ce qui est la base de la théorie quantitative.

S'ils ressentent un état de pénurie monétaire, les détenteurs de monnaie voudront et pourront corriger cet inconvénient en réduisant les prix qu'ils acceptent de payer pour les biens : la valeur de la monnaie haussera donc. L'inverse se passera en cas de pléthore monétaire.

La principale critique de Wicksell à l'égard la théorie quantitative est qu'elle met trop largement l'accent sur l'influence DIRECTE des encaisses sur les prix. Wicksell insiste sur

[1] Wicksell [385], t2 p. 20

[2] Les 4 lettres de l'équation de l'échange M,V,P,T sont une création de Fisher, dont le livre est postérieur à celui de Wicksell ; nous nous en servirons néanmoins pour la clarté de l'exposé.

l'influence que la masse monétaire peut avoir sur les prix par un autre canal, via le taux d'intérêt. Ces deux types d'influence sont parfois appelés le *mécanisme direct* et le *mécanisme indirect*.

Un accroissement de *M* rend plus abondant l'argent susceptible de devenir du capital ; ce faisant, il pousse l'intérêt à la baisse, ce qui stimule la demande de biens et donc aussi l'inflation. Cette idée, fort ancienne, avait donné lieu à d'âpres débats lors de la controverse sur *le currency principle*.

Le concept central de la théorie wicksellienne est le *taux d'intérêt normal*. Wicksell insiste sur la distinction entre :
- le rendement réel du capital investi dans les entreprises (real rate) et
- le *taux d'intérêt monétaire*, qui est payé par les emprunteurs aux prêteurs ou aux banques (*loan rate* ou *money rate*).

Quelle est la relation entre eux ? « It is a well recognized principle that in the last analysis the money rate of interest depends upon the supply of and demand for real capital or, as Adam Smith and later Ricardo, expressed it, that the rate of interest is regulated by the employment of capital itself and not by the number or quality of the pieces of metal which facilitate the turnover of its products »[1]. A partir de là, Wicksell donne cette définition: « The loan rate, which is a direct expression of the real rate, we call the normal rate »[2]. C'est à ce niveau que le taux d'intérêt équilibre l'épargne et l'investissement ; aucune entreprise n'emprunterait à un taux plus élevé et toutes voudraient emprunter plus s'il était plus bas. Le *taux normal* dépend de la productivité du capital et du désir d'épargner du public.

Le taux d'intérêt normal assure l'équilibre sur le marché des prêts, mais également sur le marché des biens et services. Il induit la juste répartition du revenu entre consommation et épargne, qui permet l'équilibre des marchés de biens de consommation et des biens d'équipement. Cette théorie est connue sous le nom de *théorie des fonds prêtables*.

Comme nous l'avons vu, Wicksell pressentait la mutation du système monétaire vers une pure monnaie de crédit et il s'en félicitait. Toutefois, il relève une faiblesses du système. La quantité de monnaie que les banques sont susceptibles de prêter est illimitée car la conversion des billets en or n'existe plus pour faire contrepoids. En acceptant de baisser leur taux d'intérêt, les banques peuvent satisfaire toute demande. Wicksell écrit : « in our days, demand and supply of money have become about the same thing, the demand to a large extend creating its own supply »[3].

Dans un tel système, le taux d'intérêt s'écarte facilement du taux normal. Supposons que les banques prêtent à un taux inférieur au taux normal. Ce taux trop bas va stimuler la demande des facteurs de production et donc leur revenu. La baisse du taux décourage simultanément l'épargne, ce qui stimule la demande de biens de consommation. Comme les économistes de son temps, Wicksell raisonne toujours avec l'hypothèse du plein emploi. Le niveau général des prix doit donc augmenter puisque la production est quasiment bloquée[4].

Les prix continueront à augmenter tant que la cause fondamentale demeure, c'est-à-dire tant que le taux d'intérêt est en dessous de son niveau normal. C'est justement cette hausse des prix

[1] Wicksell [385], t2 p. 190
[2] Wicksell [385], t2 p. 192
[3] Wicksell [384] § W8
[4] L'accroissement éventuel de la production n'arriverait de toute façon que passé un délai.

qui amènera la réconciliation entre le taux naturel et le taux bancaire, parce qu'avec des prix plus élevés, la quantité de monnaie disponible cessera d'être pléthorique. Le taux d'intérêt remontera à son niveau normal, mais le niveau des prix restera à son nouveau niveau plus élevé ; aucun mécanisme n'existe pour le faire tendre vers un quelconque équilibre.

Contre la crainte bullioniste de voir des taux d'intérêt bas stimuler l'inflation, Tooke opposait la constatation empirique (correcte selon Wicksell) que les périodes où les taux d'intérêt sont élevés correspondent justement à celles où les prix sont élevés ou en croissance. Wicksell propose une explication du phénomène qui neutralise l'objection de Tooke. Les choses se passent à l'inverse de ce qu'on pourrait croire. L'écart entre le taux normal et le taux bancaire n'est pas dû aux variations du second mais à celles du premier. L'évolution économique générale, avec ses composantes technologiques et autres, fait que le rendement du capital réel évolue sans cesse ; ce sont ces fluctuations qui créent l'écart avec le taux bancaire car les banques ne peuvent ou ne veulent pas adapter ce dernier en permanence. Lorsque le taux bancaire s'est fait dépasser par le taux normal, cela stimule la demande de biens d'équipement et pousse leurs prix à la hausse ; il en résulte un besoin de capitaux plus élevé, ce qui fait hausser le taux d'intérêt jusqu'à la réconciliation avec son niveau normal. Autrement dit, si les banques n'adaptent pas de leur propre chef leur taux d'intérêt au rendement réel du capital, ce sont les fluctuations des prix qui s'en chargeront.

En conséquence, « the only possibility of a rational control of the price level must lie in (…) the proper regulation of the interest policy of the banks »[1].

4.3.3. Demande d'encaisse et théorie quantitative

Chez Wicksell et Fisher, la monnaie se voit attribuer un statut particulier la distinguant des autres biens, ce qui en fait un peu un OVNI dans le monde néoclassique focalisé sur les choix opérés par les agents. Voyons maintenant deux théories, qui restent quantitatives, mais qui considèrent la monnaie comme un objet désirable en soi, qu'il y a lieu d'étudier par l'analyse de l'offre et la demande. La théorie walrassienne est celle qui rapproche le plus la monnaie des autres biens. Celle des Cambridgiens (Marshall, Pigou) est à mi-chemin.

LE COEFFICIENT *K* DES CAMBRIDGIENS

En 1917, Arthur Cecil Pigou publie « The Value of Money », où apparaît le fameux coefficient *k*. Cet article analyse successivement la demande de monnaie, l'offre et l'équilibre monétaire, mais manifestement, c'est la demande de monnaie qui est au cœur de l'étude.

Voici l'explication qu'en donne Pigou : « In the ordinary course of life, people are continually needing to make payments in discharge of obligations contracted in terms of legal tender money (…) Besides the flow of obligations that are thus continually maturing against them, most people have also a flow of claims that are similarly maturing in their favor. But the obligations and the claims that become due at any moment seldom exactly cancel one another, and the difference has to be met by the transfer of titles into legal tender (…) If a person is unable to meet his obligations from these sources when they fall due, he will certainly be inconvenienced and will possibly be rendered bankrupt. Hence, everybody is anxious to hold enough of his resources in the form of titles to legal tender both to enable him to effect the ordinary transactions of life and to secure him against unexpected demands »[2].

[1] Wicksell [385], t2 p. 216
[2] Pigou [286] p. 164

Comment est déterminée la quantité de monnaie que l'agent veut conserver ? A travers la monnaie, c'est une certaine quantité de pouvoir d'achat réel qu'on désire détenir. On peut l'exprimer en une quantité de n'importe laquelle des marchandises susceptibles d'être achetée, par exemple le blé. « Let *R* be the total resources, expressed in terms of wheat, that are enjoyed by the community; *k* the proportion of these resources that it chooses to keep in the form of titles to legal tender »[1]. Si *D* est la demande totale d'encaisse (exprimée en blé), on a :

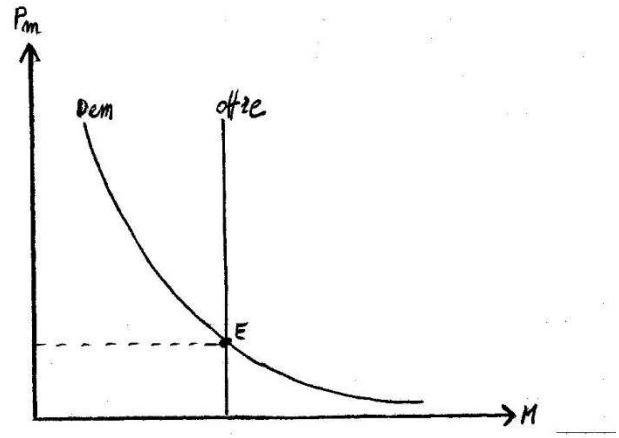
$$D = k.R \qquad (4.25)$$

La courbe de demande de monnaie donne en ordonnée le prix (exprimé en blé) d'une unité monétaire et en abscisse le nombre d'unités monétaires. Si P_M est le prix d'une unité monétaire et *M* le nombre de celles-ci, la formule est :

$$P_M = (k.R) / M \qquad (4.26)$$

Lorsque *k* et *R* sont constants (autrement dit, indépendants de *M*), et à cette condition uniquement, la courbe de demande est une hyperbole équilatère (élasticité égale à 1). Ne confondons pas cette courbe avec celle de Wicksell (graphique 4.21), qui n'est pas une courbe de demande et qui est TOUJOURS une hyperbole équilatère.

Graphique 4.22 : l'offre et la demande de monnaie chez Pigou



Pigou analyse ensuite les facteurs qui déterminent *k* :

1. pour les firmes : la comparaison des productivité marginales respectives des outils de production et de l'encaisse.
2. l'intervalle de temps entre deux recettes.
3. la possibilité plus ou moins grande d'effectuer des paiements par compensation ou endossement.
4. le degré de spécialisation des firmes : plus de firmes signifie plus de paiements et un besoin accru d'encaisses.
5. les prévisions quant au prix des biens. L'anticipation d'une baisse des prix des biens incitera à des reports d'achats et gonflera les encaisses dans l'intervalle.

On remarquera que Pigou ne cite pas le taux d'intérêt parmi ces facteurs.

Pigou compare son équation (4.26) à l'équation de l'échange de Fisher. Il les considère comme deux façons différentes de dire la même chose. Comme P_M vaut 1/P, on a :

[1] Pigou [286] p. 165. L'introduction du blé dans l'explication ne doit pas troubler. Il est simplement pris comme numéraire. Pigou exprime le revenu en numéraire pour insister sur l'indépendance de la demande de monnaie par rapport aux facteurs purement monétaires.

$$\frac{kR}{M} = \frac{T}{MV} => kV = \frac{T}{R} \qquad (4.27)$$

Comme le *T* de Fisher et le *R* de Pigou sont des grandeurs assez similaires[1], on peut en conclure que k.V est une constante. *V* est en quelque sorte l'inverse de *k*. « Any increase in the ready purchasing power that people choose to keep will diminish proportionately rapidity of circulation, and vice versa »[2].

Mais l'essentiel est que P_M est le facteur d'ajustement, comme le *P* de Fisher.

Concernant l'offre de monnaie, Pigou considère plusieurs systèmes monétaires. Le seul qui nous intéresse aujourd'hui est la monnaie inconvertible. Dans ce cas, l'offre est simplement une droite verticale, dont la place sur l'abscisse est fixée « by arbitrary decision of the government ». Le prix P_M d'équilibre dépend de l'intersection entre les deux courbes. Pigou reconnaît que la réalité est plus compliquée que la théorie : « in the real world, we cannot always hope to meet only with causes that act either on demand or on supply alone. The same cause may easily act upon both »[3].

L'ENCAISSE DÉSIRÉE SUIVANT WALRAS

Bien qu'elle soit plus ancienne, nous avons réservé la théorie walrassienne pour la fin, d'abord vu son originalité, ensuite de par sa complexité : les commentaires très variés dont elle a fait l'objet montrent qu'elle est certainement ambiguë. Dans le système walrassien, la frontière entre la micro- et la macroéconomie cesse d'être nette, puisque la monnaie, l'objet macroéconomique par excellence, se mêle aux équations de la théorie de la valeur. Au sous-chapitre 3.1.3, nous nous étions arrêtés au seuil de cette intégration ; franchissons-le.

Walras impose maintenant à son système que tous les achats-ventes de services producteurs et de biens ainsi que l'épargne s'effectuent par l'intermédiaire de la monnaie. Il envisage également l'aspect cyclique de la production. Tant que le produit final n'est pas vendu, l'entrepreneur n'a pas de quoi payer les services producteurs et les matières premières. Pendant le processus de production, le capital circulant s'accumule. Walras divise le capital circulant en deux parts : l'encaisse et les stocks, ignorant la troisième : l'encours de production. Comme à son habitude, Walras construit une fiction censée représenter la réalité. Les ménages disposent d'une dotation comportant de la monnaie et des biens entrant dans le capital circulant. Eux aussi ont besoin de stocker certains biens (le garde-manger) et ont besoin de monnaie pour payer leurs achats dans l'intervalle entre deux versements de leur revenu. Ils prêtent aux entreprises la part qu'ils n'utilisent pas eux-mêmes. Dans ce schéma, la monnaie offre ce que Walras appelle le « service d'approvisionnement », c'est-à-dire qu'elle permet aux ménages et aux entreprises de s'approvisionner en dépit de la non-synchronisation des recettes et des dépenses.

Dans la fiction walrassienne, il y a un marché pour chaque capital circulant, où la demande provient des entreprises et l'offre vient des particuliers. La monnaie est un de ces éléments. Dans les équations, les biens A, B, C…U se voient affecter le signe PRIME lorsqu'ils servent comme capital circulant : on a donc A', B', C'…U'. Les prix d'équilibre correspondent à la valeur de ces stocks multipliée par le taux d'intérêt, ce qui signifie que le taux d'intérêt influe sur l'offre et la demande de ces capitaux.

---

[1] Le *T* de Fisher est manifestement un flux. Le *R* de Pigou semble être un stock, mais son auteur ne le définit qu'assez vaguement.
[2] Pigou [286] p. 174.
[3] Pigou [286] p. 178.

$$p_{a'} = i \; ; \quad p_{b'} = p_b.i \; ; \ldots ; \quad p_{u'} = p_u.i \qquad (4.28)$$

La manière dont Walras conçoit la demande de monnaie tant par les particuliers que par les entreprises est symptomatique : la monnaie apparaît comme une anticipation du capital circulant réel. Par exemple, il y aura une double demande de A' : sous forme réelle et sous forme monétaire. Des coefficients techniques sont ajoutés pour l'input des différents capitaux circulants sous chacune de ces deux formes.

La prise en compte des capitaux circulants en nature génère de nouvelles équations et augmente le nombre d'inconnues. Mais contentons-nous ici d'examiner le cas de la monnaie. L'intégration de celle-ci dans les équations se fait différemment selon qu'elle est un bien spécifique, un bien normal ayant une double fonction (monétaire et non monétaire) ou plusieurs biens (bimétallisme), selon que la monnaie est simultanément le numéraire ou non. Walras explore ces différents cas. Nous la considérerons ici comme un bien spécifique (*U*) avec *A* qui reste le numéraire.

Walras désigne par $\alpha, \beta, \gamma \ldots \varepsilon$ les quantités demandées des biens A', B', C'…E' sous forme de monnaie par les ménages. Mais comme ceux-ci sont offreurs de monnaie, cette demande agit comme une réduction de l'offre. L'offre de capital circulant et de monnaie fonctionne comme celle des services producteurs, explicitée par l'équation (3.15), on a donc :

$$\varphi_\alpha(\alpha) = p_{a'}.\, \varphi_a(d_a) \qquad (4.29)$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$\varphi_\varepsilon(\varepsilon) = p_{a'}.\, \varphi_a(d_a)$$

Dont on déduit :

$$\alpha = f_\alpha\,(p_t, p_p, p_k, \ldots p_b, p_c, p_d, \ldots p_{a'}, p_{b'}, p_{c'} \ldots p_{u'}) \qquad (4.30)$$
$$::$$
$$\varepsilon = f_\varepsilon\,(p_t, p_p, p_k, \ldots p_b, p_c, p_d, \ldots p_{a'}, p_{b'}, p_{c'} \ldots p_{u'})$$

Sommant pour l'ensemble des ménages, les valeurs $\alpha, \beta, \gamma \ldots \varepsilon$, Walras obtient $d_\alpha, d_\beta, d_\gamma \ldots d_\varepsilon$. Si Q_u désigne le stock monétaire total dont sont dotés les ménages, l'offre de monnaie (O_u) est exprimée par la formule :

$$O_u = Q_u - (d_\alpha.p_{a'} + d_\beta.p_{b'} + d_\gamma.p_{c'} + \ldots + d_\varepsilon.p_{a'})/p_{u'} \qquad (4.31)[1]$$

Walras examine ensuite la demande de monnaie par les entreprises de façon analogue. Partant des coefficients techniques et des quantités produites, il détermine les grandeurs $\delta_\alpha, \delta_\beta, \delta_\gamma \ldots$ qui sont respectivement les quantités totales de A', B', C' que les entreprises souhaitent détenir sous forme de monnaie. La demande de monnaie totale, D_u est donnée par l'équation :

$$D_u = (\delta_\alpha.p_{a'} + \delta_\beta.p_{b'} + \delta_\gamma.p_{c'} + \ldots)/p_{u'} \qquad (4.32)$$

A l'équilibre, on a évidemment $O_u = D_u$. J'ai trop simplifié l'exposé de Walras pour pouvoir ici dénombrer les équations et les inconnues, mais le lecteur voudra bien me croire que l'égalité est maintenue ; l'équilibre général reste donc possible avec le capital circulant et la monnaie. Un concept central dans le paradigme walrassien est celui d'*encaisse désirée* : en numéraire, elle vaut la somme de $(\delta_\alpha.p_{a'} + \delta_\beta.p_{b'} + \delta_\gamma.p_{c'} + \ldots)$ et de $(d_\alpha.p_{a'} + d_\beta.p_{b'} + d_\gamma.p_{c'} + \ldots + d_\varepsilon.p_{a'})$; désignons la par H_α. L'équilibre ($O_u = D_u$) requiert :

$$Q_u = H_\alpha/p_{u'} \qquad (4.33)$$

[1] Q_u, O_u et D_u sont exprimés en unités monétaires et non en numéraire.

Walras constate : « Le prix du service de la monnaie s'établit par hausse ou par baisse suivant que l'encaisse désirée est supérieure ou inférieure à la quantité de monnaie »[1].

Comme $p_u = p_u \cdot /i$, les ajustements de la valeur unitaire de la monnaie, p_u, constituent le facteur d'équilibre monétaire.

La similitude entre l'équation (4.34) et l'équation (4.26) des Cambridgiens est évidente. L'équation (4.34) relève manifestement de la théorie quantitative. Si on augmente Q_u dans une certaine proportion, il faut que p_u qui représente le pouvoir d'achat de la monnaie varie en proportion inverse. Le quantitativisme est néanmoins un peu moins parfait ici, car H_α est presque indépendant de p_u, mais pas tout à fait. La variation minime de H_α, causée par celle de Q_u, affectera légèrement celle de p_u.

4.3.4. Caractéristiques essentielles de la macroéconomie pré-keynésienne

LA DICHOTOMIE ET LA NEUTRALITE DE LA MONNAIE

Une caractéristique importante de la théorie néoclassique est la dichotomie : l'économie se compose de deux secteurs indépendants l'un de l'autre : le secteur « réel », gouverné par les lois microéconomiques que nous avons étudiées et le secteur monétaire. Le revenu réel a pour déterminants principaux la technologie et les goûts des consommateurs. Le volume de *M*, de nature macroéconomique, n'influence pas le revenu réel. Par contre, il lui revient de déterminer le niveau général des prix. Si l'économie pouvait fonctionner sans la monnaie, le fait de la retirer ne changerait pas la quantité produite des différents biens ni le rapport entre leurs prix.

Comme l'explique Fisher, les prix relatifs sont régis par la loi de l'offre et de la demande, à l'opposé du niveau général des prix qui est déterminé par l'équation de l'échange. Si l'on considère une courbe d'offre ou de demande d'un bien exprimée en monnaie, elle présuppose déjà inconsciemment un niveau général des prix donné. « The price of sugar in dollars depends partly on sugar, partly on dollar »[2]. Sans référence à une valeur du dollar donnée de façon exogène, le prix du sucre en dollars est une inconnue et il ne peut servir de base aux courbes d'offre et de demande.

Wicksell écrit : « there is nothing in the act of exchange as such which can determine the value of money or concrete commodity prices (…) At bottom, it is only goods which are exchanged against each other. »[3]. La monnaie n'est donc qu'un voile[4].

En conséquence, une variation de *M* est incapable d'affecter le revenu réel, de même que les prix relatifs ; elle ne peut toucher que *P*, le niveau général des prix. C'est ce qu'on a appelé la neutralité de la monnaie. Les concepts de dichotomie et de neutralité de la monnaie sont très proches mais néanmoins distinct. Le premier est théorique : les forces qui gouvernent le revenu réel et la monnaie sont deux mondes qui ne s'interpénètrent pas. Le second est pratique : la politique monétaire visant à soutenir la croissance économique est condamnée à être inefficace.

---

[1] Walras [380] p. 465.
[2] Fisher [92] p. 177
[3] Wicksell [385] p. 22
[4] Cette conception existait déjà chez les classiques. Mill écrit : « There cannot, in short, be intrinsically a more insignificant thing, in the economy of society, than money (…) It is a machine for doing quickly and commodiously, what would be done, though less quickly and commodiously, without it (…) The introduction of money does not interfere with the operation of any of the Laws of Value laid down in the preceding chapters » ([258] § 3.7.8 et 3.7.9).

L'IMPOSSIBILITE DU CHOMAGE INVOLONTAIRE

La conception walrassienne de l'équilibre s'est imposée, au point que la définition première de l'équilibre (l'absence d'incitation au changement) fut éclipsée par la conception walrassienne de l'égalisation entre l'offre et la demande tous les marchés grâce à la flexibilité des prix. Ceci vaut évidemment pour le marché du travail et le salaire qui sont inclus dans l'équilibre général walrassien. N'oublions pas que l'égalité de l'offre avec la demande sur un marché signifie qu'au prix du marché, toute demandeur qui est d'accord de payer ce prix trouve une offre correspondante et que tout offreur qui se satisfait du prix du marché trouve une demande correspondante. Sur le marché du travail, ceci implique l'absence de chômage involontaire. Les inactifs sont ceux qui choisissent de ne pas travailler pour un salaire qui est inférieur à leur *salaire de réserve*[1]. Si du chômage involontaire devait exister, il ne durerait pas, car le système de prix s'ajusterait. Le plein emploi est assuré par la flexibilité du salaire. Il se trouvera toujours un salaire suffisamment bas pour inciter un employeur au recrutement.

On sait que dans l'économie, il y a en permanence des emplois qui disparaissent parce que des firmes périclitent ou déclinent et des emplois qui se créent parce que des firmes naissent ou se développent. On appelle *chômage frictionnel* la situation des chercheurs d'emploi, que seul sépare de leur emploi le délai normal de recherche ; ils chôment, non parce qu'il y a peu d'emplois, mais parce que même lorsqu'il y en a suffisamment, chercher le meilleur emploi prend du temps. Le chômage frictionnel est donc forcément de courte durée. Pour cette raison, la théorie économique le considère comme négligeable. Ce sont les formes de chômage involontaire plus graves que le chômage frictionnel, qui sont niées par les économistes néoclassiques.

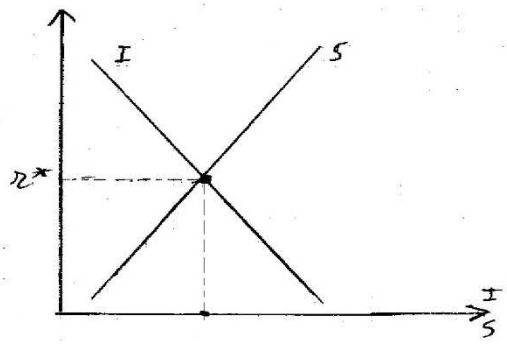
LA THEORIE DES FONDS PRETABLES

Depuis Ricardo, toute une tradition considère qu'en équilibre à long terme, le taux d'intérêt correspond au taux de rendement du capital dans la sphère réelle. Pour Fisher, il s'égalise à la fois avec le taux de rendement au dessus du coût et avec le taux d'impatience. Pour Wicksell, si le taux d'intérêt est inférieur à son taux naturel, le processus cumulatif s'enclenche pour rétablir dans la douleur l'égalité entre le taux d'intérêt et le rendement réel de la production. Robertson affirme également cette conception : « If we ask what ultimately governs the judgements of wealth owners as to why the rate of interest should be different in the future from what it is to-day, we are surely led straight back to the fundamental phenomena of Productivity and Thrift »[2].

[1] Un concept fondamental de la théorie néoclassique est sous-jacent dans cet argument : *le salaire de réserve*. L'offre de travail repose sur le choix entre du loisir et du revenu. Le salaire de réserve, propre à chaque individu, est le salaire minimum pour le décider à offrir sa force de travail.

[2] Robertson [298] p. 25

Graphique 4.23 : la théorie des fonds prêtables



Le taux d'intérêt r égalise l'offre et la demande de fonds prêtables.*

Il y a une courbe de demande de capitaux (I) déclinante puisque la productivité réelle diminue avec la taille du capital et une courbe d'offre de capital (S) croissante, car l'épargne est sensible au taux d'intérêt.

4.4- LES FONCTIONS DE PRODUCTION AGREGEES

Nous avons vu au chapitre 3.2 le contraste entre les théories de Clark et de Wicksteed à propos de la production. Prenant progressivement de l'ascendant, les économistes clarkiens vont s'approprier la fonction de production en la dénaturant par rapport à sa forme wicksteetienne. Ils en arrivent à ce qu'on appelle la *fonction de production agrégée*, qui se présente comme suit :

$$P_F = F(K,L) \qquad (4.34)$$

Le volume de la production totale dans l'économie (P_F) est fonction de la quantité de travail (L) et de la quantité de capital (K). Wicksteed s'intéressait lui à la relation entre une production particulière et les quantités de chacun des nombreux facteurs qui y intervenaient. On assiste donc une double agrégation : les divers facteurs sont rassemblés en deux groupes et les unités de production sont conglomérés au niveau d'une économie ou de l'industrie dans son ensemble.

La fonction de production agrégée la plus connue est celle dite de ***Cobb-Douglas***, dont la formule est :

$$P_F = b.K^{\alpha}.L^{\beta} \qquad (4.35)$$

Où α, β et *b* sont trois paramètres affectant les variables *K* et *L*. Il se fait que les exposants α et β correspondent à la part du produit P_F gagnée par le facteur *K* ou *L* qui leur sert de base. Montrons-le avec *K* : la dérivation partielle de (4.35) par rapport à *K* donne $\partial P_F/\partial K = \alpha.P/K$. Multipliant cette rémunération unitaire par la quantité de capital *K*, on obtient que la rémunération totale du capital égale $\alpha.P_F$.

Cette fonction fut « inventée » par Wicksell lors du traitement du problème d'Akerman, mais c'est l'économiste Paul Douglas qui la rendit célèbre en la testant empiriquement avec l'aide du mathématicien Charles Cobb, dans l'article « A Theory of Production » qu'ils publièrent en 1928 et que nous allons résumer.

Cette fonction est homogène de degré un lorsque $\alpha+\beta = 1$. Voulant défendre la théorie de la productivité marginale comme explication de la répartition du revenu, nos auteurs privilégient logiquement l'hypothèse des rendements d'échelle constants, nécessaire à l'épuisement du produit[1]. En estimant les paramètres α, β et b par la méthode des moindres carrés, Cobb et Douglas arrivent à l'équation :

$$P_F = 1{,}01.K^{1/4}.L^{3/4} \qquad (4.36)$$

L'étude porte sur les statistiques de la production manufacturière américaine au niveau macroéconomique, entre 1899 et 1922. Cobb et Douglas travaillent, non pas avec les valeurs absolues de *P*, *K* et *L* mais avec leurs indices annuels où l'année 1899 sert de base (100)[2].

Le premier test concerne la corrélation entre *P* (l'indice de la production suivant les statistiques) et P_F (la production théorique). Le coefficient de corrélation obtenu est de 0,94 (après avoir neutralisé l'influence du trend croissant, source d'une corrélation trop facile). Les auteurs se félicitent de ce résultat, dont ils déduisent que « the equation $P_F = 1{,}01.L^{3/4}.K^{1/4}$ describes in a

[1] Si $\alpha+\beta > 1$, on aurait des rendements croissants ; si $\alpha+\beta < 1$, on aurait des rendements décroissants.

[2] Les valeurs absolues servirent évidemment de point de départ à l'établissement des indices. Les statistiques étant moins élaborées à l'époque, Cobb et Douglas durent pas mal manipuler les données pour les rendre assimilables par leur exercice. Concernant le capital, ils partirent de la valeur des bâtiments et des machines qu'ils déflatèrent à l'aide d'un indice des prix.

fairly accurate manner the actual processes of production in manufacturing during this period »[1]. Ils constatent avec satisfaction que P et P_F suivent tous deux le cycle conjoncturel. En fait, les fluctuations de P_F sont moins amples que celles de *P*, ce qui s'explique par une double déficience des données statistiques entrantes :

- elles ne tiennent pas compte du fait qu'une partie de la capacité de production (K) demeure inemployée pendant les dépressions
- elles ignorent les heures supplémentaires du personnel, plus nombreuses pendant les phases d'expansion, car la main-d'œuvre (L) est simplement comptée en hommes et non en hommes/heures.

Cobb et Douglas montrent également que :

- la productivité marginale du travail égale 0,75.P/L
- la productivité marginale du capital égale 0,25.P/K
- la masse salariale totale égale 0,75.P
- le profit total égale 0,25.P
- la répartition du revenu est ¾ pour le travail et ¼ pour le capital. Ces parts sont donc indépendantes du rapport K/L (intensité capitalistique)[2].

Le deuxième test porte sur la corrélation entre la masse salariale totale telle que calculée ci-dessus (toujours en indice) et l'indice des salaires réels sortant directement des statistiques. Le coefficient de corrélation n'est cette fois plus que de 0,69 ; les mouvements à court terme ne sont quasiment pas corrélés, contrairement aux tendances[3]. Mais Cobb et Douglas estiment ce résultat suffisant pour conforter la théorie. Ils font d'ailleurs état d'autres études statistiques qui estiment la part des salaires à trois quarts.

Cobb et Douglas reconnaissent que de nombreuses améliorations pourraient être apportées à leur exercice statistique mais ne doutent pas que la théorie de la productivité marginale en sort confortée. A tel point qu'ils se sentent obligés de mettre en garde contre la conclusion hâtive qu'on pourrait en tirer quant à la justification éthique de l'ordre social existant. Ils écrivent : « …even if there were precise correspondence, it would not furnish any light upon the question as to whether capital for example should be privately owned to the degree to which it is in our society. For while capital may be "productive", it does not follow that the capitalist always is. »[4]. Même un communiste devrait pouvoir croire en la théorie de la productivité marginale.

Comme nous le verrons plus loin, la macroéconomie moderne recourt intensivement à la fonction de Cobb-Douglas. Elle a toutefois fait l'objet de critiques. Ainsi, Blaug s'étonne de ce miracle qu'une fonction de Cobb-Douglas homogène de degré un règle la rémunération des facteurs dans nos économies, « malgré les monopoles, les impôts, les secteurs nationalisés et les syndicats de travailleurs »[5]. Il s'interroge également sur le principe de l'agrégation des fonctions de production. Dans leur article, Cobb et Douglas ne semblent pas s'en préoccuper. Certains auteurs tendent à considérer que la fonction de Cobb-Douglas rend aussi compte de la production dans les entreprises individuelles. Or, rappelle Blaug, il a été démontré

[1] Cobb et Douglas [62] p. 159.

[2] L'*élasticité de substitution* de la fonction de Cobb-Douglas est toujours uniformément égale à l'unité. Puisque les parts respectives des deux facteurs restent constantes lorsque le rapport K/L augmente, on en conclut que la baisse du rapport R/W (profit/salaire) qui accompagne la hausse du rapport K/L doit la compenser exactement. Rappelons que le concept de l'élasticité de substitution date de 1932 (quatre ans après l'article de Cobb et Douglas).

[3] La corrélation entre les moyennes mobiles sur sept ans monte à 0,89.

[4] Cobb et Douglas [62] p. 164

[5] Blaug [35] p. 554

mathématiquement que l'agrégation additive de fonctions de Cobb-Douglas ne donne pas une fonction de Cobb-Douglas.

TITRE V- LES CRITIQUES CONTRE LA THEORIE NEOCLASSIQUE

5.1- LA REVOLUTION KEYNESIENNE

A partir des années trente, l'université de Cambridge (Grande Bretagne) devient le foyer de contestation de la théorie néoclassique dominante.

KEYNES ET LA *THEORIE GENERALE*

Dans les années trente, le monde connaît la plus grande dépression de son histoire : le chômage bat tous les records. Or l'économie politique tant classique que néoclassique donne une description des rouages de l'économie qui exclut ce genre de crises. Elle enseigne que grâce notamment à la flexibilité des prix (en ce compris les salaires et le taux d'intérêt), tout déséquilibre entraîne un ajustement qui fait revenir le système à un nouvel équilibre. Or le chômage involontaire n'est rien d'autre qu'un déséquilibre sur le marché du travail, en l'occurrence un excès de l'offre sur la demande. En pareil cas, les économistes néoclassiques estimaient qu'une baisse du salaire suffisait à rétablir l'équilibre général et le plein emploi.

C'est dans ce contexte que l'économie politique connaît la plus grande secousse de son histoire. La « General Theory of Employment, Interest and Money » que John Maynard Keynes publie en 1936 est un coup de tonnerre dans le ciel de l'économie. Keynes fustige les économistes qui l'ont précédé, qu'il appelle les « classiques », mais qui comprennent aussi bien les néoclassiques que les classiques[1]. Ce livre comporte deux aspects qui sont totalement imbriqués : d'une part, il est une critique impitoyable de l'économie classique ; d'autre part, il propose un nouveau modèle macroéconomique.

Agé de cinquante-trois ans lorsque parut la Théorie Générale, Keynes avait déjà un curriculum glorieux. Il avait été élevé dans la tradition marshallienne, régnait sur Cambridge avec Pigou[2] et était devenu l'une des voix les plus écoutées dans le monde de l'économie politique. La *Théorie Générale* innove par rapport aux écrits antérieurs de Keynes, principalement son « Treatise on Money » (1930) vis-à-vis duquel il prit ses distances.

La Théorie Générale suscita immédiatement un débat intense parmi les économistes, débat qui sera d'autant plus passionné que ses implications sont politiques. En effet, Keynes condamne le *laissez faire* cher aux classique ; lorsque le marché ne parvient pas à amener le plein emploi, il revient aux pouvoirs publics de soutenir la demande globale en dépensant plus, même au prix d'un déficit du budget de l'Etat. Keynes se tint personnellement en retrait de ces discussions, si ce n'est par quelques articles parus en 1937 (cf. infra) ; ensuite sa santé déclina. L'essentiel du débat se déroula dans les années cinquante à soixante-dix ; nous lui consacrons le titre VII.

5.1.1. Les mécanismes keynésiens

1- Le marché de l'emploi et la rigidité des salaires nominaux

Vers la fin de son ouvrage, Keynes écrit : « Nous avons démontré que, en cas d'insuffisance de la demande effective, la main d'œuvre est sous-employée en ce sens qu'il y a des hommes sans emploi désireux de travailler pour moins que le salaire réel en vigueur »[1]. Le *chômage involontaire* est au centre de l'ouvrage. Keynes en exclut le *chômage frictionnel*. Concernant la

[1] Bien que je considère la distinction des écoles classique et néoclassique comme essentielle, lorsque je traiterai de Keynes, j'utiliserai sa propre terminologie qui regroupe les deux écoles sous l'appellation « classiques ».

[2] La *Théorie Générale* prit un ouvrage de Pigou consacré au chômage (1933) comme modèle de l'économie classique, objet de ses critiques. La relation entre les deux hommes en souffrit.

thèse classique que le chômage ne serait pas involontaire, Keynes écrit : « Le volume de l'emploi connaît d'amples variations, sans qu'il y ait des changements apparents, ni dans les salaires minima exigés par la main-d'œuvre ni dans sa productivité. L'ouvrier n'est pas plus intransigeant en période de dépression qu'en période d'essor »[2]. Que le lecteur ne se méprenne pas : il ne s'agit pas de savoir s'il peut ou non y avoir du chômage volontaire mais s'il peut y avoir ou non du chômage involontaire.

Pour les classiques, le marché du travail se caractérise par une offre correspondant à la désutilité croissante du travail qui croise une demande correspondant à la productivité marginale décroissante du travail. L'égalisation de cette offre et cette demande détermine à la fois le niveau de l'emploi et le salaire. Le salaire ainsi déterminé est évidemment le *salaire réel*. Le *salaire nominal* n'est rien d'autre que le salaire réel multiplié par le niveau général des prix et la distinction entre les deux ne joue aucun rôle dans la théorie classique. Si à la suite d'une baisse de la productivité marginale du travail, un chômage devait advenir, la baisse du salaire nominal négocié entre les employeurs et les employés, réduirait le salaire réel jusqu'à ce qu'il atteigne le nouveau niveau compatible avec l'équilibre. Le chômage involontaire ne peut donc pas être durable.

Pour Keynes, la distinction entre le salaire nominal et le salaire réel est essentielle. Il considère que les classiques font erreur lorsqu'ils pensent que les négociations salariales sont capables de déterminer le salaire réel. Il considère que le critère de l'offreur de travail n'est pas le salaire réel mais le salaire nominal : « Alors que la main d'œuvre résiste ordinairement à la baisse des salaires nominaux, il n'est pas dans ses habitudes de réduire son travail à chaque hausse du prix des biens de consommation »[3]. En voici la raison : les baisses du salaire nominal se négocieraient, secteur par secteur, entreprise par entreprise, salarié par salarié ; elles modifieraient donc toute l'échelle des salaires les uns par rapport aux autres. Une hausse des prix, quant à elle, érode tous les salaires de façon égale et suscitera donc moins de tensions. « En d'autres termes, la compétition autour des salaires nominaux influe surtout sur la répartition du salaire réel global entre les groupes de travailleurs et non sur le montant moyen par unité de travail, lequel dépend comme nous le verrons plus tard, d'une autre série de facteurs »[4].

Il est donc illusoire de compter sur une baisse des salaires nominaux pour contrer le chômage.

[1] Keynes [186] p. 291
[2] Keynes [186] p. 35
[3] Keynes [186] p. 34
[4] Keynes [186] p. 40

2- La demande effective et la fonction d'emploi

L'idée novatrice de Keynes est que le volume de l'emploi ne se détermine pas sur le marché du travail mais sur les marchés des produits et services. L'emploi est déterminé par le revenu national et le revenu national résulte de l'équilibre entre l'*offre globale* et la *demande globale*, c'est-à-dire l'offre et la demande de l'ensemble des biens. Mais plutôt que de traiter ces variables comme des grandeurs monétaires qu'il convertirait ensuite en emplois, Keynes préfère exprimer l'offre globale, la demande globale et le revenu national directement en unités d'emploi. Certes, tout niveau d'emploi pourrait correspondre à de multiples niveaux du revenu national en fonction de la nature et du volume du capital utilisé. Mais Keynes raisonne dans le court terme marshallien où le travail est le seul facteur variable, le capital étant donné.

Le graphique 5.1 expose l'*offre globale* (courbe *Z*) et la *demande globale* (courbe *D*). Suivant la logique marshallienne, l'abscisse indique une quantité et l'ordonnée un prix. La quantité en abscisse est exprimée en unités d'emploi [1] et non en quantité produite ou en unités monétaires.

Le prix d'offre globale peut être défini comme « le produit attendu qui est juste suffisant pour qu'aux yeux des entrepreneurs, il vaille la peine d'offrir ce volume d'emploi »[2]. De même, pour un emploi donné, le prix de demande est le produit[3] que les entrepreneurs peuvent raisonnablement espérer tirer de la production correspondant à ce niveau d'emploi.

Nous avons donc : $Z = \varphi(N)$ et $D = f(N)$ où *N* est le niveau de l'emploi. « Si pour un certain volume de l'emploi *N*, le produit attendu est supérieur au prix de l'offre globale, c'est-à-dire si *D* est supérieur à *Z*, il y aura un mobile qui incitera les entrepreneurs à accroître l'emploi et s'il le faut, à élever les coûts en se disputant les uns aux autres les facteurs de production, jusqu'à ce que l'emploi ait atteint le volume qui rétablit l'égalité entre *Z* et *D*. Ainsi le volume de l'emploi est déterminé par le point d'intersection de la courbe de la demande globale et de la courbe d'offre globale ; car c'est à ce point que la prévision de profit des entrepreneurs est maximum. Nous appellerons *demande effective,* le montant du produit attendu *D* au point de la courbe de demande globale où elle est coupée par celle de l'offre globale ».[4]

Sur le graphique 5.1, N_D est l'emploi réel, tel que déterminé par la demande effective. Il est déterminé indépendamment du plein emploi N_P qui égalise la productivité marginale avec la désutilité marginale du travail selon le schéma classique. Par hasard, les deux peuvent correspondre, mais ce n'est nullement assuré. Lorsque $N_P > N_D$, il y a du chômage involontaire. Keynes exclut que N_P puisse être inférieur à N_D (salaire inférieur à la désutilité marginale)[5].

---

[1] La variable *emploi* tient compte des différences de salaire. Soient, w_1, w_2… les différents salaires existants et S_1, S_2… les rémunération globales de ces catégories de main d'œuvre. L'emploi (N) vaudra $S_1/w_1+S_2/w_2+\ldots$

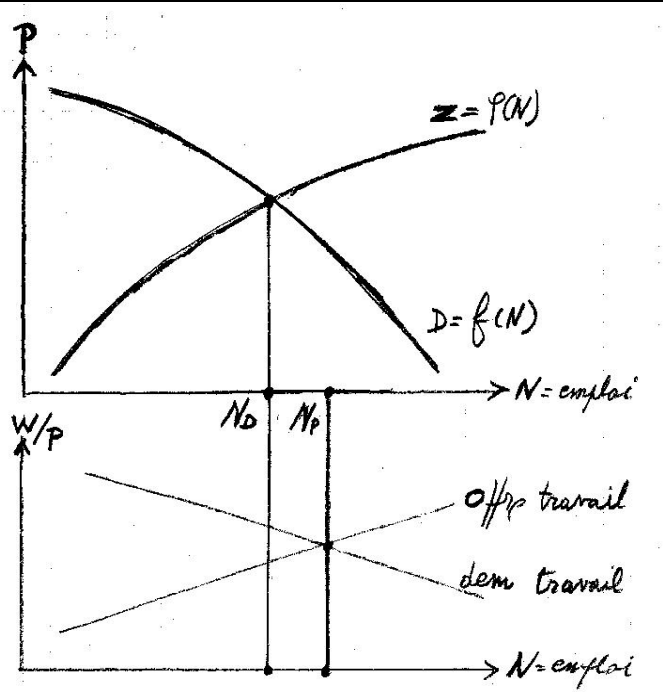
[2] Keynes [186] p. 48

[3] Le terme « produit » qui explique l'ordonnée des deux courbes se limite ici à la *valeur ajoutée* par l'entreprise. Il est net de coût d'usage parce que selon Keynes, ce dernier est influencé par nombre de facteurs extérieurs à la problématique.

[4] Keynes [186] pp. 49-50

[5] Dans un ouvrage ultérieur, il envisagera cette possibilité génératrice d'inflation (cf. sous-chapitre 7.3.3).

Graphique 5.1 : la détermination de l'emploi par la demande effective



Keynes reproche a l'économie politique classique de n'avoir pas une théorie de la production « as a whole », de négliger l'offre et la demande GLOBALES.

La demande globale se compose de la consommation plus l'investissement ; étudions maintenant comment se déterminent leurs volumes respectifs, en gardant à l'esprit que les facteurs qui les déterminent sont en dernière instance les déterminants de l'emploi.

3- La consommation et l'épargne

Les agents partagent leur revenu entre la consommation et l'épargne. Par définition, l'épargne n'est rien d'autre que de la non-consommation.

La théorie keynésienne de la consommation est assez simple : la communauté (ou l'individu) dépense pour sa consommation une fraction de son revenu déterminée par des facteurs tant objectifs que subjectifs. Keynes appelle *propension à consommer* la relation fonctionnelle entre la consommation et le revenu. Il passe en revue les facteurs qui constituent la propension à consommer. Cette énumération rappelle les considérations de Fisher sur la préférence pour le présent. Une hausse de l'impatience doit normalement coïncider avec une hausse de la propension à consommer. Keynes considère que la propension à consommer est assez stable dans le temps.

La *propension marginale à consommer* est la part consommée d'un accroissement marginal du revenu. Si une augmentation du revenu de 100 entraîne une augmentation de la consommation de 80, la propension marginale à consommer vaut 0.8. D'après Keynes, la propension marginale à consommer est positive mais décroissante. Plus le revenu augmente, plus la part épargnée tend à croître. Les classes aisées épargnent une plus grande part de leur revenu que les moins nantis. Keynes pense que l'enrichissement d'un pays freine la propension à consommer de ses habitants.

Rappelons que dans la vision classique, c'est le taux d'intérêt qui détermine les parts respectives de la consommation et de l'épargne ; une hausse du taux d'intérêt augmente l'attrait de l'épargne et réduit la consommation. Keynes conteste la présence de pareils comportements.

Selon lui, la propension à consommer est indépendante du taux d'intérêt. « Rares sont les personnes qui modifient leur train de vie parce que l'intérêt baisse de 5 à 4%, si leur revenu global reste le même. Les effets indirects peuvent être plus nombreux bien qu'ils ne soient pas tous dans le même sens »[1].

4- L'incitation à investir

Keynes met au centre de son raisonnement, un concept qu'il nomme *efficacité marginale du capital* (MEC pour *marginal efficiency of capital*). Il la définit comme le « taux d'escompte qui, appliqué à la série d'annuités constituée par les rendements escomptés de ce capital pendant son existence entière, rend la valeur actuelle des annuités égale au prix d'offre de ce capital ». Et ce dernier est défini comme « le prix juste suffisant pour décider un fabricant à produire une unité supplémentaire de ce capital, c'est-à-dire ce que l'on appelle parfois son coût de remplacement ». Chaque bien capital a donc sa propre MEC ; on peut considérer qu'en sélectionnant les biens capitaux dont la MEC est le plus élevée, on définit la MEC du capital en général. Keynes estime que sa MEC est un concept fort proche du RROC de Fisher. Comme le fera remarquer Lerner, l'appellation *efficacité marginale de l'INVESTISSEMENT* (MEI) eût été plus appropriée, car le concept s'applique au capital additionnel et non au stock total de capital.

A mesure que l'investissement augmente, la MEC diminue de par la conjonction de ces deux causes :
- le rendement du capital diminue lorsque sa quantité augmente
- l'accroissement de la demande de biens capitaux entraînera leur renchérissement.

Le graphique 3.20 traduit parfaitement la pensée de Keynes. Seront réalisés les investissements dont l'efficacité marginale est au moins égale au taux d'intérêt. Dans la *théorie des fonds prêtables* (cf. sous-chapitre 4.3.4), l'intersection des courbes d'offre et de demande de capitaux déterminaient simultanément le taux d'intérêt et le montant de l'investissement. Pour Keynes, le taux d'intérêt est déterminé ailleurs que sur le marché des fonds prêtables (cf. infra).

Les revenus futurs procurés par un investissement, qui servent à calculer sa MEC, sont évidemment **incertains**. Ils font l'objet d'une évaluation par l'entrepreneur qu'on appelle *anticipation*. Cette incertitude occupe une place centrale dans la théorie keynésienne. Dans un article de 1937, Keynes énonce les principales différences entre les classiques et lui et il place le rôle dévolu à l'incertitude en première position. Avant lui, déplore Keynes, soit les économistes traitaient les choix économiques comme s'ils se présentaient dans un monde certain, soit ils neutralisaient l'incertitude avec le calcul des probabilités.

Selon Keynes, « Actually, we have, as a rule, only the vaguest idea of any but the most direct consequence of our acts »[2]. Quand nous ne disposons pas des informations nous permettant d'agir selon les canons de la théorie orthodoxe, nous devons nous contenter de comportements qu'elle jugerait irrationnels. Keynes cite trois comportements :
- L'attitude en retrait, l'inaction, le non-investissement, la thésaurisation, le refuge dans un actif sans revenu, la monnaie. « Partly on reasonable and partly on instinctive grounds, our desire to hold money as a store of wealth is a barometer of the degree of our distrust of our own calculations and conventions concerning the future ».
- La marche forcée en avant à propos de laquelle on préfère ne pas soupeser le pour et le contre. Le besoin d'agir, d'entreprendre prend le dessus. L'entrepreneur repose ses espoirs sur sa bonne étoile ou son ange gardien. C'est ce que Keynes qualifie d'« animal spirit ».

[1] Keynes [186] p. 111
[2] Keynes [183] p. 213

- La copie du jugement majoritaire. « Knowing that our own individual judgment is worthless, we endeavour to fall back on the judgment of the rest of the world which is perhaps better informed »[1]. Mais si un grand nombre d'investisseurs en bourse pensent selon le marché, l'édifice général devient très fragile. Il donne lieu à des corrections brutales vers le haut ou vers le bas, le marché étant pris d'une fièvre d'optimisme ou de pessimisme suivant le cas. Keynes dresse un tableau peu flatteur de la bourse qu'il compare à un casino. Les investisseurs plus qualifiés et plus compétents utilisent leur avantage, non pour indiquer le droit chemin vers les entreprises méritantes, mais pour tirer profit à court terme des erreurs de la majorité.

Selon Keynes, chacun de ces comportements est susceptible de créer de l'instabilité économique en empêchant l'équilibre classique de se réaliser.

5- L'ajustement par le revenu et le multiplicateur

Rappelons-nous la théorie des *fonds prêtables*, hégémoniste avant la *Théorie générale*. Le taux d'intérêt s'ajustait de façon à égaliser l'épargne et l'investissement et donc aussi l'offre et la demande globales. Keynes attaque cette conception. Selon lui, l'investissement et l'épargne sont égaux PAR DEFINITION. L'équation (5.1-C) résulte tout simplement des deux premières qui sont tautologiques :

Composition du revenu :	$Y = C + I$	(5.1-A)[2]
Définition de l'épargne :	$S = Y - C$	(5.1-B)
donc :	$S = I$	(5.1-C)

où *Y* est le revenu national, *C* est la consommation nationale, *S* est l'épargne nationale et *I* est l'investissement national.

L'épargne et l'investissement de la société n'ont donc pas besoin des variations du taux d'intérêt pour s'égaliser. Certes, un individu peut faire varier son investissement et son épargne indépendamment l'un de l'autre. Mais ce fait ne peut être généralisé à la société dans son ensemble, car les dépenses des uns sont les revenus des autres. Si tous les agents veulent beaucoup épargner et peu investir, ils ne pourront réaliser cet objectif simultanément : l'économie subira une baisse du revenu. La faiblesse de *I* poussera *S* à la baisse et donc également *Y*. Quand le désir d'épargner et celui d'investir divergent, c'est le revenu qui s'ajuste pour maintenir l'égalité, et non le taux d'intérêt. Keynes ne considère pas l'épargne d'un œil très favorable et pense que des mesures favorables à la consommation sont parfois nécessaires pour relancer une économie en crise.

[1] Keynes [183] p. 214

[2] Le revenu est ici apprécié sous l'angle de la dépense globale, qui se compose de l'investissement et de la consommation. Les dépenses publiques et le solde de la balance des paiements sont ignorés par souci de simplification.

Illustrons la différence entre les deux théories par un exemple chiffré schématique. Soit une économie dont le revenu est 100, qui se partage entre 80 de consommation et 20 d'investissement alors que le taux d'intérêt est de 5%. Imaginons que la demande d'investissement baisse pour ne plus valoir que 18. Les classiques supposeront qu'une baisse du taux d'intérêt à 4% (par exemple) rétablira l'égalité entre I et S, et que le revenu se maintiendra à 100. Non, dit Keynes, c'est le revenu qui baissera, passant d'abord à 98 mais continuant à décroître, car la propension à consommer, un facteur assez stable, est de 0,8 ; si l'épargne vaut 18, la consommation baissera jusqu'à ce que le revenu s'établisse à 90, ce qui maintient la propension marginale à consommer.

Nous avons introduit implicitement le *multiplicateur*, un concept essentiel chez Keynes. Partant de l'idée que tout investissement a son épargne correspondante, et que le montant de l'épargne totale est lié à celui du revenu national par un rapport donné, Keynes en déduit la relation qui doit exister entre la variation de l'investissement et la variation consécutive du revenu.

En vertu de l'équation (5.1), on a : $\Delta Y = \Delta C + \Delta I$. Or par définition, $\Delta C = c.\Delta Y$, où *c* est la propension marginale à consommer. On peut donc en conclure que $\Delta Y = 1/(1\text{-}c).\Delta I$. Keynes appelle ce rapport 1/(1-c) le *multiplicateur d'investissement*, désigné par la lettre *k*. On a donc :

$$\Delta Y = k.\Delta I \qquad \text{où} \qquad k = 1/(1\text{-}c) \qquad (5.2)$$

Ce multiplicateur joue tant à la hausse qu'à la baisse (cf. notre exemple numérique ci-dessus).

C'est le revenu NOMINAL qui bénéficie du multiplicateur. Lorsque le plein emploi est atteint, le revenu réel ne peut augmenter et c'est le niveau général des prix qui concentre sur lui la totalité de l'accroissement. En fait, plus l'économie est éloignée du plein emploi, plus l'emploi bénéficiera de l'accroissement de l'investissement.

La valeur de *k* résultant de la formule (5.2) représente la valeur maximum de la fourchette dans laquelle il se situera. De nombreux facteurs relatifs à la complexité de l'économie réelle contrecarreront partiellement l'action du multiplicateur. Keynes en énumère quelques-uns : baisse de la propension marginale à consommer quand le revenu augmente, importations en provenance de l'étranger…

Le multiplicateur d'investissement est un phénomène encourageant pour les gouvernements désireux d'accroître l'emploi par une politique de travaux publics. Mais l'investissement additionnel doit évidemment représenter un accroissement NET de l'investissement total. Le phénomène ne jouera pas ou pas complètement si l'investissement neuf se réalise au détriment d'autres investissements, par exemple à cause d'une hausse du taux d'intérêt. Une politique monétaire adéquate devra peut-être accompagner l'investissement public.

Le système bancaire peut produire l'illusion que ses prêts sont à même de générer des investissements sans épargne correspondante. Selon Keynes, il n'en est rien. Soit un investissement financé par un crédit bancaire qui est réellement additionnel dans la mesure où il est certain que cet investissement n'aurait pas eu lieu sans ce crédit, ni aucun autre à sa place. « On ne peut forcer personne à rester en possession de la monnaie additionnelle correspondant à ce crédit bancaire nouveau ; il faut que quelqu'un préfère, pour certaines raisons, conserver un supplément de monnaie plutôt qu'une autre forme de richesse »[1], et cela n'est possible que moyennant un accroissement du revenu, qui devra être supérieur au crédit bancaire. L'accroissement du revenu ΔY entraînera une variation de l'épargne $(1-c).\Delta Y$ qui financera l'investissement additionnel.

[1] Keynes [186] p. 103

Histoire du multiplicateur

Ce concept apparaît pour la première fois dans un article de l'économiste **Richard Kahn**, collègue cambridgien de Keynes, publié en 1931, où il défend une politique de travaux publics pour combattre la crise qui sévit. Kahn imagine un investissement, comme la construction d'une nouvelle route, financé par le système bancaire de façon à assurer qu'il augmente bien l'investissement total de la société. Il fournit à des anciens chômeurs un certain nombre d'emplois que Kahn qualifie de *primaires* car ils sont liés directement à la dépense publique initiale. Mais chaque emploi primaire entraîne lui-même la création d'un certain nombre d'emplois dits *secondaires* liés à la fabrication des biens que consommeront ces travailleurs nantis d'un nouveau pouvoir d'achat. Désignons ce nombre par la lettre *m*. Pour déterminer *m*, Kahn donne la formule :

$$m = [a.(W - U)+b.P]/(W+P+R)$$

W est le salaire des nouveaux embauchés, *U* les allocations de chômage (économisées), *P* le profit de l'entrepreneur qui fournit le travail, *R* les matières et produits intermédiaires importées pour les nouvelles productions, *a* la part du salaire qui achète des produits non importés et *b* la part du profit qui achète des produits non importés ($b < a$, à cause de l'épargne des capitalistes). Les emplois *secondaires* suscitent d'autres emplois *secondaires* de la même façon que les emplois *primaires*, ce qui donne lieu à une création en chaîne « though with diminished intensity. And so on ad inifinitum »[1]. L'emploi *secondaire* total par emploi *primaire* vaut donc $m+m^2+m^3+\ldots$, soit un total de $m/(1-m)$, que Kahn estime approximativement à 0,75 pour la Grande Bretagne à son époque. Cette formule n'est évidemment valide que moyennant l'hypothèse *toutes autres choses restant égales*, notamment les prix, les salaires, la productivité, les recettes fiscales. La courbe d'offre doit être parfaitement élastique, ce qui sera généralement le cas en période de chômage intense.

En 1932, l'économiste danois **Jens Warming** publie un article qui anticipe magistralement la conception keynésienne, où le multiplicateur trouve la forme que Keynes a retenue. La première section est intitulée « Activity creates capital », entendant que le capital naît dans la production, non dans l'épargne. L'investissement crée l'épargne qui le financera. Warming appelle *épargne latente*, celle qui serait générée par la mise au travail des chômeurs ; il ne tient qu'à la volonté des pouvoirs publics de la transformer en épargne réelle.

La banque qui accorde un crédit d'investissement travaille, non pas avec l'épargne déposée chez elle mais avec l'épargne latente. « The capital does not lie idle beforehand ». L'argent déposé à la banque n'est pas le capital, mais un reflet du capital, lui-même constitué des objets productifs ; il ne peut s'accroître qu'avec la valeur de ces objets.

L'erreur de la Loi de Say est de ne pas envisager le cas où beaucoup veulent épargner mais peu désirent emprunter. Dans ce cas, « the saving will go to waste ». Si j'épargne en période de dépression, je fais baisser la demande, donc la production, donc le revenu des autres, donc l'épargne des autres. "…then the saving will be neutalised by the consumption of property by people who have no income on account of the depression"[2].

[1] Kahn [168] p. 173
[2] Warming [381] p. 217

6- L'intérêt et la préférence pour la liquidité

Nous savons que le taux d'intérêt déterminera le montant de l'investissement (et donc du revenu national et de l'emploi) par l'égalisation avec l'efficacité marginale du capital. Mais où est déterminé le taux d'intérêt ? Sur le marché monétaire. La théorie de Keynes de la monnaie et du taux d'intérêt s'appelle *théorie de la préférence pour la liquidité* ; elle est donc la rivale de la *théorie des fonds prêtables*.

En matière d'épargne, l'agent a deux décisions successives à prendre : d'abord partager son revenu entre la consommation et l'épargne ; ensuite partager son épargne entre la détention de monnaie (thésaurisation) et l'achat de titres et créances. Comme nous l'avons vu, Keynes refuse au taux d'intérêt toute influence sur la première décision. « Il devrait être évident que le taux de l'intérêt ne peut être une récompense pour l'épargne ou l'abstinence en tant que telle. Lorsqu'un homme accumule ses épargnes sous forme d'argent liquide, il ne gagne aucun intérêt bien qu'il épargne tout autant qu'un autre »[1]. Par contre, Keynes attribue au taux d'intérêt un rôle déterminant dans la seconde décision ; son rôle économique est d'égaliser le désir de détenir des encaisses avec la masse monétaire disponible.

Les titres rapportent de l'intérêt, la monnaie n'en rapporte pas. Keynes pose donc la question : « Why should anyone outside a lunatic asylum wish to use money as a store of wealth ? »[2]. Il distingue quatre motifs :

1) Le *motif de transaction*. Une encaisse minimale est nécessaire pour effectuer les transactions courantes. Cela est vrai des entreprises et de l'Etat aussi bien que des particuliers.
2) Le *motif de précaution*. Il faut garder une réserve pour faire face à des dépenses imprévues pouvant survenir à tout moment. Conserver cette réserve sous forme de titres serait extrêmement risqué car à cause de l'incertitude sur les taux d'intérêt futurs, on ne sait jamais à quel prix ces titres pourraient être revendus.
3) Le *motif financier*[3]: Avant d'installer et de payer ses nouveaux équipements, l'entrepreneur-investisseur passe par un stade de récolte de fonds. Il met au point une méthode de financement en collaboration avec le marché des capitaux ou avec son banquier. Les fonds sont rassemblés, au moins partiellement, avant d'être dépensés ; ils cessent d'être présents chez cet entrepreneur dès qu'il a payé ses fournisseurs mais macroéconomiquement, le fonds existe en permanence et se déplace d'une entreprise à l'autre.
4) Le *motif de spéculation*. « Lorsqu'une personne pense que les valeurs futures du taux d'intérêt seront supérieures aux valeurs prévues par le marché, elle a un motif pour garder actuellement de l'argent liquide »[4].

Sans l'incertitude, les motifs deux et quatre n'existeraient pas. Si on savait quand surviendrait la dépense imprévue, il serait préférable de garder sa réserve sous la forme d'une créance de durée adéquate. De même si on savait avec certitude la survenance et la date de la variation du taux. Mais avec l'incertitude, la conservation de ces réserves sous forme de titres risque d'entraîner une perte de valeur lors de la vente du titre, ce qui justifie la détention de monnaie.

La demande de monnaie, ainsi justifiée, peut être divisée en deux compartiments :

---

[1] Keynes [186] p. 178
[2] Keynes [183] p. 216
[3] Ce motif a été ajouté dans un article publié en 1937.
[4] Keynes [186] p. 181

- la demande de monnaie de TRANSACTION, découlant des trois premiers motifs. Indépendante du taux d'intérêt, elle est fonction principalement du revenu de l'agent si on se place au niveau individuel et du revenu national si on se place au niveau macroéconomique. L'encaisse de transaction égale le revenu divisé par la vitesse de rotation de la monnaie dans les transactions, qui est une constante à court terme. Elle est notée $L_1(Y)$.
- La demande de monnaie de SPECULATION, au contraire, dépend de la relation entre le taux d'intérêt courant et l'état de la prévision. Comme l'indique le graphique 5.2, elle baisse quand le taux d'intérêt (i) augmente. L'explication de Keynes s'appuie sur la diversité des opinions des agents concernant l'évolution future des taux d'intérêt. Avec un taux d'intérêt plus bas, plus d'agents à penseront que ça vaut la peine de détenir des encaisses. Elle est notée $L_2(i)$.

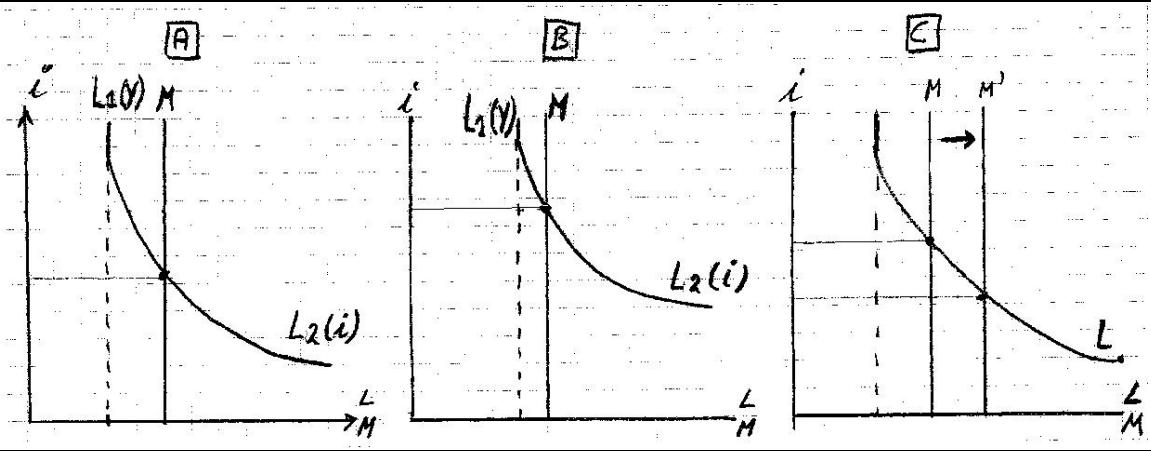
Sur le graphique 5.2-A, la demande d'encaisse de transaction est la ligne verticale pointillée, parfaitement inélastique par rapport au taux d'intérêt. A sa droite, la courbe descendante $L_2(i)$ représente la demande d'encaisse de spéculation. La somme horizontale des deux donne la demande de monnaie totale qui doit s'égaliser avec l'offre de monnaie. L'offre de monnaie est verticale parce qu'indépendante du taux d'intérêt. C'est la courbe *M*. La Théorie Générale est muette à son sujet ; sans doute résulte-t-elle du pouvoir décisionnel des autorités monétaires, comme chez les classiques. L'équilibre monétaire, correspondant à l'intersection de l'offre et de la demande de monnaie, est donné par l'équation :

$$M = L_1(Y) + L_2(i) \qquad (5.3)$$

Cet équilibre monétaire détermine le taux d'intérêt des titres, qui s'établira au niveau précis qui dissuadera les épargnants de détenir des encaisses supérieures à la masse monétaire. Si les agents, pris individuellement, se fondent sur le taux d'intérêt pour répartir leur patrimoine entre les encaisses et les titres, au niveau macroéconomique, la masse monétaire étant fixée, c'est le taux d'intérêt qui doit s'ajuster.

La figure B montre que lorsque la masse monétaire reste constante, un accroissement de la demande de monnaie élève le taux d'intérêt. La figure C montre que lorsque la demande de monnaie reste constante, un accroissement de l'offre abaisse le taux d'intérêt.

Graphique 5.2 : l'équilibre monétaire et le taux d'intérêt selon Keynes



On notera la ressemblance superficielle de ce graphique avec le graphique 4.22 de Pigou. La différence essentielle vient de ce que la variable en ordonnée est ici le taux d'intérêt et non le niveau des prix. Pour les classiques, la monnaie n'influençait pas les grandeurs réelles comme le taux d'intérêt mais seulement les prix.

Keynes analyse en détail les conséquences d'un accroissement de la masse monétaire. Il distingue trois cas :

1) Tout l'accroissement est absorbé par l'encaisse de transaction. Rappelons que Keynes réfute l'idée que les banques, en créant de la monnaie, puissent financer des investissements sans épargne correspondante des ménages. L'augmentation de *M* alimente nécessairement un accroissement du revenu des agents fournissant des services aux entreprises emprunteuses auprès des banques. L'encaisse de transaction des ménages haussera puisqu'elle est fonction de leur revenu. Les ménages affectent une fraction (donnée) du revenu supplémentaire à l'achat de titres, ce qui poussera leur valeur à la hausse et alimentera la trésorerie des entreprises. Le taux d'intérêt baisse donc[1]. Comme les agents sont confiants que cette baisse est durable, aucun d'eux n'accroîtra son encaisse de spéculation. Bref, l'encaisse de transaction a absorbé la totalité de la hausse de *M* et le taux d'intérêt a baissé. C'est le scénario idéal.
2) Très forte préférence pour la liquidité. Ce cas est exactement l'inverse du précédent. Un facteur essentiel est la confiance du public dans l'autorité monétaire. Si le public doute que celle-ci parvienne à réaliser durablement son objectif de réduire les taux d'intérêt, et s'il y a unanimité à ce propos, les agents conserveront en encaisse de spéculation toute la partie de l'accroissement monétaire qu'ils ne consomment pas directement. Ils n'achèteront pas de titres et le taux d'intérêt ne diminuera pas. Ce genre de situation a plus de chances de se produire lorsque les taux en vigueur sont déjà bas. C'est le scénario catastrophe.
3) Le taux d'intérêt baisse et l'encaisse de spéculation augmente. C'est la situation intermédiaire. Une fraction du public pense que la baisse du taux d'intérêt pourrait n'être qu'éphémère et préfère gonfler son encaisse de spéculation, alors qu'une autre fraction, plus confiante, achète des titres. Le taux d'intérêt baissera, mais moins que dans le premier cas. C'est le cas le plus probable.

7- Retour sur la question du salaire nominal

Concernant le salaire *réel*, Keynes le considère comme parfaitement flexible. Il adhère au postulat classique selon lequel il dépend de la productivité marginale du travail, et que, donc, en dehors du progrès technique, sa baisse est nécessaire pour qu'augmentent la production et l'emploi. Par contre, concernant le salaire *nominal*, au début de l'ouvrage, comme nous l'avons vu, il postulait sa rigidité. Mais vers la fin de l'ouvrage, dans le fameux chapitre 19, il prétend que cette rigidité n'était qu'une hypothèse utile à l'exposé de sa théorie, mais que sa théorie est compatible avec un salaire nominal flexible aussi bien que rigide. C'est ce chapitre qui a fait couler le plus d'encre et dont l'exposé, pas toujours très clair, a donné lieu à des interprétations divergentes.

Keynes croit utile de rappeler les bases de la théorie classique du lien entre le salaire et l'emploi, réputée pour son leitmotiv que les baisses de salaire sont à même d'enrayer le déclin de l'emploi. Il existe plusieurs versions de cette théorie, mais le raisonnement de base est que la baisse du salaire nominal se répercute sur les prix, ce qui revigore la demande globale grâce à quoi celle-ci évite la chute qui aurait contrecarré l'effet de l'augmentation de l'offre globale.

Keynes reconnaît que lorsque la demande effective (mesurée en monnaie) reste constante, une réduction des salaires nominaux s'accompagne d'un accroissement de l'emploi ; la question est précisément de savoir si elle pourra se maintenir à son niveau antérieur. Certes, une industrie particulière peut (éventuellement) réduire ses salaires sans voir sa demande affectée, mais cette situation ne peut pas être transposée à l'économie globale. Il écrit : « nous avons démontré en

[1] Celui qui ACHETE des titres OFFRE du capital. Le taux d'intérêt est le prix du capital et non des titres ; il diminue quand le prix des titres augmente.

effet que le volume de l'emploi est associé par une relation biunivoque au montant de la demande effective mesurée en unités de salaire et que la demande effective, étant la somme de la consommation et de l'investissement, ne peut pas varier si la propension à consommer, la courbe d'efficacité marginale du capital et le taux d'intérêt demeurent tous inchangés »[1]. Selon son influence sur ces trois facteurs, la baisse du salaire nominal affaiblira la demande globale ou ne l'affaiblira pas.

Décortiquant cette influence, Keynes attribue les effets uivants à la baisse du salaire nominal :

- Un premier effet négatif : le transfert de pouvoir d'achat des salariés vers des classes plus aisées va réduire la propension à consommer
- Un deuxième effet négatif : la baisse des prix va défavoriser les débiteurs et donc les entreprises.
- Un effet sur l'efficacité marginale du capital, qui sera positif si la baisse du salaire est perçue comme ponctuelle et qui sera négatif si elle fait anticiper des baisses futures en cascade.
- Un effet positif : la baisse des prix réduit l'encaisse nécessaire aux transactions, libérant des liquidités pour l'encaisse de spéculation. Le taux d'intérêt baissera, ce qui favorisera l'investissement.

La conclusion est donc : « A la lumière de ces considérations, nous estimons que le maintien de la stabilité du niveau général des salaires nominaux constitue, tout bien pesé, la politique la plus sage pour un système fermé »[2].

A propos du dernier effet cité ci-dessus, les adversaires de Keynes s'en empareront comme d'une arme favorable à leur théorie, qui sera paradoxalement baptisée *effet Keynes*[3].

8- L'équation quantitative

Keynes s'intéresse à la théorie quantitative résumée par l'équation de Fisher : M.V = P.T où, *T* étant considéré comme fixe, les variations de *M* entraînent des variations proportionnelles de *P*. La vision de Keynes diffère de ce schéma. A priori, on pourrait s'attendre à ce qu'en-dessous du plein emploi, ce soit *T* qui varie proportionnellement à *M* et qu'au plein emploi, *P* prenne le relai et varie en proportion de *M*.

La réalité est plus nuancée, A cause d'un ensemble de facteurs, l'augmentation de *M* entraînera une hausse conjointe de *T* et de *P* bien avant le plein emploi. Plus on se rapprochera du plein emploi, plus l'effet sur *P* l'emportera sur l'effet sur *T*. Comme le dit Keynes, les prix « montent progressivement, à mesure que l'emploi augmente ». Les facteurs favorisant la hausse des prix avant le plein emploi sont :

- La variation de la demande effective pourra être plus faible que celle de la masse monétaire. Ce facteur dépend largement de la variation du taux d'intérêt.
- Les rendements d'échelle décroissants nécessiteront que toute hausse de la production s'accompagne d'une hausse de prix.

[1] Keynes [186] p. 264
[2] Keynes [186] p. 273
[3] Au sujet de cet effet, Keynes estime qu'on peut arriver au même résultat avec une monnaie fixe et des salaires nominaux en baisse ou avec une croissance de la masse monétaire et des salaires nominaux rigides. Il ajoute : « Il faudrait être insensible à l'équité pour préférer une politique de salaires souple à une politique souple de la monnaie, si l'on ne peut invoquer en faveur de la première aucun avantage qui ne puisse être obtenu de la seconde » (Keynes [186] p. 272).

- Les ressources nécessaires à la production des différentes branches ne sont pas interchangeables, d'où des goulots d'étranglements dans certains secteurs quand d'autres sont encore loin du plein emploi.
- Le salaire nominal tend à augmenter avant le plein emploi.

5.1.2. Le statut de la monnaie

Le système classique était dichotomique. La monnaie n'avait aucun impact réel ; un accroissement de *M* ne pouvait aboutir qu'à une augmentation de *P*. En déplaçant la détermination du taux d'intérêt de la sphère réelle vers la sphère monétaire, Keynes a mis la monnaie au centre du système économique et rompu la dichotomie classique. La monnaie cesse d'être neutre : la préférence pour la liquidité peut être source de dépression, mais l'économie peut également être relancée par la monnaie.

Après avoir reconnu à la monnaie un rôle essentiel, Keynes, paradoxalement, doute de l'efficacité de la politique monétaire. C'est principalement le « scénario catastrophe » du §6 qui l'inquiète : « il se peut que, une fois le taux d'intérêt tombé à un certain niveau, la préférence pour la liquidité devienne virtuellement absolue, en ce sens que presque tout le monde préfère l'argent liquide à la détention d'une créance qui rapporte un taux d'intérêt aussi faible. L'autorité monétaire perd alors la direction effective du taux de l'intérêt »[1]. Les successeurs de Keynes appelleront cette situation la *trappe des liquidités*.

*

Le chapitre le plus mystérieux de la Théorie Générale est le chapitre 17 intitulé « Les propriétés essentielles de l'intérêt et de la monnaie ». Keynes semble y jeter les bases d'une nouvelle théorie de la monnaie distincte de celle que nous avons exposée au §6 ; distincte mais non contradictoire.

Le taux d'intérêt propre

Au début des années trente, viennent de paraître « Treatise on Money » de Keynes et « Prices and Production » de Hayek. Ces deux ouvrages apportent des explications opposées sur les fluctuations économiques. Hayek attaque l'ouvrage de Keynes. Celui-ci demande à Sraffa de riposter contre l'ouvrage de Hayek. L'une des idées maîtresses de Hayek imputait les fluctuations économiques aux banques qui prêtent la monnaie aux entreprises à des taux d'intérêt s'écartant du taux naturel (cf. Wicksell). Selon Hayek, dans une économie non monétaire, le taux d'intérêt refléterait toujours le taux naturel et l'économie serait ancrée de façon stable au plein emploi. Dans un article paru en 1932, Sraffa objecte que dans une économie non monétaire, chaque bien susceptible d'être prêté aurait son taux d'intérêt propre et qu'il y aurait des écarts entre eux. Ils ne seraient égaux qu'à l'équilibre. Le taux d'intérêt propre d'un bien dépend de l'écart entre son prix comptant et son prix à terme. L'abondance relative des biens influence ces écarts. Dans ces conditions, un taux d'intérêt naturel unique est une illusion. Ainsi naît le concept de *taux d'intérêt propre*, qui retrouvera une deuxième vie chez les néo-walrassiens, comme nous le verrons au sous-chapitre 6.4.1.

Le taux d'intérêt de la monnaie n'est pas un phénomène extraordinaire. Il n'est rien d'autre que le pourcentage d'excès d'une somme de monnaie livrable à terme par rapport à sa valeur au comptant. Mais un tel concept, « un taux d'intérêt mesuré au moyen de cette richesse elle-même prise comme étalon », est dérivable pour toute richesse durable susceptible d'être prêtée. Comme l'écart entre le prix à terme et le prix comptant est très variable d'une marchandise à l'autre, il ne faut pas s'attendre à ce que le *taux d'intérêt propre* soit le même pour le cuivre, le blé, les machines, la monnaie…

[1] Keynes [186] p. 215

Quelles caractéristiques d'un bien tendent à donner un excès à son prix à terme par rapport à son prix comptant ? Keynes en mentionne trois :
1- ***q*** : le rendement en production ou l'utilité à l'utilisation.
2- ***c*** : le coût de conservation pour contrecarrer un processus naturel de détérioration. Il vient en négatif.
3- ***l*** : la prime qui récompense la liquidité de l'actif.
Par exemple, une maison aura *q* élevé et *c* et *l* faibles. Des bijoux auront *c* élevé et *q* et *l* faibles. La monnaie aura *l* élevé et *c* et *q* faibles.

Le taux d'intérêt propre vaudra q – c + l. En outre, Keynes représente par la lettre *a* l'appréciation du prix d'un bien quelconque par rapport à la monnaie. Comme les agents chercheront à faire entrer dans leur patrimoine en priorité les biens dont q – c + l est élevé, à l'équilibre, la détention des différents actifs apportera un avantage égal. Il y aura donc égalisation des différentes valeurs (a+q-c+l).

Les biens qui seront mis en production sont ceux dont l'efficacité marginale dépasse le taux d'intérêt propre (les deux grandeurs étant calculées dans le même étalon). Keynes suppose une situation où nombre de biens satisfont cette condition et où la société accroît progressivement sa production de ces biens, ce qui réduit leur taux d'intérêt propre et plus encore leur efficacité marginale.

Keynes attribue à la monnaie la propriété que son taux d'intérêt décline très lentement, plus lentement que tous les autres. Il arrivera un moment où le taux d'intérêt de la monnaie sera plus élevé que toutes les efficacités marginales. Le taux d'intérêt trop élevé de la monnaie devient alors un obstacle à l'accroissement de la production.

Les caractéristiques de la monnaie qui freinent la décroissance de son taux d'intérêt sont :
1- Son *élasticité de production* est proche de zéro. Quand la valeur du cuivre augmente, l'industrie tente d'affecter plus de ressources en travail et en capital à sa production. Quand la valeur de la monnaie augmente, un scénario semblable est impossible.
2- Son *élasticité de substitution* est proche de zéro. Quand le prix du blé augmente, les agents tenteront de lui substituer d'autres produits, ce qui en diminuera la demande. Quand la valeur de la monnaie augmente, elle est plus chère à acquérir mais également plus désirable car son pouvoir d'achat augmente. Il n'y a donc pas de tendance à la baisse de sa demande.
3- La liquidité de la monnaie ne diminue pas avec son volume et son coût de conservation augmente peu.

Quand le public demande des autos, on met au travail des ouvriers dans l'industrie automobile. Quand il déplace sa demande vers la monnaie, par un accroissement de la préférence pour la liquidité, il démobilise un pan de l'appareil productif : « le chômage se développe parce qu'on demande la lune. Les hommes ne peuvent être employés lorsque l'objet de leur désir (i.e. la monnaie) est une chose qu'il n'est pas possible de produire et dont la demande ne peut être facilement restreinte »[1].

5.1.3. Est-il possible de combattre le chômage et comment ?

La demande globale se compose de la consommation et de l'investissement. Ses soubresauts funestes trouvent leur cause dans les vicissitudes du second plutôt que dans celles de la consommation. Celle-ci est déterminée mécaniquement à partir du revenu et de la propension à

[1] Keynes [186] p. 242

consommer. Par contre, la hauteur de l'investissement dépend de l'efficacité marginale du capital et du taux d'intérêt, deux facteurs très aléatoires. Le premier est sujet à l'incertitude générale de l'économie et le second subit la volatilité de la préférence pour la liquidité. Keynes met principalement l'accent sur l'efficacité marginale du capital, plus que taux d'intérêt. Il écrit : « Le mieux à notre avis est de considérer le cycle économique comme la conséquence d'une variation cyclique de l'efficacité marginale du capital »[1]. Il ajoute : « la baisse de l'efficacité marginale du capital peut être si profonde qu'aucune réduction possible du taux d'intérêt ne suffirait à la contrebalancer »[2].

Constatant l'énorme fébrilité des marchés financiers, leur instabilité dans la manière de digérer l'information économique, leur « esprit capricieux », il conclut : « Dans un régime de laissez-faire, la suppression des larges fluctuations de l'emploi exigerait un profond changement des conditions psychologiques des marchés financiers, qu'il n'y a aucune raison de prévoir. Nous concluons qu'on ne peut sans inconvénient abandonner à l'initiative privée le soin de régler le flux courant de l'investissement »[3].

Keynes préconise les investissements publics et le soutien aux investissements privés. Grâce au multiplicateur, tout investissement additionnel se ressent positivement tant dans l'industrie des biens d'équipement que dans celle des biens de consommation. En situation de sous-emploi, on ferait construire d'inutiles pyramides, dit Keynes, que le bien-être augmenterait quand même grâce au multiplicateur. Cette adaptation des achats publics à la situation économique s'appelle la *politique budgétaire.* La principale alternative est la *politique monétaire* qui consiste à agir sur le taux d'intérêt par le contrôle de la masse monétaire. Comme nous l'avons vu, Keynes se montre circonspect quant à l'efficacité de cette dernière ; il ne la rejette pas mais la considère comme un second choix.

Keynes s'oppose à l'idée d'une politique contracyclique du taux d'intérêt. Selon lui, la hausse des taux, qui pourrait tenter les autorités en situation de tension conjoncturelle, consisterait à s'attaquer au malade plutôt qu'à la maladie. La tension conjoncturelle est due aux excès de spéculation et aux illusions des marchés financiers qui orientent mal les investissements, là où l'Etat, plus raisonnable, garderait la tête froide et viserait l'intérêt de la communauté. En cas de boom, la meilleure politique consiste à renforcer l'attrait de la consommation par rapport à celui de l'investissement, par exemple en augmentant les revenus des classes moins favorisées.

Quant au taux d'intérêt, il doit rester le plus bas possible en toutes circonstances. Keynes souhaite d'ailleurs la disparition de la classe des rentiers. Il se dit préoccupé par les deux vices du monde économique moderne : le chômage et le manque d'équité dans la répartition de la fortune et du revenu. « Pour notre part, nous pensons qu'on peut justifier par des raisons sociales et psychologiques de notables inégalités dans les revenus et les fortunes, mais non des disproportions aussi marquées qu'à l'heure actuelle »[4]. Le taux d'intérêt bas n'empêcherait pas la formation de l'épargne nécessaire aux investissements, car celle-ci est indépendante du taux d'intérêt.

Keynes reste attaché à l'initiative et à la propriété privées, mais attend de l'Etat qu'il encadre l'investissement. « Hors la nécessité d'un contrôle central pour maintenir en équilibre la propension à consommer et l'incitation à investir, il n'y a pas plus de raison qu'auparavant de

[1] Keynes [186] p. 311
[2] Keynes [186] p. 314
[3] Keynes [186] p. 317
[4] Keynes [186] p. 367

socialiser la vie économique »[1]. En fait, la pensée politique de Keynes, très modérée, associait des idées de gauche et de droite ; on peut le qualifier de centriste. Il est peut-être éclairant de savoir qu'il adhérait au *Liberal party*.

5.1.4. La controverse Ohlin - Keynes : *ex-ante* ou *ex-post*

En 1937, Bertil Ohlin publie un article intitulé « Some Notes on the Stockholm Theory of Savings and Investment », qui, comme son nom ne l'indique pas, est une réponse à la Théorie Générale. L'article s'articule sur deux idées : d'abord attirer l'attention des économistes sur l'existence d'une école méconnue qui, sans attendre la révolution keynésienne, avait quitté les chemins de la stricte orthodoxie pour analyser les problèmes macroéconomiques. « There are surprising similarities as well as striking differences between that apparatus and the conclusions reached in Sweden on the one hand and Mr. Keynes' General Theory »[2]. Ensuite, l'article jette un regard critique sur la nouvelle théorie keynésienne.

Les principaux économistes de Stockholm sont Lindahl, Myrdal et Ohlin lui-même. L'Anglais Robertson est dans leur sillage. Wicksell est leur père spirituel, mais la génération suédoise de la grande crise va bien au-delà de ses théories.

La vision suédoise de l'économie se veut dynamique. Lindahl est le pionnier d'une méthode qui fera florès, consistant à découper du temps en une succession de périodes, pour analyser comment les grandeurs économiques évoluent. Dans cette optique, la distinction entre les variables *ex-post* et les variables *ex-ante* est centrale. L'analyse *ex-post* est un peu comme le livre comptable enregistrant les valeurs prises effectivement par les variables au cours d'une période (forcément terminée). L'analyse *ex-ante* est plus comportementaliste et « forward-looking » : elle porte sur les plans d'action que les agents économiques élaborent pour la période en cours ou les périodes futures. Ces plans sont basés sur la connaissance du passé et l'anticipation du futur.

Comment s'imbriquent les facteurs ex-ante et ex-post dans le domaine de la relation épargne-investissement ? Notons par I_a l'investissement planifié et S_a l'épargne planifiée (données ex-ante) ; « There is no reason for assuming that they should be equal ». I_p et S_p sont respectivement l'investissement et l'épargne ex-post pour la même période ; ils sont nécessairement égaux entre eux. Comment passe-t-on d'une possible inégalité à une nécessaire égalité ? « The answer is that the inequality of S_a and I_a sets in motion a process which makes realized income differ from expected income, realized savings from planned savings and realized new investment differ from the corresponding plan »[3]. Ce sont le revenu, l'épargne et l'investissement inattendus (ou inintentionnels) qui font la différence. Ils sont notés Y_u, S_u, I_u. On a :

$$S_a \neq I_a \qquad (5.4)$$
$$[S_a + S_u = S_p] \text{ et } [I_a + I_u = I_p]$$
$$S_p = I_p$$

Au cours d'une période *n*, on peut assister à l'augmentation de l'investissement ex-post par rapport à la période *n-1*, même si l'épargne planifiée de la période *n* est au même niveau que l'épargne ex-post de la période *n-1*.

---

[1] Keynes [186] p. 372
[2] Ohlin [271] p. 53
[3] Ohlin [271] p. 65

Ohlin récuse la thèse classique selon laquelle c'est le taux d'intérêt qui égalise l'épargne et l'investissement. D'une part, l'épargne et l'investissement ex-ante peuvent diverger. D'autre part, l'épargne et l'investissement ex-post s'équivalent, quel que soit le taux d'intérêt. Ohlin rejette également la théorie wicksellienne du taux d'intérêt naturel. Quelle est alors l'explication de la détermination du taux d'intérêt ? « The rate of interest is simply the price of credit, and it is therefore governed by the supply of and demand for credit »[1].

La théorie suédoise conteste également le principe orthodoxe liant le salaire réel et l'emploi dans le sens que toute augmentation de l'un postule ou entraîne la diminution de l'autre. Ohlin remarque sarcastiquement qu'on le retrouve même dans la *Théorie Générale*. Selon lui, aucune théorie statique ne peut expliquer le chômage conjoncturel.

Venons-en maintenant aux commentaires d'Ohlin à l'égard de la Théorie Générale. La même année 1937, Vassili Leontief avait déploré que Keynes s'écartait de la tradition macroéconomique en ne respectant pas le *postulat d'homogénéité*, un concept que Leontief mit en avant pour désigner une hypothèse commune mais implicite jusqu'alors : que l'offre et la demande des différents biens particuliers sont insensibles au niveau des prix ABSOLUS ; seuls comptent les prix RELATIFS. Or l'offre de travail keynésienne dépend explicitement du salaire NOMINAL. Ohlin fait le même constat… avec satisfaction ; Keynes s'est heureusement débarrassé d'un ensemble de postulats tacites de l'économie classique, parmi lesquels le postulat d'homogénéité, qui mettent hors jeu les variations cycliques de la production et les processus d'inflation et de déflation. Une théorie macroéconomique digne de ce nom considère les effets réels et les effets nominaux d'un événement comme deux objets complémentaires dignes d'attention. L'analyse keynésienne satisfait cette exigence et peut dès lors servir de base à l'étude du cycle conjoncturel.

Ces éloges sont toutefois assortis de critiques. Comme celle des classiques, la théorie keynésienne reste une théorie d'équilibre, ce qui la rend « old fashioned ». Face à ce statisme, Ohlin préconise une vision dynamique, seule capable d'expliquer le cycle conjoncturel[2].

Rapidement, il apparaît que la principale pierre d'achoppement est la supposée incapacité de Keynes à distinguer l'ex-ante et l'ex-post. Plus précisément, à reconnaître l'existence de l'épargne et l'investissement ex ante. Effectivement, Keynes s'y refuse ; dans sa réponse, il rappelle son hostilité envers toute théorie de l'intérêt fondée sur l'offre et la demande de fonds prêtables et s'en tient à la thèse de l'inévitable égalité entre I et S. Il conteste par ailleurs que la conception de Ohlin innove par rapport à la théorie des fonds prêtables traditionnelle.

L'épargne ex-ante semble à Keynes être un concept vide de sens : avant de percevoir un revenu et de s'abstenir du consommer, l'épargnant n'effectue aucune démarche particulière. Par contre, l'investissement ex-ante est un facteur très important ; Keynes dit maintenant regretter de l'avoir négligé précédemment. Avant d'installer et de payer ses nouveaux équipements, l'entrepreneur-investisseur passe par un stade de récolte de fonds. C'est à cette occasion que Keynes introduit le **motif financier** de demande de monnaie, absent dans la Théorie générale. Comme toute demande de monnaie, celle-ci fait monter le taux d'intérêt si la masse monétaire ne suit pas le mouvement. D'où l'illusion, fausse, que l'augmentation de l'investissement fait hausser le taux d'intérêt parce qu'il est une demande de fonds prêtables.

[1] Ohlin [271] p. 221

[2] Comme Ohlin, Leontief s'était également étonné du caractère statique de la théorie keynésienne.

Ohlin accepte l'idée keynésienne que le taux d'intérêt égalise l'utilité marginale de la détention de titres avec celle de la détention de monnaie. Mais il estime que ce rôle n'est pas contradictoire avec celui que le taux d'intérêt joue pour égaliser l'offre et de la demande de fonds prêtables ; il lui est complémentaire. Ohlin conçoit ces deux ajustements dans le cadre d'un équilibre général où comme les autres prix, le taux d'intérêt influence les diverses offres et demandes. Il rejette la conception keynésienne qui met le marché de la monnaie sur un piédestal, au-dessus de l'équilibre général, agissant comme unique déterminant du taux d'intérêt. Mais Keynes réinsiste sur le caractère unique du marché de la monnaie : « The investment market can become congested through shortage of cash. It can never become congested through shortage of saving. This is the most fundamental of my conclusions within this field »[1].

5.1.5. Le salaire et la marge face au cycle conjoncturel

Comme nous l'avons vu au sous-chapitre 5.1.1., Keynes adhérait à la conception dominante de son époque concernant l'évolution du salaire réel. En basse conjoncture, la baisse de la production fait croître la productivité marginale et donc le salaire réel, même si le salaire nominal est en baisse. A l'opposé du salaire nominal, le salaire réel évolue contracycliquement. Marshall semble être à l'origine de cette idée. Peu après la parution de la Théorie Générale, deux articles publiés l'un par Dunlop et l'autre par Tarshis relèvent une corrélation statistique positive entre le salaire nominal et le salaire réel et remettent en cause l'ancienne conception. En 1939, Keynes consacre à cette question un article intitulé « Relative Mouvements of Real Wages and Output ». Il reconnaît avoir eu tort d'adhérer à l'opinion dominante, mais – ajoute-t-il- la nouvelle vérité est plutôt favorable à la Théorie Générale, car elle montre que l'amélioration de l'emploi n'implique pas nécessairement une baisse du salaire réel. Diffusés plus tôt, ces éléments neufs auraient permis à la Théorie Générale de simplifier son argumentation.

La suite de l'article de Keynes est exploratoire. En l'absence de données empiriques complémentaires, il est difficile de déterminer avec précision les fautes de raisonnement de l'ancienne théorie, mais Keynes se hasarde à certaines hypothèses qu'il souhaiterait voir testées. Son analyse se situe à la frontière entre la micro et la macroéconomie, une zone arpentée par de nombreux économistes face à la crise économique persistante. Keynes se demande si l'hypothèse de la courbe de coût marginal réel croissante n'est pas en cause. En tout cas – constate-t-il- un économiste comme Kalecki préfère postuler une courbe horizontale. Keynes émet également des doutes sur la théorie qui lie le prix au coût marginal ; ne serait-il pas plutôt guidé par le coût moyen à long terme ? Ses réflexions à ce sujet l'amènent à des idées proches de la thèse hétérodoxe du *full cost* qui sera analysée au chapitre 5.5.

Keynes se réfère également à des statistiques montrant que la part des salaires dans le revenu national est étonnamment stable dans le temps, par-delà les fluctuations conjoncturelles. S'en trouve affaiblie la thèse couramment acceptée que la marge de profit incluse dans le prix évolue procycliquement. Une thèse qu'on retrouvait notamment dans l'ouvrage de Harrod « The Trade Cycle » (1936). Il y développe sa théorie de *l'élasticité de la demande décroissante* : lorsque leur revenu augmente, les agents sont moins sensibles à l'avantage d'acheter au meilleur prix ; la demande est donc plus élastique pendant la dépression que pendant le boom conjoncturel ; la marge bénéficiaire y est donc supérieure, comme l'indique l'équation (4.3). A peine publiée, cette théorie est combattue par Joan Robinson et par Kalecki. Tous deux avancent des

[1] Keynes [182] p. 669

arguments favorables à une évolution contracyclique de la marge bénéficiaire[1]. La concurrence devient plus rude lors du boom, d'une part parce que la survie de la firme est moins menacée, d'autre part parce que l'expansion des marchés suscite l'entrée de nouvelles firmes. Cette concurrence accrue y comprime les marges. Avec beaucoup de précautions, Keynes prend le parti de Kalecki et Robinson.

Kalecki et les parts du travail et du capital dans le revenu national

L'article de Kalecki « The Determinants of Distribution of National Income » (1938) s'intéresse également à ce fait empirique de la constance des parts des classes sociales dans le revenu national. Selon Kalecki, la part du capital est une fonction croissante de deux variables : le degré de monopole de Lerner lié à la part de la marge dans le prix et le rapport du prix des matières premières sur les salaires. Comme nous l'avons vu, le premier est supposé contracyclique alors que le second est procyclique. La constance des parts découle de la compensation entre ces effets opposés. Keynes salue l'intérêt de cette analyse mais reproche à Kalecki de ne pas expliquer pourquoi les deux effets se compensent presque parfaitement.

[1] Encore aujourd'hui, les économistes sont divisés sur la forme cyclique des marges. Même les données empiriques avancées par les uns et les autres ne concordent pas. On pourrait en conclure que la forme cyclique des marges est sensible aux circonstances, qu'elle peut varier selon le lieu et le temps. Dans l'article « Market Structure and Macroeconomic Fluctuations » (1986), Robert Hall insiste sur un autre phénomène, dont la constatation empirique est assez largement acceptée : le caractère procyclique de la productivité. Il écrit : « The fact that drives all my results is well known and uncontroversial : for many industries and all broad aggregates, output can rise substantially with only a modest increase in measured labor input » ([129] p. 304). Il en impute la cause au *pouvoir de marché*, très répandu, qui autorise certaines pratiques telles la *rétention de main d'œuvre* (labor hoarding), consistant à garder une parte de l'effectif sous-utilisé lorsque l'activité est réduite. Ce phénomène devrait favoriser une évolution procyclique des marges.

5.2- SRAFFA REMET RICARDO A L'HONNEUR

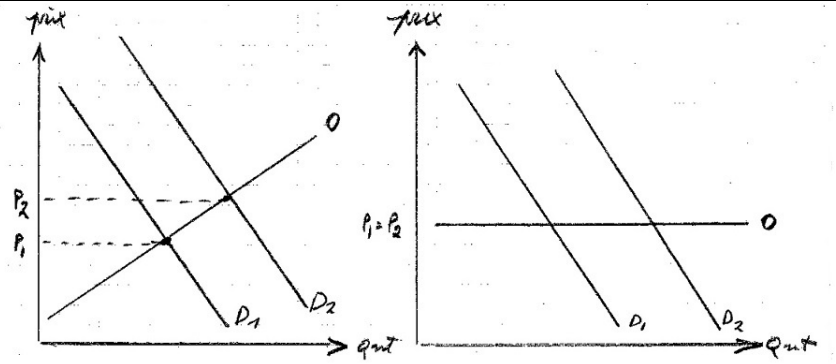
L'Italien Piero Sraffa rallie l'Angleterre au début des années vingt. Dès cette époque, il entre en contact avec Keynes et il deviendra avec lui un pilier de l'école cambridgienne.

5.2.1. Le parcours de Sraffa

Le premier fait d'arme de Sraffa est sa critique de la théorie marshallienne de l'offre, exposée dans deux articles publiés en 1925 et 1926.

La théorie marshallienne associe la courbe d'offre à celle du coût marginal. Cette courbe est croissante parce que la productivité marginale décroît avec la quantité produite. Sraffa conteste cette optique : « the idea of interdependence between quantity produced and the cost of production of a commodity produced under competitive conditions is not suggested by experience at all and could not arise spontaneously »[1]. D'ailleurs, les auteurs classiques considéraient le coût comme indépendant de la quantité et Sraffa se range à leur point de vue. En fait la théorie néoclassique a besoin de cette relation fonctionnelle pour prouver que l'utilité et la demande influencent le prix. Comme le montre le graphique 5.3, lorsque le coût unitaire est constant et la courbe d'offre plate, le coût est seul à déterminer le prix, sans interférence de la demande.

Graphique 5.3 : le prix selon que le coût soit croissant ou constant



Sraffa critique abondamment la théorie de la productivité marginale et s'attaque à celle des économies d'échelle, deux piliers de l'économie marshallienne.

Marshall s'était déjà rendu compte que les économies d'échelle internes étaient incompatibles à long terme avec la concurrence. Pour concilier ce fait avec l'expérience qui révèle des rendements d'échelle souvent croissants, il inventa la notion d'*économies externes*, un phénomène compatible avec la concurrence. Sraffa doute que la réalité offre beaucoup d'exemples de telles économies. Ce concept lui semble parfaitement artificiel.

Sraffa critique également à la méthode marshallienne de l'*équilibre partiel*. Les courbes d'offre et de demande impliquent l'hypothèse TOUTES AUTRES CHOSES RESTANT EGALES. Le prix d'un produit donné est censé ne dépendre que de son marché et devrait donc être insensible aux variations de l'offre et de la demande des autres produits. Ces restrictions sont incompatibles avec les coûts croissants. Par exemple, l'offre croissante d'un produit fait hausser le prix d'un facteur qui y intervient mais qui intervient également dans un autre secteur d'activité. La hausse du coût dans ce deuxième secteur y élève le prix. Si ce secteur produit un substitut du premier

[1] Sraffa [344] p. 3

produit, la demande de celui-ci sera affectée. Le coût peut être croissant, constant ou décroissant. Sraffa préconise l'hypothèse méthodologique d'un coût constant, la seule qui permette de théoriser sur la valeur en tenant compte de l'interdépendance générale.

L'article de 1926 révèle une certaine évolution de la pensée de Sraffa ; il préconise maintenant de renoncer à l'hypothèse de la concurrence parfaite. Son analyse anticipe le principe de la concurrence monopolistique qui sera développée par Chamberlin (cf. sous-chapitre 4.1.3).

Cette opinion n'était que provisoire. Rapidement, et cette fois définitivement, il opta pour le renoncement à l'équilibre partiel. Un équilibre général avait déjà été construit par Walras. Mais dès cette époque, Sraffa était convaincu que l'économie néoclassique était une régression par rapport à l'économie classique. Il lui faudrait donc élaborer lui-même un EQUILIBRE GENERAL RICARDIEN. Dès la fin des années vingt, Sraffa avait une idée relativement précise des équations de l'équilibre général ricardien, mais la gestation de l'ouvrage où il fut exposé durera jusqu'en 1960. C'est l'objet de la section suivante.

Entre-temps, Sraffa était engagé dans une tâche gigantesque : l'édition de l'œuvre complète de Ricardo et de toute sa correspondance. Le premier volume ne sortit qu'en 1951. Dans son introduction, Sraffa inaugure une interprétation novatrice de l'œuvre de Ricardo, qui permet à celle-ci, ainsi qu'à toute l'économie classique, de retrouver une deuxième jeunesse. Une nouvelle école naît à partir de la lecture sraffienne de Ricardo : les néo-ricardiens.

5.2.2. La production de marchandises par les marchandises

En 1960, Sraffa publie l'œuvre de sa vie : « Production of Commodities by Commodities » avec le sous-titre « Prelude to a Critic of Economic Theory ». Contrairement à ce que le sous-titre suggère, cet ouvrage vise plus à ébaucher une théorie alternative qu'à attaquer la théorie néoclassique.

L'EQUILIBRE GENERAL RICARDIEN

Que signifie le titre du livre ? Il traduit une caractéristique de l'économie classique : la production est considérée comme un flux circulaire. Le produit est divisé en deux parties :

- celle qui fait une boucle pour revenir dans la production : ce sont les marchandises qui produisent les marchandises (pensons aux biens d'investissement et aux matières premières). Ce processus est peu présent dans les modèles néoclassiques que nous avons examinés ci-dessus.
- le « surplus », qui sort du processus productif pour être consommée improductivement. Comme il s'agit d'un modèle stationnaire, la part du surplus qui est réinvestie n'est pas envisagée.

Dans la vision des économistes classiques du XIXe siècle, le surplus ne comportait que les produits de luxe et se retrouvait intégralement dans le profit capitaliste ou la rente foncière. Les biens de consommation ouvrière étaient comptés dans les marchandises qui produisaient des marchandises, comme de l'essence qu'on mettrait dans un moteur. Il est frappant que les écrits du XIXe siècle reprennent souvent le blé parmi les biens entrant dans le capital. A l'époque où le salaire a dépassé le minimum vital, au moins une part celui-ci devrait être comptée dans le surplus. Pour simplifier, c'est la totalité du salaire que Sraffa intègre dans le surplus où il rejoint le profit.

Suivant une méthode pédagogique courante, Sraffa expose d'abord un modèle comportant de nombreuses hypothèses simplificatrices ; par la suite, il les lèvera une à une. Commençons par le modèle simplifié, dont voici les hypothèses :

- Les biens entrent dans la production exclusivement comme capital CIRCULANT.
- Il n'y a pas de *production jointe*.
- Le travail est le seul facteur primaire, la terre étant ignorée.
- Une seule technique de production est possible pour chaque bien (coefficients techniques fixes, qui excluent la substitution).

Aucune hypothèse n'est posée quant aux rendements d'échelle. Enfin, le modèle suppose que les salaires sont payés après le processus de production, qu'ils ne doivent donc pas être avancés par le capital.

Construisons maintenant le système d'équations. Il y a *k* biens numérotés de *a* à *k* ; A,B… sont les quantités produites annuellement de *a*,*b*..., qui sont données ; Les coefficients de production, qui sont donnés, ne sont pas unitaires comme chez Walras ; par exemple, K_a est la quantité de *k* absorbée par la production annuelle de *a*. L_a est la quantité de travail absorbée par la production annuelle de *a*[1] ; p_a est le prix de *a* ; *w* et *r* sont respectivement le salaire réel et le taux de profit, tous deux uniformisés par la concurrence.

Pour chacun des *k* biens, on a une équation expliquant la valeur totale de sa production comme précisé ci-dessus :

$$(A_a.p_a + B_a.p_b + ... + K_a.p_k)(1+r) + L_a.w = A.p_a \qquad (5.5)$$
$$\vdots$$
$$(A_k.p_a + B_k.p_b + ... + K_k.p_k)(1+r) + L_k.w = K.p_k$$

A ces *k* équations, Sraffa en ajoute une (k+1)ème :

$$[A-(A_a+A_b...+A_k)].p_a + ...+[K-(K_a+K_b...+K_k)].p_k = 1 \qquad (5.6)$$

Le revenu national vaut l'unité. C'est une manière de normer le système. Le numéraire n'est pas tel ou tel bien mais l'ensemble de la production sociale.

Il y a (k+1) inconnues : les *k* prix ainsi que *r* ; *w* est considéré comme connu[2]. Les équations sont également au nombre de (k+1) [3]. Le système a donc une solution.

Nous constatons que :

1- Seuls les prix sont des inconnues : les quantités produites sont données de façon exogène. Nous reviendrons plus loin sur cette différence importante avec le modèle walrassien.
2- Le taux de profit est une inconnue sur le même pied que les prix des marchandises. Il ne peut être déterminé avant le prix des biens, puisque la détermination du taux de profit nécessite que le capital (hétérogène) ait une valeur, ce qui est impossible en l'absence d'un système de prix. Parallèlement, le prix des biens ne peut être déterminé sans que soit connu le taux de profit puisque ce taux affecte le coût de production. Le génie de Sraffa fut donc de lier le sort des prix relatifs des biens et du taux de profit dans un système d'équations simultanées.

[1] Plus précisément, L_a est la part du travail total consacrée à la marchandise a : $L_a + \ldots + L_k = 1$.

[2] Mathématiquement, on peut considérer indifféremment que *r* est connu et *w* inconnu ou l'inverse. Economiquement, il est plus sensé de considérer *r* comme inconnu, car *r* dépend de la valeur du capital qui elle-même dépend des prix qui sont inconnus.

[3] Ce n'est pas comme chez Walras : tous les prix sont des inconnues car aucun bien ne sert de numéraire. De plus, les présentes équations concernent les coûts de production dans les *k* industries et sont indépendantes, contrairement à celles qui égalisent l'offre et la demande, unies par la contrainte budgétaire.

La deuxième remarque est importante car elle montre que Sraffa a corrigé la principale erreur de Marx. Revoyons le système d'équations (2.16) de Marx : le taux de profit y est déterminé avant les prix, sur base du taux de plus-value dont nous savons le manque de consistance. Le système d'équations sraffien (5.5) est une balance carrée. Il assure que les prix des composants constituant le coût sont réglés suivant la règle générale de formation des prix. Cette précaution est absente du système (2.16) de Marx ; les évaluations du capital constant de *A*, *B* et *C*, respectivement 80, 90 et 70 sortent du néant ; quelle règle respectent-elles ?

Si l'on regarde le système d'équations (5.5), on constate que tous les biens n'y jouissent pas de la même influence. Sraffa opère cette distinction :

- Les biens *non fondamentaux* : ils ne servent pas dans la production des autres biens. (Les produits de luxe par exemple). Si *b* est un produit de ce type, $B_a \ldots B_k$ vaudront zéro ; si un progrès technique divisait par deux les inputs pour produire *b*, p_b serait divisé par deux mais tous les autres prix resteraient inchangés.
- Les biens *fondamentaux* entrent dans la production des autres biens ; une modification de leurs conditions de production affecte tous les prix et le taux d'intérêt.

En fait, Sraffa compte parmi les biens non fondamentaux également ceux qui entrent dans la production des autres biens non fondamentaux. Si on supprime un bien qui entre dans la fabrication de trois autres, le système perd quatre inconnues et quatre équations mais reste déterminé. Il en découle que sont fondamentaux, les biens qui interviennent dans la production de TOUS les biens, directement ou indirectement. Si on supprime un bien fondamental, il n'y a plus de système d'équations.

PROFIT, SALAIRE ET ETALON INVARIANT

Le revenu national est partagé entre le profit[1] π et le salaire *w*. Si $w = 1$, $\pi = 0$ et donc r est également nul ; si $w = 0$, $\pi = 1$ et *r* vaudra le maximum possible que nous désignons par *R* et sur lequel nous reviendrons. Si nous partons d'un salaire nul et que nous l'augmentons progressivement, outre la baisse régulière de *r* et π, il en découlera un chamboulement du système de prix. Ricardo l'avait déjà remarqué : les produits à haute intensité de main d'œuvre verront leur prix augmenter et ceux à haute intensité capitalistique verront le leur décliner. Mais plus précis que son mentor, Sraffa spécifie qu'il faut tenir compte de l'intensité en main d'œuvre dans sa globalité, qui intègre la production du bien lui-même et celle de ses inputs. Un bien à faible intensité de main d'œuvre pourrait subir une hausse du prix si ses principaux inputs sont produits avec une forte proportion de main d'œuvre.

Arrivé à ce point, Sraffa n'est pas encore satisfait car il veut résoudre le problème de la mesure invariante de la valeur, qui avait eu raison de Ricardo. Il voudrait pouvoir exprimer l'ensemble des prix des biens et le salaire dans une unité qui elle-même ne serait pas affectée par un changement dans le rapport entre le salaire et le profit. Suivons Sraffa dans son raisonnement, à vrai dire assez sinueux.

L'astuce de Sraffa est la suivante : entre les haussiers et les baissiers, il doit forcément exister un rapport entre le capital et le travail, que Sraffa appelle « rapport d'équilibre », pour lequel la hausse du salaire laisserait le prix constant. Si un bien satisfaisait à cette condition, on le prendrait comme numéraire et on résoudrait par la même occasion le problème ricardien de la mesure invariante de la valeur. Attention, les couches successives d'inputs devraient répondre à cette même condition.

[1] Le profit est ici le rendement exigé par le marché des capitaux, c'est-à-dire le profit à son taux normal, qui est qualifié d' « intérêt » par d'autres auteurs.

Comment déterminer le rapport critique entre le capital et le travail ? Sraffa démontre que le bien qui satisferait à ce rapport d'équilibre satisferait à une autre condition : son rapport output/ input vaudrait (1+R) où *R* est le taux maximum de profit.

Il y a peu de chances de trouver dans tout le système économique un bien qui satisfasse à notre condition. L'idée de Sraffa est de le « créer », en l'occurrence, en constituant une marchandise composite, comprenant différents biens dans des proportions choisies, comme lorsqu'on compose le panier des biens d'un indice des prix.

Un exemple chiffré fera comprendre la procédure. Partons du système productif ou d'un extrait qui le réplique en miniature :

$$\begin{array}{l} 90\text{ T fer} + 120\text{ T charbon} + 60\text{ qx blé} + 3/16\text{ travail} \rightarrow 180\text{ T fer} \qquad (5.7) \\ 50\text{ T fer} + 125\text{ T charbon} + 150\text{ qx blé} + 5/16\text{ travail} \rightarrow 450\text{ T charbon} \\ \underline{40}\text{ T fer} + \underline{40}\text{ T charbon} + \underline{200}\text{ qx blé} + \underline{8/16}\text{ travail} \rightarrow 480\text{ qx blé} \\ 180 \qquad 285 \qquad 410 \qquad 16/16 \end{array}$$

Le rapport output/input du fer est de 180/180 ; celui du charbon est de 450/285 et celui du blé est de 480/410 ; tous ces rapports sont différents. En arrangeant les quantités produites des différents biens tout en respectant les coefficients techniques, on peut toutefois les égaliser :

$$\begin{array}{l} 90\text{ T fer} + 120\text{ T charbon} + 60\text{ qx blé} + 3/16\text{ travail} \rightarrow 180\text{ T fer} \qquad (5.8) \\ 30\text{ T fer} + 75\text{ T charbon} + 90\text{ qx blé} + 3/16\text{ travail} \rightarrow 270\text{ T charbon} \\ \underline{30}\text{ T fer} + \underline{30}\text{ T charbon} + \underline{150}\text{ qx blé} + \underline{6/16}\text{ travail} \rightarrow 360\text{ qx blé} \\ 150 \qquad 225 \qquad 300 \qquad 12/16 \end{array}$$

Chaque bien est produit dans la proportion qui égalise les rapports de chacun comme output et comme input : 180/150 = 270/225 = 360/300 = 1,2. Cet « extrait » de l'économie que nous avons isolé, que Sraffa nomme le *système étalon*, donne donc bien la marchandise étalon qui comprend 1T de fer + 1,5 T de charbon + 2 qx de blé. Quant au taux maximum de profit, *R*, il vaut 20% puisque tel est pour chaque bien le surplus de l'output sur l'input.

Comment avons-nous constitué le système étalon ? Nous avons appliqué des multiplicateurs aux équations : 1 pour celle du fer, 3/5 pour celle du charbon et ¾ pour celle du blé. Sraffa démontre mathématiquement que pour tout système d'équations du type (5.5), il est toujours possible de déterminer un et un seul ensemble de multiplicateurs $c_a \ldots c_k$ permettant d'arriver au résultat voulu. Toute économie a donc un système étalon et une marchandise composite pouvant servir d'étalon invariant.

Une de ses caractéristiques est qu'elle ne comporte que des biens fondamentaux. Les bien non fondamentaux n'apparaissent jamais du côté gauche des équations et le dénominateur de leur rapport output/ input est donc toujours indéterminé.

<u>A quoi sert cet étalon</u> ? D'abord, si on l'utilise dans le système d'équations (5.5), on a la certitude que la variations d'un prix provient effectivement de modifications inhérents au marché du bien concerné et non de changements dans les conditions de production de la marchandise étalon. Pour ce faire, il faut modifier l'équation (5.6) de façon à rendre unitaire, non le revenu national réel mais le *revenu national étalon* (celui qui prévaudrait si toute la force de travail était utilisée à produire la marchandise étalon).

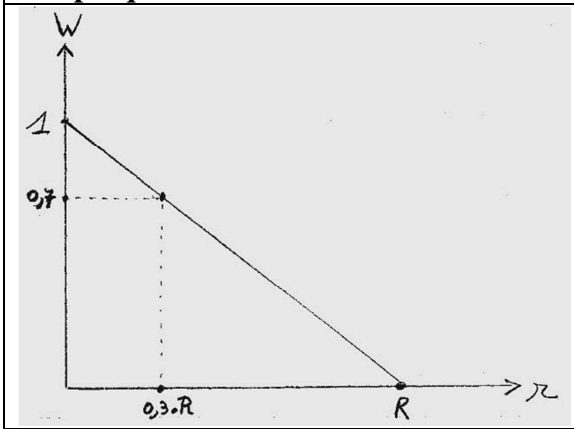
Cet avantage est évidemment très théorique. Pratiquement, il est impossible de construire la marchandise étalon d'une économie développée comportant une multitude de biens fondamentaux.

Mais le système étalon a une autre utilité ; il permet d'établir une relation LINEAIRE entre les salaires et les profits. Si *w* est exprimé en unités du système étalon, la relation entre *w* et *r* est donnée par la formule suivante :

$$r = R.(1 - w) \qquad (5.9)[1]$$

Nous avons donc défini une relation linéairement décroissante entre le salaire et le profit. Dans l'exemple ci-dessus où R valait 20%, supposons que w = 0,75 (autrement dit, ¾ du RN étalon vont aux salaires) ; le taux de profit vaudra le quart de 20% soit 5%.

Graphique 5.4 : la relation linéaire entre le salaire et le taux de profit



L'avantage de la relation (5.9) n'est pas sa linéarité, qui- de toute façon dépend de l'hypothèse que les salaires sont payés en fin de production : sans cette hypothèse, le graphique 5.4 montrerait une hyperbole. L'important est que la relation w/r est indépendante du système de prix. Quels que soient le rapport de prix, la relation w/r est celle donnée par la formule (5.9) et le graphique 5.4. Cette propriété se vérifie, peu importe à quel moment de la production, le salaire est payé.

CAS PARTICULIERS

Sraffa montre ensuite comment son système survit lorsqu'on lui retire ses hypothèses simplificatrices, principalement lorsqu'on introduit d'une part la *production jointe* et d'autre part, la terre en tant que deuxième input non produit outre le travail. L'ampleur du présent ouvrage ne permet pas d'exposer ces développements.

Jetons toutefois un rapide coup d'œil sur l'introduction du **capital fixe** dans le système. Astucieusement, Sraffa en fait un type particulier de *production jointe*. La machine d'âge initial apparaît parmi les inputs ; cette même machine dont l'âge a augmenté d'une période se retrouve dans le membre de droite, en tant que résultat collatéral du processus productif[2]. La machine *m* va apparaître dans les équations avec différents indices représentant son « âge ». Par exemple, avec une machine neuve, on aurait :

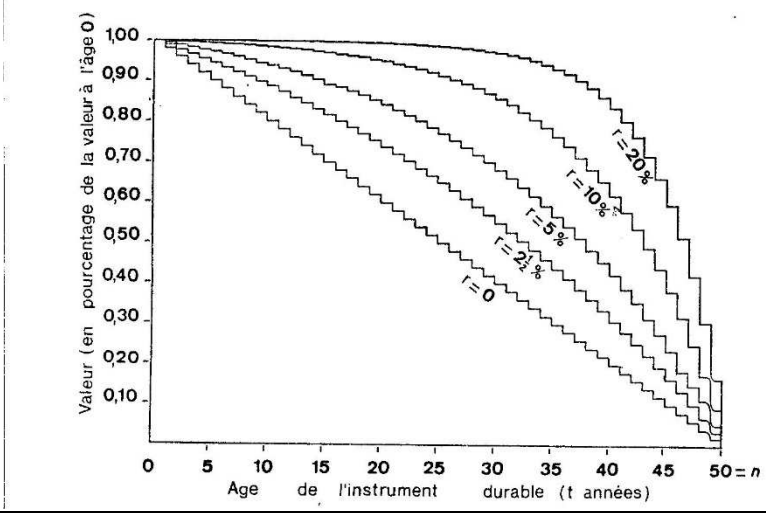
[1] On pourrait aussi dire : w = 1 – r/R

[2] Cette procédure se distingue d'une méthode plus courante consistant à n'intégrer dans l'input que la dépréciation ; dans ce cas, le capital fixe est évidemment tout à fait absent de l'output.

$$(M_0.p_{m0} + A_a.p_a + ... + K_a.p_k)(1+r) + L_a.w = A.p_a + M_1.p_{m1} \quad (5.10)$$

La difficulté consiste à déterminer la valeur des machines non neuves d'une façon qui tienne effectivement compte de la dépréciation physique tout en rapportant le taux moyen de profit. Sraffa dessine des courbes montrant la valeur résiduelle diminuant avec l'âge ; lorsque la taux d'intérêt est positif, elles sont concaves (la dépréciation va en s'accélérant) et dépendent du taux d'intérêt. La concavité est plus marquée lorsque le taux d'intérêt est plus élevé.

Graphique 5.5: la vitesse de dépréciation du capital fixe

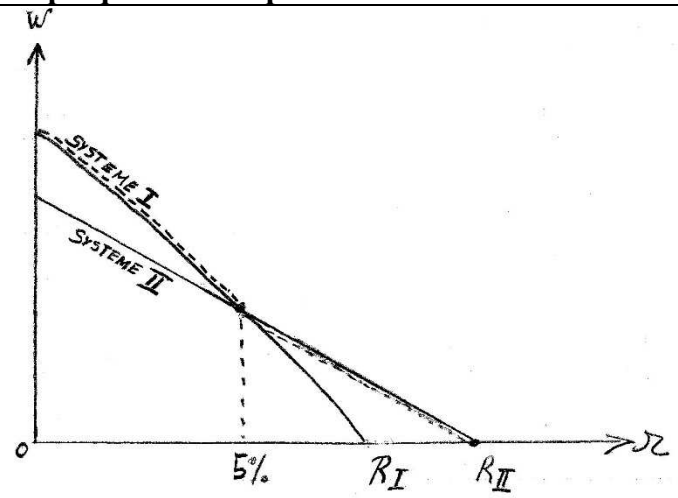


LE CHANGEMENT DES METHODES DE PRODUCTION

Imaginons que dans le système d'équations (5.5), le bien *a* puisse être produit par deux méthodes, notées I et II. Les coefficients A_a, B_a…K_a auront une valeur différente suivant qu'on est dans la méthode I ou la méthode II, ce qui revient à considérer deux équations différentes pour ce bien. L'entrepreneur, confronté au choix technologique, recourra à la méthode qui minimise le coût de production. Remarquons immédiatement cette évidence de l'algèbre : ce ne sera pas nécessairement la même méthode qui sera la plus rentable pour des rapports différents entre le taux de profit et le salaire. Par exemple, la méthode I serait plus rentable si $r < 5\%$ et la méthode II si $r \geq 5\%$. Sraffa remarque que peut survenir ce cas curieux que la méthode I serait plus efficace pour les taux inférieurs à 2% ou supérieurs à 6% et la méthode II serait plus efficace pour les taux compris entre 2 et 6%.

S'il s'agit d'un bien fondamental, nous aurons alors deux systèmes d'équations de type (5.5) distincts, les systèmes I et II, avec leur solution (prix relatifs et rémunération des facteurs) propre, leur marchandise étalon propre et leur taux de profit maximum (R) propre. Comment Sraffa s'y prend-il pour effectuer la comparaison entre leurs efficacités respectives ? Il choisit une unité de mesure unique, qui est la marchandise étalon d'un des deux systèmes, disons le II. Sur le graphique 5.6, nous avons regroupé les courbes de relation entre le salaire et le profit relatives aux deux méthodes.

Graphique 5.6: comparaison de deux méthodes de production d'un bien



w est exprimé en marchandise étalon II

Comme le salaire du système I est exprimé dans une unité de mesure qui lui est étrangère, la fonction liant le profit et le salaire dans ce système n'est pas une droite. Les courbes de relation entre *w* et *r* diffèrent selon que l'étalon soit choisi dans le système I ou le système II, mais Sraffa démontre que l'ordre de rentabilité entre les deux méthodes n'est pas affecté par le choix de l'étalon. Le taux de profit de 5% sur le graphique est donc un point d'aiguillage absolu, indépendant de l'étalon choisi.

La méthode à préférer (qui minimise le coût) est celle qui donne le meilleur taux de profit pour un salaire donné ou le meilleur salaire pour un taux de profit donné. Comme en dessous de 5%, l'entrepreneur utilise la méthode I et au-delà la méthode II, la véritable fonction liant *r* et *w* est la « courbe enveloppe » (en trait accentué sur le graphique). Cette courbe conserve le caractère essentiel qu'elle avait lorsque les variations des techniques étaient ignorées : bien qu'elle ne soit plus linéaire, elle reste absolument décroissante. Les taux de rémunération r et *w* ne peuvent varier qu'en sens inverse[1].

LA REDUCTION A DES QUANTITES DE TRAVAIL DE PERIODES DIFFERENTES

Sraffa tente une opération qui n'aurait pas déplu à Böhm Bawerk ou Wicksell: il va décomposer le prix d'un bien en une série de quantités de travail datées[2]. Prenons par exemple l'équation de la marchandise *A* dans le système d'équations (5.5).

Remplaçons-y les prix p_a, p_b …, qui figurent dans le membre de gauche, par leur propre équation (5.5) (divisée par la quantité produite). Répétons cette opération de nombreuses fois ; nous remontons ainsi dans le temps en décomposant chaque fois le capital de la période en du travail de cette période combiné avec du capital de la période précédente ; la part du capital non réduit à du travail diminue à chaque stade et nous arrêtons le processus lorsqu'elle atteint un niveau négligeable. A ce moment, additionnons les quantités de travail d'âge identique, et nous obtenons une série :

$$L_{a0}.w + L_{a1}.w.(1+r) + L_{a2}.w.(1+r)^2 + ... + L_{an}.w.(1+r)^n + ... = A.p_a. \qquad (5.11)$$

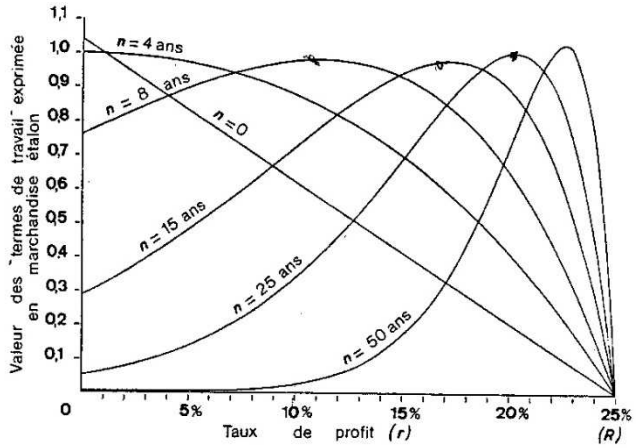
[1] Comme chez Wicksell, il faut interpréter ce modèle en statique comparative : le long de la courbe enveloppe se logent des économies distinctes qui diffèrent par leur taux d'intérêt et éventuellement par leurs méthodes de production.

[2] N'oublions pas que Ricardo est le père de la période de production ; dans la masse des différences entre le ricardisme et la conception autrichienne du capital nichent quelques similitudes.

Chaque terme, qu'on appelle le *travail réduit d'âge n*, correspond à la valeur ACQUISE de la quantité de travail *L* dont l'indice indique l'âge, exprimée en unités de la marchandise étalon.

Sraffa met en évidence l'effet de la répartition du revenu entre le capital et le travail sur les termes $L_{an}.w.(1+r)^n$. Si l'on fait augmenter *r* progressivement de 0 à *R*, l'effet sera double : le profit inclus dans le *travail réduit* augmente mais la hausse de *r* fait baisser *w*. La balance entre ces deux effets opposés diffère selon l'âge du *travail réduit*. Sraffa l'illustre par un exemple chiffré synthétisé dans le graphique 5.7. L'ordonnée est la valeur du travail réduit sur base de son âge et du taux d'intérêt.

Graphique 5.7: valeur des termes d'âge différent suivant le taux d'intérêt



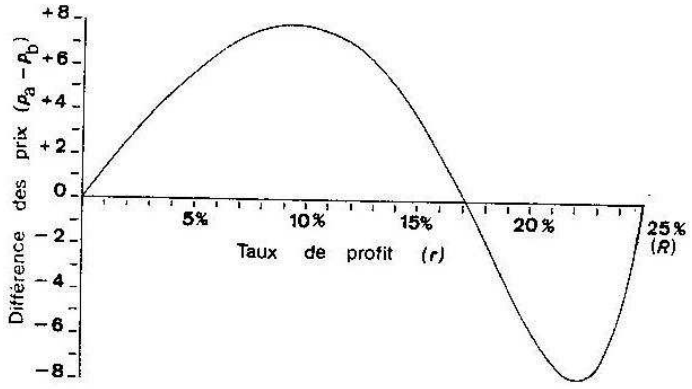
Les paramètres choisis par Sraffa pour cet exemple chiffré sont R = 25% ; L_0 = 1,04 ; L_4 = 1 ; L_8 = 0,76 ; L_{15} = 0,29 ; L_{25} = 0,0525 ; L_{50} = 0,0004.

On constate que la valeur acquise des termes d'âge récent (où $n < 1/R$) décline monotoniquement lorsque *r* augmente. Par contre, les termes plus anciens commencent par augmenter en valeur avec la hausse de *r*, jusqu'à un maximum après lequel la hausse de *r* les fait diminuer.

Passons maintenant à l'effet de ces variations du taux d'intérêt sur le prix de la marchandise, obtenu en sommant les *travaux réduits* des divers âges (formule 5.11). Pour le mettre en évidence, Sraffa compare les prix de deux marchandises *a* et *b* en tous points identiques, sauf que la distribution temporelle des L_n[1] diffère. L'écart (p_a - p_b) apparaît sur le graphique 5.8.

[1] Non précisée ici.

Graphique 5.8: effet des variations du taux d'intérêt sur les prix relatifs



On constate dans cet exemple que pour *r* supérieur à 17%, la marchandise *b* coûte plus cher que la marchandise *a* alors que c'est l'inverse pour les taux d'intérêt inférieurs. On peut parfaitement décomposer cet écart en parts imputables aux termes de travail des divers âges.

Il est donc illusoire d'espérer classer les intensités capitalistiques de différentes productions indépendamment de la répartition du revenu, comme l'économie néoclassique le pratique couramment.

5.3- LA CONTROVERSE DU CAPITAL

Cambridge est vraiment le lieu d'où part la contestation de l'économie néoclassique, le courant dominant. Des économistes cambridgiens suivront les pas de Keynes et de Sraffa ; comme Sraffa avait remis Ricardo à l'honneur, on les appelle les néo-ricardiens. Les principaux noms sont Joan Robinson, Nicholas Kaldor, Luigi Pasinetti.

C'est un article de Robinson, publié en 1954, qui déclenche la *controverse du capital*, un épisode épique de l'histoire de la science économique, qui durera une grosse dizaine d'années. Face aux néo-ricardiens qui attaquent la théorie néoclassique de la production et de la distribution, sa défense est assurée principalement par deux économistes du MIT, Paul Samuelson et Robert Solow [1]. Paradoxalement, ces deux économistes appartiennent aussi à un courant se réclamant de Keynes, que nous étudierons au chapitre 7.1.

La production littéraire dans le cadre de cette controverse est abondante et bouillonne d'idées diverses, dont certaines éphémères et d'autres « historiques ». Nous nous contenterons de la survoler.

5.3.1. La critique des fonctions de production agrégées

Dans sa première phase, l'attaque cambridgienne porte principalement sur les fonctions agrégées. L'article « The Production Function and the Theory of Capital » (1954) de **Robinson** commence ainsi : « Moreover, the production function has been a powerful instrument of miseducation. The student of economic theory is taught to write O = f(L,C) where *L* is a quantity of labour, *C* a quantity of capital and *O* a rate of output of commodities. He is instructed to assume all workers alike, and to measure *L* in man-hours of labour ; he is told something about a the index number problem involved in choosing a unit of output ; and he is hurried on the next question, in the hope that he will forget to ask in what units *C* is measured. Before ever he does ask, he has become a professor and so sloppy habits of thought are handed on from one generation to the next »[2].

L'évaluation du capital est un casse-tête. « Should capital be valued according to its future earning power or its past costs? »[3]. Le plus logique serait d'escompter le flux de revenus futurs au taux d'intérêt, mais pour cela, on doit déjà connaître le taux d'intérêt, alors que celui-ci est censé dépendre de la productivité marginale de ce qu'on doit préalablement mesurer. Les comptables enregistrent le capital à sa valeur d'achat, mais cette valeur ne résiste pas à l'érosion du pouvoir d'achat de la monnaie. On pourrait lui préférer le coût de remplacement, mais le même matériel n'est peut-être plus sur le marché. En outre, comment tenir compte de l'âge des biens d'équipement ? Il y a encore la valeur de revente de la firme ou sa valeur boursière. « The price at which (a plant) could be sold as an integral whole has not much significance, as the market for such transactions is narrow ». Les méthodes d'évaluation sont multiples mais aucune n'est totalement satisfaisante.

Pour mieux l'attaquer, Robinson synthétise la *fonction de production agrégée*, d'une manière novatrice, qui sera appelée *fonction de production intensive*. Elle relie la production par tête au capital par tête. Le graphique 5.9 la présente comme Joan Robinson ne la dessinera pas, pour une raison qui apparaîtra bientôt, mais telle que la produira Robert Solow deux ans plus tard.

---

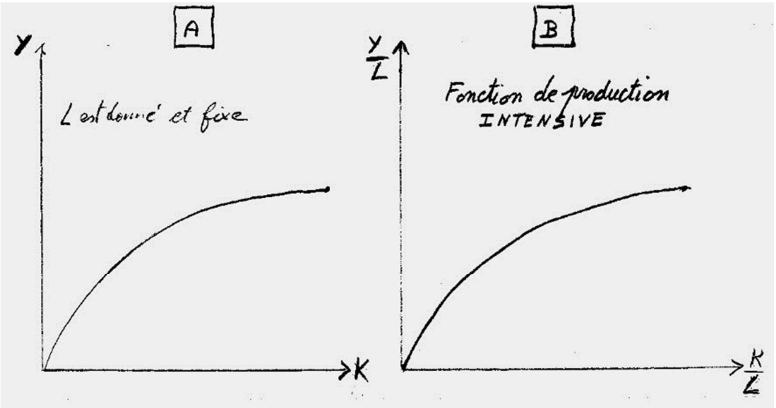
[1] On a aussi appelé cette controverse « the Two Cambridge Controversy », car elle opposait Cambridge (GB) à Cambridge (E-U).

[2] Robinson [303] p. 81

[3] Robinson [303] p. 81

On peut la rapprocher du graphique 4.1-A, dans l'hypothèse où le capital est le facteur variable et le travail le facteur fixe et où les secteurs I et III sont ignorés.

Graphique 5.9 : la fonction de production intensive dans sa version néo-classique



La pente de la courbe 5.9 donne la productivité marginale du capital dont se déduit le taux d'intérêt. La simplicité de cette fonction de production est frappante, mais selon Robinson, elle ne s'obtient qu'en faisant l'impasse sur nombre de difficultés. Robinson lui oppose une fonction intensive alternative, infiniment plus complexe malgré un tir groupé d'hypothèses simplificatrices. Dans l'article, Robinson la présente comme expérimentale, servant principalement à mettre en évidence le côté « conte de fées » de la fonction clarkienne.

MODELE SIMPLIFIE AVEC LE CAPITAL ET LE TRAVAIL

Le but est de redessiner le graphique 5.9-B de façon moins naïve. Le modèle n'est pas simple ; mais l'exposé de la théorie de Sraffa ci-avant facilitera la compréhension.

Voyons d'abord les hypothèses. Il y a un seul facteur primaire : le travail (considéré comme homogène). Les rendements d'échelle sont constants. La rémunération des deux facteurs, le salaire et l'intérêt épuise donc le produit : il n'y a pas de profit net. La composition de la production ne change pas avec les variations du revenu ou de sa distribution, ce qui permet de traiter la production comme si elle était composée d'un seul produit. Le salaire est exprimé en unités de ce produit. La masse de main d'œuvre disponible est fixe et l'équilibre du système postule qu'elle soit entièrement employée.

Quelle version du capital figure en abscisse ? Le *capital physique*, collection hétérogène de biens, est inutilisable en économie. Les économistes quantifient généralement le capital en unités du produit (ce que nos hypothèses rendent possible). Mais cette version ne rend pas correctement compte du rapport entre les facteurs ; pensons à *l'effet Wicksell* par lequel un salaire plus élevé réduit la capacité mécanique du capital. Pour y remédier, Robinson crée le concept de *capital réel*, qui n'est autre que le capital exprimé en unités de salaire, autrement dit le capital en unités du produit divisé par le salaire unitaire. Le capital réel par tête lui semble le moins mauvais indice de l'intensité capitalistique de la production. Comme l'a montré l'opération de réduction de Sraffa, les deux ingrédients du capital réel sont des doses de travail (d'âges divers) et de l'intérêt composé sur ces doses (dans l'hypothèse simplificatrice d'un seul facteur primaire).

Comment se détermine la quantité de capital en activité ? Robinson ne croit pas à la théorie néoclassique de l'équilibre à long terme[1] selon laquelle l'accumulation vise le point où le rendement du capital égale le « supply price of waiting », un concept qu'elle juge « treacherous ». Son point de vue consiste "to postulate that the stock of capital in existence at any moment is the amount that has been accumulated up to date, and that the reason why it is not larger is that it takes time to grow"[2].

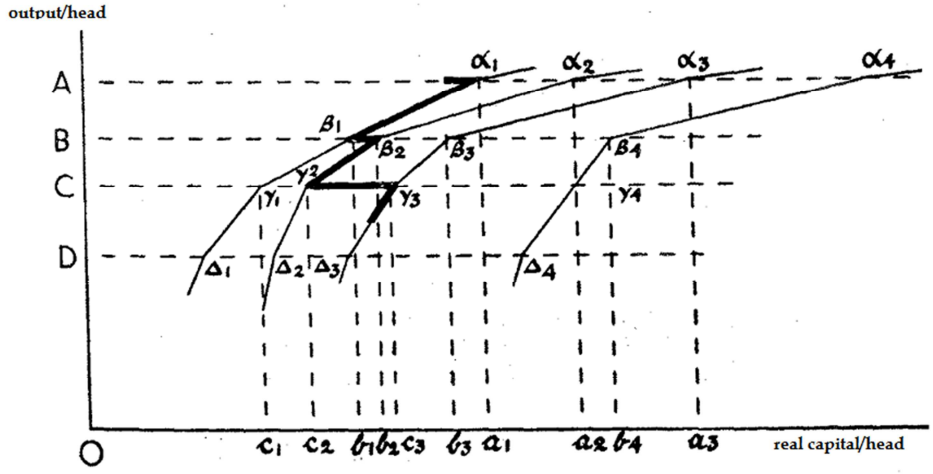
Robinson récuse l'existence souvent postulée d'une infinité de techniques. Elle envisage un nombre limité de techniques où le passage de l'une à l'autre a la caractéristique d'un saut. Ces techniques sont notées dans l'ordre de décroissant de *degré de mécanisation*. Le degré de mécanisation plus élevé signifie une production par tête plus élevée grâce à des outils plus performants. Si α est plus productif que β, la contrepartie est que cette technique est également plus gourmande en capital. Robinson imagine la construction d'une route dans plusieurs pays : en alpha, quelques hommes armés de bulldozers font le travail ; « in gamma, thousands of men are working with wooden shovels and little baskets to remove the soil ». Entre des techniques différentes, le capital diffère donc qualitativement autant que quantitativement.

Bien qu'elle en reconnaisse le caractère abstrait pour ne pas dire arbitraire, Robinson adopte la distinction courante entre le progrès technique et la technologie. Les techniques α, β… sont dans le domaine public, à la disposition des firmes qui souhaitent les appliquer et qui disposent du capital nécessaire. Le progrès technique est analysé dans l'appendice.

Sur le graphique 5.10, Robinson relie les différentes technologies α, β, γ, δ…., formant ce qu'elle appelle une *courbe de productivité*. La technologie plus mécanisée se situe plus à droite car elle exige plus de capital réel par tête et plus haut puisqu'elle produit plus[3]. Il semble logique qu'à un niveau de connaissance technique donné, la productivité tend vers une limite. La courbe a donc une asymptote horizontale et est globalement concave, ce qui exprime la productivité marginale décroissante du capital. Mais on ne peut exclure que deux segments consécutifs puissent former une portion de courbe non concave.

[1] Cf. sous-chapitre 3.2.8 : l'investissement selon Ramsey.
[2] Robinson [303] p. 87.
[3] Aux altitudes A, B, C et D, les techniques α, β, γ et δ sont utilisés de façon générale. Par contre sur les segments entre ces altitudes, les entreprises utilisent l'une ou l'autre des deux techniques aux extrémités ; au fur et à mesure que l'on se déplace vers le haut et la droite du segment, la part des entreprises optant pour la mécanisation supérieure s'élève ; au niveau de la société, cela représente donc une accumulation de capital

Graphique 5.10 : la fonction de production alternative de Robinson



Il y a une courbe de productivité par taux d'intérêt et donc, ce qui revient au même, par répartition du revenu entre le salaire et l'intérêt. Cette répartition a en effet un double effet sur le capital :

- Un effet du passé : plus le taux d'intérêt est élevé, plus le capital réel s'est chargé d'intérêt au cours de sa formation et plus il est à droite sur le graphique. C'est ce que Robinson appelle l'*effet d'intérêt*.
- Un effet du présent : quand le salaire est élevé et le taux d'intérêt bas, le degré de mécanisation élevé coûte moins cher à l'entreprise. Dans le cas contraire, le faible degré de mécanisation est plus économique.

Pour ces deux raisons, un taux d'intérêt plus élevé correspond à une courbe de productivité plus à droite ($r_4>r_3>r_2>r_1$).

Malgré leur ressemblance avec la courbe 5.9, les courbes de productivité ne sont pas encore la fonction de production recherchée. Faisons maintenant intervenir le phénomène mis en lumière par Ricardo. Les entrepreneurs choisiront la technique qui minimise leur coût ; ils élèvent leur degré de mécanisation lorsque le salaire s'élève et le taux d'intérêt baisse. On peut sélectionner sur chaque courbe de productivité le segment significatif et les relier entre eux. Le résultat est la courbe en gras que Robinson appelle *factor ratio curve*[1]. Avec ces zigzags sur le graphique, nous avons enfin notre fonction de production, que Robinson appelle *factor ratio curve*. C'est la droite brisée α_1-β_1-β_2-γ_2-γ_3 sautant d'une courbe de productivité à l'autre. Comme le remarque Robinson, « it has a somewhat bizarre appearance compared to the smooth sweep of the usual text-book production function »[2].

Robinson pratique la *statique comparative* ; en conséquence, il n'est pas question de variations du taux d'intérêt mais d'économies pratiquant des taux d'intérêts différents constants dans le passé, le présent et le futur. Dans une économie à taux supérieur, le capital réel s'est donc gonflé de plus d'intérêt.

Les pays industrialisés se caractérisent par l'abondance de capital ; les pays en développement par l'abondance de travail. Ce n'est évidemment pas l'effet du hasard si dans les premiers prédomine une technique de type α et dans les seconds une technique comme γ. Si un pays en

[1] Au taux d'intérêt r1, il y a indifférence de profitabilité entre les techniques α et β ; au taux r2, indifférence entre β et γ ; au taux r3, indifférence entre γ et δ. Les segments en gras comportent donc chaque fois une mixité entre deux techniques.

[2] Robinson [303] p. 93

développement voulait employer toute sa main d'œuvre avec la technique α, cela l'obligerait à fournir un effort d'accumulation insoutenable. Son abondance rend la main d'œuvre bon marché. Les entrepreneurs de ce pays nuiraient à leur profit en appliquant les techniques α, à moins que les écarts de productivité soient prodigieux. Robinson insiste sur l'importance du salaire réel en tant que déterminant du degré de mécanisation.

En conclusion, la substitution néoclassique entre le capital et le travail s'avère être la résultante de forces plus élémentaires que cette théorie n'explore pas et devant lesquelles elle fait paravent. La rétribution des facteurs influence le rapport capital/travail via trois canaux :

- *L'effet Wicksell* par lequel une hausse du salaire réel fait baisser le nombre d'ouvriers qu'un capital donné peut engager.
- *L'effet d'intérêt,* par lequel une hausse du taux d'intérêt fait hausser la somme d'intérêt comprise dans le capital réel.
- *L'effet Ricardo*, par lequel, une hausse du salaire réel rend plus rentables les techniques de production plus intenses en capital et inversement.

LA REPLIQUE DE SOLOW

Dans une réponse publiée rapidement, **Solow** défend l'agrégation. Il critique la *factor curve* de Robinson, ce qui ne suffit pas évidemment pas à « sauver » la fonction de production néoclassique. Il donne d'abord l'impression de relever le défi robinsonien : « If I write Q = f(L,C), I simply assume that there exists only one kind of physically homogeneous capital good, and *C* like *L* is measured in unambiguous physical units. Of course, it is not true that only one kind of capital goods exists, but then there is also more than one kind of labor »[1].

Mais après ces premiers mots ambitieux, Solow se contente de répondre à cette question : « Suppose we have a production function which relates the output of a single commodity to inputs of labour (assumed homogeneous) and the services of several capital goods. When if ever can the various capital inputs be summed up in a single index figure, so that the production function can be 'collapsed' to give output as a function of inputs and 'capital-in-general'? »[2]

Il faut arriver au résultat qu'il y a équivalence entre l'équation (5.12) (fonction non agrégée où C_1 et C_2 sont deux biens capitaux) et (5.13) (fonction agrégée où *K* est un index du capital global) :

$$Q = f(L,C1,C2) \qquad (5.12)$$
$$Q = F(L,K) \text{ où } K = \varphi\,(C1,C2) \qquad (5.13)$$

Solow propose tout simplement une fonction de production en deux stades: d'abord, obtenir *K* à partir de C_1 et C_2 seuls, par la fonction φ, et ensuite produire *Q* à partir de *K* et *L*, par la fonction *F*. Cela tient mathématiquement puisque la seule condition est que *L* n'interfère pas dans l'agrégation de C_1 et C_2.

Solow montre ensuite que la notion de « prix du capital en général » est sensée : on peut déterminer un p_K qui répond à la condition suivante :

$$p_K = p_1.C_1 + p_2.C_2 \quad \text{où } K = \varphi(C_1,C_2) \qquad (5.14)$$

[1] Solow [337] p. 101
[2] Solow [337] p. 102

On peut donc imaginer que le processus de minimisation du coût se passe en deux stades: d'abord déterminer le rapport K/L idéal sur base de p_K et p_L selon la fonction *F* et ensuite le rapport C_1/C_2 idéal sur base de p_1 et p_2 selon la fonction φ. Le résultat sera le même que si on avait directement déterminé L,C_1,C_2 à partir de leurs prix respectifs en une seule opération selon la fonction *f*.

Solow montre encore que moyennant les hypothèses de son modèle, le rapport de prix des facteurs p_K/p_L est une fonction inverse de l'intensité en capital K/L, toute belle comme dans les manuels décriés par Robinson. Mais comme le fera remarquer celle-ci, la réponse de Solow ignore sa principale préoccupation, l'effet du rapport p_K/p_L sur la valeur de *K*, car il ne fait pas intervenir la période de production des biens capitaux, leur âge et leur durée de vie. Dans la discussion qui s'ensuivra, Solow produira des modèles tenant compte de l'aspect « temps » du capital, mais les protagonistes ne parviendront pas à se mettre d'accord.

Une phrase du début de l'article est prémonitoire : « For many purposes, it is remarkably useful to assume that there exists only one physical good which can either be consumed or used in the production of more of itself »[1]. Comme cela apparaîtra dans la phase suivante, la vraie question est effectivement de voir si un modèle à un seul bien est représentatif de toutes les difficultés rencontrées dans un modèle avec plusieurs biens.

5.3.2. La *frontière des prix des facteurs de production*

Parmi les faits d'arme mémorables, il faut mentionner un article de **Samuelson** paru en 1962. Il y présente un nouvel outil d'analyse, la *frontière du prix des facteurs*. Etudions le d'abord et voyons ensuite à quel titre il apparaît dans la controverse.

Le but de Samuelson n'est plus de contredire la critique de Robinson, mais de prouver que, contrairement aux apparences, la fonction de production néoclassique englobe la *factor ration curve*. Son modèle fixe donc un cadre assez proche de celui de Robinson. Soient une série de techniques de production (peu importe le nombre), α, β, γ… dans l'ordre décroissant d'intensité en capital. Chaque technique exige un équipement qui lui est propre et un nombre déterminé fixe de travailleurs pour le faire fonctionner. Les technologies considérées produisent un même bien de consommation, mais chacune doit en outre assurer le remplacement du capital spécifique usagé. Ainsi, le processus α produit donc une certaine quantité du bien de consommation et une certaine quantité du bien d'équipement K_α ; celui-ci est donc à la fois input et output. A ce propos, Samuelson présume que la production du bien de consommation et celle de K_α consomment la même ptroportion des inputs, « with full warning this is a drastically simplifying assumption whose limitation will be commented on later »[2]. Le rythme du renouvellement est fixé de façon à permettre que la structure d'âge du capital reste constante.

Pour chaque technique α, β, γ…, on peut construire une fonction qui donne le salaire pour un intérêt donné et vice versa. Les rendements d'échelle sont constants et la rémunération des deux facteurs, capital et travail, épuise le produit net (c'est-à-dire déduction faite du remplacement du matériel). On a donc l'équation :

$$Y = r.K + L.w \qquad (5.15)$$

Y et *K* sont des valeurs et non des grandeurs physiques. Divisons l'équation (5.15) par *L* pour obtenir des grandeurs par tête. Soient *k* = le capital par tête (K/L) et *y* = le revenu par tête (Y/L). Nous obtenons la très importante équation :

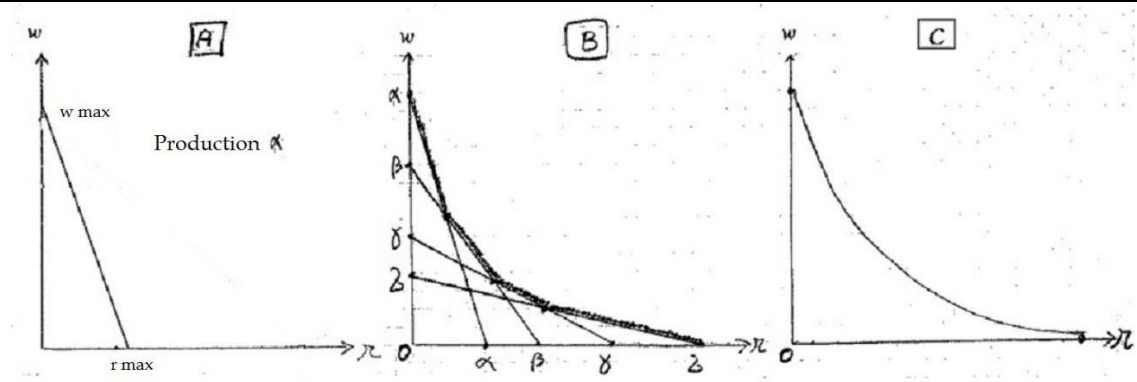
[1] Solow [337] p. 101
[2] Samuelson [319] p. 197

$$w = y - r.k \qquad (5.16)$$[1]

La situation est comparable au partage d'un gâteau : plus *r.k* est élevé, plus bas sera le salaire réel *w*. Comme le dit Samuelson : « There is always a trade-off between the wage and profit level : in the absence of innovation, both cannot go up and (…) both cannot go down »[2], une affirmation parfaitement ricardienne. C'est ce qu'indique la figure 5.11-A, qui n'est pas sans rappeler le graphique 5.4 de Sraffa[3]. La frontière est rectiligne de par l'hypothèse de la proportion identique des facteurs produisant la bien final et K_α.

Dessinons maintenant cette même fonction pour chacune des techniques α, β … et superposons-les sur la figure 5.15-B. Les différences de pente s'expliquent par le fait qu'une même hausse du salaire provoquera une baisse du taux d'intérêt plus sensible pour une technique moins intense en capital, car le salaire y représente une part du coût plus élevée.

Graphique 5.11 : la frontière du prix des facteurs



Face à cette gamme de techniques, quel sera le comportement du producteur ? Il choisira évidemment celle qui minimise son coût. Comment la reconnaître sur ce graphique ? On peut démontrer que celle qui minimise le coût est également celle qui maximise le salaire pour un taux d'intérêt donné ou – ce qui revient au même- maximise le taux d'intérêt pour un salaire donné. Partons du taux d'intérêt : ce n'est pas la même technique qui sera choisie quel que soit le taux d'intérêt ; une technique plus intense en capital sera préférée lorsque le salaire est plus élevé et vice versa. Sur le graphique 5.11, la ligne brisée en caractère gras, la courbe « enveloppe », donne le salaire pour un intérêt donné (ou vice versa) en intégrant ce choix des techniques. Samuelson l'appelle la « frontière du prix des facteurs » et les Cambridgiens la « frontière salaire-profit ». Elle n'est pas sans rappeler le graphique 5.6 de Sraffa. La décroissance de cette courbe montre que le choix entre plusieurs techniques n'affecte pas le principe selon lequel salaire et taux d'intérêt ne varient qu'en sens inverse[4].

La figure 5.15-C correspond à la figure B lorsque le nombre de techniques possibles est infini. La frontière devient continue.

La frontière du prix des facteurs renseigne non seulement sur le salaire et le taux d'intérêt mais également sur la part de chacun des facteurs de production dans le produit net total. L'élasticité

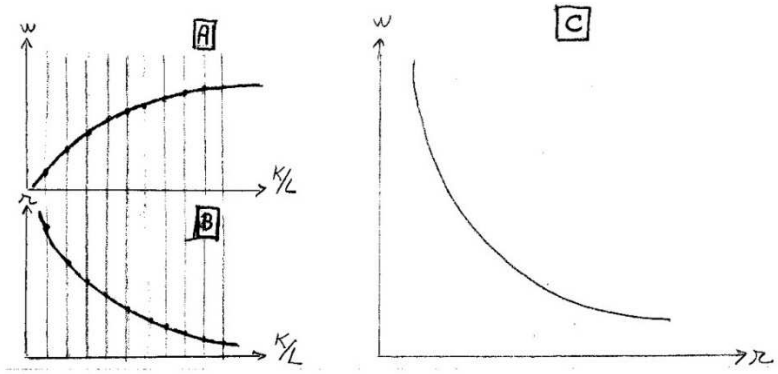
[1] Ce qui implique que r = (y-w)/k
[2] Samuelson [319] p. 196
[3] Ces graphiques ont la même signification ; la différence est dans le détail. L'unité dans laquelle est mesurée *w* diffère. Chez Sraffa, le partage portait sur le revenu national étalon dont *w* était une fraction. Ici, nous travaillons avec des grandeurs moins abstraites ; *w* est un salaire horaire quelconque.
[4] Comprenons le verbe « varier » au sens de la statique comparative.

de cette courbe en un point est l'indicateur de la répartition. Si elle égale trois, cela signifie que le travail capte une part du revenu trois fois supérieure à celle du capital. Une élasticité unitaire signifierait le partage en deux parts égales.

Pourquoi Samuelson fait-il atterrir la *frontière du prix des facteurs* dans la controverse du capital ? Tout simplement, parce qu'elle émane d'un raisonnement tenant compte de l'hétérogénéité du capital, qui devrait donc séduire les Cambridgiens. Samuelson s'efforce de montrer que le « conte de fées » de Clark, la fonction de production agrégée, est à même d'engendrer cet outil que les Cambridgiens ne peuvent réfuter.

Prenons un procès de production clarkien, synthétisé dans la figure 5.9, où le produit homogène sert à la fois de bien final et de capital. Un tel capital a tout pour déplaire aux Cambridgiens. D'une telle fonction de production, on peut aisément déduire les figure 5.16-A et B. Et de celles-ci la courbe 5.16-C. Quand elle est déduite de la fonction de production clarkienne et non construite à partir d'une analyse « réaliste », Samuelson appelle la courbe 5.16-C, la *fonction de production subrogée*.

Graphique 5.12: la frontière du prix des facteurs subrogée



Il est même possible de valoriser objectivement le capital issu de ce système clarkien. Samuelson montre que la pente de la courbe, dw/dr, vaut –K/L[1]. Donc, en chaque point, *K* vaut -(dw/dr)L.

5.3.3. Le double aiguillage et la marche arrière du capital[2]

La frontière salaire-profit est le tournant qui nous amène à la dernière phase de la controverse, d'où va émerger le concept décisif de *double aiguillage*. Sans le vouloir, Samuelson a donné à ses adversaires l'arme qui fera le plus mal. Samuelson pouvait se permettre de considérer les courbes d'arbitrage salaire-profit des techniques α,β… comme des droites, grâce à son hypothèse d'une intensité K/L identique pour produire le bien de consommation et pour remplacer du capital. Mais que se passe-t-il si l'on abandonne cette hypothèse ?

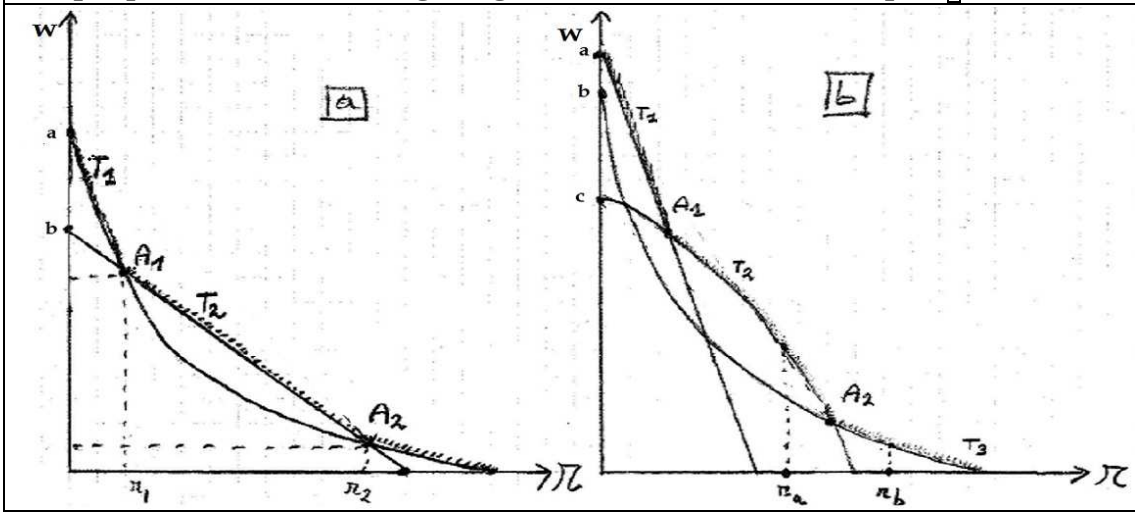
Si on permet aux industries des biens de consommation et de production d'avoir des intensités capitalistiques différentes, la courbe d'arbitrage salaire-profit[3] peut être convexe par rapport à l'origine (K/L supérieur pour le bien de consommation) ou concave (K/L supérieur pour le bien

[1] Il en découle que l'intensité capitalistique K/L = -dw/dr, qui est positive puisque *dw* est négatif.
[2] Cet exposé est inspiré avec quelque liberté de Hausman [136].
[3] Pour clarifier mon exposé, je réserve l'appellation *frontière* à la courbe enveloppe multi-techniques et j'appelle *courbe d'arbitrage* les fonctions propres à une technique.

capital)[1]. Notons que le caractère déclinant reste de rigueur. Lorsque les deux industries n'ont pas la même intensité en capital, la variation de la rémunération des facteurs induit une modification du rapport entre la valeur du bien de consommation et celle du bien de production qui affecte la valeur du capital et modifie tout le système de choix de la technique la plus efficace. Comme Robinson le pressentait depuis le début, la sensibilité de la valeur du capital à la rémunération des facteurs mine la fonction de production.

Les effets de l'incurvation des courbes d'arbitrage sur la frontière salaire-profit sont majeurs. Sur la frontière salaire-profit, on appelle *aiguillages* les points où deux technologies sont également rentables, qui sont ceux où l'on passe d'une technologie à l'autre lorsque change le rapport entre le salaire et le profit. Sur le graphique 5.11-B, lorsqu'on se déplaçait le long de la frontière vers l'est (*r* croissant), les points d'aiguillage menaient toujours à une technique moins intense en capital, ce qui confortait la vision traditionnelle du lien entre le capital et l'intérêt. Mais avec des courbes d'arbitrages non rectilignes, il n'est pas exclu que deux mêmes technologies aient plusieurs points d'aiguillage entre elles. Les Cambridgiens ont baptisé ce phénomène le *double aiguillage*.

Graphique 5.13 : le double aiguillage et la marche arrière du capital[2]



Le graphique 5.13-A illustre ce phénomène : les points d'aiguillage, A1 et A2, impliquent les deux mêmes techniques, T1 et T2. Pour un taux d'intérêt en dessous de r_1 ou au-dessus de r_2, la technique T1 est plus profitable que la technique T2 alors qu'entre r_1 et r_2, c'est la technique T2 qui doit être préférée. Une technologie devient moins intéressante qu'une autre quand le taux d'intérêt baisse mais redevient plus intéressante quand il baisse encore plus.

L'aiguillage A2 est d'autant plus perturbant que le passage de T1 à T2 représente une réduction dans l'intensité capitalistique alors que le taux d'intérêt décroît[3]. Ce phénomène a été baptisé *marche arrière du capital* par les Cambridgiens. Il contredit le lien inverse entre l'intensité en capital et le taux d'intérêt postulé par toute l'économie néoclassique, de Clark à Wicksell.

[1] Ceci n'invalide pas la courbe rectiligne de Sraffa au graphique 5.4. Là, *w* était justement exprimé dans LE numéraire qu'il fallait pour neutraliser l'effet des variations dans la répartition entre *w* et *r*.

[2] Hausman [136] pp. 75-76

[3] Comment évaluer l'intensité capitalistique d'une courbe d'arbitrage non rectiligne ? Pour un taux d'intérêt nul, le salaire capte la totalité du revenu ; le long de l'ordonnée, on a donc $w = y$. L'intersection d'une courbe d'arbitrage avec l'ordonnée indique donc la quantité produite par tête, qui est toujours supérieure lorsque l'intensité capitalistique s'élève.

Ironie du sort, Samuelson avait introduit cette frontière dans le débat, justement pour montrer qu'elle confirmait la thèse néoclassique traditionnelle.

En cas de double aiguillage, l'un des deux aiguillages (le plus à droite) implique toujours une *marche arrière du capital*. La figure 5.13-B montre que ce phénomène peut également survenir en l'absence de double aiguillage ; il n'y a pas de double aiguillage car ce ne sont pas les mêmes courbes qui se coupent en A1 (T1 et T2) et en A2 (T2 et T3). En A2, il y a marche arrière du capital, car quand le taux d'intérêt diminue de r_b à r_a, on passe d'une technique plus intense en capital (T3) à une technique moins intense en capital (T2).

Dans l'équation (5.16), *w*, *y* et *k* sont des valeurs exprimées dans un numéraire. Soit le bien de consommation, ce numéraire, dont le prix égale donc l'unité. Mais si les différentes techniques dont le graphique 5.13 analyse les points d'aiguillage ont des intensités capitalistiques différentes on doit s'attendre, en fonction de la loi de Ricardo, à ce que le rapport entre le prix du bien de consommation et le capital diffère selon que ce soit l'une ou une autre qui est en vigueur. On place sur un même graphique des courbes $w = y - k.r$ où ces variables sont exprimées dans des unités différentes. Est-ce défendable ? La démonstration mathématique a été apportée que les variations du numéraire n'avaient pas d'influence sur la détermination des points d'aiguillage et la sélection de la méthode de production la plus rentable. La frontière salaire-profit qui est la courbe enveloppe entre plusieurs courbes d'arbitrage est donc parfaitement légitime

Dans l'article « A summing up » (1966), Samuelson, forcé de reconnaître la réalité du double aiguillage, expose le problème de façon synthétique et pédagogique. Il donne l'exemple suivant : soient deux techniques pour obtenir du champagne, dont l'application du travail se distribue comme suit, entre les trois années qui précèdent la mise en vente :

	t = -3	**t = -2**	**t = -1**	**t = 0**
Méthode a	0	7	0	1 bouteille de ch.
Méthode b	2	0	6	1 bouteille de ch.

Cet exemple ne comporte que du capital circulant et montre donc que la production de biens capitaux n'est même pas nécessaire pour obtenir ce résultat.
C'est la méthode la moins coûteuse qui sera utilisée : le coût est donné par les équations :

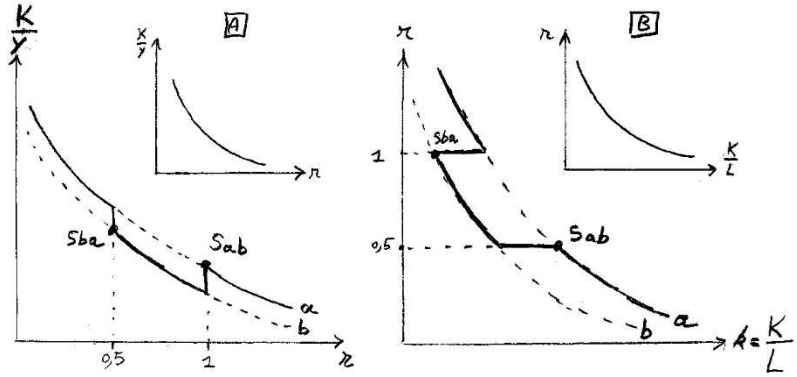
- en technique a : $7w.(1+r)^2$
- en technique b : $2w.(1+r)^3 + 6w.(1+r)$

Les coûts s'égalisent pour r = 1 (100%) et r = 0,5 (50%), qui sont les deux points d'aiguillage. Entre r = 0,5 et r = 1, la technique *b* est plus intéressante ; pour r > 1 ou r < 0,5, la technique *a* est la moins coûteuse.

La figure 5.14-A dessine la courbe du rapport capital / production pour cet exemple chiffré. Elle comporte une encoche qui la démarque de la courbe orthodoxe.

Reprenant le même exemple chiffré, les auteurs néo-ricardiens Cohen et Harcourt, dessinent la courbe de demande de capital (par tête) : c'est la figure B, qui comporte forcément la même encoche.

Graphique 5.14: la demande de capital et le double aiguillage



Supposons que r diminue. Au point Sab (switch de a vers b), on passe à une technique moins intensive en capital. Il y a marche arrière du capital.

Le double aiguillage se trouve son explication dans le fait que les différentes technologies ont des structures d'âge de leur capital très disparates. Comme Sraffa l'avait montré, l'effet des ces structures d'âge sur les prix des biens est une fonction du taux d'intérêt qui peut se révéler surprenante. L'école autrichienne se contentait de mesurer la période de production MOYENNE. Il faudrait également se préoccuper de l'écart-type voire même de la forme de la distribution des âges des doses de travail.

Les économistes néoclassiques ont admis la possibilité théorique du double aiguillage et de la marche arrière du capital mais ils ont tenté d'en réduire la portée. Toute une discussion s'est engagée sur la question de la fréquence de ces phénomènes. Il a été démontré que les conditions nécessaires à l'évitement du double aiguillage sont en fait très restrictives. Pour certains économistes néoclassiques, cette démonstration théorique ne suffit pas. Il faudrait en outre prouver empiriquement que le double aiguillage et la marche arrière du capital sont fréquents. Un débat s'est engagé sur la probabilité de la marche arrière du capital.

Pour certains Cambridgiens, le fond du problème n'est pas là. Ils estiment avoir mis en évidence que l'école néoclassique est incapable d'expliquer rigoureusement le lien entre le capital et l'intérêt, puisque des cas tout à fait rationnels d'un point de vue économique peuvent être construits où leur théorie est prise en défaut.

5.4- NEO-RICARDIENS ET NEOCLASSIQUES

5.4.1. La réponse néoclassique

Samuelson concluait ainsi son « Summing Up » : « Pasinetti, Morishima, Bruno-Burmeister-Sheshinski, Garegani merit our gratitude for demonstrating that reswitching is a logical possibility in any technology, indecomposable or decomposable (…) If all this causes headache for those nostalgic for the old time parables of neoclassical writing, we must remind ourselves that scholars are not born to live an easy existence »[1].

Cette attitude n'a pas eu d'écho durable. Le double aiguillage et la marche arrière du capital sont presque totalement absentes de la littérature néoclassique moderne. Pour l'expliquer, il faut anticiper sur le titre VI, consacré à l'avènement d'un nouveau courant néoclassique qui évincera ses prédécesseurs. Les tenants de cette nouvelle version, dite *néo-walrassienne*, prétendent que le double aiguillage et la marche arrière du capital disqualifient effectivement les théories de Clark et Wicksell car celles-ci tentaient d'agréger le capital ; la théorie néo-walrassienne dénuée de cette tendance à l'agrégation, est par contre immunisée contre cette critique. Ce que conteste Pasinetti : selon lui, le double aiguillage et la marche arrière du capital ne sont pas causées par le processus d'agrégation et peuvent donc se produire également dans un contexte néo-walrassien.

La thèse néo-walrassienne est affirmée avec force par **Frank Hahn** dans un article très remarqué, « The Neo-Ricardians », paru en 1982. Dans cet article, Hahn s'attaque à la théorie de Sraffa et surtout à une interprétation simpliste de celle-ci qui aurait fait flores.

Dans l'équilibre général néo-walrassien, le salaire (w) et l'intérêt (r) sont parfaitement déterminés tout comme l'intensité capitalistique de la production et d'ailleurs toutes les quantités et tous les prix de l'économie : un seul point de la frontière du prix des facteurs est donc compatible avec l'équilibre général[2]. Comme *w* et *r* valent la productivité marginale de leur facteur, on peut dire qu'ils dépendent de la technologie. Si l'on regarde le graphique 5.4 de Sraffa et son équation (5.9), on a au contraire l'impression que toutes les combinaisons de *w* et *r* se trouvant sur la frontière sont compatibles avec un équilibre général ricardien[3]. La combinaison *w-r* coulisse avec une aisance totale le long d'une droite. Certains disciples de Sraffa que Hahn attaque violemment en auraient conclu que la rémunération des facteurs de production est indépendante de la technologie et n'est qu'une question de rapport de forces. Ces disciples auraient le tort de s'être focalisés sur le graphique 5.4 de Sraffa plutôt que sa figure 5.6. Cette dernière montre que pour faire coulisser la combinaison *w-r* le long de la frontière, à certains points, il faut changer la technologie, ce qui n'est pas rien. Sraffa en était conscient, mais Hahn lui reproche d'avoir reporté ce problème à la fin de son ouvrage, après l'analyse du lien entre le salaire et le profit.

Selon Hahn, Sraffa n'a pas de théorie du taux d'intérêt. Le fait de prendre le taux d'intérêt comme donné[4] sans l'expliquer n'autorise pas ses disciples à prétendre qu'il ne dépend pas de la technologie. Dans la théorie néoclassique également, « if we know technology and nothing else, we don't have enough information to say what prices and what rate of profit will

[1] Samuelson [320] p. 583

[2] Hormis le cas de l'équilibre multiple (cf. infra)

[3] Pour chaque combinaison de *w* et *r*, les prix relatifs d'équilibre entre les biens seraient néanmoins différents

[4] Rappelons que dans le système de Sraffa, il est indifférent de prendre le taux d'intérêt ou le salaire comme variable indépendante.

prevail »[1]. Au sous-chapitre 6.4.1, nous examinerons la théorie néo-walrassienne de l'intérêt. Hahn montre qu'en partant de celle-ci et en y ajoutant une série d'hypothèses tellement restrictives qu'elles en deviennent ridicules, on arrive au résultat sraffien.

5.4.2. L'économie néo-ricardienne

Le procédé de Hahn pour ridiculiser l'économie néo-ricardienne n'est pas très honnête. Les manières néoclassique et néo-ricardienne d'aborder l'économie diffèrent à la base. Si on analyse l'une à l'aune de l'autre, on arrivera nécessairement à un résultat absurde. Une critique sérieuse doit rechercher leurs contradictions internes ou les confronter à la réalité économique.

L'économiste **Heinz Kurz**, qui adhère à la pensée (néo-)ricardienne, la caractérise par les propriétés suivantes :

1- La distinction (smithienne) entre les valeurs réelles et les valeurs normales des variables économiques. La valeur normale est celle de « longue période », qui s'établirait si le temps était laissé à tous toutes les forces en présence de s'ajuster. Tout particulièrement, le profit et le salaire tendent vers un taux uniforme entre tous les secteurs de l'économie (abstraction faite de facteurs objectifs tels que le risque).

2- Le prix tend à s'égaliser avec le coût de production ; le coût est objectif, physique ; il se manifeste principalement par la destruction (en une fois ou progressive) de biens dans le processus de production. Chez les néoclassiques, le coût, assimilé à la désutilité, est principalement psychique.

3- La description d'une économie à court terme comporte ces quatre facteurs exogènes :
   - Un ensemble de techniques alternatives à la disposition de producteurs qui visent à minimiser leur coût.
   - Le volume et la composition de la production sociale, reflétant les besoins et les goûts des différentes classes de la société ainsi que les exigences de la production et de l'accumulation du capital.
   - Le taux de salaire en vigueur. Comme nous l'avons vu, on pourrait aussi bien prendre le taux d'intérêt comme variable indépendante et le salaire comme variable dépendante. Il y a de toute façon traitement « asymétrique » entre les facteurs de production, par opposition à la symétrie qui règne dans la théorie néoclassique.
   - La quantité et la qualité de la terre et des ressources minérales.

   Pour comparaison, dans la théorie walrassienne, les variables indépendantes sont :
   - L'ensemble des techniques alternatives à la disposition des producteurs visant à minimiser leur coût.
   - Les préférences des consommateurs.
   - L'allocation initiale de ressources productives aux ménages : travail, terre et capitaux.

   Alors que dans le modèle walrassien, les quantités produites sont des inconnues au même titre que les prix, le modèle ricardien considère qu'à un moment donné, le volume et la composition de la production sociale sont une caractéristique de la société qui s'impose à l'économiste. Dans les équations à court terme, ils figurent comme variable indépendante. C'est à la théorie de la croissance qu'il revient d'expliquer les

[1] Hahn [127] p.362

quantités produites. La même distinction entre l'analyse à court terme et la théorie de la croissance existe également chez beaucoup d'auteurs néoclassiques ; mais les modèles néoclassiques font de la quantité produite une variable endogène aussi bien à court terme qu'à long terme.

Pour les néoclassiques, le salaire et le profit sont des prix comme les autres, déterminés dans le cadre de l'équilibre général. Le traitement des deux rémunérations est symétrique. Cette symétrie n'existe pas chez le néo-ricardiens. La raison en est la relation particulière qui lie le salaire et le taux de profit du fait de la frontière du prix des facteurs. L'une des rémunération (peu importe laquelle) est exogène et l'autre est endogène. La *frontière salaire-profit* concerne la DEMANDE des facteurs. Les néoclassiques intègrent également l'OFFRE des facteurs dans leurs équations. Les fonctions d'utilité des ménages en sont le moteur. Les néo-ricardiens rejettent cette explication subjectiviste et considèrent que ces faits relèvent plutôt de la sociologie ou de l'histoire économique.

4- La production est considérée comme un *flux circulaire*, comme l'indique le titre de l'ouvrage de Sraffa : « la production des marchandises par les marchandises ». Le système d'équations (5.5) voit le membre de droite (outputs) largement recyclé dans les ingrédients du membre de gauche (inputs). « In contradistinction, in Austrian and in much of neoclassical analysis, production is conceived of as a linear flow which leads from the services of the primary factors of production to final goods »[1].

5.4.3. Les théorèmes de non-substitution

A côté des modèles d'équilibre général walrassien et sraffien, deux auteurs ont présenté leur modèle d'équilibre général propre : von Neumann en 1937 et Leontief en 1941. Ces modèles affichent des similitudes importantes envers celui de Sraffa. Comme chez celui-ci, les prix relatifs n'influencent pas la répartition de la consommation ; mais en outre, des variations dans les goûts des consommateurs laissent le système des prix inchangé.

Ces deux auteurs ne s'inscrivent pas dans une école et sont difficilement classables. Le cas de **von Neumann** se singularise en ceci que son héritage est revendiqué, et même à bon droit, aussi bien par les néo-ricardiens que par les néoclassiques. Son modèle sera exposé assez longuement au sous-chapitre 6.1.2 au vu du rôle qu'il aura joué dans le développement de la théorie néowalrassienne.

Examinons brièvement le modèle de **Leontief**. Dans l'ouvrage « The Structure of American Industry » (1941), déjà précédé d'un article en 1936 et suivi d'une multitude d'articles et de livres, c'est une nouvelle branche de la science économique qu'ouvre Leontief : les modèles d'*input-output*. Son œuvre allie une construction théorique fortement empreinte de mathématiques avec une préoccupation empirique rare et un travail statistique impressionnant.

Un tableau d'input-output est un tableau à double entrée où les différents secteurs de l'économie font chacun l'objet d'une ligne et d'une colonne. Le nombre de secteurs de production dépend du niveau de différenciation souhaité. Appelons le *m*, auquel on en ajoute un énième qui représente les ménages. Sur chaque ligne *i* sont alignés en colonnes les fournitures du secteur *i* aux autres secteurs (output). Les différentes lignes d'une colonne *j* indiquent donc l'input que reçoit le secteur *j* de chacun des autres. Les éléments du tableau sont les x_{ij},

[1] Kurz [201] p. 379. Les ricardiens partagent avec les autrichiens la préoccupation pour le temps de la production. Mais le temps CYCLIQUE s'oppose au temps LINEAIRE.

désignant les ventes du secteur *j* au secteur *i*, exprimées en unités physiques. La production totale du secteur *i* est notée X_i.

En divisant x_{i1}, x_{i2}…x_{in} par X_i, on obtient les coefficients de production a_{i1}, a_{i2}…a_{in}, quantifiant les ingrédients de la production d'une unité de *i*. Leontief considère les coefficients de production comme fixes (absence de substitution). De plus, les rendements d'échelle sont constants.

On pourrait construire ainsi un système d'équations où toutes les quantités et tous les prix sont des variables endogènes et donc des inconnues. C'est ce que Leontief appelle un système *fermé*. Mais il lui préfère un système *ouvert*. Les ménages occupent une place particulière. La consommation des ménages ventilée par biens (les x_{ni}) est imposée de façon exogène, comme chez Sraffa[1]. Le système d'équations détermine les *m* prix ainsi que les quantités produites et échangées (non seulement les *m* X_i mais aussi les x_{ij} où *i* et *j* = 1…*m*). Dans les équations, les a_{ij} sont des données mais dans les traitements statistiques, les a_{ij} sont déduits des x_{ij}.

Il y a deux systèmes d'équations dont le premier est lié aux quantités :

$$X_1 - a_{21}.X_2 - \ldots - a_{m1}.X_m = x_{n1} \qquad (5.17)$$
$$::$$
$$-a_{1m}.X_1 - a_{2m}.X_2 - \ldots + \quad X_m = x_{nm}$$

Les x_{in}, qui sont les prestations des ménages pour la production, se calculent en multipliant le X_i par les a_{in} qui sont donnés comme les autres coefficients. Ici le modèle de Leontief pèche au moins par ambiguïté. Dans certains passages, Leontief considère que les x_{in} correspondent à l'emploi sectoriel. Mais contrairement à celles de Sraffa, les équations de Leontief ne multiplient pas les inputs matériels par (1+r). La place de la rémunération du capital est peu claire. Normalement, elle devrait être incluse dans les x_{in} et c'est ainsi que Leontief l'entend dans un passage où il assimile les x_{in} à la *valeur ajoutée*. Quoi qu'il en soit, le modèle de Leontief se soucie beaucoup moins que celui de Sraffa de la relation entre le salaire et le profit. A noter que Leontief intègre également la formation nette de capital dans les x_{in}.

Un deuxième groupe d'équations concerne les prix égalisés avec le coût, où R_i est la *valeur ajoutée* dans le secteur *i* (a_{in}), qui est donnée :

$$a_{11}.P_1 + a_{12}.P_2 + \ldots + a_{1m}.P_m + R_1 = P_1 \qquad (5.18)$$
$$::$$
$$a_{m1}.P_1 + a_{m2}.P_2 + \ldots + a_{mm}.P_m + R_m = P_m$$

La citation suivante fera office de conclusion : « En bref, la structure d'intrant-extrant d'un système ouvert définit une relation unique entre a) l'ensemble de tous les prix, b) l'ensemble de tous les taux de salaires, c) l'ensemble de tous les taux de profit unitaires. Etant donnés deux de ces éléments, le troisième peut être calculé indirectement »[2].

[1] Potentiellement, chaque bien est un input des *m*-1 autres et de la consommation. Il n'y a donc pas de distinction a priori entre biens de production et biens de consommation. Mais x_{ni}, exogène, a un statut différent de x_{1i}…x_{mi} qui sont endogènes. Chez Sraffa, la consommation n'est qu'une donnée implicite et secondaire ; c'est la répartition de la production totale qui est exogène.

[2] Leontief [220] p. 200. A noter que les termes *input* et *output* n'apparaissent pas une fois dans la traduction française ; ils sont systématiquement traduits par *intrant* et *extrant*.

Le modèle de Leontief a intrigué les économistes néoclassiques qui ont eux-mêmes recherché les conditions nécessaires pour obtenir un tel résultat en partant des postulats néoclassiques. Arrow (1951), Samuelson (1951), Stiglitz (1970) et d'autres ont produit ce qu'on appelle des *théorèmes de non-substitution*, que Bliss définit ainsi : « a non-substitution theorem is a result which shows that under certain specified conditions, an economy will have one particular price structure, regardless of the pattern of final demand, this price structure corresponding to the costs of production of a particular set of activities »[1].

Dans le modèle de Samuelson, la connaissance des fonctions de production suffit à déterminer le système de prix relatifs. Si l'on fait varier la répartition de la consommation ou de la quantité de travail disponible, le système de prix ne sera pas affecté. Leontief travaillait avec des coefficients de production fixes, mais Samuelson, le premier, montre que les théorèmes de non substitution sont également compatibles avec les coefficients variables.

D'après ces modèles, les principales conditions pour obtenir un tel résultat sont :
1. les rendements d'échelle constants dans toutes les productions
2. absence de production jointe
3. un seul facteur primaire (la terre ou le travail).

Les deux premières conditions sont assez évidentes.
- Si les rendements d'échelle d'un bien X sont croissants et ceux d'un bien Y constants, il est évident que les coûts de production de X et de Y évolueront différemment selon qu'on doive produire 10X et 10Y ou 10X et 15Y.
- Comme la laine et la viande du mouton sont issues du même processus de production, aucun facteur technologique ne peut déterminer le rapport de prix entre elles ; seules leurs demandes respectives le peuvent.

[1] Bliss [38] p 251.

5.5- LES DEBATS SUR LE MARGINALISME[1]

Passons maintenant à une autre controverse, un peu plus ancienne, qui partit de l'université d'Oxford. Elle a duré de 1939 au début des années cinquante, touchant à la fois l'Angleterre et aux Etats-Unis. Le marginalisme fut attaqué et confronté à un nouveau principe : le *full cost pricing*. En marge se déroulèrent des controverses connexes portant sur la forme des courbes de coût, ainsi que l'emploi et le salaire.

5.5.1. Le « Full cost pricing »

Vers la fin des années trente, un groupe d'économistes de l'Université d'Oxford[2], mené par Roy Harrod, sondent les entrepreneurs quant à leur politique de prix et de production, dans le but de peaufiner la théorie du cycle conjoncturel. A leur étonnement, les réponses des businessmen sont en porte à faux avec la théorie économique, c'est-à-dire avec la thèse marginaliste enrichie par le récent exposé de Joan Robinson. L'article « Price Theory and Business Behaviour »de Hall et Hitch (1939) résume leurs découvertes.

En fait, les firmes dont les réponses sont exploitables ne sont que trente-huit. Les auteurs se doivent donc d'apporter la mise en garde que « In the light of the smallness and the biased character of the sample, no significance can be attached to the precise percentages of firms behaving in particular ways. But on some questions, the replies are so nearly unanimous that it is impossible to ignore their implications »[3].

D'après la doctrine officielle, les firmes doivent pousser leur production jusqu'au point où la recette marginale est égalisée avec le coût marginal. Les enquêtes montrent qu'une majorité de firmes ne se reconnaissent pas dans ce schéma ; au lieu de cela, « in pricing policy they try to apply a rule of thumb which we shall call "full cost", and maximum profits, if they result at all from the application of this rule, do so as an accidental (or possibly evolutionary) by-product »[4].

La règle du *full cost,* est-elle une théorie alternative au marginalisme ou seulement un argument empirique pour contester celui-ci ? La réponse des auteurs n'est pas évidente. Voyons comment cette règle fonctionne. Elle consiste à déterminer le prix sur base d'un coût moyen complet incluant une marge pour le profit. La formule du coût complet varie en détail d'une firme à l'autre, mais on la résume correctement comme suit : coût direct plus pourcentage pour couvrir les coûts indirects[5] plus pourcentage pour la marge bénéficiaire. Mais la concurrence rabote les écarts de prix entre les firmes. Soit, c'est le privilège du leader du marché d'appliquer « son » full cost et les autres suivent, soit un accord doit se dégager inconsciemment.

Quels arguments les entrepreneurs avancent-ils pour justifier leur inclination pour le *full cost pricing* plutôt que pour la méthode marginaliste ?

1- sentiment que le full cost est le seul prix irréprochable du point de vue des concurrents et des acheteurs.

---

[1] Je ne cache pas que ce débat constitue à mes yeux un des épisodes les plus intéressants de l'histoire de la science économique.

[2] L'Oxford Economic Research Group (OERG).

[3] Hall & Hitch [128] p. 13

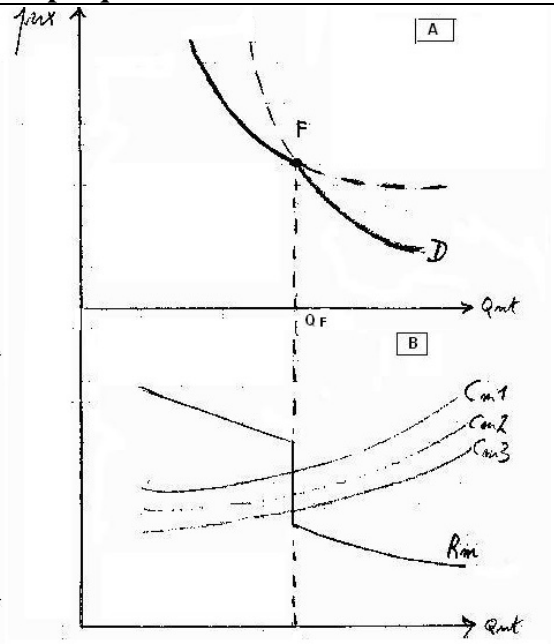
[4] Hall & Hitch [128] pp. 18-19

[5] Dans le cas d'une firme multiproduits, les coûts indirects sont ceux qui ne peuvent être imputés directement à un produit, car ils en concernent plusieurs.

2- l'instabilité des prix comporte de nombreux inconvénients, qui font préférer une politique maintenant des prix conventionnels à une politique de maximisation du profit à court terme qui impliquerait de fréquentes variations de prix.
3- les firmes ne connaissent pas leurs courbes de demande, de recette marginale et de coût marginal. Dans le cas de l'oligopole, où elle dépend des réactions des concurrents, la courbe de demande est indéterminée. Lorsqu'elle n'est pas indéterminée, elle n'en est pas pour autant connue des firmes. L'élasticité de la demande ne semble même pas importer aux entrepreneurs.
4- peur d'augmenter le prix, car les concurrents pourraient ne pas suivre ; et parallèlement peur de baisser le prix, parce que les concurrents pourraient suivre. Cette double crainte tient au fait que beaucoup de firmes œuvrent dans un environnement oligopolistique, un fait largement sous-estimé par l'économie politique dominante, selon Hall et Hitch.
5- impression que l'élasticité de la demande est insuffisante pour justifier une baisse du prix. Il s'agit ici de la demande du marché et non de la demande à la firme.
6- un prix supérieur au *full cost* pourrait entraîner l'arrivée de nouveaux offreurs sur le marché.

Pour illustrer le quatrième argument ci-dessus, Hall et Hitch exposent la fameuse théorie de la *courbe de demande brisée*, caractéristique de l'oligopole. Ce cas est illustré par le graphique 5.15. La figure A montre la demande à l'entreprise, dont la production actuelle vaut QF. F est un point de brisure. A gauche de F, la courbe D est très élastique ; la hausse de prix, non suivie par les concurrents, réduirait gravement la demande. A droite de F, la courbe D est peu élastique ; la baisse du prix, imitée par les concurrents, accroîtrait peu la demande.

Graphique 5.15 : la courbe de demande brisée



La figure B montre la courbe de recette marginale correspondante confrontée au coût marginal. Pour la quantité correspondant à la brisure, la courbe Rm comporte un segment vertical. En conséquence, la probabilité est élevée que la courbe Cm la coupe dans ce segment. Par exemple, que le coût marginal soit donné par Cm1, Cm2 ou Cm3, elle coupe toujours Rm pour une même quantité Q_F.

Quel est l'effet de la politique des firmes sur la situation du marché ? Hall et Hitch distinguent deux cas : la stabilité des prix et l'instabilité des prix :

- « In cases of relative stability each firm adheres as closely as it is possible to its own formula. Where costs do not differ widely within an industry all firms will charge similar prices, and the consumers will be distributed among them according to the factors which make the market imperfect, such as the proximity or attachment of customers to particular firms »[1]… « The price is likely to be changed without a departure from price stability if there is a change in costs which will affect all firms together, such as a change in wages or the price of materials »[2].
- En cas de dépression, le prix peut se retrouver sous le full cost, du moins si celui-ci est calculé sur base de la production réelle (donc avec une hausse du coût fixe unitaire). Mais on ne tombe dans l'instabilité des prix que si les firmes se lancent individuellement dans des réductions pour attirer les clients.
- Les autres causes d'instabilité des prix sont l'arrivée d'une nouvelle firme, qui veut obtenir directement une part de marché importante ; ou la conviction d'une firme d'avoir des méthodes plus productives que les concurrents. D'une façon générale, les prix seront moins flexibles que dans la théorie marginaliste, car seuls des événements à long terme décident les entreprises à modifier leur prix.

5.5.2. La productivité marginale, l'emploi et le salaire

En 1946, l'article de Lester, « Shortcomings of Marginal Analysis for Wage-Employment Problems », se base sur les réponses écrites à son questionnaire, que lui ont renvoyées 58 entreprises industrielles du sud des Etats-Unis. Dans les secteurs où elles opéraient, existait un différentiel salarial avec le nord (salaires plus élevés dans le nord).

A la demande de Lester, elles ont classé par ordre d'importance les cinq facteurs susceptibles d'influencer l'emploi, de la liste suivante :

a- La demande présente et prospective du bien fabriqué
b- Le niveau du salaire et ses changements
c- Le coût des facteurs autres que le travail
d- Les variations du profit
e- Les nouvelles techniques de production, les nouveaux équipements

Le facteur *a* l'emporte très largement. « Factor *b* (wages and changes in wage levels) and factor *d* are rated surprisingly low by executive officers of these 56 firms in view of the emphasis placed on those two factors by marginal analysis. On the other hand, the relative stress placed on materials and other non-labor costs as a factor in determining the firm's volume of employment is surprisingly high »[3]. Outre la faiblesse générale du facteur *b*, Lester remarque la maigre corrélation entre l'importance qui y est attachée et la part du travail dans le coût global.

D'après Lester, la faiblesse du facteur *b* est corroborée par cette autre constatation, que c'est justement dans le sud, où les salaires ont le plus augmenté, comme suite à la fixation légale de salaires minimaux, que l'emploi s'est le plus accru.

Lester interrogea également les firmes sur la perception qu'elles ont de leurs coûts. « In reply to the question 'At what level of operations are your profits generally greatest under peacetime conditions ?' 42 firms answered 100 per cent of plant capacity »[4]. Des réponses ressort l'image

[1] Hall & Hitch [128] p. 27
[2] Hall & Hitch [128] p. 28
[3] Lester [230] pp. 65-66
[4] Lester [230] p. 68

d'une courbe de coût moyen (et marginal) décroissante jusqu'à la *capacité*[1]. Ce n'est pas le coût fixe unitaire qui est en cause, car Lester parle bien du coût variable moyen. Sur un équipement donné, la productivité semble augmenter avec la quantité produite. Très peu des firmes ont répondu positivement quant à la possibilité de réduire le coût de production unitaire en réduisant la production. La théorie des courbes en U ne trouve pas confirmation auprès des entrepreneurs. Or si le coût marginal ne remonte pas, son égalisation avec la recette marginale peut se révéler impossible.

Lester explique le peu d'intérêt manifesté par les firmes envers les concepts de productivité marginale et de coût marginal : les entreprises « see the extreme difficulty of calculating marginal variable costs and the marginal productivity of factors, especially in multiprocess industries and under present accounting methods »[2].

D'après la théorie marginaliste, les entreprises devraient sans cesse faire varier la proportion entre la quantité de travail occupée et les équipements, pour répondre aux variations du rapport de prix entre les facteurs. Ce comportement ne semble pas se vérifier. Supposant une augmentation salariale, Lester demanda aux dirigeants d'entreprise de classer par ordre d'importance ces cinq réactions possibles :

a- Installation de machines économisant du travail.
b- Amélioration de l'organisation et de la gestion.
c- Changement du prix ou de la qualité du produit.
d- Augmentation de l'effort de vente pour accroître la production.
e- Réduction de la production pour diminuer l'emploi

La réaction *b* semble la plus courante, devant la *a* ; suivent la *c* et la *d*. La *e*, la plus en phase avec la théorie marginaliste selon Lester, ferme la marche.

Le peu de succès de la réponse *e* s'explique :

- d'une part par l'accroissement du coût unitaire (Lester voit la courbe de coût moyen entièrement décroissante)
- d'autre part, parce qu'un équipement donné ne peut fonctionner qu'avec un effectif donné. Lester met en doute la toute-puissance de la substitution entre les facteurs affirmée par le marginalisme. « Most industrial plants are designed and equipped for a certain output, requiring a certain work force. Often, effective operation of the plant requires a work force of a given size »[3]. La variation de l'effectif n'a de sens que dans le cadre d'un rééquipement. Mais modifier l'intensité capitalistique en réponse à une augmentation de salaire est une opération risquée et complexe. Lester cite une étude montrant que le changement technologique est très peu connecté aux variations du salaire.

Lester a également demandé aux firmes qui possédaient des usines à la fois dans le nord et dans le sud, si le différentiel de salaires entre les deux régions les conduisait à des méthodes de production plus intenses en travail dans le sud où les salaires sont inférieurs. Très majoritairement, les firmes concernées répondirent par la négative. Lester constate que les

[1] Les courbes de coût en U sont assez largement contestées par les économistes non marginalistes. C'est notamment le cas de Eiteman en 1952. Lui aussi procéda à un sondage auprès d'entreprises, leur soumettant huit formes différentes que pourrait revêtir la courbe de coût moyen, plus précisément le coût moyen global (fixe plus variable) à court terme. Les firmes devaient simplement voter pour la forme qui leur semblait la plus représentative de leur réalité. Plus de la moitié des réponses mirent en avant une courbe décroissante sur toute sa longueur jusqu'à ce que soit atteinte la capacité.

[2] Lester [230] p. 72

[3] Lester [230] p. 72

firmes du sud investissent souvent dans les équipements les plus modernes, les plus performants qui sont disponibles sur le marché.

Terminons avec cette conclusion : « The practical problems involved in applying marginal analysis to the multi-process operations of a modern plant seem insuperable, and business executives rightly consider marginalism impractical as an operating principle in such manufacturing establishments »[1].

5.5.3. La contre-attaque marginaliste

Tout au long des années quarante, un débat opposa des partisans du *full cost* et des économistes fidèles à la théorie de la concurrence imparfaite. Ces résistants au full cost se recrutèrent principalement à l'Université de Cambridge[2] où officiait Joan Robinson. Le paradigme cambridgien est fondé sur l'égalisation orthodoxe de la recette marginale avec le coût marginal, qu'il combine avec le *mark up*, c'est-à-dire que les entrepreneurs fixent les prix en appliquant une marge sur le coût marginal, pour couvrir les coûts indirects et générer du profit. Ainsi, la théorie marginaliste explique le niveau des prix, mais que la théorie du full cost lui est associée pour décrire les procédures que les entreprises appliquent pour arriver à ce résultat. Le full cost est mis au service du marginalisme et ramené au statut de « rule of thumb ». Une façon de désamorcer la bombe du full cost.

En 1946, l'économiste austro-américain Fritz **Machlup** mène une double charge, à la fois contre la position de Hall et Hitch et contre celle de Lester. L'argumentation comporte deux volets. D'abord, un recadrage des principes marginalistes qui auraient été interprétés fallacieusement ; ensuite, une critique de la méthode empirique qui fut appliquée par ses opposants.

Première mise au point : « Instead of giving a complete explanation of the "determination" of output, prices, and employment by the firm, marginal analysis really intends to explain the effects which certain *changes*, in conditions may have upon the actions of the firm »[3]. Sans prendre position sur le fond de cet argument, on constate que les modèles marginalistes sont rarement exposés de cette façon par leurs promoteurs.

Les courbes de recette et de coût de l'entreprise, affirme ensuite Machlup, n'ont pas pour vocation de relater le niveau REEL de ces grandeurs mais la perception SUBJECTIVE, bonne ou mauvaise, qu'en ont les gestionnaires. En outre, « anticipations alone are the relevant variables in the marginal calculus of the firm »[4]. Machlup veut également éclaircir un malentendu concernant la distinction entre le court et le long terme. Pendant le court terme, la quantité utilisée de certains facteurs de production n'est pas nécessairement optimale. C'est à tort qu'on impute couramment cette imperfection au délai que requiert un ajustement complet aux conditions du marché. C'est un autre délai qu'il faut prendre en compte : la durée des nouvelles conditions, est-elle suffisante pour justifier l'ajustement de l'appareil de production ? Des investissements ne sont jugés rentables que si les variations de la demande ne sont pas perçues comme transitoires.

---

[1] Lester [230] pp. 82

[2] L'Université de Cambridge abritait pas mal d'économistes de gauche, qui appréciaient la théorie de la concurrence imparfaite, car elle montrait que dans cette structure de marché, les forces du marché menaient à une situation sous-optimale, ce qui justifiait l'intervention de l'Etat sur ces marchés.

[3] Machlup [242] p. 521. Dans le même ordre d'idées, Machlup raille (gentiment) l'habitude de dessiner des graphiques où les grandeurs en abscisse et en ordonnée varient de zéro à un multiple du niveau réel.

[4] Machlup [242] p. 523

Machlup est également agacé par les allusions à « the extreme difficulty of calculating ». Non, argumente-t-il, les businessmen ne doivent pas estimer la recette et le coût marginaux pour toute quantité produite et calculer l'écart pour déterminer à quelle échelle il s'annule. Grâce à leur « ability to size up a situation without reducing its dimensions to definite numerical values »[1]. Il fait l'analogie avec un automobiliste qui doit décider s'il dépasse le camion qui le ralentit sur la route. Un théoricien de la conduite automobile, qui voudrait modéliser son problème, devrait établir des équations qui prennent en compte des variables quantifiables comme la vitesse du camion, la distance qui sépare l'auto du camion, ses possibilités d'accélération, la distance entre son auto et celle qui approche dans l'autre sens… Clairement, les automobilistes ne font pas ces calculs, ce qui ne les empêche pas de conduire. Le fait que les entrepreneurs prétendent ne pas connaître l'élasticité de leur demande, ne pas se soucier du coût et de la recette marginaux ne disqualifie pas une explication théorique recourant à ces concepts, qui doit nécessairement se placer à un niveau d'abstraction plus grand.

Le vocabulaire de l'économiste diffère de celui du businessman, qui ne connaît pas la *recette marginale*, le *coût marginal*… Ceci ne compromet d'aucune manière la théorie marginaliste. Le vocabulaire scientifique qui analyse les actions de tel ou tel types est généralement étranger aux individus dont les actions sont expliquées. La phraséologie des entrepreneurs est empreinte du coût moyen. Selon Machlup, il faut gratter le vernis terminologique ; si l'entrepreneur envisage des coûts moyens différents pour des volumes de production différents, il fait du marginalisme sans s'en rendre compte. Machlup martèle sa conviction que les actions des entrepreneurs ne sont pas guidées par des considérations de coût moyen mais basées sur « the profitableness of the added business ». D'une façon générale, Machlup reproche aux deux enquêtes étudiées ci-dessus de n'avoir pas pris de recul par rapport aux réponses des entrepreneurs, de les avoir prises pour argent comptant au lieu de les passer au crible. En se contentant des réponses écrites sans les compléter par des interviews (contrairement à Hall et Hitch), Lester s'est coupé de la possibilité de décoder les réponses des businessmen.

Parmi les allégations des entrepreneurs qui provoquent sa suspicion, Machlup relève les réticences « quasi-morales » à vendre en dessous du coût complet conventionnel. Il doute que l'adhésion à ce prix conventionnel résisterait à l'opportunité de maximiser le profit en vendant moins cher. Contre ces allégations, il assène cet autre argument que l'adhésion au prix conventionnel relève du comportement de cartel ; or celui-ci est un cas particulier intégrable dans l'explication marginaliste : il se singularise par une courbe de demande à la firme qui retombe verticalement sur l'axe de l'abscisse à partir du prix conventionnel ; que la courbe de demande soit ainsi ou autrement, la règle de maximisation du profit demeure.

Machlup épingle la reconnaissance par Hall et Hitch, que la marge de profit incluse dans le coût complet est sujette à variation. Il en conclut : « These differences and variations strongly suggest that the firms consult other data besides or instead of their average costs »[2]. D'un examen critique de leurs réponses ressort l'impression qu'elles se préoccupent de l'élasticité de leur demande (même si cela se passe inconsciemment).

Machlup réserve la suite de ses critiques à l'article de Lester. Il s'en prend notamment à la liste des cinq critères susceptibles d'influencer l'emploi que les entrepreneurs sondés avaient à classer. Selon lui, cette liste pèche en ceci que les facteurs proposés ne sont pas complètement indépendants entre eux. Il considère que toutes les propositions (sauf la *d*-, trop vague) sont des variables du principe marginaliste, qui n'est donc nullement compromis par le succès électoral

[1] Machlup [242] p. 525
[2] Machlup [242] p. 545

du facteur *a-* (perspectives de ventes). Il reproche à Lester de vouloir contrer le marginalisme avec une théorie qui ferait reposer la détermination du volume de l'emploi sur un seul facteur : les perspectives de demande du produit fabriqué.

Les foudres machlupiennes s'abattent ensuite sur la courbe de coût variable moyen qui serait décroissante jusqu'à la *capacité*. Machlup déplore que le sens du terme *capacité* n'est pas précisé et suspecte que tous les entrepreneurs ne le comprennent pas identiquement. Lui-même envisage deux définitions possibles : le volume de production correspondant, soit au minimum de la courbe de coût moyen variable, soit au minimum de la courbe de coût moyen total à court terme. Avec ces définitions de la capacité, il conclut forcément que Lester enfonce une porte ouverte. Il y a ici un malentendu terminologique. Pour les antimarginalistes, la capacité est une frontière au-delà de laquelle les courbes de coût perdent leur signification. Pour Machlup, une telle frontière n'existe pas en tant que telle : la production peut toujours être augmentée ; si cela s'avère très difficile, la courbe de coût augmentera très fort. Il y a donc toujours une phase croissante dans la partie droite de cette courbe.

Machlup défend également le rôle central de la substitution entre les facteurs au sujet de laquelle Lester émettait de sérieux doutes. Elle est plus praticable que les critiques du marginalisme ne l'admettent. « Substitution between capital and labor does not have to take the form of changes in the machinery of the plant »[1]. Il donne l'exemple de tâches peu qualifiées ayant pour objet de garder les machines en bon état de marche ou de réduire les gaspillages : des recrutements à ces postes affectent l'intensité capitalistique et peuvent être rentables dans certaines circonstances.

Lester estime que le calcul marginaliste serait trop complexe dans une firme multiproduits. En réponse, Machlup prétend exactement l'inverse : contrairement au coût moyen qui nécessite une répartition des coûts indirects, le coût marginal est le seul qui est aisément séparable et déterminable.

De la vision marginaliste de la maximisation du profit, le portrait du businessman qui ressort est celui d'un homme « in a continuous state of alert »[2]. Toujours à l'écoute du moindre signal, il est en permanence prêt à modifier son prix ou le volume de sa production. Dans sa réponse à Machlup, l'anti-marginaliste **Oliver** estime que cette description s'écarte de la réalité. Un mélange de conservatisme et de prudence pousse au contraire les hommes d'affaires à maintenir le cap tant que les événements ne démontrent pas la nécessité de s'en départir. Un prix conventionnel est rassurant, tant sont nombreuses les inconnues qui rendent aléatoire les calculs de prix individuels.

C'est le dogme de la maximisation du profit qui est questionné ici. Maximisation du profit qui est quasiment synonyme de rationalité dans la conception marginaliste[3].

5.5.4. La polémique reprend de plus belle … et puis s'éteint

ANDREWS ET LE *NORMAL COST*

En 1949 paraît un des plus beaux exemples de manuel d'économie politique hétérodoxe, « Manufacturing Business » de l'économiste oxfordien Philip **Andrews**. La microéconomie y

[1] Machlup [243] pp. 150-151
[2] Oliver [272] p. 381
[3] Herbert Simon consacrera l'essentiel de sa carrière à introduire l'analyse des comportements à rationalité limitée dans la science économique.

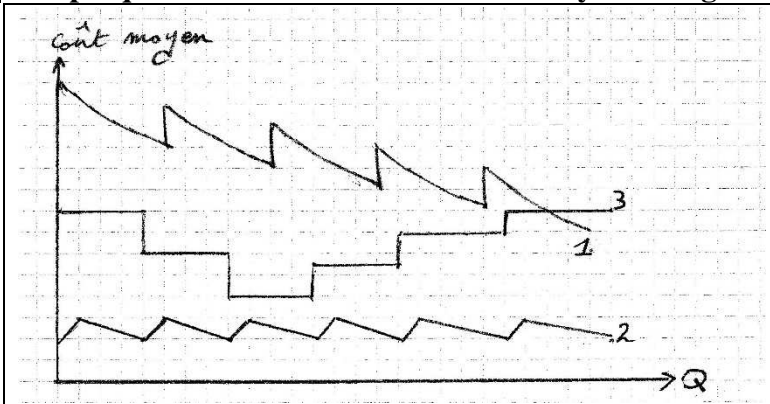
est réécrite, mais sans prétention, car l'auteur a aussi une approche empirique. Il décrit le comportement des entrepreneurs, non en vertu d'hypothèses et de principes comme la maximisation du profit, mais sur base de ses constatations sur le terrain. A de nombreuses reprises, il critique la théorie marginaliste qu'il juge éloignée de cette réalité.

Deux sujets retiennent particulièrement son attention : la courbe de coût moyen à LONG terme et la politique de prix de la firme.

Le coût moyen à long terme (décomposé sur le graphique 5.16) comprend le *coût technique de production* et le *coût managérial de production*. Le **coût technique de production** est le coût des facteurs techniques lorsqu'ils sont utilisés avec une efficacité optimale. Il a la forme en dents de scie. La longueur d'une dent correspond à celle de l'indivisibilité qui doit être renouvelée lorsqu'elle arrive à sa limite. Comme il y a des économies d'échelle à exploiter à tous les niveaux, la courbe a un trend légèrement déclinant.

Le **coût managérial de production** se compose du *coût des facteurs managériaux* et du *coût managérial technique*. Andrews considère que le volume des *facteurs managériaux* à mettre en œuvre évolue par paliers. Au début de chaque palier, le volume du management par unité produite augmente puis il diminue. Andrews critique abondamment la *théorie de l'inefficacité croissante du management* qui justifie la croissance finale de la courbe marginaliste de coût moyen à long terme. Un principe ad hoc, suspecte-t-il, car cette phase ascendante est indispensable à la détermination d'une taille optimale. Un principe jamais démontré, estime-t-il. Les techniques modernes de gestion des ressources humaines permettent de compenser l'éloignement de l'entrepreneur du terrain des opérations. Le *coût managérial technique* est le surplus du *coût technique réel* (c'-à-d réellement constaté) sur le *coût technique de production* tel que défini ci-dessus ; cet écart peut être considéré comme un indice inversé de l'efficacité de la gestion. Sur chaque palier, elle est horizontale. D'un palier à l'autre, elle est d'abord décroissante, puis croissante mais de moins en moins. Cette troisième forme n'est que faiblement argumentée et ne corrobore qu'imparfaitement les critiques d'Andrews contre la théorie de l'inefficacité croissante du management.

Graphique 5.16 : les éléments du coût moyen à long terme selon Andrews



Coût technique de production (1), coût des facteurs manageriaux (2) et coût managerial technique (3). Sur la présente figure, l'altitude relative de ces 3 courbes n'est pas significative.

La conclusion d'Andrews est qu'à partir d'une échelle importante, le coût de production global unitaire tend à devenir constant.

Pour la politique de prix, Andrews se réfère à une classification des coûts à court terme : les *coûts directs* sont ceux qui sont directement incorporés dans le produit.
La formule du ***prix demandé*** est : ***coût direct moyen plus marge de coût***.

La *marge de coût* doit permettre de couvrir tant les *coûts indirects* que le profit[1].

Tant le calcul du coût direct moyen que celui de la marge de coût impliquent le recours à un volume de production budgété. Si la production réelle dépasse cette quantité, il en découlera un bénéfice ; dans le cas contraire, c'est une perte qui s'ensuivra.

Voilà pour le prix. Concernant la quantité produite, Andrews réfute la formule d'égalité entre la recette marginale et le coût marginal. L'entreprise fournit au marché tout ce que celui-ci accepte de prendre au prix déterminé selon la règle du coût normal expliquée ci-dessus. L'entreprise épuise la totalité de sa demande ; l'autolimitation contenue dans l'égalisation de *Rm* avec *Cm* est contraire à ses intérêts.

Andrews critique vertement la théorie marginaliste de la concurrence sur les prix, qui prête aux entrepreneurs une vision se limitant au court terme. Alors qu'en réalité, les considérations à long terme sont déterminantes. « In general, the market of an individual business will expand only because it proves to be more efficient than its rivals »[2]. Espérer accroître sa part de marché en diminuant le prix sans diminution du coût est contraire aux intérêts de l'entreprise. « The typical business (…) will change its price only to the extent to which an actual or clearly foreseen change in cost would justify it ».

L'ouvrage d'Andrews eut à subir la critique d'**Austin Robinson**, le mari de Joan. Celui-ci reproche notamment à Andrews de prêter à la théorie marginaliste l'idée que l'entreprise maintient son profit à son niveau maximum EN RESTREIGNANT SA PRODUCTION. Une caricature, selon Robinson ! Robinson semble ne pas comprendre le propos d'Andrews. La restriction dont il est question ici ne résulte pas d'une volonté de créer la pénurie ; simplement l'égalisation entre Rm et Cm fait produire moins que l'épuisement de la demande au normal cost.

Austin Robinson épingle également cette contradiction chez Andrews : il met sans cesse en avant la concurrence, mais il nie la maximisation du profit traditionnellement considérée comme un comportement de survie dans un univers concurrentiel.

Andrews se voit aussi reprocher une autre contradiction, celle-ci relative à la *marge de coût*. D'une part, la détermination du prix à la façon d'Andrews privilégie totalement le coût par rapport à la demande. D'autre part, la discussion de cas particuliers amène fréquemment Andrews à faire dépendre implicitement la marge de coût de cette même demande. Robinson vise précisément à prouver qu'une théorie des prix ne peut pas se passer de la demande.

HARROD ET LA PREVENTION DE L'ENTREE

En 1952 paraissent les « Economic Essays » de Harrod. L'un d'eux, intitulé « Theory of Imperfect Competition Revisited » nous intéresse directement. Il s'attaque à ce qu'il appelle la « doctrine of excess capacity », c'est-à-dire l'idée initiée par Joan Robinson et Chamberlin que la production d'équilibre s'établit à un niveau plus bas que la capacité, comme l'indique la figure de droite du graphique 4.7. Ce phénomène est d'autant plus marqué lorsque l'accès au marché est libre et que les courbes RM et Rm ont subi des poussées vers la gauche à la suite de l'arrivée de nouvelles firmes sur le marché.

---

[1] Certains spécialistes réservent l'appellation *full cost pricing* à l'addition de deux marges séparées au coût direct, respectivement pour les coûts indirects et pour le profit. L'application d'une marge unique pour couvrir ces deux éléments est alors qualifiée de *normal cost pricing*. Fondamentalement, les deux reviennent au même.

[2] Andrews [9] p. 268

Il est absurde de la part d'une entreprise de réaliser des investissements qui bâtissent une capacité de production de *x* si la production optimale est *x-y*. Pour éviter ce gaspillage, autant construire directement une capacité de *x - y*. Mais comment éviter alors que la production optimale passe à *x - y - z* ? La solution proposée par Harrod évite ce cercle vicieux.

L'entrée potentielle de nouveaux concurrents constitue le nœud du raisonnement tout au long de l'essai. Harrod expose très explicitement une thèse qui sera encore développée par les *full costers* ultérieurement : la *prévention de l'entrée*. Les entrepreneurs ne sont pas stupides au point de vouloir accaparer un profit à court terme appelé à s'évanouir et qui, de plus, réduira leur part de marché à moyen terme. Comment parer à la menace ? « If (the firm) is to entrench itself firmly in available markets before new competitors arrive, it must confine itself to a normal rate of profit »[1]. La provision pour un profit normal est incluse dans le coût moyen, donc dans le full cost. Dans cette optique, le full cost pricing devient un instrument stratégique de la firme pour défendre son marché. Ce prix raisonnable lui assurera un marché plus étendu tout en rapportant le profit escompté. « If the representative entrepreneur is so actuated, new competitors will not normally appear. Thus, no excess capacity will be created »[2].

Agissant ainsi, les firmes font reculer l'incertitude. A contrario, « by charging the high price, (the firm) forgoes the present opportunity of establishing itself in a somewhat larger market, and thus deliberately makes its position weaker for the time when it has to face the incursion of new entrants »[3].

Selon Harrod qui a participé aux interviews de l'OERG, les entrepreneurs connaissent le niveau de prix à partir duquel la menace de nouveaux concurrents devient sérieuse.

Que l'entrée potentielle de nouveaux offreurs affecte ainsi les comportements justifierait que les économistes revoient leur classification des marchés. Harrod déplore que l'élasticité de la demande de la firme est érigée en critère suprême pour les départager et que, donc, tous ceux qui ont une courbe descendante ont tendance à être mis dans la même case, celle de la non-concurrence. L'accès ouvert ou fermé au marché lui paraît au moins aussi important. La concurrence imparfaite avec accès ouvert mériterait une appellation propre. Concernant cette lacune terminologique, il écrit : "We lack a vocabulary that is both well established and appropriate". Et il propose: "Free competition is an expression often used in popular literature, and it might be convenient to adapt this for technical purposes. It would be natural to use it for all cases where there is unrestricted (or relatively unrestricted) entry, and these would be divided into those of free competition with a perfect market and those of free competition with an imperfect market (downward-sloping short-period particular demand curve"[4].

L'équilibre du producteur œuvrant en *libre concurrence avec marché imparfait* est basé sur des considérations à long terme. Sa première décision consiste à déterminer la taille de son équipement. Il la calibrera de telle façon que le coût de production sera minimisé pour la production que ses études de marché l'amènent à anticiper. Désignons cette taille par x°. Le full cost y est un prix d'équilibre. Harrod montre que si le marché réel s'écarte légèrement des prévisions de l'entrepreneur (au-dessus ou en dessous de x°), le full cost sera quand même le prix le plus approprié.

---

[1] Harrod [134] p. 147.
[2] Harrod [134] p. 152
[3] Harrod [134] p. 147
[4] Harrod [134] p. 171 et 179.

Par rapport à la figure 4.7-B, l'équilibre est donné par la tangence entre une courbe de *demande à long terme* et la courbe de full cost. La courbe de *demande à long terme* intègre la vulnérabilité de la firme qui fixe un prix rapportant un profit à court terme ; elle est forcément beaucoup moins pentue que la demande à court terme ; elle est plate à gauche de x° et reste en-dessous de la courbe de coût moyen à droite de x°. La quantité d'équilibre est x°. Harrod semble satisfait de constater que son modèle sauvegarde l'égalité traditionnelle entre *Rm* et *Cm*, corollaire de la tangence des deux autres courbes. Il faut toutefois souligner que la signification de *Rm* a changé.

HEFLEBOWER ET LA FIN DE LA CONTROVERSE

Une conférence intitulée « Business Concentration and Price Policy » se tint aux Etats-Unis en 1953 : les économistes orthodoxes tentèrent de dégager des conclusions du débat, dans un sens qui leur est favorable, mais moins enclines à la polémique que les articles de Machlup. L'article « Full cost, Cost Changes, and Prices » d'Heflebower peut en être considéré comme la synthèse. En voici les lignes directrices.

D'abord, l'auteur admet le fait empirique de la courbe de coût marginal plate. « The conclusion seems clear that there is a substantial volume range within which marginal cost, particularly as viewed by management, are approximately constant, given constant factor prices ».[1]. Cette horizontalité caractérise l'intervalle d'environ 30% qui précède la capacité.

Heflebower estime que la pratique de *full cost* règne couramment en cas de collusion sur les prix, en cas de *price leadership* et est une hypothèse plausible pour la construction d'une théorie de l'oligopole. Par contre, sa présence est loin d'être démontrée dans les autres cas.

Certaines situations rendent hasardeuse la fixation des prix par le full cost. D'une part, les entreprises qui travaillent à la demande doivent fixer un prix spécifique à chaque commande, ce qui est peu compatible avec la relative rigidité affirmée par Hall et Hitch. Lorsque la part du coût direct est faible, l'usage du full cost tend à engendrer une diversité de prix. L'émergence d'un prix de marché est également difficile lorsque les coûts diffèrent sensiblement entre les producteurs ; les solutions proposées à ce problème divergent.

Heflebower reconnaît que l'usage de formules de prix basées sur les coûts est très répandu dans le monde de l'entreprise. Mais, estime-t-il, ces prix sont soumis au test du marché. « The present writer has found few exceptions to the rule that transaction prices vary to some degree relative to quoted prices when the latter do not move in response to demand »[2].

Heflebower retient deux apports positifs de la contestation de la doctrine orthodoxe :
- L'incertitude peut conduire les entreprises à attacher plus d'importance au coût qu'à la demande
- La maximisation du profit ne peut ignorer certaines considérations à long terme relatives au coût et à la demande.

On en revient à la thèse déjà évoquée que le full cost est une simple « rule of thumb » mais que le prix est expliqué par l'analyse marginaliste. Pressés de tourner la page, la majorité des

[1] Heflebower [148] p. 372
[2] Heflebower [148] p. 390

économistes se rallièrent à cette conclusion et l'intervention d'Heflebower constitue le point final de cette controverse qui se termina donc par une victoire marginaliste.

Les arguments des différents économistes « full costers » ne constituent pas un ensemble homogène ou même cohérent. Leur position générale a été déforcée par :
- la mise en cause de la maximisation du profit, inacceptable pour beaucoup d'économistes.
- la théorie de la demande brisée, peu convaincante
- l'absence d'explication algébrique de la marge de profit incluse dans le full cost
- l'absence de pont entre le coût de la firme individuelle et le prix du marché, si ce n'est la solution de facilité consistant à invoquer le *price leadership*[1].

Malgré la clôture du débat théorique après l'intervention de Heflebower, de très nombreuses études empiriques ont été menées au long du dernier tiers du XXe siècle et encore après ; elles confirment généralement les conclusions empiriques de l'OERG[2], à savoir le très large recours au full cost.

Je m'en voudrais de terminer ce chapitre sans avoir mentionné deux contributions de William **Baumol** qui n'adoptent pas la thèse du full cost mais attaquent la théorie orthodoxe en empruntant des chemins qui lui sont parallèles :
- la thèse selon laquelle les firmes oligopolistiques maximisent non leur profit mais leur chiffre d'affaires moyennant la contrainte d'un taux de profit minimum (1958).
- la théorie des *marchés contestables* (1982) voulant que lorsque l'entrée et la sortie de firmes sur un marché sont parfaitement libres et aisées, ce marché acquiert les qualités de la concurrence parfaite, même s'il est de nature oligopolistique. La raison en est le risque permanent qu'un outsider opère une incursion temporaire pour rafler le profit.

5.5.5. Approfondissement du *full cost pricing*

Le milieu et la fin des années cinquante voient diverses contributions compléter la théorie du full cost et combler certaines lacunes de sa première génération. Les progrès empruntent trois directions :
1- les modèles de prévention de l'entrée
2- le taux de profit cible
3- le passage du full cost de la firme individuelle au prix du marché.

Ces avancées se passent en dehors des projecteurs. Pour les économistes néoclassiques, le dossier est clos.

LA PREVENTION DE L'ENTREE

Des trois thèmes, c'est celui qui aura suscité la littérature la plus abondante. Chez Harrod qui a initié le sujet, le prix doit être fixé de telle façon qu'aucun profit net n'est généré, faute de quoi ce profit serait annihilé par l'arrivée de nouveaux concurrents. Le prix ainsi établi égale donc le coût de production puisqu'il ne peut inclure de profit. Cette conception n'est pas unanime. "Much of Andrews's writings suggests, however, that some premium in the form of a profit margin in excess of normal profits (…) can be secured where entry is not easy"[3]. Cette idée

[1] Ce concept, initié par Stackelberg, désigne la situation où une firme, généralement la plus importante, fixe le prix et où ses concurrents suivent comme s'ils acceptaient que ces prix leur soient imposés. Beaucoup de modèles de l'oligopole s'articulent sur ce concept.
[2] Le lecteur trouvera plus de précisions chez Nubbemeyer [269] et Lucas [235].
[3] Bhagwati [33] p. 302

sera développée par **Bain** et **Sylos-Labini** dans plusieurs ouvrages et articles ainsi que par **Modigliani** (1958) qui formalisera leur approche.

Le concept central, dans cette optique est celui que Modigliani appelle “entry preventing price” et Bain “limit price” : il s’agit du prix le plus élevé que les firmes en place pensent pouvoir demander sans susciter l’arrivée d’entrants.

Le modèle de Modigliani défend le plus ardemment l’existence de la prime précitée et analyse mathématiquement les facteurs qui l’influencent. Il repose sur deux idées-forces :

1- En concurrence imparfaite où les offreurs représentent plus que des éléments négligeables par leur taille ; l’arrivé d’un nouvel offreur réduira le prix du marché. C’est ce prix réduit qui devra permettre au nouvel arrivant d’atteindre la rentabilité, situation qu’il anticipera normalement. La compréhension de ce jeu offre donc aux firmes en place une marge de sécurité.
2- L’hypothèse de base du modèle de Modigliani est ce qu’il appelle le “Sylos postulate”. Sylos-Labini, analysant l’arrivée de nouvelles firmes, considère que les firmes existantes ne modifient pas leur production et se soumettent à la baisse de prix découlant de l’élargissement de l’offre[1].

Modigliani arrive à la formule suivante qui donne une approximation (en l’occurrence le maximum) du prix préventif d’entrée (P_0) :

$$P_0 \approx P_c\,(1+1/\varepsilon.S) \qquad (5.19)$$

Le prix P_0 égale le prix de concurrence parfaite P_c (correspondant au minimum du coût moyen) multiplié par un coefficient inversement proportionnel à la taille du marché (S)[2] et à l’élasticité de la demande (ε) aux environs de (P_c). P_0 sera également plus élevé si la courbe de coût moyen est plus pentue.

Bhagwati estime que la prime telle qu’elle ressort du modèle SBM (Sylos-Bain-Modigliani) est surestimée. D’une part, l’existence de cette prime rend l’entrée plus attractive et stimule donc les tentatives, ce qui a pour effet de réduire cette prime. Le risque d’échec sera plus facilement accepté si le profit potentiel est supérieur. D’autre part, il ne faut pas présumer avec trop de confiance que les firmes existantes survivront en priorité. Les pertes initiales sont souvent considérées comme inhérentes à l’investissement. Comme les entrants sont souvent des firmes multiproduits, elles disposent généralement de réserves leur permettant de les supporter. Bhagwati mentionne aussi le cas où la demande sur le marché est en croissance comme favorable à la réduction de la prime.

Sylos-Labini et Modigliani trouvent naturel de n’attribuer au nouvel arrivant quasiment aucune part de la demande préexistante. **Edwards** avance la rationalisation suivante : le goodwill attache les clients à leur fournisseur habituel, surtout dans le secteur des biens de production. Il reconnaît toutefois que si l’entrant fait découvrir à ces clients qu’ils étaient « exploités » avant son arrivée, il n’aura pas de peine à s’accaparer la plus grande part de la demande.

Edwards fait remarquer qu’outre le profit, l’inefficience des firmes en place est également un incitant à l’entrée. Si l’entrant est plus efficace, son succès est assuré.

---

[1] Il justifie ce comportement par la volonté de décourager les arrivants et la crainte de voir le coût unitaire croître si la production diminue.

[2] *S* est le ratio de la quantité absorbable par le marché au prix P_c sur la quantité à produire par la firme pour minimiser le coût moyen.

LE TAUX DE PROFIT CIBLE

En 1958, Robert **Lanzillotti** renoue avec la technique d'investigation consistant à interviewer les entrepreneurs quant à leur pratique. Le but de l'article "Pricing Objectives in Large Companies" est de remédier à l' "inadequate state of knowledge of the price-making process"[1]. Vingt entreprises ont été questionnées parmi les 200 plus grandes entreprises industrielles, qui faisaient face à des contextes concurrentiels variés. Les questions visaient à dégager l'objectif qui sert de guide dans la fixation des prix mais également de comprendre les procédures en vigueur.

Les objectifs les plus fréquemment cités sont :
1. pricing to achieve a target return on investment
2. stabilization of price and margin
3. pricing to realize a target market share
4. pricing to meet or prevent competition

Le taux de profit cible vient en tête. Une stratégie courante consiste à accepter des variations du taux de profit d'une année à l'autre, tant que l'objectif est atteint sur une période plus longue. Le calcul du prix se base sur un volume d'activité standard de façon à éviter que des fluctuations du taux d'activité réel n'affectent indûment le prix.

Lanzillotti assimile le deuxième objectif aux méthodes dites « cost-plus » dont fait partie le full cost pricing. Il estime que la distinction entre les deux premiers objectifs est difficile à définir et conclut : "Cost-plus, therefore, may be viewed as one step on a road to return-on-investment as a guide"[2].

Lanzillotti estime également que "target-return pricing implies a policy of stable or rigid pricing, even though exceptions are found within particular product lines"[3].

Sylos-Labini quant énonce une formule de la détermination de la marge de profit et du prix. Il en a donné plusieurs versions et a continué à la perfectionner après la publication de son fameux ouvrage "Oligopoly and Technical Progress". Telle qu'elle apparaît dans l'article « La théorie des prix en régime d'oligopole et la théorie du développement » (1971), elle s'inscrit clairement dans l'optique du taux de profit cible :

$$p = v + k/x_n + rK/x_n \qquad (5.20)$$[4]

p est le prix, *v* est le coût unitaire variable, *k* est le total des coûts fixes, x_n est la production annuelle, *r* est l'objectif de taux de profit et *K* est le capital engagé. Comme Lanzillotti, Sylos estime que « les grandes entreprises qui jouent le rôle de 'price leaders' ont l'intention de réaliser un taux de profit non dans chaque année isolée mais sur une série d'un certain nombre d'années »[5].

---

[1] Lanzillotti [212] p. 921
[2] Lanzillotti [212] p. 932
[3] Lanzillotti [212] p. 940
[4] En réalité, Sylos s'intéresse plus à la dynamique des variations du prix dans le temps qu'à la statique de la formation des prix. Selon lui, les fluctuations de la demande se traduisent dans des variations de la production et celles du coût donnent lieu à des adaptations du prix. Mais il croit que le prix des facteurs de production fluctue cycliquement. Grâce à leur caractère cyclique, les variations du coût peuvent n'être répercutées sur les prix que de façon amortie, sans que soit menacée l'obtention du taux de profit cible sur le long terme.
[5] Sylos-Labini [354] p. 250

D'après Sylos, le taux de profit cible est en quelque sorte « le taux permis par les barrières de protection dont jouit l'entreprise. Et puisque celles-ci sont différentes entre les divers marchés et même à l'intérieur de chaque marché, les taux de profit pris comme objectif seront différents »[1].

LE PRIX DU MARCHE

Chaque entreprise calcule et applique son full cost pour déterminer son prix d'offre, mais si les acheteurs sont rationnels, il ne peut y avoir qu'un seul prix de marché. Comment celui-ci s'établit-il ?

Traditionnellement, les full costers recourent au concept de "Price Leadership" pour dénouer ce problème. Andrews, suivi en cela par Sylos-Labini, en établit le paradigme comme suit : La firme qui a la taille la plus élevée a un coût de production plus bas, ce qui fait d'elle le price leader, car inévitablement elle tentera d'imposer un prix basé sur ce coût.

Dans un article intitulé "Competitive Price Leadership : A Critique of Price Leadership Models" paru en 1957, **Lanzillotti** montre les carences des modèles traditionnels du price leadership et tente d'en promouvoir une nouvelle conception.

Selon les modèles traditionnels, le *price leader* est clairement identifié et les rôles sont codifiés : il sonne la cloche et les suiveurs appliquent automatiquement les hausses et baisses de prix qu'il estime justifiées. Lanzillotti reproche à ces modèles leur caractère statique. Les éléments dynamiques d'où émanent ces comportements sont ignorés. "Moreover, the models appear to be based largely on highly institutionalized structures wherein interfirm price relationships are essentially settled, under which circumstances price leadership emerges as a type of collusion with the ringleader clearly identifiable"[2]. D'après Lanzillotti, les marchés de l'économie réelle se caractérisent plutôt par leur instabilité ; la faiblesse des modèles traditionnels pour rendre compte de leur fonctionnement est patente. Ils ne s'avèrent pertinents que pour un type de cas très particulier.

Lanzillotti propose un nouveau modèle qu'il appelle le "Competitive Price Leadership", du fait que "the prices set are those which materialize from the operation of competitive forces"[3], ce qui contraste avec les comportements collusifs des modèles antérieurs. Il donne une description des caractéristiques de ces marchés qu'il m'est impossible de rapporter en peu de lignes. Résumons-la ainsi : une concurrence véritable, mais avec un nombre restreint de firmes.

A mon avis, ce que Lanzillotti décrit est tout simplement la concurrence telle qu'elle existe, non dans les ouvrages d'économie politique mais dans la réalité. L'attribution de l'étiquette "price leadership" me paraît inutile et même confondante. Nonobstant cette remarque, le modèle de Lanzillotti est une avancée remarquable.

Comment fonctionne le "competitive price leadership" ? Farm, un auteur contemporain inspiré par Lanzillotti, répond : "In this model, the market price goes down if and only if a price cut appears profitable for a firm even if its competitors follow suit, while the market price goes up

[1] Sylos-Labini [354] p. 256
[2] Lanzillotti [211] p. 55
[3] Lanzillotti [211] p. 55

if and only if a higher market price is profitable for every firm. Thus, the market price is determined by the lowest market price preferred by a firm…”[1].

Cette règle a été appliquée implicitement par Sylos-Labini. Je dis « implicitement » parce que le modèle de Sylos fonctionne avec ses propres préoccupations et des hypothèse particulières : y opèrent sur le marché, trois groupes de firmes, les petites, les moyennes et les grandes, ayant des coûts de production décroissants dans cet ordre. Le problème consiste à déterminer vers quel prix et quelle répartition des entreprises entre ces trois groupes on s'oriente, compte tenu de la demande totale, de la situation initiale et des différentes structures de coût. Bien évidemment, les grandes firmes ont le privilège de fixer le prix et déterminent donc selon leur avantage quelles catégories de firmes survivront à leurs côtés. La problématique de l'élimination de concurrents se combine avec celle de la prévention de l'entrée. La conclusion de Sylos est celle-ci : “the price tends to settle at a level immediately above the entry preventing price of the least efficient firms which it is to the advantage of the largest and most efficient firms to let live”[2].

5.5.6. Théorie évolutionniste de la firme

Milton **Friedman** publie en 1953 un essai consacré à la méthodologie en science économique, qui fit grand bruit. En voici la thèse centrale : « the only relevant test of the validity of a hypothesis is comparison of its predictions with experience. The hypothesis is rejected if its predictions are contradicted (“frequently” or more often than predictions of alternative theories) »[3]. Certes la confirmation des prédictions par l'expérience ne garantit pas la validité de la théorie ; mais elle renforce son crédit puisqu'une occasion de l'infirmer vient d'être manquée. Puisqu'il s'agit là du seul critère pertinent pour apprécier une théorie, sa critique sur base du REALISME DE SES HYPOTHESE est une erreur fondamentale. La théorie a besoin des hypothèses pour isoler les relations importantes de leur magma contextuel. Selon Friedman, si on imposait aux hypothèses la condition du réalisme, elles ne serviraient à rien. Par exemple, si on devait décrire réalistement le marché du blé, on noierait le lecteur avec une masse de données inutiles comme l'âge des intervenants, le nombre de leurs enfants… Théoriser deviendrait impossible ; or nous ne percevons la réalité que grâce aux théories.[4]

Comme toujours, une affirmation a ses motivations. Ce qui dérange Friedman, ce sont les critiques sempiternelles contre le défaut de réalisme des hypothèses de la théorie néoclassique. Friedman prend le débat sur le marginalisme comme exemple. Que firent Hall et Hitch, Lester et Oliver, sinon précisément contester le réalisme de l'hypothèse marginaliste de la maximisation du profit, que Friedman reformule ainsi : « … under a wide range of circumstances, individual firms behave *as if* they were seeking rationally to maximize their expected returns (generally if misleadingly called profits) and had full knowledge of the data needed to succeed in this attempt ;… »[5]. Avec ce que cela implique de connaissance des fonctions de demande et de coût, de calcul marginal... Friedman souligne la clause *as if*, la base de toute hypothèse. Selon Friedman, la confirmation des prédictions basées sur cette hypothèse s'accumule depuis des décennies ; malheureusement, il n'étaye pas cette affirmation.

---

[1] Farm [89] p. 1
[2] Sylos-Labini [353] p. 50
[3] Friedman [98] pp. 8-9
[4] Cette argumentation est sophistique. Une bonne hypothèse simplifie la réalité, mais une mauvaise hypothèse la travestit. Une hypothèse comporte toujours une part d'interprétation de la réalité et rien n'empêche les interprétations douteuses.
[5] Friedman [98] p. 21

Friedman introduit dans son discours un argument qui sort du cadre de la méthodologie qui semblait en former la trame : « Unless the behavior of businessmen in some way or other approximated behavior consistent with the maximisation of returns, it seems unlikely that they would remain in business for long (…) the process of "natural selection" thus helps us to validate the hypothesis… »[1].

Nous avons vu que les (ou DES) full costers s'étaient montrés ambigus voire carrément sceptiques quant à la maximisation du profit. L'argument de la sélection naturelle les vise certainement. La sélection naturelle peut se révéler un argument efficace contre la négation de la maximisation du profit, mais il semble évident qu'il ne pèse pas lourd contre l'objection selon laquelle l'égalisation de la recette marginale avec le coût marginal est une mauvaise voie pour maximiser le profit à long terme. L'argument de la sélection naturelle pourrait même être renversé au profit du full cost pour expliquer la prépondérance persistante de celui-ci dans les études empiriques. La prévention de l'entrée et la réduction des capacités inutilisées favorisent le profit à long terme.

Friedman a repris le thème de la sélection naturelle chez **Alchian**, un économiste proche de lui, qui l'avait exposé trois ans plus tôt dans un article intitulé « Uncertainty, Evolution and Economic Theory ». Résumons cet article, qu'on considère généralement comme le point de départ des théories évolutionnistes en économie.

L'incertitude rend vaine toute axiomatisation de la maximisation du profit, puisque les rendements des différentes actions entre lesquelles il faut choisir sont inconnus[2]. Alchian constate que les manuels d'économie ont pour habitude d'exposer une règle précise pour la maximisation du profit, et qu'ensuite ils assortissent le théorème d'un commentaire succinct réduisant sa portée en remarquant que l'incertitude écarte la réalité de ce schéma théorique. Il serait plus logique que la théorie prenne en compte directement l'incertitude, quitte à donner un empire plus modeste à ses théorèmes. Tel est l'objet de sa démarche.

Alchian insiste sur la distinction entre l'INTENTION d'une action et son RESULTAT. A propos de l'intention, on peut ergoter sans fin ; le résultat, par contre, est objectivable. Or c'est le résultat qui fait le succès d'une action et cela tout à fait indépendamment de l'intention. En économie, sur une longue période, le succès se marque par la durée de vie de l'entreprise. Et celle-ci est assurée par la réalisation de profits. La maximisation du profit n'est pas indispensable. L'essentiel est de faire mieux que les concurrents. « Even in a world of stupid men, there would still be profits »[3].

Que peut et doit faire une théorie économique dans ces conditions ? « With a knowledge of the economy's realized requisites for survival and by a comparison of alternative conditions, (the economist) can state which types of firms or behavior relative to other possible types will be more viable, even though the firms themselves may not know the conditions or even try to achieve them by readjusting »[4]. Alchian relève toutefois un facteur tendant à induire le comportement adéquat : la tendance à imiter les comportements couronnés de succès.

---

[1] Friedman [98] p. 22

[2] On pourrait certes sélectionner les actions sur base de l'espérance mathématique du rendement, mais il faut alors recourir à une fonction d'utilité pour départager l'espérance de gain et le risque qui lui est associé. Bref, la maximisation du profit, comme telle, se révèle insuffisante.

[3] Alchian [3] p. 213

[4] Alchian [3] p. 216

5.5.7. Le *mark-up* de Kalecki

On ne peut clôturer ce chapitre sur le *full cost pricing* sans remarquer la similitude avec la théorie du *mark-up* de Kalecki exposée dans son article « Costs and Prices » (1943). Avec les Cambridgiens, Kalecki est resté méfiant à l'encontre des full costers, mais à mon sens, cette attitude résulte plus de l'instinct de groupe animant des milieux rivaux que d'une véritable opposition de fond.

Kalecki distingue deux groupes de produits selon le principe qui régit la fixation des prix.

- Pour les matières premières, la DEMANDE est déterminante, du fait de l'inélasticité de l'offre.
- Le prix des produits finis est déterminé par leur COUT, car l'augmentation de la demande suscite l'accroissement de l'offre. C'est à ce deuxième type de produit que l'article s'intéresse.

Considérerons d'abord une firme prise individuellement. Supposons qu'elle dispose d'une réserve de capacité et que sa courbe de *coût direct moyen* est horizontale. Kalecki ajoute: « In view of the uncertainties faced in the process of price fixing, it will not be assumed that the firm attempts to maximize its profits in any precise sort of manner »[1].

La firme œuvre dans un environnement de concurrence imparfaite. Elle fixe son prix selon la formule suivante :

$$p = m.u + n.\bar{p} \qquad (5.21)$$

où *u* est le coût direct unitaire ;
$\bar{p}$ est le prix moyen pondéré demandé par les firmes sur ce marché [2];
m est un multiplicateur de marge, toujours positif ;
n représente la sensibilité de la firme aux prix des concurrents. On a $0 < n < 1$.

Kalecki démontre que le rapport m/(1-n) peut être considéré comme un indice du degré de pouvoir monopolistique.

L'influence des prix constatés chez les concurrents est la particularité la plus originale du modèle de Kalecki.

Comme on le voit- et Kalecki insiste sur ce point-, le coût indirect (overhead) n'influence pas directement le prix. Comme chez Andrews (mais pas Hall et Hitch), la marge de profit est censée le couvrir.

Venons-en à l'émergence du prix de marché. Additionnons les équations (5.21) des différents offreurs, pondérées par leur part de marché. Nous obtenons :

$$\bar{p} = \frac{\bar{m}}{(1-\bar{n})} .\bar{u} \qquad (5.22)$$

La barre sur une lettre indique qu'on a affaire à la moyenne des firmes individuelles.

---

[1] Kalecki [180] p. 44
[2] L'intervention de cette variable est l'une des particularités du modèle de Kalecki qui le distingue des full costers

Envisageons le cas de la firme représentative, dont *m* et *n* sont au niveau moyen. La formule liant son prix et son coût direct est :

$$\mathrm{p/u} = \overline{m} + \overline{n}.\overline{p}/u \qquad (5.21/5.23)$$

"If *m* and *n* are constant, the long-run changes in prices will reflect only the long-run changes in unit prime costs. Technological progress will tend to reduce the unit prime cost *u*. But the RELATIONS between prices and unit prime cost can be affected by changes in equipment and technique only to the extent to which they influence the degree of monopoly"[1].

Kalecki examine les facteurs qui accroissent le degré de monopole sur la longue période :

- La concentration industrielle formant des firmes géantes conscientes de l'influence de leur *p* sur $\bar{p}$ et donc enclines à le faire monter.
- L'importance de la préférence de marque (publicité etc.)
- Le poids des coûts indirects, car comprimant le profit, ils incitent à la conclusion d'accords visant à le préserver. Kalecki leur reconnaît donc une influence INDIRECTE sur le prix. Comme en dépression, le coût direct subit une diminution plus marquée que l'overhead, cette phase du cycle est la plus propice à l'élévation du degré de monopole (cf. 5.1.5).
- La force des syndicats. Elle incite les entreprises à modérer leur marge pour éviter la spirale revendicatrice entre les prix et les salaires.

Il est couramment présumé qu'à long terme, l'accroissement tendanciel de l'intensité capitalistique (rapport capital fixe/production*)* pousse à la hausse le rapport *p/u*. Les statistiques de la période 1879-1937 ne confirment une hausse du rapport prix/coût que pour de courtes sous-périodes (1879-1889 et 1923-1929). Kalecki l'explique ainsi : la hausse du rapport capital fixe/production peut être compensée par la baisse du rapport (profit + overhead) /capital fixe. En conclusion, la hausse de l'intensité capitalistique n'entraîne pas nécessairement celle du degré de pouvoir monopolistique.

[1] Kalecki [180] p. 52

TITRE VI- LES NEO-WALRASSIENS

Au cours des dernières décennies du vingtième siècle, l'économie politique a changé de visage. Les composantes marshallienne, wicksellienne ou parétienne de l'économie néoclassique se sont progressivement effacées au profit de l'école dite néo-walrassienne. C'est le grand retour au modèle d'équilibre général exposé par Walras à la fin du XIXe siècle et perfectionné au XXe.

Cette émergence est le fruit d'une lente évolution au long du XXe siècle, que nous allons exposer maintenant, ce qui nous forcera à un retour en arrière, alors que jusqu'à présent nous avions plutôt bien respecté l'ordre chronologique.

Il convient de préciser que la « nouvelle économie » atteint un haut niveau d'abstraction et que l'utilisation des mathématiques y est fort poussée. Notamment, les vecteurs et les matrices interviennent dans la plupart des modèles. Précisons encore que nous nous conformerons à l'habitude courante de noter les matrices par une lettre majuscule en caractère gras et les vecteurs par une lettre minuscule en caractère gras.

6.1- WALRAS : UN NOUVEAU DEPART

6.1.1. Le modèle de walras-cassel

CASSEL, CONTRE LES THEORIES DE LA VALEUR

En 1918, l'économiste suédois Gustav Cassel publie en langue allemande son traité d'économie politique « Theoretische Sozialökonomie ». Cet ouvrage couvre l'ensemble des questions économiques, mais nous limiterons notre analyse à sa partie microéconomique.

Comme dans beaucoup de traités, les premiers chapitres sont consacrés à l'échange. L'échange est réglé par ce que Cassel appelle le *principe de rareté*, qu'il définit ainsi : « The fixing of prices thus has the social economic task of so restricting the demands for goods that they can be met with the available means. »[1]. Cette définition se rapporte a l'échange pur, mais Cassel l'élargit a la production : "The demand for commodities is indirectly a demand for factors of production (…) it represents a conflict between consumers for the factors of production, a conflict which can be adjusted only by placing on the factors of production prices high enough to effect coincidence of demand with the available quantities of the factors of production"[2].

Cassel attaque le concept de VALEUR, parfaitement inutile selon lui. Ce qui compte, ce sont les PRIX, qui sont des faits observables. Il critique donc les deux principales théories de la valeur que sont la théorie de la valeur-travail et la théorie de l'utilité marginale et principalement cette dernière, car Cassel estime qu'elle pollue la tradition néoclassique à laquelle il adhère. Cette théorie est à la fois superflue et erronée. Le cerveau du consommateur n'est pas capable de déterminer un niveau d'utilité pour chaque degré de satisfaction de chaque besoin et encore moins de déterminer in abstracto un prix acceptable pour chaque bien sur pareille base. « This whole scale of reckoning is thus necessarily bound with existing prices »[3]. Le consommateur peut se positionner par rapport à des variations du système de prix existant, pas en concevoir un ex nihilo

Cassel enchaîne avec la relation entre le prix et le coût, dans laquelle il fait intervenir quatre principes. Du *principe de rareté*, Cassel déduit le *principe du coût* : « The pricing process which we have here described means that every finished commodity receives a price which corresponds to its cost of production, or, more generally, that every demand shall bear the whole cost of satisfying it »[4]. Le principe du coût doit être complété par le *principe différentiel* voulant que ce soit le coût du fournisseur le moins performant qui détermine le prix (unique) du marché et par le *principe de substitution* selon lequel parmi les méthodes de production accessibles, c'est la moins chère qui détermine le prix.

Comme Cassel rejette le concept de valeur, sa théorie explique les prix directement en monnaie. Il ironise contre l'habitude des économistes de concevoir l'économie comme un grand troc sur lequel on vient greffer la monnaie par la suite[5]. Dans sa critique, Wicksell estime au contraire

[1] Cassel [46] p. 66
[2] Cassel [46] p. 90
[3] Cassel [46] p. 81
[4] Cassel [46] p. 93
[5] Au risque de se contredire, il prétend ailleurs que la demande n'est fonction que des prix relatifs et que les prix absolus ne peuvent être expliqués sans une théorie de la monnaie.

que cette abstraction consciente de la monnaie « was the decisive step which first gave economics a truly scientific character »[1].

Après avoir analysé l'échange des biens, Cassel s'intéresse aux marchés des facteurs de production et donc à la répartition du revenu entre les classes sociales.

Concernant l'entrepreneur, Cassel distingue le *profit brut* et le *profit net*. Le profit brut est la rémunération que perçoit l'entrepreneur. Si celui-ci travaille dans son entreprise, une partie du profit brut est un salaire caché pour son travail. S'il y a investi du capital, une autre partie est un intérêt caché sur ce capital, en ce compris une prime de risque. Le profit net est le solde après déduction de ces éléments. Le niveau de ce profit est fort aléatoire et échappe à toute rationalisation.

Comme la théorie de l'utilité marginale ci-avant, Cassel conteste la théorie de la productivité marginale. Il invoque de nouveau cet argument : la productivité marginale dépend autant du prix qu'elle ne l'influence. Il reproche à cette théorie de surestimer la variabilité des facteurs de production. Néanmoins, son *principe de substitution* implique une certaine variabilité. La hausse du prix relatif d'un facteur stimulera sa substitution par un autre facteur, qui à son tour modérera la hausse du prix du facteur. Cela dit, la substitution entre les facteurs joue dans la formation de leurs prix un rôle secondaire en comparaison avec leur rareté. Concernant la substitution entre travail et capital, Cassel remarque qu'en cas de hausse généralisée des salaires, le prix des biens d'équipement s'en trouvera haussé, ce qui freinera la substitution.

En conséquence, les revenus payés pour les facteurs de production dépendent plus de leurs raretés relatives (relation entre l'offre et la demande) que des conditions techniques. Si la collectivité veut infléchir la répartition des revenus dans un sens moins inégalitaire, elle aurait tort de violenter le principe de rareté. Il est préférable de jouer avec les déterminants de l'offre et de la demande. Par exemple, par l'éducation, accroître l'offre de travail dans les qualifications plus productives et mieux payées.

Venons-en à la théorie du capital et de l'intérêt. Dans le système des prix, l'intérêt est un prix comme un autre, mû par le principe de rareté. Mais de quoi est-il le prix ? Non pas du capital, encore moins des biens-capitaux, mais de ce que Cassel appelle l'*attente* ou la *disposition du capital*. L'unité de mesure de la quantité d'attente est *M.T*, c'est-à-dire le montant en monnaie multiplié par le temps.

La façon dont Cassel explique la nature et la fonction de l'*attente* rappelle le rôle du temps dans la théorie autrichienne. Mais il y a une différence fondamentale. La théorie autrichienne considère le capital comme un facteur de production secondaire constitué par incorporation dans le temps des deux facteurs primaires que sont le travail et la terre. Pour Cassel, il y a trois facteurs primaires sur pied d'égalité : le travail, la terre et l'attente[2].

Cassel critique d'ailleurs abondamment la conception de l'intérêt de Jevons, Böhm Bawerk et Fisher. La sous-estimation des jouissances futures par rapport aux présentes est un facteur inutile et partiellement erroné. Quant à la thèse que l'intérêt correspondrait à la productivité marginale de l'allongement de la période de production, Cassel considère qu'elle rend mal

[1] Wicksell [385] p. 223

[2] Wicksell, généralement peu complaisant à l'égard de son collègue compatriote, reproche à Cassel cette conception hybride où le temps intervient mais où son rôle n'est pas distinct de celui des facteurs primaires. Selon lui, il fallait jouer soit la carte Walras (conception synchronique), soit la carte autrichienne.

compte du capital fixe et il lui paraît possible d'étendre la production en augmentant le capital sans augmenter la période de production[1].

Comme tout prix, l'intérêt est animé par le principe de rareté. Il sert donc à restreindre la demande au niveau de l'offre. Le fait qu'une épargne considérable existerait déjà si le taux d'intérêt était nul importe peu. D'autant plus que Cassel est convaincu qu'avec un taux nul, la demande de capital tendrait vers l'infini. Cassel pense également que le taux d'intérêt devrait rester positif dans une économie stationnaire, car il faudrait toujours prévenir la consommation du capital.

Cassel reconnaît trois mobiles à l'épargne :
- le souci de son avenir qui anime chaque individu
- dans le chef des entrepreneurs, le désir de renforcer leur entreprise
- dans le chef des riches capitalistes, le prestige de la grande fortune

Cassel ne contredit pas catégoriquement l'idée généralement admise que l'épargne croît avec sa rémunération, mais il la tempère. Analysant les réflexes des intervenants mentionnés ci-dessus, il arrive à la conclusion que des variations du taux au sein d'une fourchette normale influeront peu sur le volume de l'épargne ; par contre, une baisse prolongée du taux qui l'amènerait en dessous de la limite inférieure de la fourchette entraînerait une chute brutale de l'épargne. L'analyse à court terme n'aurait donc pas tort de concevoir l'offre de capital comme donnée.

Entre les chapitres sur l'échange des produits et sur l'échange des facteurs, Cassel intercale un chapitre qui met l'équilibre général de l'économie en équations. Ce chapitre est bizarrement placé, car Cassel est obligé de prendre des hypothèses conservatoires simplificatrices concernant les facteurs de production et après l'analyse de ceux-ci, il ne revient plus sur ses équations[2]. Ces équations rappellent évidemment celles de Walras que Cassel se permet de ne pas citer. Elles s'en écartent toutefois, notamment par son refus de la notion d'utilité. A vrai dire, alors que les équations constituent la charpente des *Eléments* de Walras, celles de Cassel sont très sommaires en comparaison avec le contenu de la *Sozialökonomie*. Elles ont pourtant joué un rôle important dans l'histoire de la science économique, ce qui justifie la présence de Cassel dans ce chapitre.

Il y a quatre grands groupes d'équations dans lesquels interviennent *n* produits de consommation (indicés *i*) et *r* facteurs de production (indicés *j*). Les variables sont les *p* (prix des produits de consommation), les *q* (prix des facteurs), les *a* (coefficients techniques), les *N* (demande des produits), les *S* (offre des produits) et les *R* (offre de facteurs).

Le prix des produits s'égalise avec leur coût de production :

$$a_{i1}.q_1 + a_{i2}.q_2 + \ldots + a_{ir}.q_r = p_i \qquad \text{pour } i = (1\ldots n) \qquad (6.1)$$

Connaissant ces prix, on peut en déduire les demandes :

$$D_i = F_i\,(p_1, p_2, \ldots, p_n) \qquad \text{pour } i = (1\ldots n) \qquad (6.2)$$

Ces demandes doivent être égales aux offres :

$$D_i = S_i \qquad \text{pour } i = (1\ldots n) \qquad (6.3)$$

Connaissant l'offre des produits, on peut en déduire la demande qui sera faite des différents facteurs de production, et celle-ci doit s'adapter à leur offre :

[1] Cette question sera débattue par Hayek et Knight (cf. sous-chapitre 9.2.1).
[2] Wicksell se demande comment Cassel intégrerait sa théorie du capital et de l'intérêt dans ses équations.

$$R_j = a_{1j}.S_1 + a_{2j}.S_2 + ... + a_{nj}.S_n \qquad \text{pour } j = (1 ...r) \qquad (6.4)$$

Sur base de cet enchaînement, Cassel conclut à l'existence de l'équilibre, sans avoir à procéder à un comptage global des équations et des inconnues[1].

Les hypothèses simplificatrices sont :
- la consommation des ménages est indépendante de la vente des facteurs
- les coefficients techniques a_{ij} sont fixes (malgré le principe de substitution)
- l'offre des facteurs (les R_j) est fixe. Les conclusions mitigées du chapitre consacré aux facteurs ne démentent pas cette hypothèse mais ne la confortent pas non plus.

Cassel introduit ensuite dans son système un taux de *croissance uniforme*[2] (dans le temps et entre les secteurs) de c%. En fait, il se contente d'une croissance sans progrès technique et sans accroissement de la richesse par individu. La quantité de tous les facteurs et les quantités produites augmentent uniformément de c%. Les prix d'équilibre restent invariables.

Des équations (6.1) à (6.4), Cassel tire les conclusions suivantes :
- La variation de la demande d'un produit n'affecte son prix que si elle entraîne un changement dans la demande des facteurs ou dans la méthode de production. Dans le cas des rendements croissants ou des rendements décroissants, le prix subira l'effet du changement d'échelle. D'une façon générale, les variations du coût de production constituent le principal levier des variations du prix.
- Les prix subissent l'influence de facteurs OBJECTIFS (les R_j et les a_{ij}) et de facteurs SUBJECTIFS (les F_i). Isolées, la théorie du coût et celle de l'utilité marginale souffrent donc d'une insuffisance constitutive.
- La répartition des revenus fait intrinsèquement partie de la problématique de la formation des prix ; c'est une erreur de la considérer comme un chapitre à part de la théorie économique. Ici, Cassel se contredit, car il fait suivre ses équations d'un chapitre consacré aux facteurs de production et les équations sont loin de rendre compte des subtilités exposées dans ce chapitre.

LE « COLLOQUE VIENNOIS » ET LE MODELE DE WALD

Le livre de Cassel eut un certain retentissement ; c'est par lui que la pensée walrassienne s'introduisit à Vienne. Dans la capitale autrichienne entre les deux guerres, de jeunes mathématiciens d'Europe centrale. Karl Menger (le fils de Carl Menger), Karl Schlesinger, Hans Neisser, Abraham Wald et d'autres se mobilisèrent sur la problématique de l'équilibre général et tentèrent d'améliorer le modèle Walras-Cassel.

Première avancée, la démonstration de l'existence de l'équilibre général par Walras n'était pas satisfaisante d'un point de vue mathématique. L'égalité entre le nombre d'équations et le nombre d'inconnues n'est ni une condition suffisante, ni une condition nécessaire pour l'existence d'une solution d'équilibre. Wald apportera une nouvelle démonstration de cette existence, fondée sur la *programmation linéaire*, encore en développement à cette époque.

---

[1] Le nombre d'équations (3n+r) est certes égal à celui des inconnues (p_i, D_i, S_i, q_j), mais Cassel ne justifie pas l'existence de l'équilibre par cette égalité.

[2] Ce type de croissance stable apparaîtra dans de nombreux modèle. Je la désignerai dans la suite par le qualificatif *semi-stationnaire* que j'emprunte à Bliss qui la caractérise ainsi : « the development of the economy through time is assumed to exhibit a repetitive aspect so that what happens in each week is the same as what happens in every other week, except perhaps for a change of scale » ([38] p. 62). La terminologie "steady state" est courante chez nombre d'économistes.

Deuxième avancée : « Professor Neisser remarked that the casselian system might have negative values of prices or quantities as solution. Negative quantities are clearly meaningless and at least in the case of capital and labor, negative prices cannot be regarded as acceptable solutions since the supply at those prices would be zero »[1]. Comment nos mathématiciens ont-ils résolu ce problème ?

Rappelons-nous deux égalités fondamentales du modèle Walras-Cassel :
1. l'égalité (6.4) entre l'offre (la quantité disponible) d'un facteur et la demande de ce facteur.
2. l'égalité (6.1) entre le prix des produits et leur coût de production.

La solution consiste à remplacer ces égalités par des inégalités, moyennant l'ajout d'une contrainte collatérale. Le prix doit être inférieur ou égal au coût de production, mais s'il est inférieur, la quantité sera nulle ; l'offre du facteur doit être supérieure ou égale à la demande mais si elle est supérieure, le prix doit être nul (bien libre).

Le modèle de Wald de 1936

Abraham Wald a démontré deux théorèmes énonçant l'existence de l'équilibre général. Le premier revient aux fondements walrassiens, mais il réduit son champ d'analyse à l'échange pur (absence de facteurs de production). Les agents offrent et demandent des biens de façon à maximiser leur utilité en respectant une contrainte budgétaire. S'y ajoute une contrainte d'égalité de l'offre et la demande au niveau du marché. Comme celles de Walras, les fonctions d'utilité de Wald postulent que l'utilité marginale de chaque bien est indépendante de la quantité que l'agent détient des autres biens.[2]

Nous n'analysons ici que l'autre modèle, qui est d'inspiration cassellienne. Wald adapte le système d'équations de son prédécesseur. Présentons le en gardant la notation de Cassel :

$$R_j = a_{1j}.S_1 + a_{2j}.S_2 + ... + a_{nj}.S_n + u_j \qquad (j = 1...r) \qquad (6.5)$$

$$u_j.q_j = 0 \qquad (j = 1...r) \qquad (6.6)$$

$$p_i = \Sigma_{j=1,}{}^{r}\, a_{ij}.q_j \qquad (i = 1...n) \qquad (6.7)$$

$$p_i = f_i(S_{1...}S_n) \qquad (i = 1...n) \qquad (6.8)$$

- L'équation (6.5) correspond à (6.4) avec en plus le terme u_j qui représente à la part non utilisée du facteur j, qui sera nulle si le facteur est rare. C'est une manière de transformer la première égalité cassellienne en inégalité (cf. supra).
- L'équation (6.6) dit que soit le prix, soit la quantité non utilisée du facteur doit être nul. C'est la contrainte collatérale mentionnée ci-dessus. Walras et Cassel avaient simplement éliminé les facteurs non rares de leur modèle ; Wald estime qu'un facteur n'est pas rare par nature, mais que sa rareté est elle-même une variable endogène. « Whether factors are free or scarce cannot be considered, a priori, a datum of the economy; it can only be determined on the basis of the production equations »[3].
- L'équation (6.7) correspond à (6.1).
- Enfin, (6.8) cumule (6.2) et (6.3), mais on remarque avec étonnement que Wald inverse la relation entre le prix (p_i) et la quantité échangée (S_i) généralement admise.

[1] Arrow-Debreu [17] p. 288
[2] C'est un retour en arrière par rapport à Edgeworth et Pareto.
[3] Wald [378] p. 371

Les inconnues sont les u_j, q_j, p_i, s_i,. Wald a démontré l'existence d'une solution non négative unique pour le système d'équations (6.5 à 6.8), pour autant que soient satisfaites les conditions suivantes :

1- Il n'y a pas d'épargne
2- une seule technique de production est possible par produit (coefficients fixes)
3- $R_i > 0$: existence d'une quantité positive de chaque facteur
4- $a_{ij} \geq 0$: la quantité d'un facteur utilisée dans la production d'un bien ne peut être négative
5- pour chaque bien, au moins un facteur est nécessaire à sa production
6- La relation entre le prix et la quantité demandée est continue
7- La demande d'un bien ne s'annule que lorsque son prix devient infiniment grand.
8- Si l'on observe les équations (6.7), les a_{ij} constituent une matrice de *r* colonnes et *n* lignes. Elle doit être de rang *r*.
9- Last but not least, Wald impose une restriction sur la fonction de demande, qui revient à formaliser la condition que le comportement des consommateurs doit être rationnel. Examinons-la de plus près.

Supposons qu'une économie passe d'un système de prix **p** à un système **p'**[1]. Considérons la fonction de demande générale, c'est-à-dire multiproduits[2] que nous notons **d.** Le changement la fait passer à **d'**[3]. La demande (partielle) de certains biens augmente alors que d'autres biens voient la leur baisser. Valorisons les vecteurs **d** et **d'** à l'ancien système de prix **p** et supposons que la dépense totale **d'.p** ne soit pas supérieure à **d.p**. Selon le postulat de Wald, on peut en inférer que valorisée au nouveau système de prix **p'**, la dépense **d'.p'** doit être inférieure à **d.p'**. En effet, si lorsque le système de prix valait **p**, le lot de biens **d** était préféré à **d'**, il doit rester préféré à **d'** lorsque le système de prix s'établit à **p'** et seule une motivation budgétaire peut expliquer le déplacement de la demande de **d** à **d'**.

Cette neuvième condition est importante ; elle sera plus tard appelée ***axiome faible de la préférence révélée***, et deviendra un pilier de la théorie moderne du consommateur.

Jusqu'ici, Wald considérait la demande du marché, mais *l'axiome de la préférence révélée* peut aussi convenir pour caractériser la demande d'un consommateur individuel. Le fait que cette condition soit satisfaite au niveau du marché n'implique nullement qu'elle le soit au niveau des individus qui le composent. Par contre, si elle est vérifiée au niveau de tous les consommateurs individuels, il est statistiquement peu probable (mais non impossible) qu'elle ne le soit pas au niveau du marché.

Wald remarque que son système produit, non des prix relatifs comme on aurait pu s'y attendre et comme il eût été logique, mais des prix absolus. L'équation (6.8) norme en fait implicitement le système en liant les quantités aux prix.

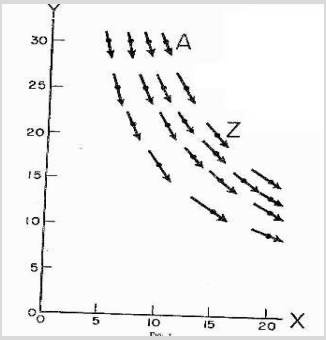
---

[1] **p** est le vecteur $(p_1,p_2,\ldots p_n)$ et **p'** est le vecteur $(p'_1,p'_2,\ldots p'_n)$, dont les éléments sont les prix des *n* biens.
[2] Cette fonction de demande, non dessinable, fait correspondre une combinaison de quantités de l'ensemble des biens pour chaque combinaison des prix de l'ensemble des biens. Cette conception de la fonction de demande est la seule qui soit rigoureuse scientifiquement.
[3] **d** et **d'** sont également des vecteurs à *n* éléments.

Samuelson et l'axiome de la préférence révélée

Indépendamment de Wald, Samuelson a usé de ce même postulat dans deux articles (1938, 1948) pour simplifier la théorie du consommateur. Rappelons que l'aspect non observable de l'utilité marginale décroissante avait conduit les économistes à lui préférer la théorie du taux marginal de substitution croissant. Mais Samuelson avoue ne pas encore être satisfait : « For just as we do not claim to know by introspection the behaviour of utility, many will argue we cannot know the behaviour of ratios of marginal utilities or of indifference directions. Why should one believe in the *increasing rate of marginal substitution*, except in so far as it leads to the type of demand functions in the market which seem plausible? »[1]. Samuelson ne se résout pas à voir la théorie économique dépendre de concepts découlant de l'introspection psychologique ; il veut fonder la théorie du consommateur sur des faits observables : « I propose, therefore, that we start a new in direct attack upon the problem, dropping off the last vestiges of utility analysis ».

A l'aide d'un modèle simplifié à deux biens (consommés en quantités x et y), Samuelson montre que de la seule observation des triplets $(x, y, p_x/p_y)$, amène à des déductions qu'on peut rapprocher de la théorie existante du comportement du consommateur[2]. La présomption fondamentale que si un duo x,y a été constaté pour un rapport de prix et un revenus donnés, c'est qu'il a été préféré à tous les autres.



Chaque point de l'espace x,y est une combinaison des quantités consommées, à laquelle correspond un rapport de prix d'équilibre (par hypothèse un seul) représentable par un segment de la droite de budget. On a en chaque point : $px/py = f(x,y)$, fonction observable présumée continue et différentiable. Considérons en chaque point une portion infinitésimale des ces droites de prix (dont la pente égale dy/dx). On a donc l'équation différentielle $dy/dx = -f(x,y)$; elle définit un ensemble de courbes dans le champ (x,y) semblables aux courbes d'indifférence bien connues. Samuelson démontre mathématiquement que par rapport à un point, ceux qui sont au-dessus de la courbe passant par lui sont révélés supérieurs et ceux qui sont en-dessous sont révélés inférieurs. La préférence révélée nous ramène aux résultats de la théorie traditionnelle.

6.1.2. von Neumann : le modèle de production linéaire

Un des esprits les plus brillants du XXe siècle, à la fois mathématicien, logicien, physicien et économiste. John von Neumann, né à Budapest et qui avait adhéré au colloque viennois, rejoignit les Etats-Unis en 1932 après un passage par Berlin.

[1] Samuelson [309] p. 61

[2] Samuelson suppose que les données empiriques relatives à ces triplets sont disponibles en quantité suffisante, une hypothèse qui a été critiquée. L'autre critique principale portée contre la théorie de la préférence révélée est que les variations entre les demandes constatées à des prix différents peuvent être polluées par un facteur non observable, en l'occurrence un changement de goût du consommateur entre les différents relevés.

Son article **« A Model of General Economic Equilibrium » (1937)** fait avancer la problématique qui nous occupe à de nombreux égards. Comme son titre l'indique, il s'agit toujours de démontrer l'existence d'un équilibre général. Par rapport aux modèles précédents, principalement celui de Wald, de nombreuses améliorations sont apportées :

1- L'input de la production comporte non seulement des facteurs primaires mais également des biens produits. A la manière de Sraffa, les marchandises sont produites par des marchandises (processus dit « circulaire »). Un bien peut être un input de sa propre production, des biens différents peuvent se servir d'input mutuellement.

2- Wald avait remplacé l'égalité (6.4) du modèle de Walras-Cassel par une inégalité. von Neumann reprend l'inégalité de Wald et fait de même pour la égalité (6.1). Du fait de la concurrence qui rabote le profit, il reste exclu que le prix soit supérieur au coût de production. Par contre, le prix d'équilibre peut être inférieur au coût de production ; dans ce cas, on impose la condition que la quantité produite soit nulle, car la production ne serait pas rentable.

3- von Neumann se départit des coefficients fixes de Cassel. Un même bien peut être produit par plusieurs combinaisons différentes d'inputs, que nous appellerons *activités*. Le nombre de combinaisons produisant un bien est fini, ce qui nous éloigne des isoquantes qui comportaient une infinité de combinaisons entre les facteurs de production.

4- Le modèle est dynamique. Le temps est divisé en périodes. A chaque période, l'output de la période précédente devient l'input de la nouvelle période. La croissance économique est présumée. Comme chez Cassel, le concept est celui d'une *croissance semi-stationnaire*. La répartition de la production ne change pas mais seulement son échelle, qui croît à un rythme régulier.

von Neumann a réellement inventé un nouveau type de fonction de production. Voyons comment elle se présente.

Une économie produit n biens (indice j) par l'exercice de m activités (indice i). Dans chaque activité, différents biens interviennent comme input et comme output. Lorsqu'une activité comporte plusieurs outputs, on a affaire à une production jointe.

Pour chaque activité, on définit un peu arbitrairement un niveau d'activité unitaire. Le niveau unitaire de l'activité ne correspond pas nécessairement à une unité du produit, parce que précisément une activité peut avoir plusieurs produits différents dans son output. Ce niveau unitaire est une abstraction.

Pour décrire la production d'une économie, von Neumann a besoin de deux matrices et un vecteur :

- La matrice **A** donne les inputs des différents biens dans les différentes activités. Ses composants sont les a_{ij}, qui représentent la consommation du bien j par l'activité i lorsqu'elle est à son niveau unitaire. Si le bien j n'entre pas dans l'activité i, a_{ij} sera nul.
- La matrice **B** donne les outputs des différents biens dans les différentes activités. Ses composants, les b_{ij}, indiquent la quantité du bien j qui sort de l'activité i lorsqu'elle est conduite à son niveau unitaire. Si le bien j ne sort pas de l'activité i, b_{ij} sera nul.

Matrice **A**

a_{11}	…	a_{1j}	…	a_{1n}
…	…		…	
a_{i1}	…	a_{ij}	…	a_{in}
…	…		…	
a_{m1}	…	a_{mj}	…	a_{mn}

Matrice **B**

b_{11}	…	b_{1j}	…	b_{1n}
…	…	…	…	
b_{i1}	…	b_{ij}	…	b_{in}
…	…	…	…	
b_{m1}	…	b_{mj}	…	b_{mn}

- Comme les matrices **A** et **B** indiquent la circulation des biens quand les différentes activités sont à leur niveau unitaire, il faut préciser à quel niveau (combien d'unités abstraites) sont conduites les différentes activités. Soit x_i, le niveau de l'activité *i*. Le vecteur **x**, à *m* éléments $x_1 \ldots x_m$, synthétise ces informations. Comme aux différentes échelles des activités, les coefficients a_{ij} et b_{ij} restent identiques, les rendements d'échelle sont constants.

L'une des simplifications excessives de ce modèle est que la consommation n'y apparaît pas en tant que telle. La consommation est considérée comme l'input des fournisseurs de facteurs primaires. Vu sous cet angle, le beurre est un input de l'acier car il est nécessaire pour faire vivre ceux qui travaillent dans cette branche et que le travail n'intervient pas en tant que tel dans le modèle.

Voyons maintenant le modèle. Les ingrédients en sont :
- les coefficients a_{ij} et $b_{ij.}$
- le vecteur **x** de l'échelle de production des différentes activités, dont on peut déduire la quantité produite des différents biens.
- le vecteur **p** des prix des *n* biens.
- *r*, le taux d'intérêt d'une période à la suivante
- *g*, le taux de croissance de la production d'une période à l'autre.

Les inconnues sont **x, p**, r et g. Dans la solution d'équilibre, **x** comportera des valeurs nulles pour toutes les activités non rentables.

Le modèle s'articule autour de deux inéquations fondamentales :

$$(1+g).\Sigma_{i=1}^{m} a_{ij}.x_i \leq \Sigma_{i=1}^{m} b_{ij}.x_i \qquad (6.9)$$

En quantité, l'output vaut l'input multiplié par (1+g). Plus précisément, on accepte que le membre de gauche puisse être inférieur mais alors certains facteurs seront libres et leur prix sera nul. C'est l'inégalité de Wald.

$$(1+r).\Sigma_{j=1}^{n} a_{ij}.p_j \geq \Sigma_{j=1}^{n} b_{ij}.p_j \qquad (6.10)$$

En valeur, l'output vaut l'input multiplié par (1+r). Plus précisément, on accepte que le membre de gauche soit supérieur, mais alors la production n'est pas rentable et la quantité produite (et donc x_i) sera nulle. C'est l'inégalité introduite par von Neumann.

von Neumann démontre que le problème a une solution en faisant appel à des progrès mathématiques récents (notamment le théorème du point fixe de Brouwer):

A l'arrivée, von Neumann obtient le résultat suivant :

$$r = g \qquad \text{avec } 1+r = 1+g = \phi(\mathbf{x},\mathbf{p}) \qquad (6.11)$$

où

$$\phi(\mathbf{x},\mathbf{p}) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} b_{ij}.x_i.p_j \,/\, \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} a_{ij}.x_i.p_j \qquad (6.12)$$

$\phi(\mathbf{x},\mathbf{p})$ correspond au rapport de la valeur totale de l'output sur la valeur totale de l'input. Tout en restant prudent car les deux approches diffèrent, on peut rapprocher le $\phi(x,p)$ de von Neumann avec le (1+R) de Sraffa.

L'équation (6.11) exprime que l'équilibre de von Neumann implique l'égalité entre *r* (le taux de profit) et *g* (la croissance de la production). Cette égalité se retrouvera plus tard fréquemment dans des modèles de croissance, en tant que condition de la maximisation de la consommation ; elle est connue sous le nom de la *règle d'or de la croissance*.

Kurz met en évidence certaines similitudes entre von Neumann et Sraffa :
1. L'équilibre général est un équilibre à long terme basé sur un taux de profit uniforme.
2. La production est conçue comme un flux circulaire. Chaque bien est susceptible d'apparaître non seulement comme output mais également comme input dans la production.
3. La relation entre le salaire et l'intérêt est telle que l'une des variables est endogène et l'autre exogène. Chez von Neumann, c'est nécessairement le salaire (ou plutôt son fantôme) qui est exogène ; chez Sraffa, c'est indistinctement l'un ou l'autre.
4. von Neumann utilise la même astuce que Sraffa pour le capital fixe. Le bien capital d'âge *t* apparaît parmi les inputs et le bien capital d'âge t+1 du côté des outputs.

*

En 1944, von Neumann se mettra de nouveau en valeur en publiant avec Oskar Morgenstern un ouvrage intitulé **« Theory of Games and Economic Behaviour »**. Bien que cet ouvrage sorte de notre objet -l'équilibre général-, on ne peut passer sous silence ses deux apports fondamentaux.

Premièrement, il expose la *Théorie des jeux*. La théorie des jeux, contrairement à ce que son appellation pourrait laisser supposer, n'est pas une théorie en soi mais une technique applicable dans la plupart des domaines de la science économique et même éventuellement dans d'autres disciplines. Lorsqu'une théorie économique met aux prises plusieurs agents qui ont chacun un intérêt à défendre sous la forme d'une variable (profit, utilité, coût) à maximiser ou minimiser, l'économiste peut modéliser leur comportement sous la forme d'un jeu. Les actions des uns influencent la situation des autres et chacun élabore une stratégie d'optimisation de sa situation qui tient compte de cette interaction. Cette stratégie consiste en des modifications que l'agent apporte à certains paramètres qui sont sous son contrôle.

Un peu comme dans une partie d'échecs, chaque joueur répond aux actions de l'autre. L'intérêt d'un tel modèle est d'examiner comment évoluent à terme les grandes variables lorsqu'on formule telle ou telle hypothèse concernant les stratégies des joueurs en présence.

Cette technique a fait l'objet de perfectionnements depuis lors et est utilisée couramment dans de nombreux modèles économiques.

Pour expliquer les stratégies des agents, von Neumann et Morgenstern ont besoin d'un indice de l'avantage que les agents retirent de leurs choix, indice qui est opérationnel non seulement quand les gains et pertes sont connus à l'avance mais aussi quand ils sont incertains. A cette fin, ils axiomatisent la théorie des choix en situation d'incertitude. Leurs axiomes reviennent à énoncer la *théorie de l'utilité espérée*, qui sera étudiée au chapitre 9.4 du présent ouvrage. Selon cette théorie, l'agent choisira la solution qui maximise l'espérance mathématique, non du gain lui-même mais de l'utilité qu'il procure (et qui est sujette à la loi de l'utilité marginale décroissante).

Ils démontrent mathématiquement qu'un indice de l'utilité peut être construit, qu'il prend la forme d'une fonction mathématique dont l'utilité est la variable dépendante et dont les variables indépendantes sont des probabilités et des possibilités de gain caractérisant les différentes alternatives. Pour jouer son rôle, la fonction doit toujours prendre une valeur supérieure pour une utilité espérée supérieure. Bâtie pour les alternatives incertaines, cette fonction d'utilité vaut également pour les événements certains, qui peuvent être en considérés comme un cas particulier.

Etonnant retour des choses : Pareto avait gommé la colline du plaisir pour ne garder que les courbes d'indifférence, afin de ne pas être piégé par la mesure de l'utilité. Voici que la colline revient. En fait, elle ne revient pas complètement. Car la fonction d'utilité ainsi conçue n'est pas unique. Si U est une fonction d'utilité de ce type, alors toutes les transformations linéaires, du type a.U + b, le sont également. En fait, c'est la mesure ordinale de l'utilité qui prend une forme cardinale.

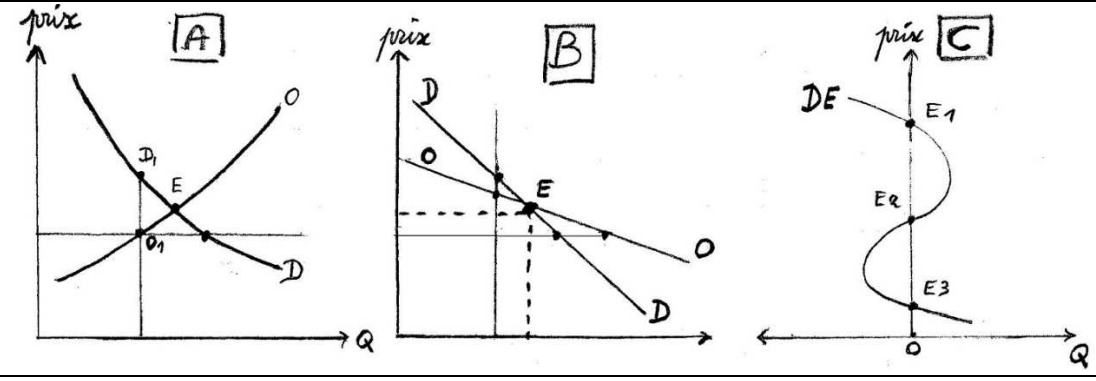
6.2- HICKS : STABILITE ET EQUILIBRE TEMPORAIRE

En 1939, Hicks publie « Value and Capital », l'un des ouvrages les plus importants de la science économique. Ce livre réussit le tour de force d'être à la fois walrassien, parétien, keynésien et autrichien. On y trouve deux grandes avancées : la question de la stabilité de l'équilibre et la prise en compte de l'aspect temporel dans la détermination de l'équilibre général.

LA STABILITE DE L'EQUILIBRE GENERAL

Le concept de stabilité d'un équilibre est important aussi bien en économie qu'en physique. « Pour que l'équilibre soit *stable*, il faut qu'un léger mouvement, à partir de la position d'équilibre, fasse intervenir d'autres forces qui tendront à rétablir l'équilibre »[1]. Les premiers à s'être intéressés à la *stabilité* de l'équilibre furent Walras et Marshall. Ce dernier écrivait : « it will be found to be a characteristic of stable equilibria that in them, the demand price is greater than the supply price for amounts just less than the equilibrium amount and vice versa. »[2]. L'équilibre E sur les figures A et B répond à cette condition.

Graphique 6.1 : la stabilité de l'équilibre sur un marché



Hicks n'apprécie guère la définition de Marshall ; admettre que l'équilibre de la figure B est stable ne lui convient pas. Hicks exprime ainsi la condition de stabilité : « une hausse de prix rend l'offre plus importante que la demande, une baisse de prix rend la demande plus importante que l'offre »[3]. Le point E de la figure B ne répond pas à la condition hicksienne alors qu'il satisfaisait la condition marshallienne : la baisse du prix ne fait qu'accroître le surcroit de l'offre sur la demande, ce qui devrait déclencher une spirale.

En outre, Hicks est le premier à s'intéresser à la stabilité d'un ensemble de marchés liés ; les ajustements du prix sur un marché influencent l'offre et la demande sur ceux des substituts et des biens complémentaires ; cette interaction rend le problème beaucoup plus complexe.

Pour faciliter l'exposé, Hicks introduit le concept de *demande excédentaire*, définie comme le surplus de la demande sur l'offre à tout prix donné ; elle sera donc négative quand l'offre est supérieure à la demande. Le marché est en équilibre lorsque la demande excédentaire est nulle, ce qu'exprimait l'équation (3.12) de Walras.

---

[1] Hicks [154] p. 55
[2] Marshall [253] § 8.3.20-21
[3] Hicks [154] p. 55. Cela revient à s'attacher à l'écart horizontal entre l'offre et la demande plutôt qu'à l'écart vertical.

Revenons un instant à la stabilité d'un équilibre partiel. La condition de stabilité peut être reformulée ainsi : il faut que la demande excédentaire soit décroissante par rapport au prix. La figure 6.1-C montre un cas d'équilibres multiples : la demande excédentaire (courbe DE) s'annule pour trois prix différents (les courbes d'offre et de demande non dessinées se coupent trois fois !). E1 et E3 sont des équilibres stables car quand le prix baisse, la demande excédentaire augmente. Par contre, E2 est instable[1].

Comment une baisse des prix peut-elle réduire la demande excédentaire et ainsi provoquer l'instabilité ? La baisse des prix suscite un double *effet de substitution*, qui élève la demande et réduit l'offre, ce qui va dans le sens de la stabilité. Elle crée également un double *effet de revenu* ; le revenu accru des acheteurs stimule leur demande, ce qui favorise la stabilité et le revenu amoindri des vendeurs les pousse à offrir plus, ce qui la défavorise. L'effet de revenu est donc ambivalent. En définitive, l'équilibre sera instable si l'effet revenu est plus fort chez les offreurs que chez les demandeurs et si cet écart est supérieur à l'effet de substitution. Hicks en conclut qu'un équilibre instable sera exceptionnel.

Passons maintenant à l'interaction entre les marchés. Deux questions se posent :

1) Un marché dont l'équilibre est stable en interne, peut-il être rendu instable par les répercussions des autres marchés ?
2) Un marché dont l'équilibre est instable en interne, peut-il être rendu stable par les répercussions des autres marchés[2] ?

Hicks estime pouvoir apporter une réponse négative à ces deux questions. Même la complémentarité entre le bien dont l'équilibre est stable en interne et d'autres biens ne suffit pas à produire une instabilité d'ensemble. « La seule source possible d'instabilité, en fin de compte, réside en une forte asymétrie des effets de revenu. Mais un degré modéré de substituabilité parmi l'ensemble des biens suffit à empêcher ce facteur de jouer »[3].

Ces considérations sont sous-tendues par des développements mathématiques, que nous nous contenterons d'effleurer. Chaque bien voit sa demande excédentaire affectée par les variations des prix de tous les autres biens et du sien propre. Chacune de ces influences, qui constituent un chassé-croisé, peut être mesurée par un rapport dz_i/dp_j, c'est-à-dire le rapport entre la variation de la demande excédentaire du bien *i* et une variation marginale du prix du bien *j* qui en est la cause ; nous noterons ce ratio a_{ij}. Le coefficient a_{ii} exprime donc l'influence du prix du bien *i* sur sa propre demande excédentaire (i = j). Considérons la matrice des facteurs a_{ij}, notée **A,** dans une économie à *n* biens :

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

[1] Si la courbe demande excédentaire avait eu la forme d'un S non inversé, c'eût été l'équilibre du milieu qui eût été stable et les deux autres instables.

[2] Hicks considère qu'un équilibre qui ne serait stable que grâce aux répercussions des autres marchés serait moins solide : pour cette raison, il qualifie sa stabilité d'*imparfaite*. Pour que la stabilité soit *parfaite*, il faut donc qu'elle soit assurée même si on impose la condition que les autres prix ne peuvent pas s'ajuster (prix rigides).

[3] Hicks [154] p. 64

Selon Hicks, la règle mathématique qui conditionne la stabilité parfaite est celle-ci: les déterminants des mineurs principaux de la matrice **A** d'ordre croissant sont alternativement de signe négatif et positif.

La théorie exposée ci-dessus est innovante, mais elle devint rapidement obsolète, car dépassée par le nouveau concept de la *stabilité dynamique*. Ses promoteurs sont Samuelson (1941, 1944), Lange (1944) et Metzler (1945) ; elle sera évoquée au sous-chapitre 6.3.2.

EQUILIBRE GENERAL ET TEMPORALITE

Les modèles d'équilibre général que nous avons vus jusqu'ici étaient statiques en ce sens que le temps n'y intervenait pas. L'idée d'une vision dynamique de l'équilibre remonte à Pareto, qui, le premier, estimait que le même bien disponible à des moments différents devait être traité comme autant des biens différents ; mais Pareto n'a pas appliqué son idée ; c'est à Hicks que revient le mérite du premier modèle d'équilibre général intertemporel[1].

Comment le facteur temps est-il introduit ? La ligne du temps est découpée en périodes discrètes, numérotées de zéro à *T* (où *T* vaut l'infini si le modèle ne place pas d'horizon temporel). La période *zéro* est le temps présent[2]. Les numéros des périodes servent à indicer les événements. Par exemple, p_{z4} sera le prix du bien *z* à la période quatre.

Les périodes sont purement abstraites. C'est pour user d'un langage imagé que Hicks les appelle *semaines* ; ceci ne préjuge en rien de la durée réelle de la période. L'idée sous-jacente est qu'elle soit suffisamment courte pour que les variations de prix puissent être négligées dans cet intervalle.

L'imputation des actions économiques à des périodes revêt plusieurs aspects. Commençons par clarifier la question avant d'exposer la théorie de Hicks. Pour simplifier le problème, prenons l'hypothèse (parfaitement irréaliste) qu'il n'y a que deux périodes : t0 et t1.

Le moment de la livraison est un critère distinctif du bien au même titre que sa nature et le lieu de sa livraison. Soient x_0 = une tonne d'anthracite en t0 et x_1 = une tonne d'anthracite en t1. Les prix sont respectivement p_0 et p_1. Sans distinction du moment de la livraison, il n'y a pas de modèle intertemporel.

Une conséquence nécessaire de cette dissociation des biens est la prise en compte des anticipations des agents. La prévision d'un agent quant à l'évolution de p_1 par apport à p_0 va nécessairement affecter les quantités de x_0 et x_1 qu'il offrira ou qu'il demandera. Même les décisions présentes sont donc affectées par les prix futurs.

L'économie comportera certainement en t_0, un marché sur lequel échanger x_0 et en t1 un marché sur lequel s'échange x_1. Mais peut-être comportera-t-elle également un marché pour échanger x_1 en t0. Si c'est le cas, on a affaire à un *marché futur*. Les *marchés à terme*[3] sont l'exemple même du marché futur. Le modèle économique, tout comme l'économie réelle, peut en comporter ou ne pas en comporter.

---

[1] Il faut toutefois noter des travaux antérieurs de Lindahl et de Hayek.

[2] Un manque de standardisation fait que ce peut être la période *un* chez d'autres auteurs.

[3] Sur le marché à terme, l'acheteur et le vendeur conviennent aujourd'hui du prix d'une transaction qui sera livrée et payée lors d'une période ultérieure.

Le marché est le lieu de la confrontation synchrone de l'offre et de la demande, avec détermination du prix d'équilibre. Cet exercice de confrontation (tâtonnement walrassien) peut se tenir en t0 puis être répété en t1. On appelle *économie séquentielle* une théorie économique qui prévoit que le marché se tient à chaque période, que chaque fois recommence le processus de tâtonnement. L'économie réelle est nécessairement séquentielle, mais- comme nous le verrons plus loin, certaines théories économiques restreignent la tenue du marché à la seule période *zéro*.

Le tableau suivant résume le propos :

	Bien x_0	**Bien x_1**
Marché en t0	Marché A (présent)	Marché B (futur)
Marché en t1	Cas impossible	Marché C (présent)

Les marchés A et B seront influencés par le prix p_0 et l'anticipation de p_1, alors que le marché C sera influencé uniquement par p_1. Dans une économie séquentielle, les livraisons de biens en t1 se fonderont sur des contrats conclus tant sur le marché B que sur le marché C. Dans une économie non séquentielle, le marché C n'existe pas et les seules livraisons en t1 sont celle décidées sur le marché B.

Après ces préliminaires, oublions l'hypothèse des deux périodes et examinons le modèle hicksien. Celui-ci est une économie *séquentielle*. Pour reprendre l'image de Hicks, le lundi se tient le marché (à la walrassienne : aucune transaction n'a lieu en dehors de l'équilibre) ; le reste de la semaine, on produit, on livre et on consomme sur base des contrats du lundi. La semaine suivante, ce processus se répète. Chaque semaine, sur base des anticipations passées, une nouvelle offre et une nouvelle demande fixent les prix, éventuellement différents de ceux d'avant.

Les entreprises et les particuliers élaborent des *plans* qui prévoient leurs actions dans la semaine en cours ainsi que celles qui suivent ; ces plans portent sur les achats et ventes des différents biens et facteurs de production, l'épargne, l'emprunt, le prêt, la thésaurisation… Ils sont influencés par les prix courants et les prix futurs anticipés. Sur les marchés présents, la confrontation réelle de l'offre et la demande produit un équilibre. Les anticipations des agents les amènent à déjà prévoir les quantités offertes et demandées en t_1, t_2… Evidemment pour les nombreux biens sans marché futur, rien ne garantit que les prévisions d'offre et les prévisions de demande s'équilibrent. Quand arrivera la période t_1, l'offre et la demande devront s'équilibrer sur le marché présent de cette période, ce qui amènera certains agents à réviser leurs plans.

Les agents hicksiens peuvent se tromper dans leurs prévisions des prix et de leurs besoins futurs. Ces erreurs entraîneront un gaspillage de ressources (investissements mal orientés).

Qui dit « futur » et « prévision » dit « incertitude ». L'incertitude peut affecter les actions en les rendant moins franches. Par exemple, l'offreur, anticipant un prix *p* mais conservant un doute à ce sujet, pourrait établir son offre comme si le prix était légèrement inférieur à *p*. Non par conviction mais pour éviter ce type de complication, Hicks pose une hypothèse forte : une fois la prévision faite concernant les prix et les conditions du marché, les agents établiront leurs plans d'achat et de vente comme s'ils tenaient leurs prévisions pour certaines.

L'usage généralisé des marchés à terme permettrait de s'assurer dès aujourd'hui que les plans des consommateurs seront compatibles avec ceux des producteurs au cours des périodes futures. Mais ce ne serait qu'une illusion, si on ne formule pas simultanément l'hypothèse que les agents sont capables de prévoir correctement le futur. En effet, les semaines suivantes, ils

devraient quand même procéder à des achats imprévus sur les marchés courants, pour corriger les erreurs de prévision qui ont présidé à leurs transactions sur les marchés à terme. Hicks refuse l'hypothèse des anticipations parfaites et il préfère laisser de côté les marchés à terme ; ce faisant, il estime ne pas trop trahir la réalité. D'ailleurs, ajoute-t-il, ces marchés sont fréquentés par des offreurs qui veulent se couvrir contre les variations de prix et ils bien plus nombreux que les demandeurs, ce qui fausse les prix, généralement inférieurs au prix comptant[1] ; les autres intervenants sont des spéculateurs tentant de tirer du profit en achetant à terme lorsqu'ils prévoient un prix comptant plus élevé dans le futur que le prix à terme.

S'il ne compte pas de marchés à terme, le modèle hicksien comporte par contre un marché qui leur ressemble un peu : celui des prêts. On y échange de la monnaie livrée aujourd'hui contre de la monnaie livrée ultérieurement[2]. L'intérêt est le prix sur le marché du prêt.

La réalité économique comporte une panoplie très large de prêts assortis de taux d'intérêt différents. Pour simplifier son exposé, Hicks ramène le crédit à un seul marché des prêts et un seul taux d'intérêt. Ceci résulte de la conjonction de plusieurs simplifications.

1. Les prêts non monétaires sont ignorés : c'est une hypothèse faible car une économie monétaire pourrait fonctionner avec ce seul type de prêt.
2. Il faut également ignorer les différences de risque : on peut considérer que des prêts ne divergeant que par le niveau de risque produisent un intérêt pur égal, associé à une prime de risque variable.
3. Le cas des prêts de durées différentes est plus complexe. Prenons un prêt courant d'aujourd'hui (semaine t_0) à la semaine t_3. Hicks considère que derrière cette situation simple, il y a implicitement une série de trois prêts d'une semaine qui s'enchaînent : de t_0 à t_1, de t_1 à t_2 et de t_2 à t_3. Si $\hat{i}_3$ est le taux à long terme (ici 3 semaines), on a :

$$(1+\hat{i}_3)^3 = [(1+i_1).(1+i_2).(1+i_3)] \qquad (6.13)$$

4. Toutes autres choses restant égales, le deuxième de ces trois prêts est plus risqué que le premier et le troisième l'est encore plus : en conséquence le taux d'intérêt i_3 sera supérieur à i_2 qui sera supérieur à i_1. Ce n'est que si on prévoit une baisse des taux à court terme dans le futur que le taux i_3 pourrait être inférieur au taux comptant i_1. Le même principe vaut pour les taux à long terme, puisqu'ils représentent la moyenne des taux à court terme sur la période considérée. Le taux à long terme doit être supérieur au taux comptant à court terme, sauf si une baisse des taux à court terme est prévue, ce qui a plus de chances de se produire si le taux à long terme est très élevé.
5. Pour réaliser son souhait de ne considérer qu'un taux d'intérêt unique, Hicks, selon les besoins de son analyse, impose la simplification que son économie ne fonctionne qu'avec des prêts d'une semaine ou qu'avec des obligations perpétuelles.

Pour démontrer qu'à chaque *semaine*, un équilibre général des marchés peut être trouvé, Hicks recourt au comptage des équations à la walrassienne[3]. Il y a n+1 marchés : la monnaie, les prêts et n-1 biens et facteurs. Vu que la monnaie est prise comme numéraire, il y a *n* prix à déterminer. Nous avons pour chaque marché une équation du type (3.12) de Walras, égalisant

[1] On appelle « déport » la différence de prix quand le prix à terme est inférieur (le plus souvent) et « report » quand il est supérieur (quand on prévoit une hausse du prix dans le futur)

[2] Sur les marchés à terme de l'économie réelle, on échange un bien réel livré ultérieurement contre de la monnaie elle aussi livrée ultérieurement.

[3] Certains y ont vu une marche arrière après le travail de Wald et von Neumann.

l'offre et la demande, toutes deux fonctions du système de prix[1]. De ces (n+1) équations, seules *n* sont indépendantes. Le système comporte donc une solution.

Le rejet de la dichotomie est évident : la monnaie n'est pas séparée des biens réels. Les prix des biens réels sont directement exprimés en monnaie. Il n'y a pas un système de prix relatifs indépendant de la monnaie qui serait ensuite traduit en prix monétaires à l'aide d'une équation propre à la monnaie.

Après ces considérations générales sur l'équilibre, Hicks approfondit l'étude du comportement des firmes et des consommateurs sur les marchés lorsqu'ils optimisent leur situation non seulement à l'instant même mais sur une durée de plusieurs semaines.

La **firme** établit un *plan de production* à un terme de *T* périodes, correspondant au flux d'entrée des différents facteurs et au flux de sortie des différents produits, période par période. Notons par la lettre *A* le coût des entrées et par *X* la recette des ventes. Le problème de la firme consiste à déterminer le flux des excédents (différence entre le flux des ventes et celui des achats) qui maximise la valeur actualisée de ce flux, au taux d'intérêt en vigueur. Il s'agit donc de maximiser :

$$(X_0\text{-}A_0) + (X_1\text{-}A_1).\frac{1}{(1+i)} + \ldots + (X_T\text{-}A_T).\frac{1}{(1+i)^T} \qquad (6.14)$$

Pour y parvenir, les règles parétiennes exposées au sous-chapitre 4.2.2 restent d'application. Hicks avait fixé trois conditions à respecter par la firme multiproduits (cf. supra). Ces conditions demeurent mais prennent en outre une dimension temporelle. Prenons le cas de la proportion entre les facteurs qui doit égaliser le TMST avec le rapport de prix. L'astuce consiste à appliquer cette règle en considérant un même facteur à deux périodes comme deux facteurs différents, avec en plus la nécessité d'escompter les prix anticipés au taux d'intérêt. La firme devra opérer une substitution intertemporelle entre les facteurs, c'-à-d moduler le flux d'entrée de chacun d'eux dans le temps en respectant l'équation (6.15), relative au facteur *L* entre les périodes *t* et *t'*.

$$\mathrm{TMST}(L,t/t') = P_{L,t}.\frac{1}{(1+i)^t} \,/\, P_{L,t'}.\frac{1}{(1+i)^{t'}} \qquad (6.15)$$

Une autre équation de type (6.15) vaut pour le facteur *K* entre t et t' ; la même égalité doit en outre lier le facteurs *K* et *L* entre eux.

Si le taux d'intérêt *i* diminue, la valeur actuelle des achats et ventes lointains augmente par rapport à celle des achats et ventes rapprochés ; il apparaît clairsemant de l'équation (6.14) que la firme augmentera la valeur actuelle de son plan si elle avance ses achats de facteurs et recule ses ventes de produits. Dans le langage de la théorie autrichienne, on dirait que la période moyenne de production augmente. Hicks reprend à son compte la démarche de Böhm Bawerk, mais conscient des lacunes de celle-ci, il élabore sa propre formule de la *durée moyenne du plan de production* :

$$P = \sum_{t=0}^{T} t.\frac{1}{(1+i)^t}.\,(X_t\text{-}A_t) \,/\, \sum_{t=0}^{T} \frac{1}{(1+i)^t}.\,(X_t\text{-}A_t) \qquad (6.16)$$

[1] Le recours au système de type (3.12) est une simplification par rapport à Walras, chez qui ce système se rapportait à l'échange pur. Or la production est présente dans le modèle de Hicks. Celui-ci se contente d'ajouter les marchés des facteurs dans le comptage des marchés à équilibrer. Rappelons que Walras avait intégré l'égalisation du coût de production et du prix dans son système d'équations de la production, un aspect entièrement ignoré par Hicks.

Le but est de quantifier la variation dans le flux des revenus quand le taux d'intérêt passe de *i* à *i'*. Il faut donc comparer la valeur de *P* aux deux taux ; mais attention : les deux valeurs de *P* doivent être calculées avec le même taux *i* dans la formule (6.16) ; ce qui importe, c'est que les valeurs de *X* et de *A* diffèrent selon que prévaut *i* ou *i'*.

Hicks estime que sa formule est immunisée contre les reproches qui furent adressés à celle de Böhm Bawerk.

Le principe est le même pour le **consommateur**. Celui-ci établit un plan de consommation et d'épargne, certes beaucoup plus informel et moins précis que le plan de production des entreprises, mais il ne peut se permettre de dépenser à l'aveugle. Le but du ménage est de maximiser son utilité intertemporelle sur la durée de ce plan. L'optimum se détermine comme au sous-chapitre 4.2.1 (équilibre statique) avec la même correction que celle indiquée ci-dessus pour les entreprises. Il faut pratiquer pour tous les biens (X) et toutes les périodes l'égalisation :

$$TMS(X,t/t') = P_{X,t}.\frac{1}{(1+i)^t} / P_{X,t'}.\frac{1}{(1+i)^{t'}} \qquad (6.17)$$

La plupart des semaines voient se constituer un excédent, positif ou négatif, des rentrées sur les dépenses. S'il est positif, cet excédent entraînera une demande de monnaie ou une demande de titres (qui correspond à une offre de capital). Dans ce système, la monnaie joue comme un substitut proche des titres. Les motifs d'en détenir expliqués par Keynes font qu'on peut préférer garder une encaisse malgré la perte d'intérêt qui s'ensuit, mais plus l'intérêt est élevé, moins le désir de conserver une encaisse se fera sentir.

Comme pour la firme, Hicks étudie les effets d'une variation du taux d'intérêt. Une hausse se traduirait par une réduction de la valeur actuelle des recettes et dépenses les plus éloignées par rapport à celles plus rapprochées. Il en résultera un *effet de substitution* et un *effet de revenu*. L'effet de substitution consiste à stimuler le report de dépenses et l'avancement de l'épargne. L'effet de revenu enrichira les agents épargnant beaucoup les semaines rapprochées et appauvrira les agents empruntant beaucoup les semaines rapprochées. Les deux effets peuvent donc se conforter ou s'opposer.

Nous pouvons conclure ce chapitre. Chaque semaine, un équilibre général s'établit, qui est propre à cette semaine, mais qui tient compte des anticipations relatives au futur et est influencé par les anticipations passées. Comme la capacité de prévision parfaite n'existe pas, cet équilibre a de fortes chances de varier de semaine en semaine. Pour cette raison, Hicks l'appelle l'*équilibre temporaire*. Après Hicks, d'autres économistes produiront des modèles d'*équilibre temporaire*.

La suite de l'ouvrage prend une orientation macroéconomique ; si l'équilibre temporaire se révèle instable, il peut en résulter des soubresauts économiques malvenus. Nous y reviendrons au sous-chapitre 7.1.1.

6.3- EMERGENCE DE LA THEORIE ACTUELLE DE L'EQUILIBRE GENERAL

6.3.1. Progrès mathématiques et la « commission Cowles »

Dans les années trente est créée aux Etats-Unis la « Commission Cowles »[1] qui a pour objectif de promouvoir l'économie mathématique. Elle rassemble de nombreux économistes, dont beaucoup d'Européens (notamment ceux du Colloque viennois), qui
franchissent l'Atlantique devant l'avance nazie. Les travaux de Wald et von Neumann sont le point de départ de la réflexion. Des congrès sont organisés. Les progrès dans l'économie néo-walrassienne ne sont plus le fruit d'initiatives isolées mais le résultat d'une activité coordonnée.

L'économie néowalrassienne entre dans une nouvelle étape, qui, ne révise pas seulement la démonstration walrassienne de l'existence de l'équilibre général mais englobe la problématique parétienne de l'optimum du consommateur, de celui du consommateur et du lien entre équilibre et optimum. Comme celle de Walras, les démonstrations néo-parétiennes comportaient des imperfections :

1- « It is obvious to everyday observation that for each individual there are some (indeed many) commodities of which he consumes nothing. Similarly, for every firm, there are some commodities which are neither inputs to nor inputs of it. But then, the argument that the marginal rate of substitution must equal the price ratio for each individual breaks down. »[2].
2- La fonction de production doit être différentiable, ce qui n'est pas toujours le cas dans la réalité, notamment avec les coefficients fixes. Lorsque les isoquantes comportent des « coins », une infinité de droites (de prix) peuvent être tangentes au coin : il y aurait indétermination de l'équilibre.
3- Le modèle parétien ignore les produits qui servent comme inputs pour produire d'autres produits.

Pour résoudre ces problèmes, l'économie néo-walrassienne s'est trouvée de nouveaux outils mathématiques mais la perspective économique de l'école de Lausanne demeure. C'est l'outil qui change, pas la théorie elle-même.

CONVEXITE ET THEORIE DU PRODUCTEUR

La nouvelle théorie de la production fut initiée principalement par Tjalling **Koopmans** qui quitta sa Hollande natale pendant la guerre pour rejoindre la commission Cowles dont il devint rapidement le président.

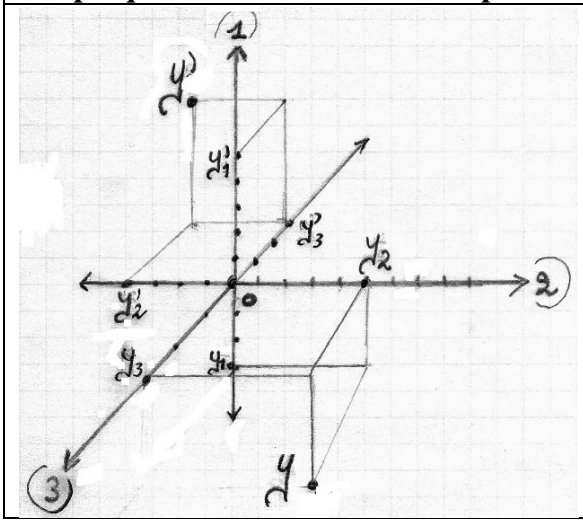
Soit une économie comportant l biens ; Koopmans reprend de von Neumann le concept d'*activité*, mais le formalise différemment. Les différentes activités de production réelles ou potentielles sont représentées par des points dans l'ensemble vectoriel à l dimensions. Un point dans l'espace vectoriel se caractérise par des coordonnées qui correspondent à un nombre réel sur chacun des l axes. Considérons que ces coordonnées sur les axes représentent les quantités physiques des biens, qui interviennent dans l'activité productive (en tonnes, mètres, unités diverses). Conventionnellement, Koopmans représente cette quantité par un nombre négatif pour les inputs et par un nombre positif pour les outputs. Par exemple, l'activité y^a représentée par le vecteur (-1,0,-4,0,0,3,1,-9) est celle où on produit trois unités du bien six et une unité du

[1] D'*Alfred Cowles*, homme d'affaire à l'origine de sa création
[2] Arrow [15] p. 259

bien sept à partir d'une unité du bien un plus quatre unités du bien trois plus neuf unités du bien huit ; les biens deux, quatre et cinq n'interviennent pas dans cette activité.

La représentation graphique est évidemment limitée à la production faisant intervenir un maximum de trois biens. Le graphique 6.2 nous montre un ensemble de deux activités : d'une part, **y** qui produit 5 unités de (2) et 3 unités de (3) à partir de 3 unités de (1) et d'autre part, **y'** qui produit 5 unités de (1) à partir de 3 unités de (3) et 4 unités de (2).

Graphique 6.2 : activités dans l'espace vectoriel



La représentation des activités par des vecteurs a l'avantage de mettre à disposition de l'économiste les ressources du calcul vectoriel, mais la lecture du présent chapitre ne nécessite presque aucune connaissance dans ce domaine. Les mathématiciens distinguent les SCALAIRES, c'est-à-dire les nombres et les VECTEURS, qui sont multidimensionnels. Le concept de vecteur inclut celui d'une direction, puisque les coordonnées du vecteur déterminent une position par rapport à l'origine (le zéro sur chaque axe). L'origine représente l'absence de production et l'éloignement d'un point par rapport à celle-ci reflète l'importance des flux que la production met en jeu. Si deux vecteurs distincts ont la même direction, ils représentent des activités où les biens entrent et sortent en proportion identique mais avec une amplitude différente.

Je me conformerai à une pratique courante consistant à distinguer les vecteurs des scalaires en les désignant par des lettres notées en caractère gras. La lettre minuscule **y** désigne généralement les activités. Les activités peuvent être définies au niveau de la firme ou de l'économie globale. L'indice *f* contenu dans $\mathbf{y_f}$ stipule qu'on a affaire à une activité de la firme *f*. L'absence d'indice signifie qu'on a affaire à des activités propres à l'économie entière. Le vecteur relatif à l'économie globale a, comme éléments, des quantités produites qui sont pour chaque bien la somme des éléments correspondants des vecteurs des entreprises.

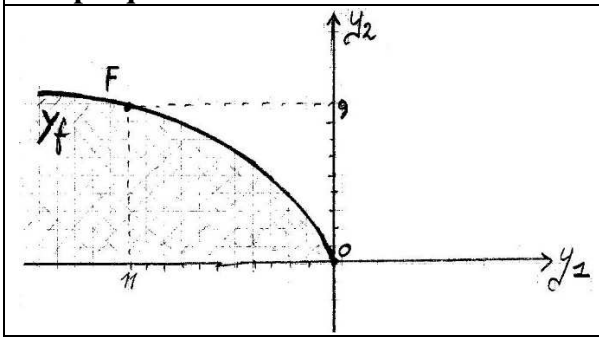
Ce n'est pas la totalité de l'espace vectoriel ainsi défini qui nous intéresse. Une activité produisant un kilo de chocolat à partir de 3 mètres de tissus et un baril de pétrole est évidemment impossible. Nous allons donc constituer dans l'espace vectoriel un sous-ensemble : l'ensemble des *activités possibles*, c'est-à-dire les activités qui sont technologiquement réalisables[1]. On utilise la lettre majuscule *Y* pour désigner l'ensemble des activités possibles. Un ensemble comme *Y* ou Y_f est une espèce de nuage dans l'espace vectoriel, qui enveloppe

[1] Les économistes font parfois la distinction entre les activités POSSIBLES et les activités FAISABLES qui, outre le fait d'être possibles, ont une consommation des inputs qui n'excède pas la disponibilité (principalement pour les facteurs primaires).

l'origine. La forme du nuage n'est pas aléatoire. Plus élevés sont les inputs, plus élevés tendront à être les outputs. Grâce à cette relation, on peut supposer que la forme du nuage ne sera pas trop irrégulière.

Pour examiner plus en détail l'ensemble *Y* ou Y_f, simplifions la situation pour avoir un graphique en deux dimensions : les activités ont un input y_1 pour un output y_2, comme sur le graphique 6.3. Supposons que Y_f soit la surface hachurée. Pour chaque valeur de l'input, l'output peut prendre les diverses valeurs comprises entre l'abscisse et la courbe *F*. Les points sur cette courbe indiquent la production maximale pour un input donné ; par exemple, elle vaut 9 lorsque l'input vaut 11. On appelle *fonction d'efficience* (F), l'ensemble des points de Y_f où la production est maximale. Elle dépend de critères technologiques. Elle constitue la frontière supérieure de l'ensemble Y_f. Normalement, la firme, motivée par la recherche du profit, devrait ignorer les activités de Y_f qui ne figurent pas sur la fonction d'efficience.

Graphique 6.3 : la frontière d'efficience



L'origine (output et input nuls) fait partie de Y_f car elle est évidemment faisable : elle se trouve sur la frontière d'efficience, car sans input, l'output maximal est nul.

L'une des caractéristiques les plus importantes, que peut avoir un ensemble comme Y_f, est la *convexité*. Prenons une paire de points (de vecteurs) quelconque de l'ensemble ; celui-ci est dit *convexe* si tous les points de la droite qui les relie font aussi partie de cet ensemble. Par exemple, l'ensemble comporte notamment l'activité **y** qui tire 20 produits de 10 heures de travail plus 4 tonnes de matière ainsi que l'activité **y'** qui tire 18 produits de 12 heures de travail et 3 tonnes de matières. Dans ce cas , l'ensemle Y_f ne sera convexe que s'il comporte également l'activité **y**° qui tire 18,8 produits de 11h20' de travail et 3,4 tonnes de matière. Mathématiquement, la convexité s'exprime ainsi :

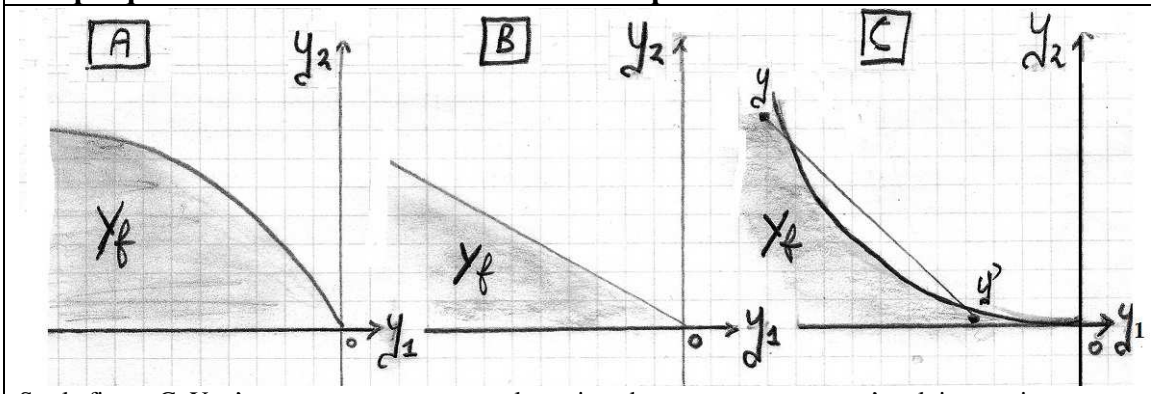
$$\text{Si } \mathbf{y} \text{ et } \mathbf{y'} \in Y_f \text{ et } 0 \leq \lambda \leq 1, \text{ alors } [\lambda.\mathbf{y} + (1-\lambda).\mathbf{y'}] \in Y_f \qquad (6.18)$$

Une activité dont certains inputs ou outputs sont indivisibles est également indivisible. Si l'activité **y** est indivisible, toutes les activités $\lambda.\mathbf{y}$ où $0 \leq \lambda \leq 1$ ne sont pas faisables. Il s'ensuit que des activités indivisibles rendent l'ensemble non convexe.

Le graphique 6.4 montre trois ensembles Y_f dont la frontière d'efficience est respectivement concave, droite et convexe. On constate que les deux premiers ensembles sont convexes et le troisième non[1].

[1] Attention au piège terminologique : l'ensemble convexe a une frontière d'efficience concave.

Graphique 6.4 : la convexité des ensembles de production faisable



Sur la figure C, Y_f n'est pas convexe, car tous les points du segment entre y et y' ne lui appartiennent pas.

La forme de la frontière d'efficience révèle le type de rendements d'échelles : à gauche, l'output croît moins vite que l'input : ce sont les rendements décroissants ; au milieu les rendements sont constants et à droite croissants. L'ensemble Y_f ne sera convexe que si les rendements ne sont pas croissants.

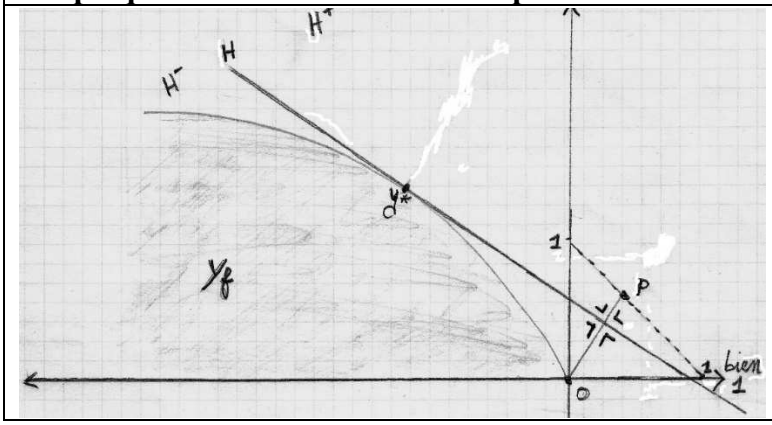
Jusqu'à présent, nous n'avons considéré que les quantités physiques. Introduisons maintenant les prix et les valeurs. Tout comme nous avons défini une infinité de vecteurs **y** dans l'espace à *l* dimensions, nous pouvons définir des vecteurs **p** dans ce même espace à *l* dimensions qui sont chacun une combinaison des prix des différents biens. Ce sont les prix relatifs qui nous intéressent ; dans un exemple à trois biens, les vecteurs de prix (4,12,5) et (8,24,10) font double emploi. Pour éliminer les vecteurs qui répètent un vecteur déjà connu, une solution fréquente consiste à ne retenir que les vecteurs de prix dont la somme a une valeur déterminée, par exemple l'unité. Seuls les vecteurs ne comportant pas de prix négatifs et au moins un prix positif sont économiquement significatifs.

Dans un exemple à trois biens (1,2 et 3) dont les prix et les quantités sont respectivement p_1 p_2 p_3 et y_1 y_2 y_3, le produit vectoriel **p.y** égale $p_1.y_1 + p_2.y_2 + p_3.y_3$. Le produit de deux vecteurs donne un scalaire et non un vecteur. Comme y_i est positif pour l'output et négatif pour l'input, le résultat de ce produit vectoriel n'est rien d'autre que le profit de la firme pour les prix considérés. Nous le noterons par la lettre π.

Voyons la figure 6.5. Le vecteur **p** (normé à 1) est dans le cadran positif. Il se fait que les vecteurs **y** qui multipliés par **p** donnent une même valeur de π forment ensemble une droite perpendiculaire au vecteur **p**. Chaque valeur de π définit donc une telle droite et ces droites sont parallèles entre elles pour un même vecteur **p**. Plus π est élevé, plus la droite est située vers le nord-est.

En concurrence parfaite, la firme considère le prix comme une donnée et s'y adapte en choisissant le vecteur **y** qui maximise son profit **p.y**. La question est donc de savoir s'il existe dans Y_f un vecteur **y*** qui a cette propriété.

Graphique 6.5 : la maximisation du profit



Introduisons la notion d'*hyperplan*. Dans l'espace des réels à *l* dimensions, un hyperplan est un ensemble de points **h** linéaire (**p.h** = c). L'hyperplan a une dimension de moins que son espace et il coupe celui-ci en deux parties : celle où **p.h** $\leq$ c (notée H^-) et celle où **p.h** $\geq$ c (notée H^+). Dans l'espace à deux dimensions du graphique 6.5, la droite *H* est un hyperplan. Les droites de profit dont il est question ci-dessus sont des hyperplans dans l'espace à 2 dimensions.

Un théorème, celui de Minkowski, dit ceci : soit un ensemble convexe *Y* dans R^l et un point **y** dans *Y*. Si et seulement si **y** n'est pas intérieur à *Y* (autrement dit s'il est sur la frontière d'efficience), il existe un hyperplan *H* passant par **y** et laissant *Y* tout entier d'un même côté.

L'hyperplan *H* relie les points pour lesquels le profit égale une valeur π^*, lorsque le vecteur de prix (non nul) est **p**. Au nord-est, le profit est supérieur à π^* et au sud-ouest, il est inférieur. Le point **y*** sur la frontière d'efficience est un point de profit maximum puisque les points de Y_f qui ne sont pas sur l'hyperplan rapportent nécessairement un profit inférieur. Ceux qui pourraient rapporter plus ne sont pas faisables.

L'application du théorème de Minkowski à notre problème implique donc l'existence d'une activité maximisant le profit pour un vecteur de prix non nul, pour autant que Y_f soit convexe. Il ressort à l'évidence du graphique 6.4-C que si les rendements d'échelle sont croissants, il n'y a pas de plan de production, maximisant le profit. Mais la convexité ne suffit pas ; si les rendements d'échelle sont constants, la maximisation du profit consiste à produire toujours plus. Ce n'est donc que si les rendements d'échelle sont décroissants que la firme peut maximiser son profit.

Tout comme les scalaires, les vecteurs peuvent entrer comme variable indépendante ou comme variable dépendante dans une fonction. Prenons le cas des vecteurs de prix **p**, pour lesquels la maximisation du profit est déterminée. On peut considérer que les vecteurs **y*** qui maximisent le profit sont une fonction **y(p)**. Le profit correspondant, π (qui est un scalaire), est également une fonction de **p** : c'est la « fonction de profit », notée $\pi(\mathbf{p})$. La fonction **y(p)** peut être considérée comme la fonction d'offre de la firme (une offre élargie toutefois, car elle inclut la demande des inputs). L'offre résulte d'un comportement optimisateur ; la fonction d'offre n'est définie que si un plan optimum peut l'être, c'est-à-dire si les rendements sont décroissants. La fonction **y(p)** est alors continue.

CONVEXITE ET OPTIMUM DU CONSOMMATEUR

La manière d'aborder la consommation est similaire, mais le problème diffère légèrement. Le consommateur doit lui aussi choisir une combinaison de biens représentée par un vecteur, qui

est noté **x**. Le travail fait partie des *l* biens. Les biens offerts, et le travail en est un, viennent en négatif. Dans l'espace euclidien à *l* dimension, on peut délimiter l'ensemble *X* des consommations faisables en éliminant celles qui n'apportent pas le minimum vital et celles qui offrent un travail au-delà des limites physiologiques. Alors que l'ensemble des productions faisables avait sa frontière vers le haut, celui des consommations faisables est borné vers le bas. L'ensemble *X* est présumé convexe.

Les consommateurs disposent souvent d'une allocation initiale (épargne passée, héritages…). Soit **e** le vecteur à *l* dimensions de cette allocation initiale.

Les consommateurs sont soumis à une contrainte budgétaire, qui les empêche de dépenser plus que leur revenu en ce compris l'allocation initiale. La contrainte de revenu, qui délimite les consommations payables, dépend de l'allocation initiale et du vecteur de prix. On peut caractériser ainsi les vecteurs de consommation, satisfaisant à la contrainte budgétaire :

$$\{\mathbf{x} \in X \mid \mathbf{p.x} \leq \mathbf{p.e}\} \qquad (6.19)$$

Le travail venant en négatif dans le vecteur **x**, les revenus du travail viennent en négatif dans le produit **p.x** ; il est donc possible que **p.x** soit plus petit que **p.e** même lorsque ce dernier est nul.

Dans le cas de la production, c'était la technologie qui délimitait les plans de production faisables. Compte tenu de la technologie, la maximisation du profit était purement algébrique et objective. Pour la consommation, nous devons introduire une notion supplémentaire, de nature subjective : la préférence du consommateur. Le consommateur individuel peut toujours comparer deux vecteurs de consommation $\mathbf{x_1}$ et $\mathbf{x_2}$ et le résultat sera nécessairement une de ces trois possibilités : $\mathbf{x_1}$ est préféré à $\mathbf{x_2}$; il lui est indifférent ou bien $\mathbf{x_2}$ lui est préféré.

Mathématiquement, il y a deux manières de présenter la préférence du consommateur. La première consiste à imposer un *préordre* dans l'ensemble *X*. Les différents composants de *X* sont classés en une série de « surfaces d'indifférences ».

Les économistes attribuent généralement diverses propriétés au système de préférence, dont les deux principales sont :

- la *non-satiété* : pour tout vecteur de consommation, il est possible d'en trouver un qui lui est préféré. La satiété peut être atteinte pour un ou quelques biens mais pas pour tous les biens ensemble
- la convexité : si $\mathbf{x_1} \succcurlyeq \mathbf{x_2}$[1], alors, $\forall\ 0 < \lambda < 1$, on a : $[\lambda.\mathbf{x_1} + (1-\lambda)].\mathbf{x_2} \succcurlyeq \mathbf{x_2}$. Si $\mathbf{x_1}$ est préféré à $\mathbf{x_2}$, une combinaison des deux sera également préférée. Il ne faut pas confondre la convexité des préférences avec la convexité de l'ensemble *X* mentionnée ci-avant.

L'autre manière de représenter les préférences est de faire appel à une *fonction d'utilité*, comme suite aux travaux de von Neumann et Morgenstern (cf. supra). A chaque vecteur **x**, on fait correspondre une valeur réelle u(**x**) telle que cette valeur est nécessairement plus élevée pour un vecteur de bien préféré et égale lorsque le consommateur est indifférent entre deux consommations. L'existence d'une telle fonction a été démontrée mathématiquement, moyennant quelques conditions légères que doit satisfaire le système de préférence, principalement la continuité.

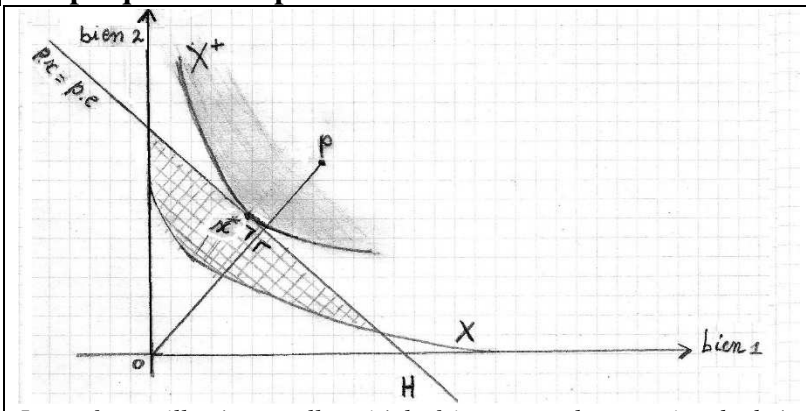
La fonction d'utilité est continue. La convexité des préférences correspond à la convexité des courbes d'indifférences, bien connue. La fonction d'utilité correspond à la colline de plaisir

[1] Ce symbole est celui des préordres : '$\succcurlyeq$' signifie « est préféré ou indifférent à … ».

dans sa dimension « hauteur » ; elle est concave lorsque les courbes d'indifférences sont convexes[1].

Le graphique 6.6 montre l'optimum du consommateur. La contrainte budgétaire définit un hyperplan (orthogonal par rapport au vecteur de prix). Le consommateur ne peut se payer que les vecteurs de consommation de *X* qui ne sont pas en dessus de cette contrainte (au nord-est). Le vecteur **x*** est optimal car il est situé sur cet hyperplan et atteint la courbe d'indifférence la plus élevée. L'ensemble des points apportant une utilité strictement supérieure à celle de **x*** (noté X^+) est situé entièrement du côté inaccessible de la contrainte budgétaire.

Graphique 6.6 : l'optimum du consommateur



*La surface grillagée est celle qui à la fois respecte la contrainte budgétaire et fait partie de X. La surface ombrée représente X^+ ; elle est strictement au-dessus de la courbe d'indifférence passant par **x*.***

Il est démontré que si **x*** est le vecteur préféré parmi ceux qui respectent la contrainte budgétaire, alors, **x*** minimise la dépense **p.x** pour l'ensemble des vecteurs de consommation **x** sur la courbe d'indifférence de **x*** ou dans la région ombrée au-dessus.

6.3.2. L'équilibre général selon Arrow et Debreu

En 1954, dans un modèle qui résonne comme l'aboutissement des dernières avancées, Kenneth Arrow et Gérard Debreu démontrent conjointement l'existence d'un équilibre général de concurrence parfaite. C'est la première démonstration vraiment satisfaisante. Elle fait appel à des mathématiques complexes. Elle jouit d'un prestige peut-être inégalé au sein de la profession économique, où beaucoup la considèrent comme le franchissement d'un seuil. Il y a l'avant- et l'après-1954. Nombre d'économistes apprécient en elle la formulation scientifique de la thèse intuitive d'Adam Smith concernant la *main invisible* du marché.

LA PREUVE DE L'EXISTENCE PAR ARROW ET DEBREU

Les ingrédients du modèle sont :

- des biens échangés (dont l'indice dans les équations est *i*), au nombre de *l*, comprenant des produits de consommation ou d'équipements ainsi que différents types de facteur travail.

- Chaque bien a un prix, qui ne peut pas être négatif. Les vecteurs **p** sont les vecteurs donnant les prix des *l* biens. Les agents prennent leurs décisions sur base des prix relatifs et non des prix absolus. Les vecteurs **p** sont normés à l'unité.

[1] Plus précisément, la fonction est *quasi-concave* car la condition de convexité des préférences n'est pas stricte.

- des firmes (indicées *f*), dont le nombre *n* «includes not only existing ones but those that might enter the market under suitable conditions »[1]. Elles ont chacune un plan de production donné par le vecteur $\mathbf{y_f}$, lorsque le vecteur de prix **p** est donné. Ce vecteur comporte l éléments y_{fi} indiquant pour chaque bien la quantité vendue comme output (alors y_{fi} est positif) ou achetée comme input (alors y_{fi} est négatif).

- des consommateurs, ménages ou institutions (indicés *h*) au nombre de *m*, qui ont chacun un vecteur de consommation et de travail $\mathbf{x_h}$, lorsque le vecteur de prix **p** est donné. Ce vecteur comporte *l* éléments x_{hi} indiquant pour chaque bien la quantité du bien *i* achetée comme bien de consommation (alors x_{hi} est positif) ou vendue comme travail (alors x_{hi} est négatif). Lorsqu'un bien *i'* n'a d'usage que dans la production où il sert comme outil ou matière première, $x_{hi'}$ sera nul pour tous les consommateurs *h* ; par contre, on verra un $y_{fi'}$ négatif chez les producteurs qui l'emploient et un $y_{fi'}$ positif chez ceux qui le produisent.

- Les consommateurs ont chacun un vecteur $\boldsymbol{\varepsilon_h}$ indiquant leur allocation initiale de chaque bien et principalement leur capacité à fournir un travail

- Ils ont également un vecteur de détention d'actions $\boldsymbol{\alpha_h}$ indiquant pour chacune des *n* firmes, quel pourcentage du profit de cette firme leur revient.

On a : $\mathbf{y} = \Sigma_j \mathbf{y_f}$ qui est le vecteur de l'offre (nette) globale des entreprises
$\mathbf{x} = \Sigma_i \mathbf{x_h}$ qui est le vecteur de la demande (nette) globale des consommateurs.

X_h est l'ensemble des vecteurs de consommation faisables du consommateur *h*, indépendamment de toute contrainte budgétaire. Les cas impossibles comme travailler plus de 24 heures par jour en sont exclus. De même, Y_f est l'ensemble des vecteurs de production faisables pour la firme *f*, compte tenu de la technologie. Il n'y a pas de plan **y** où le travail ne figure pas comme input ; le travail ne peut jamais être un output. De nombreuses hypothèses concernent les caractéristiques de X_h et Y_f ; voyons-en les plus significatives :

H1) Y_f, qui est l'ensemble de plans de production faisables ($\mathbf{y_f}$) de la firme *f* est fermé et convexe, ce qui exclut les rendements d'échelle croissants (cf. supra)[2].
H2) Une fonction d'utilité $u_h(\mathbf{x_h})$ attribue un niveau d'utilité à chaque vecteur de consommation possible pour le consommateur *h*. Cette fonction est continue sur X_h.
H3) Les préférences des consommateurs sont convexes et la saturation des besoins et désirs est exclue.

Les autres hypothèses sont :

H4) Chaque consommateur est capable d'offrir au moins un type de travail productif
H5) Les coefficients techniques ne sont pas fixes.

Les producteurs tentent de maximiser leur profit, c'est à dire $\mathbf{y_f}.\mathbf{p}$. Il s'agit ici du profit BRUT et non du profit NET puisque aucun intérêt n'est déduit[1]. Comme les producteurs ont un

[1] Arrow-Debreu [17] pp. 267-268
[2] Les rendements d'échelle CONSTANTS sont donc compatibles avec l'existence d'un équilibre général, même s'ils ne permettent pas de déterminer le montant à produire par chaque firme individuellement.

comportement de *price taker*, ils n'agissent que sur les quantités et non sur le prix ; ils choisissent le vecteur $\mathbf{y_f}$ idéal, noté $\mathbf{y_f^*}$, parmi tous les plans de production faisables.

Les entreprises distribuent intégralement leur profit aux ménages, chacun d'eux recevant une fraction proportionnelle à la part de capital qu'il détient dans cette entreprise.

Les consommateurs (qui agissent également comme des *price takers*) tentent de maximiser leur utilité, tout en respectant leur contrainte budgétaire. Le vecteur idéal est noté $\mathbf{x_h^*}$.

Le triplet formé par les trois vecteurs liés **p*, y*, x*** forme un équilibre pour autant que soit satisfaite l'égalité entre l'offre et la demande sur les différents marchés. Cette condition s'écrit :

$$\mathbf{z} \leq 0 \text{ mais pour } z_i < 0, \text{ on a } p_i = 0 \qquad (6.20)$$

Le vecteur **z** est celui de la demande excédentaire. On a : $\mathbf{z = x - y - \varepsilon}$. A l'équilibre, ce vecteur devrait être nul, c'est-à-dire composé exclusivement d'éléments nuls ; toutefois, dans le cas des biens libres, une demande excédentaire négative est admise pour autant que le prix correspondant soit nul.

Arrow et Debreu démontrent mathématiquement qu'un tel équilibre existe. La clé de voûte est, comme chez von Neumann le théorème du point fixe. La démonstration doit résoudre certaines difficultés, notamment tenir compte de l'interaction entre les consommations individuelles : étant donné les consommations de certains individus, des vecteurs deviennent impossibles pour les autres individus : pour résoudre cette difficulté, Arrow et Debreu font appel à la théorie des jeux.

Il importe de remarquer un point sur lequel le modèle d'Arrow-Debreu est en rupture par rapport à toute l'économie politique antérieure, y compris celle de Walras. Plus aucune proportionnalité n'est supposée entre le profit d'une entreprise et son capital. Le taux de profit des différentes entreprises ne tend pas vers l'égalité. L'équilibre général d'Arrow-Debreu ne peut donc pas revendiquer le statut d'équilibre à long terme.

Dans les années qui suivirent la publication de cette démonstration, Arrow et Debreu, ainsi que d'autres économistes, continuèrent à affiner la démonstration, surtout dans le but d'alléger les hypothèses. Debreu fournit ainsi une démonstration de l'existence où l'hypothèse H1 est remplacée par la convexité de *Y*, l'ensemble de la production possible de toute l'économie, ce qui est moins contraignant que la nécessité de la convexité dans chaque firme.

EQUILIBRE D'ARROW-DEBREU : CARACTERISTIQUES ET CAS PARTICULIERS

Cette démonstration constitue une étape fondamentale dans le développement de l'économie néowalrassienne, mais pas un terminus. Arrow et Debreu continuèrent à produire de nombreux théorèmes chacun de son côté. Ils publièrent également, chacun, un ouvrage de synthèse : « The Theory of Value : An Axiomatic Analysis of Economic Equilibrium » de Debreu sortit en 1959. Arrow écrivit « General competitive analysis » avec Frank Hahn en 1971.

Outre la question de l'existence de l'équilibre, Arrow, Debreu et Koopmans ont également redémontré, sur base de leur nouvelle approche mathématique, les théorèmes fondamentaux du bien-être.

---

[1] Etonnant retour des choses : les classiques utilisaient le terme « profit » pour la rémunération globale de l'entrepreneur capitaliste. Les néoclassiques ont abandonné cet usage, gardant l'exclusivité du terme pour le résidu après rémunération de tous les facteurs. La première acception effectue son retour.

Voyons maintenant quelques questions particulières abordées dans les deux ouvrages précités.

Les non-convexités

C'est le double processus de maximisation de l'utilité par les consommateurs et du profit par les firmes, qui détermine l'offre et la demande des différents biens. Or ce processus d'optimisation a pour condition, la convexité des ensembles X_i et Y_j. L'existence de l'équilibre dépend donc de l'hypothèse de convexité. Arrow et Hahn écrivent : « Convexity implies that responses of firms and households to changes in prices tend to be continuous ; even when jumps occur, every point in the whole interval between the two extremities of the jump is a permissible response so that there are no gaps in which an inequality between supply and demand can be fitted »[1]. Ils donnent l'exemple suivant (un cas d'école):

Soient deux biens *A* et *B*, le premier entrant dans la production du second et n'ayant pas d'autre utilisation. Nous nous intéressons à l'équilibre entre l'offre et la demande sur le marché du facteur *A*. Pour simplifier, supposons que l'offre de *A* (S_A) soit fixe (élasticité nulle). La demande de *A* dépendra de la production de *B*.

Commençons par le cas où l'ensemble Y_f est convexe, illustré par les compartiments A1 et A2 de la figure 6.7. Dans le compartiment A1, la fonction d'efficience révèle des rendements d'échelle d'abord constants, puis décroissants à partir d'une quantité égale à q'. Le compartiment A2 montre la courbe de demande de *A* (D_A) déduite de cette fonction de production, indiquant la quantité de *A* demandée en fonction du rapport PB/PA. Plus ce rapport de prix est élevé, plus le vecteur de prix s'approche de la verticalité.

- Si le rapport p_B/p_A est inférieur à un certain seuil *k* (donné par l'équation $y_A = -k.y_B$), son vecteur (le 3) sera fort penché vers la droite. Imaginons la figure 6.5 avec un tel vecteur ; l'impossibilité de la maximisation du profit est évidente. La firme ne peut échapper aux pertes qu'avec une production nulle ; elle choisira donc $y_A = 0$ pour ces rapports de prix.
- Si $p_B/p_A = k$ (vecteur 2), le vecteur de prix sera orthogonal au segment rectiligne de la courbe d'efficience. Le maximum de profit, égal à zéro[2], peut être obtenu pour n'importe quelle valeur de y_A comprise entre zéro et *q'*.
- Si $p_B/p_A > k$ (vecteur 1), la firme obtiendra un profit positif en consommant une quantité $y_A > q'$, qui va croissant avec p_B/p_A.

La courbe de demande de *A* ainsi construite coupe la courbe d'offre ; il y a donc un équilibre (E).

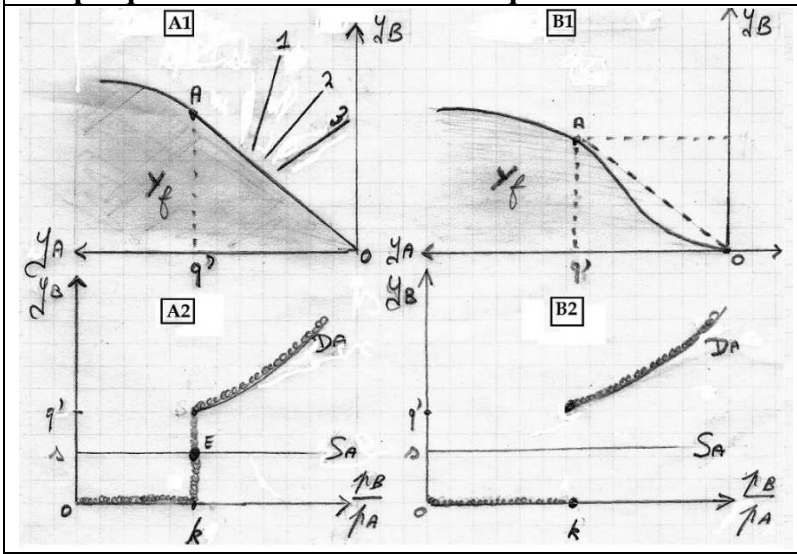
Si on conserve exactement les mêmes données du problème, sauf que le segment rectiligne de la courbe d'efficience est remplacé par un secteur à rendements croissants (Y_f non convexe), l'effet sur D_A est visible dans le compartiment B2.

- Pour les prix inférieurs ou supérieurs à *k*, rien ne change.
- Par contre, pour le prix *k*, sur la portion OA de la courbe d'efficience, aucune combinaison (y_A,y_B) autre que zéro et *q'* n'apporte un profit non négatif. Le producteur de *B* demandera y_A valant soit zéro, soit *q'* ; si l'offre de *A* est comprise entre ces bornes, l'offre et la demande ne se couperont pas et l'équilibre ne sera pas défini.

[1] Arrow & Hahn [16] p. 169

[2] Le produit de deux vecteurs orthogonaux est toujours nul.

Graphique 6.7 : non-convexités et équilibre



Arrow et Hahn s'intéressent aux économies comportant des non-convexités. Ils proposent une méthode permettant de mesurer l'ampleur absolue des non-convexités ainsi que celle du déséquilibre (les demandes excédentaires non nulles). Ils arrivent à la conclusion suivante : l'ampleur ABSOLUE du déséquilibre augmente avec celle des non-convexités mais elle reste constante lorsque varie la taille de l'économie (nombre d'agents). Par rapport à la taille de l'économie, toutes autres choses restant égales, la magnitude du déséquilibre provoqué par les non-convexités tend donc à se réduire sensiblement. L'idée sous-jacente est celle-ci : lorsque plusieurs firmes sont confrontées à un choix entre plusieurs plans de production également profitables, toutes n'opteront pas pour le même plan. Dans l'exemple ci-dessus, au prix *k*, certaines firmes produiraient *q'* et d'autres produiraient zéro. Supposons que le facteur *A* soit demandé par exactement DEUX firmes. La courbe de demande du marché, pour le prix *k*, comprendrait les trois valeurs : 0, q', 2q'. Les « trous » dans la courbe de demande gardent la magnitude *q'*, comme lorsqu'il n'y avait qu'une firme. Mais la diversité des comportements rendrait la demande agrégée assez proche de l'offre globale.

Unicité de l'équilibre

L'équilibre général d'Arrow-Debreu, est-il unique ? Equilibre unique signifie qu'une seule combinaison de prix permet à l'économie d'atteindre l'équilibre. Arrow et Hahn ont analysées les conditions de l'unicité dans l'ouvrage précité[1]. La matière est extrêmement complexe ; résumons leur analyse en deux mots.

Revenons à la matrice **A** dont il fut question lors de l'examen de la stabilité hicksienne. Hicks avait défini la caractéristique de cette matrice qui assurait la stabilité de l'équilibre (signes des déterminants des mineurs en alternance positifs et négatifs). Arrow et Hahn envisagent différents cas correspondant à des caractéristiques particulières de la matrice **A** et posent chaque fois la question : dans ce cas, l'équilibre est-il unique ? Cette analyse ne leur a pas permis de déterminer une condition nécessaire à l'unicité de l'équilibre. Par contre, diverses hypothèses relatives à cette matrice agissent comme condition suffisante. Toutefois, la plupart de ces conditions mathématiques représentent des exigences, soit qui sont irréalistes, soit auxquelles il est difficile de trouver une signification économique.

[1] Il existe toute une littérature sur les sujets de l'unicité et de la stabilité. En somme, Arrow et Hahn font une espèce de synthèse.

En définitive, une condition sort du lot, plus réaliste et sensée que les autres : la *dominance diagonale*. Ce terme un peu barbare reflète que dans la matrice **A**, l'effet du prix des biens sur leur propre demande excédentaire est donné par les éléments qui constituent la diagonale (a_{11}, a_{22}…). Et justement, cette condition stipule que la demande excédentaire de tout bien est plus affectée par une variation de son propre prix que par une variation des autres prix. Les a_{ij} où $i \neq j$ sont inférieurs aux a_{ii}. Diverses conditions secondaires, que nous passons sous silence, s'imposent en outre pour que la dominance diagonale puisse être posée comme condition nécessaire à l'unicité.

La *dominance diagonale* est une condition moins contraignante qu'une autre condition à laquelle on pourrait penser : la *substituabilité brute* entre tous les biens de l'économie. Des biens *i* et *j* sont des substituts bruts si $a_{ij} > 0$, c'est-à-dire si le renchérissement de *j* fait croître la demande excédentaire de *i*. Arrow et Hahn montrent que la dominance diagonale est impliquée par la substituabilité brute mais que l'inverse n'est pas vrai. Vu l'existence de biens complémentaires, l'hypothèse de la substituabilité brute généralisée n'est pas réaliste.

Stabilité (dynamique) de l'équilibre général

Comme nous le savons déjà, l'heure est à la *stabilité dynamique*. Elle est qualifiée ainsi car l'analyse prend en compte le facteur temps. La variable *t* représentant le temps entre dans les équations. Le temps $t = 0$ correspond à la constatation initiale d'un déséquilibre. Celui-ci entraîne un ajustement du prix : cet ajustement prend un certain délai, pendant lequel le système de prix parcourt un « chemin ». Si au bout de ce chemin, l'économie arrive à l'équilibre, on peut conclure à la stabilité du système, même s'il faut que *t* tende vers l'infini pour que ce dénouement heureux atteigne sa plénitude. Ceci peut s'écrire :

$$\lim_{t\to\infty} \boldsymbol{p}(t) = \mathbf{p^*} \qquad (6.21)$$

p(t) est le vecteur de prix au temps *t* et **p*** est le vecteur de prix d'équilibre. Dans l'analyse de la stabilité dynamique, la vitesse du processus d'ajustement est importante : elle est donnée par le ratio dp_i/dt. On a :

$$dp_i/dt = F_i(z_i(\mathbf{p})) \qquad (6.22)$$

La vitesse d'ajustement du prix du bien *i* est une fonction (croissante) de la demande excédentaire sur ce marché, qui est elle-même fonction du système de prix dans son ensemble. La fonction F_i est différente d'un marché à l'autre, certains étant plus réactifs que d'autre au besoin d'ajustement.

Après ces préliminaires, résumons très brièvement les chapitres que **Arrow et Hahn** consacrent à la stabilité, dans l'ouvrage précité qui nous sert de référence. Bien que dans un deuxième temps, ils envisageront d'autres hypothèses, leur analyse part du *tâtonnement* walrassien, qui se caractérise par :

- des prix criés par le commissaire-priseur et
- l'interdiction de toute transaction avant que l'équilibre soit atteint.

L'aspect essentiel du problème est la règle que le commissaire-priseur est censé observer, règle que nos auteurs résument comme suit : « Never change any price if the given price vector is an equilibrium ; never leave all prices unchanged if the given price vector is not an equilibrium »[1]. La variation du vecteur de prix dont il est question dans la deuxième partie de la phrase est résumée par l'équation (6.22). Au commissaire de trouver la fonction F_i qui convient le mieux. Les biens libres font toutefois exception en ce sens qu'une offre excédentaire de leur part ne doit pas inciter à une baisse de leur prix.

[1] Arrow & Hahn [16] p. 266

Concernant le vecteur **p,** Arrow et Hahn l'envisagent de deux façons : soit les prix sont exprimés dans une unité abstraite, soit l'un des biens sert de numéraire et son prix est donc invariable. Le deuxième cas se révélera plus décevant, car le choix du numéraire ne sera pas toujours sans influence sur les conclusions en matière de stabilité.

Les équations (6.22) des différents marchés forment ensemble un système d'équations différentielles simultanées. La solution de ce système est un chemin parcouru par le système de prix ; comme le chemin parcouru à un moment donné dépend de *t* et des conditions initiales **p**(0), Arrow et Hahn notent la solution **p**(t|**p**(0)). Le lien entre le chemin et le point de départ est essentiel. Arrow et Hahn considèrent **p**(t|**p**(0)) comme un fonction continue de **p**(0). Pour mener cette analyse, ils sont obligés de poser, contre leur plein gré, que le chemin partant d'un **p**(0) donné est unique : les chemins parcourus à partir de conditions initiales différentes peuvent ne pas se rencontrer, mais un seul chemin est possible à partir d'une condition initiale donnée. Nos auteurs démontrent que dans ces conditions, si le chemin converge, il ne peut converger que vers un équilibre.

Concernant la stabilité du système, Arrow et Hahn l'envisagent à deux niveaux : la stabilité de la règle du commissaire et la stabilité de l'équilibre : « The auctioneer's rule will be called *globally stable* if for every **p**(0) > 0, the solution **p**(t|**p**(0)) approaches an equilibrium (…) If the rule followed by the auctioneer is globally stable and if the economy possesses a unique equilibrium then that *equilibrium* is called *globally stable*. If the economy possesses more than one equilibrium, then no equilibrium can be globally stable »[1]. La stabilité globale d'un équilibre postule celle de la règle, mais l'inverse n'est pas vrai.

Comment savoir si dans un cas donné, la règle du commissaire est stable ? La méthode de *Lyapounov* permet de résoudre le problème. La règle sera stable[2] si le système des équations (6.22) admet l'existence d'une fonction V[**p**(t|**p**(0)] qui a pour caractéristiques :
- d'être continue
- de converger pour tous les **p**(0) admissibles
- d'être constante si et seulement si **p**(0) est un équilibre.

Après la *stabilité globale* d'un équilibre, comparable à un aimant dont le champ se fait sentir quel que soit le point de départ (en déséquilibre), voyons la *stabilité locale* d'un équilibre, dont le champ d'attraction est limité à une zone formée par des vecteurs de prix assez proches du vecteur de prix d'équilibre. Ce concept permet d'étudier l'effet de petites déviations par rapport à un équilibre : s'en éloigne-t-on ou y est-on ramené ? Dans le cas d'un équilibre multiple, si la règle du commissaire est globalement stable, certains équilibres peuvent être localement stables mais pas tous. Ceci est illustré par la troisième figure du graphique 6.1 : la règle est globalement stable, mais les équilibres sont alternativement stables et instables (localement). En outre, comme nous le savons déjà, aucun de ces équilibres ne sera globalement stable (par définition).

Armés de la méthode de Lyapounov, Arrow et Hahn tentent de déterminer les conditions économiques qui assurent la stabilité. Comme pour la stabilité hicksienne et comme pour l'unicité de l'équilibre, les propriétés de la matrice **A** jouent un rôle prépondérant. Nos deux auteurs analysent de nombreux cas. En définitive, les deux seuls, où la stabilité globale de la règle est assurée, sans qu'il soit nécessaire d'adjoindre d'autres hypothèses, sont la *substituabilité brute* et l'hypothèse que tous les consommateurs sont identiques. Même la

[1] Arrow & Hahn [16] pp. 271-272
[2] Je passe sous silence le cas de plusieurs équilibres non isolés les uns des autres.

dominance diagonale, qui assurait l'unicité de l'équilibre, ne suffit pas pour en garantir la stabilité[1].

Arrow et Hahn tentent ensuite de se défaire de l'hypothèse trop lourde du *recontrat*. Les agents sont maintenant autorisés à acheter et vendre à tout prix, même s'il ne s'agit pas d'un prix d'équilibre. Toutefois leur analyse manque d'intérêt, car elle se limite au cas de l'*échange pur* (absence de production). Il convient de préciser que cette optique ne soustrait qu'une des deux hypothèses du tâtonnement walrassien. Arrow et Hahn préfèrent conserver l'hypothèse d'un commissaire-priseur qui crie les prix, dont il est très difficile de se passer. « For example, in a production economy, if every firm faces a horizontal demand curve (or thinks that it faces such a curve), it is not easy to visualize any firm changing the price at which its product is sold »[2]. Autrement dit, la vision traditionnelle de la concurrence parfaite condamne à conserver le commissaire-priseur… A moins de prendre le chemin d'Edgeworth qui correspond à la situation inverse : le recontrat est la règle, mais il n'y a pas de commissaire-priseur.

La conjecture d'Edgeworth remise à l'honneur

La *conjecture d'Edgeworth* est remise à l'honneur principalement par Herbert Scarf et Gérard Debreu, dans les années soixante. Voici comment ils présentent le problème[3] :

- « An allocation for the economy $\mathcal{E}$ specifies the commodity vector assigned to each agent ». Dans un processus d'échange pur, elle est une redistribution de l'allocation initiale ; elle n'est donc faisable que si la quantité de chaque bien dans la nouvelle distribution ne dépasse pas la quantité initiale.
- « A coalition S of agents blocks the allocation (x_i) if its members can redistribute their own initial commodity vectors among themselves so that every one of them is at least as satisfied as he is with the allocation (x_i) and at least one of them is more satisfied. »
- « The core of the economy is formally defined as the set of attainable allocations that no coalitions of agents blocks ». C'est donc la courbe des règlements finaux d'Edgeworth. Nous traduisons par *noyau.*

Ils démontrent mathématiquement que:

1) tout équilibre général de $\mathcal{E}$ est dans le noyau de $\mathcal{E}$.
2) si une allocation (x_i) est dans le noyau de $\mathcal{E}$ pour tout nombre de participants à l'échange, elle doit être un équilibre général de cette économie.

Debreu et Scarf restent dans le schéma d'Edgeworth d'une économie multipliée : les agents sont répartis en groupes caractérisés par l'identité des goûts et des allocations initiales ; le nombre de ces groupes est quelconque mais constant : la croissance en taille de l'économie repose donc uniquement sur l'augmentation du nombre d'agents dans les groupes. Cette façon d'envisager l'extension de l'économie, certes un peu bizarre, sera abandonnée dans d'autres démonstrations de la conjecture d'Edgeworth (notamment Aumann en 1964).

L'EQUILIBRE INTERTEMPOREL ET L'INCERTITUDE

Parmi les divers raffinements qu'Arrow et Debreu apportent à leur théorie dans les années qui suivent la démonstration de l'existence de l'équilibre, il y a la prise en compte du temps et de

[1] Un système peut satisfaire la condition de *dominance diagonale* pour tel choix de numéraire et pas pour tel autre. Qu'il y ait *dominance diagonale* avec un numéraire suffisait à l'unicité de l'équilibre. Par contre, la règle du commissaire pourrait être déclarée stable avec tel numéraire et instable avec tel autre.
[2] Arrow & Hahn [16] p. 325
[3] Debreu & Scarf [74] p 286

l'incertitude. Commençons par le temps : une comparaison avec l'*équilibre temporaire* de Hicks s'impose. Voyons d'abord ce qu'il y a de commun entre les deux conceptions. Comme chez Hicks, les consommateurs et les firmes établissent des plans de consommation et de production intertemporels ; ils planifient leur production et leur consommation des différents biens pour toutes les périodes à venir. Comme chez Hicks, ces plans résultent d'une démarche d'optimisation.

A part cela, la conception d'Arrow et Debreu, souvent appelée *full intertemporal equilibrium*, s'oppose à l'*équilibre temporaire* ; d'une part, ils supposent l'existence de marchés non seulement pour les biens présents mais aussi pour TOUS les biens futurs à toutes les dates, alors que Hicks dédaignait les marchés à terme et ne retenait qu'un seul marché futur : celui des prêts de monnaie. Comme le lecteur, Arrow et Debreu savent que tous ces marchés futurs n'existent pas. Leur modèle est celui d'une économie « idéale » rendant possible toutes les optimisations. Sans un marché pour chaque date, sur lequel les agents peuvent opérer leurs transactions dès aujourd'hui, la répartition de la production et de la consommation dans le temps ne peuvent être optimisées. L'équilibre temporaire de Hicks est moins optimal mais plus réaliste.

Contrairement à celle de Hicks, l'économie d'Arrow-Debreu n'est pas *séquentielle*. Le processus de tâtonnement walrassien, la grand-messe au cours de laquelle tous les prix sont fixés et toutes les transactions conclues, se tient une seule fois : à la période initiale. Par la suite les agents n'auront plus qu'à mettre en pratique les livraisons de biens décidées initialement. Il n'y aura jamais de pénurie ou d'invendus puisque le processus de tâtonnement veille à équilibrer les marchés futurs jusqu'à l'horizon temporel. Dans un tel contexte, les livraisons ne doivent même pas s'accompagner d'un paiement monétaire. En fait le tâtonnement initial est un grand troc où tout le monde offre certaines choses présentes ou futures, en contrepartie d'autres choses présentes ou futures. Le vase de Chine que j'acquerrai en période 18 est la contrepartie du travail que j'aurai fourni pendant les périodes zéro à 17 ; le paiement, c'est tout simplement le fait que les échanges sont équilibrés grâce à un système de prix adéquat. Ceci constitue d'ailleurs une différence avec le fonctionnement réel des marchés futurs où le paiement a lieu à la livraison, bien que le prix soit fixé initialement.

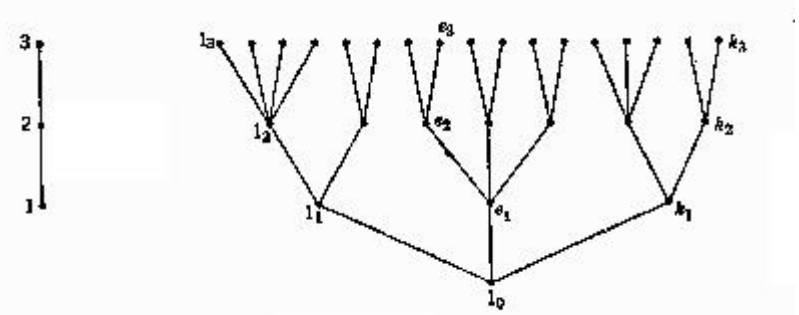
Chez Hicks comme chez Walras avant lui (rappelons-nous ses marchandises *U* et *E*), la monnaie et les titres figuraient dans la liste des biens. Dans un équilibre intertemporel complet, ils ne sont plus nécessaires. Il y a juste un grand troc de produits et de facteurs primaires. L'énoncé d'un modèle, dont la monnaie est absente, présuppose l'idée, soit que la monnaie n'affecte le mécanisme économique que superficiellement, soit que la prise en compte de la monnaie doit faire l'objet d'une étude distincte. Nous verrons au chapitre 6.5 comment des auteurs néo-walrassiens tentent d'intégrer la monnaie.

La conception de l'équilibre intertemporel complet d'Arrow et Debreu semble s'être imposée face à l'équilibre temporaire hicksien.

A première vue, il y a peu de place pour **l'incertitude** dans ce modèle. Arrow et Debreu veillent pourtant à lui en réserver une. Ils construisent une nouvelle théorie des choix en incertitude, qu'ils lient à leur équilibre général. C'est la théorie des biens à *l'état contingent*. Un parapluie en 2019 et un parapluie en 2020 sont deux biens différents. Mais un parapluie en 2020 si cette année est ensoleillée et un parapluie en 2020 si cette année est pluvieuse sont également deux biens différents. ENSOLEILLE et PLUVIEUX sont deux états alternatifs et ils sont aléatoires dans la mesure où les données climatiques de l'année 2020 sont indépendantes des actions des agents et inconnues à la période 0, lorsque se concluent les transactions.

A chaque période, quantité de situations peuvent évoluer de plusieurs façons ; des événements, soit qui concernent l'agent personnellement (sa santé, notamment le fait d'être en vie ou non, sa situation familiale…), soit qui concernent l'économie en général (tels produits qui deviennent abondants et meilleur marché ou au contraire se raréfient…). Debreu illustre cette théorie avec l'*arbre des événements*[1] :

Graphique 6.8 : l'arbre des événements



Voici un exemple sur trois périodes. Chaque état conserve la mémoire du chemin parcouru. Le nombre d'événements augmente à chaque période comme les branches d'un arbre.

Le nombre d'*états contingent* (aussi parfois appelés *événements*) est supposé être fini, mais on conçoit qu'il est astronomiquement grand. Il suffit de penser au nombre de facteurs influençant le moindre phénomène.

Voyons maintenant ce que savent et ce qu'ignorent les agents dans ce modèle. Ils ignorent le futur. Donc ils ne savent pas quel état contingent se concrétisera. Par contre, ils ont en tête la composition de l'ensemble des états possibles. Ils sont à même d'attribuer à chaque état une probabilité subjective. Et pour chaque état, dès la période initiale, ils peuvent échafauder un plan de production et de consommation optimal.

Comment ces états contingents sont-ils intégrés dans la théorie de l'équilibre général ? La définition du bien est adaptée. « Un contrat de livraison d'une marchandise spécifie maintenant, outre ses propriétés physiques, son lieu et sa date de disponibilité, un événement dont la réalisation conditionne la livraison »[2]. Dans ces conditions, le prix p_{is} est « the price paid in return for a promise to supply one unit of commodity *i* if state *s* occurs and nothing otherwise »[3]. L'indice *s* renseigne à la fois sur l'état et la période. « Le paiement est fait irrévocablement, bien que la livraison n'ait pas lieu si les événements spécifiés ne se sont pas réalisés. Un agent qui achète un quintal de blé rouge d'hiver n°2 disponible à Chicago, à la date *t*, *quel que soit l'événement*, achète en fait autant de marchandises qu'il y a d'événements à la date *t* »[4].

Le vecteur d'activité $\mathbf{y_{fs}}$ est le vecteur de production de la firme *f* si l'état *s* advient. Un exemple serait la production d'une ferme si les conditions climatiques sont favorables. La variable à maximiser devient $\sum_i \sum_s p_{is}.y_{is}$.

Le cas du consommateur est rendu plus complexe par deux difficultés :

[1] Debreu [72] p. 107
[2] Debreu [72] p. 106
[3] Arrow & Hahn [16] p. 125
[4] Debreu [72] p. 108

- « It should be noted that a preference ordering for consumption vectors in the new interpretation contains elements of judgement about the likelihoods of the different states of the world as well as elements of evaluation of tastes »[1]. Un vecteur comportant une consommation x_{is} très désirée mais improbable sera peut-être jugé inférieur à un autre vecteur proche où x_{is} est remplacé par un élément moins prisé mais plus probable. La théorie de l'*utilité espérée* (cf. 6.1.2 et 8.5) est une explication de la manière dont les probabilités subjectives affectent le jugement du consommateur. Mais elle n'est qu'une explication parmi d'autres. La théorie des biens à l'état contingent ne l'implique donc pas.
- La convexité des préférences reste une condition de l'existence d'un équilibre, mais il est démontré que les préférences ne sont convexes dans un monde incertain que si l'individu fait preuve d'aversion envers le risque.

N'est-il pas paradoxal que cette théorie attribue un caractère risqué à la situation des ménages, d'où l'appel aux probabilités, alors que la situation des entreprises en semble dénuée ? N'oublions pas que tous les contrats sont conclus à la période initiale ; la firme vend donc toute sa production dans tous les états possibles, qu'ils adviennent ou non. Au niveau des quantités vendues, il n'y a donc pas d'incertitude et le profit est connu dès la période initiale. Mais subrepticement, l'incertitude, qui est inévitable, vient se loger dans le système de prix : « Probabilities will, of course, be reflected in prices »[2]. Malheureusement, cette méthode rend plus difficile l'analyse de l'effet de l'incertitude sur les comportements. Comme on le constate, avec les *états contingents*, le modèle d'Arrow-Debreu fait un bond dans l'abstraction.

[1] Arrow & Hahn [16] p. 123
[2] Diamond [75] p. 774

6.4- DEUX DIMENSIONS SUPPLEMENTAIRES : LE TEMPS ET L'ALEA

6.4.1. Malinvaud, le capital et l'intérêt

Avec le développement de l'école néo-walrassienne, surtout avec la version de l'équilibre général intertemporel complet, le cadre analytique de la science économique a beaucoup changé. Il était « agrégatif » c'est-à-dire qu'auparavant les économistes traitaient des grandeurs agrégées telles LA consommation, LE capital, LE taux d'intérêt, LE salaire… L'économie néo-walrassienne désagrège : elle réduit l'économie à un nombre astronomique (mais fini) d'entités élémentaires, des espèces d'atomes de la matière économique. Facteurs non produits, produits de consommation, produits d'équipement, monnaie, tout n'est que BIENS. En ajoutant le temps et l'incertitude, on ne fait que multiplier le nombre de biens. Comment reconstruire une théorie du capital et de l'intérêt à partir de cette approche atomistique ? Telle est la tâche à laquelle s'attelle l'économiste français Edmond Malinvaud, d'abord dans un article publié en 1953 dans le cadre des travaux de la commission Cowles, ensuite dans ses « Leçons de théorie microéconomique » parues en 1972, dont nous résumons ci-dessous le chapitre intitulé « Economies temporelles ».

LE TAUX D'INTERET NEO-WALRASSIEN

Nous n'avons pas encore rencontré le taux d'intérêt dans la conception néowalrassienne. En fait, il se cache dans les plis du système de prix. Partons du système de prix d'Arrow-Debreu que nous simplifions en le délestant des états contingents, ce qui laisse deux dimensions : la caractéristique physique (donnée par l'indice q variant de 1 à Q) et la période (donnée par l'indice t variant de 1 à T). Lions-le aux théories de Böhm Bawerk et Fisher concernant la préférence pour le présent. Si l'agent préfère une utilisation plus rapprochée d'un bien, logiquement, le prix de marché d'une livraison rapprochée devra être supérieur à celui d'une utilisation plus éloignée. Vu que le principe d'impatience s'applique, on aura pour le bien q :

$$p_{q,1} > p_{q,2} > \ldots > p_{q,T-1} > p_{q,T} \qquad (6.23)$$

Le taux d'intérêt entre deux périodes n'est rien d'autre que l'agio entre les prix de ces périodes. Notons par ρ_t le taux d'intérêt entre t et $t+1$.

Par exemple, si $p_4 = 100$ et $p_5 = 98$, $\rho_4 = (100/98) - 1 = 2{,}04$. N'oublions pas que nous sommes dans une économie non monétaire. Les seuls prêts sont des prêts en nature. Prenons l'exemple du prêt de semences à un fermier. Comme nous le savons, le rapport des quantités est l'inverse du rapport des prix. Si de $t4$ à $t5$, le prix de la semence passe de 100 à 98, cela signifie que le fermier devra rembourser 100 quintaux de semence en t5 pour 98 empruntés en $t4$: cela fait bien un taux d'intérêt de 2,04%.

Une remarque vient directement à l'esprit ; si les différents biens physiques ont des prix qui évoluent de façon divergente, ils auront des taux d'intérêt différents. En effet, la théorie néowalrassienne ne se soucie pas d'avoir un taux d'intérêt unique. Chaque bien physique a ce qu'on appelle un « taux d'intérêt *propre* » (cf. sous-chapitre 5.1.2) et rien n'assure qu'ils sont égaux. On peut écrire la formule du taux d'intérêt propre du bien q :

$$p_{q,t+1} = p_{q,t} .1/(1+\rho_{q,t}) \qquad (6.24)$$

LE taux d'intérêt général auquel la théorie économique nous avait habitués n'est pas nécessaire dans le présent canevas[1]. A ce stade, une première attitude consiste à s'en tenir au constat que chaque bien a un taux d'intérêt propre ; le taux d'intérêt général n'est qu'un leurre. Telle est l'attitude d'économistes comme Hahn et Bliss. Ce dernier écrit : « Capital theory should be liberated from the concept of the rate of interest, meaning by that one rate »[2].

Malinvaud, pour sa part, se refuse à abandonner complètement LE taux d'intérêt.

LA THEORIE DU CAPITAL ET DE L'INTERET DE MALINVAUD

Malinvaud ne crée pas à proprement parler une nouvelle théorie du capital et de l'intérêt. Il veut précisément montrer que le modèle désagrégé permet d'obtenir les résultats des théories traditionnelles sous une forme proche. Avant d'examiner ces résultats, il nous faut prendre connaissance de deux outils que Malinvaud s'accorde pour traiter le sujet.

D'abord, comment Malinvaud ressuscite-t-il le taux d'intérêt général ou plutôt comment en crée-t-il un ersatz ? Considérant le système de prix d'Arrow-Debreu, qu'il appelle *prix actualisés*, Malinvaud institue une représentation modifiée de celui-ci, qui prévaudrait si on était dans une économie séquentielle comme celle de Hicks, avec paiement au moment de la livraison. Pour ces prix, qu'il appelle *prix non actualisés*, l'équation (6.23) ne s'applique pas puisqu'elle se justifiait par l'actualisation. Vu qu'on est dans une économie non séquentielle, un tel système de prix est évidemment purement artificiel, mais il sera utile à Malinvaud. Chaque bien a donc maintenant deux prix : le prix actualisé que nous connaissons $p_{q,t}$ et un prix non actualisé noté $p'_{q,t}$, construit artificiellement à partir du premier suivant la méthode suivante. Un des biens (disons *Q*) est choisi comme numéraire. Son prix non actualisé vaudra *un* à chaque période. Les prix non actualisés des autres biens se calculent de telle façon qu'à l'intérieur de chaque période les prix relatifs soient les mêmes dans le système de prix non actualisés et dans le système actualisé. Appelons ρ_t (sans indice de nature de bien), le taux d'intérêt propre du numéraire. Pour chaque bien, se vérifient la relation :

$$p'_{qt} = p_{qt} / p_{Qt} \qquad (6.25)$$

L'astuce consiste à mettre en exergue le *taux d'intérêt propre* du numéraire et voir si l'assimilation du taux d'intérêt général à celui-ci permettrait de tirer des conclusions théoriques le concernant. Pour que cette méthode soit fondée, il convient de choisir adéquatement le numéraire : il faut choisir un bien dont le prix évolue de façon assez neutre, à la moyenne des autres biens. Avec un bon numéraire, le niveau général des prix non actualisés est stable d'une période à l'autre[3].

Malinvaud réintroduit également la notion de *fonction de production*, qu'il reformalise à sa façon. Le vecteur **y** de Koopmans est décomposé en deux vecteurs. A chaque période *t*, la production peut être décrite à l'aide du couple de vecteurs $(\mathbf{a_t},\mathbf{b_{t+1}})$ à Q composants:

- $\mathbf{a_t}$ est le vecteur des inputs qui indique les quantités de tous les biens produits ou non produits entrant dans la production. Ses éléments sont positifs contrairement aux inputs de Koopmans.
- $\mathbf{b_{t+1}}$ est le vecteur des biens disponibles en t+1, qui résultent de la production en t.

[1] Si le taux d'intérêt peut varier d'un bien physique à l'autre, il peut aussi varier d'une période à l'autre pour un même bien physique. Selon Fisher et Hicks les taux annoncés par les intermédiaires financiers, uniques sur une durée de plusieurs périodes, ne sont que des moyennes associant des taux dissociés par période.

[2] Bliss [38] p. 10

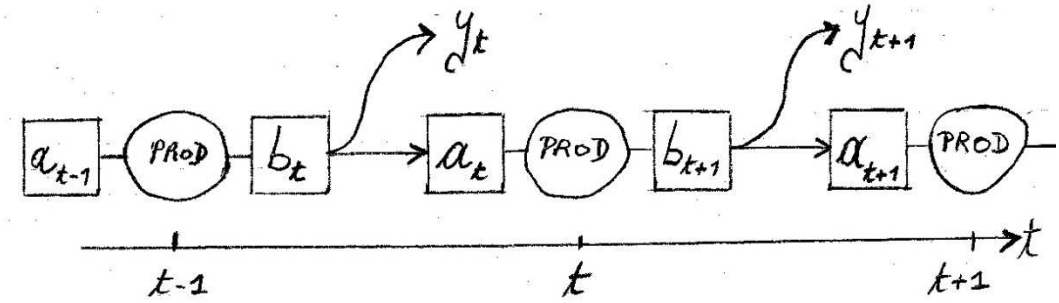
[3] Des considérations macroéconomiques comme l'inflation monétaire n'ont évidemment par leur place ici.

La formule $g(\mathbf{a_t},\mathbf{b_{t+1}}) \leq 0$ est une manière alternative d'exprimer que $\mathbf{b_{t+1}}$ est fonction de $\mathbf{a_t}$. La fonction *g* que Malinvaud suppose différentiable est la bonne vieille fonction de production telle qu'elle fut instituée par Wicksteed et qui a nourri toute l'économie néoclassique. Seule la forme change. Comme l'explique Malinvaud, lorsque Koopmans ou Arrow et Debreu considèrent qu'un plan de production fait partie de l'ensemble des plans de production faisables, ceci implique qu'il y ait une relation cohérente entre les inputs et les outputs. La cohérence d'un ensemble des plans faisables définit implicitement une fonction de production.

Malinvaud note par $y_{q,t}$ la quantité produite du bien *q* à la période *t*, qui est disponible pour la consommation, qu'il appelle *production nette*. Par définition, on a :

$$y_{q,t} = b_{q,t} - a_{q,t} \qquad (6.26)$$

Graphique 6.9: la production dans le temps selon Malinvaud



Le vecteur $\mathbf{a_t}$ est assimilé au capital à l'instant *t*. Comme Malinvaud le reconnaît, c'est une acception approximative puisque le travail est compris dans $\mathbf{a_t}$ et donc inclus dans le capital, à la façon du capital variable de Marx. Pour traiter le capital fixe, Malinvaud recourt à la même méthode que von Neumann et Sraffa ; il l'applique également pour l'encours de production (divisé en stades) : chaque âge de l'équipement, chaque stade du produit est un bien distinct[1].

Le cadre analytique étant posé, voyons l'usage qu'en fait Malinvaud.

Une première analyse porte sur la liaison entre l'intérêt et le profit[2]. Dans une perspective intertemporelle, la maximisation du profit signifie la maximisation de la valeur actuelle des profits (bruts) futurs, ce qui correspond à la valeur de l'entreprise elle-même[3]. Dans le cas d'une production unique, l'entreprise maximisera :

$$\Sigma_t \beta_t . p'_t . (b_t - a_t) \quad \Leftrightarrow \quad \Sigma_t p_t . (b_t - a_t) \qquad (6.27)$$

[1] Malinvaud précise bien : « Dans l'étude de l'équilibre, (cette exigence) implique cependant que les prix des équipements anciens et des produits en cours de fabrication soient bien définis » ([246] p. 284). Cette condition n'est quasiment jamais satisfaite dans l'économie réelle. Pour que le prix soit « défini » de façon significative dans le cadre d'un équilibre général concurrentiel, il faut que chaque élément du capital circulant et chaque âge du capital fixe fasse l'objet d'un marché où opèrent des offreurs et des demandeurs infiniment nombreux.

[2] Nous appelons profit BRUT le surplus des recettes sur les coûts autres que le coût du capital et profit NET le profit brut déduction faite du coût du capital.

[3] Il est généralement convenu que tel doit être l'objectif des gestionnaires d'une entreprise.

Où $\beta_t = p_{Qt}/p_{Q1}$; c'est le facteur d'actualisation qui vaudrait $1/(1+\rho)^t$ si ρ_t était identique à toutes les périodes. A la suite de quelques manipulations et hypothèses et en recentrant l'analyse sur une seule période (de t à t+1), Malinvaud en déduit l'équation du profit net (noté π'):

$$\pi'_t = p'_{t+1}.b_{t+1} - (1+\rho_t).p'_t.a_t \qquad (6.28)$$

Le taux de profit brut peut s'écrire :

$$r^\circ_t = (\rho_t.p'_t.a_t + \pi'_t) / p'_t.a_t \qquad (6.29)$$

Malinvaud note que la propriété de convexité de la fonction de production sera souvent associée à des rendements d'échelle constants. Lorsque c'est le cas, π'_t sera égal à zéro, car sinon sa maximisation est impossible (car elle mènerait vers l'infini, le profit croissant proportionnellement à l'échelle)[1]. Dans ce cas, la production rapporte exactement le taux d'intérêt. Alors, on a $r^\circ_t = \rho_t$.

Le *taux marginal de profit* est celui qui résulte d'un investissement supplémentaire δa_t et s'écrit :

$$r_t = (\rho_t.p'_t.\delta a_t + \delta\pi'_t) / p'_t.\delta a_t \qquad (6.30)$$

Ici, Malinvaud retombe sur la formule de Fisher affirmant que la maximisation du profit mène le MEI à égalité avec le taux d'intérêt. En effet si π'_t est maximisé, $\delta\pi'_t$ ne peut que valoir zéro. Alors, on a : $r_t = \rho_t$.

Malinvaud pose ensuite la question : la maximisation du profit sur une durée de plusieurs périodes, implique-t-elle une stratégie à long terme ou suffit-il à chaque période de maximiser son profit en ne tenant compte que des paramètres propres à cette période ? Malinvaud montre que les maximisations successives du profit de la période conduisent à la maximisation du profit à long terme, mais à une condition : il faut qu'à chaque période, la firme puisse travailler avec exactement la quantité de capital dont elle a besoin à ce moment. Conscient que cette exigence est tout à fait irréaliste, Malinvaud note : « Cette propriété suppose l'existence de marchés parfaits pour tous les biens y compris les équipements usagés et les produits en cours de fabrication. Elle implique en particulier qu'aucun coût de transaction ne grève la vente ou l'achat de matériels d'occasion (…) En somme, la propriété en cause suppose que le capital est librement transférable moyennant des prix librement définis. Dans la réalité, une grande partie de ce capital est 'fixe'. Son transfert d'un emploi à un autre a souvent un coût rédhibitoire ». Il faut donc se rabattre sur des stratégies à long terme, au sujet desquelles la théorisation est plus que difficile ; Malinvaud poursuit : « Il semble malheureusement impossible d'arriver à des résultats théoriques généraux quand on prend en compte l'irréversibilité effective des opérations d'investissement dans lesquelles les équipements déjà installés ne peuvent plus en pratique changer d'utilisation »[2].

Malinvaud examine ensuite les conditions d'optimalité (ou d'efficacité, ce qui revient au même) dans le cadre intertemporel. A cette fin, il avance le concept de *programme*, qu'il définit ainsi : c'est l'ensemble des plans des consommateurs et des plans des producteurs, s'étalant sur un intervalle de temps donné. Le programme est réputé *possible* si le plan de chaque agent est physiquement possible pour lui et si de plus pour chaque produit et chaque date, l'offre globale est égale à la demande globale.

---

[1] Ce type d'argumentation, est-il acceptable ? Peut-on écarter la recherche d'un profit net pour le fait de ses effets si aucun élément de motivation n'est dissuasif ?

[2] Malinvaud [246] pp. 295-296.

- Un programme est qualifié d'*optimum de la production* s'il est possible et s'il n'existe pas d'autres programmes possibles ayant un vecteur $\mathbf{a_t}$ de départ comparable qui donneraient une production nette globale $\sum_j y_{jqt}$ (*j* est l'indice de la firme) au moins aussi élevée pour tous les couples (q,t) et plus élevée pour l'un d'eux au moins. Assez logiquement, le critère d'efficacité porte, non sur la production brute, mais sur la production nette y = b – a, celle qui sert de support à la satisfaction des besoins.
- Un programme est qualifié d'*optimum de Pareto* s'il est possible et s'il n'existe pas de programme possible qui soit jugé au moins équivalent par tous les consommateurs et préférable par l'un d'entre eux »[1]. Cet optimum tient compte non seulement de la quantité produite mais également de l'utilité de la production.

Dans le présent chapitre, Malinvaud se contente d'examiner l'optimum de la production. Son analyse consiste à chercher les propriétés des programmes optimaux en les comparants avec d'autres programmes. Malheureusement, cette analyse doit se limiter à des programmes *stationnaires* ou *semi-stationnaires*, sinon la complexité empêche d'arriver à quelque résultat intéressant. Les vecteurs $\mathbf{a_t}$ et $\mathbf{b_t}$ d'un programme stationnaire sont parfaitement répétitifs d'une période à l'autre[2].

Moyennant l'hypothèse que l'ensemble des couples $(\mathbf{a_t},\mathbf{b_{t+1}})$ est convexe et pour autant que la fonction *g* réponde à certaines conditions, Malinvaud démontre que face à un optimum de production, il existe un vecteur de prix **p** non nul et un nombre ρ (supérieur à –1), tels que de tous les couples $(\mathbf{a_j},\mathbf{b_j})$ possibles, c'est celui de l'optimum de production qui maximise le profit net de la firme *j* donné par $\mathbf{p.b_j} - (1+\rho).\mathbf{p.a_j}$. Si les firmes cherchent à maximiser leur profit dans un environnement concurrentiel, l'optimum de production doit normalement émerger. L'une des caractéristiques intéressantes de ce système de prix lié à l'optimum est que les taux d'intérêt propres des différents biens sont égaux entre eux et que ce taux d'intérêt unique est lui-même invariable d'une période à l'autre.

Malinvaud spécialise ensuite son analyse pour la concentrer sur les croissances *semi-stationnaires* qui ont un taux de croissance α identique. Il obtient un nouveau résultat intéressant : l'*optimum* admet ici comme caractéristique la *règle d'or* par laquelle le taux de croissance α égale le taux d'intérêt (cf. von Neumann). Pour en arriver là, Malinvaud doit ajouter l'hypothèse que la fonction de production *g* manifeste des rendements d'échelle constants et il modifie légèrement le concept d'optimum de production qu'il appelle maintenant *optimum capitalistique* de telle façon qu'« il n'est tenu aucun compte explicite des situations initiales et terminales du programme »[3]. Malinvaud ne craint pas de pousser cette analyse jusqu'à sa conclusion logique : « Ainsi dans une économie stationnaire les revenus du capital disparaîtraient si le capital était suffisamment abondant, si les opérations de production étaient organisées de manière efficace et si le système de prix reflétait correctement les égalités marginales »[4].

Enfin, Malinvaud pose la grande question du lien entre l'intensité capitalistique et le taux d'intérêt. Pour l'économie néoclassique traditionnelle, aussi bien clarkienne que wicksellienne, de deux économies, celle qui est la plus intense en capital devait aussi avoir le taux d'intérêt le plus bas, un théorème invalidé lors de la *controverse du capital* avec l'argument de la *marche arrière du capital*. L'approche de Malinvaud confirme les conclusions de Cambridgiens : cela peut sembler une victoire pour ceux-ci, mais c'est également un atout pour les néoclassiques de

[1] Malinvaud [246] p. 277.
[2] Comme le lecteur l'aura remarqué, Malinvaud applique ici la méthode de la statique comparative.
[3] Malinvaud [246] p. 300.
[4] Malinvaud [246] p. 302.

disposer d'une théorie qui soit compatible avec la marche arrière du capital. Celle-ci n'est donc plus une arme disqualifiant la théorie néoclassique du capital, puisque sa version néo-walrassienne n'est pas vulnérable.

Pour arriver à ce résultat, Malinvaud adapte son modèle. Interviennent un facteur primaire utilisé en quantité z à un prix w, un bien de consommation non durable A (numéraire) et un bien durable B servant à la fois de capital et pour la consommation (prix = p). Les *productions nettes* sont respectivement y_A et y_B. La quantité de capital entrant comme input dans les productions A et B est la variable a qui sert d'indice de l'intensité capitalistique si $z = 1$. Le profit net à maximiser est :

$$\pi' = y_A + p.y_B - \rho.p.a - (1+\rho).w.z \qquad (6.31)$$

Malinvaud pratique la *statique comparative* sur des équilibres stationnaires différant par leur valeur a pour déterminer la relation entre a et ρ. Il en conclut que le travail tend à se substituer au capital lorsque le taux d'intérêt augmente, mais que cet effet normal de l'intérêt sur l'intensité capitalistique est perturbé par un autre facteur : le rapport entre les productions nettes des deux biens (y_B/y_A), qui dépend des préférences des consommateurs. Une économie X plus intense en capital que l'économie Y peut néanmoins avoir un taux d'intérêt plus élevé que celle-ci si son rapport (y_B/y_A) est nettement plus bas que celui de l'économie Y.

6.4.2. L'incertitude et les anticipations

ARROW ET LES VALEURS MOBILIERES

Nous avions l marchés dans l'équilibre général atemporel. On est passé à $l.S$ dans l'équilibre intertemporel complet où S est le nombre d'états possibles à l'horizon temporel considéré. Dans l'article « The Role of Securities in the Optimal Allocation of Risk-Bearing » (1953), Arrow propose un moyen de réduire ce nombre. Il s'agit de mettre à la disposition des agents le moyen de transférer du pouvoir d'achat d'un état contingent à un autre. Ces transferts sont nécessaires car certains états peuvent affecter négativement les revenus (invalidité) ou faire hausser les prix (sécheresse) ou les besoins (famille nombreuse). Des procédés de transfert existent. Le plus évident est la souscription de contrats d'assurance. Toutefois, ce moyen est insuffisant, car seuls une petite partie des risques potentiels sont assurables. C'est un autre procédé qu'Arrow met en avant : le placement en valeurs boursières.

Il imagine d'abord un cas d'école, certes irréaliste : que pour chaque état contingent, il existe une valeur boursière qui rend une unité de revenu si l'état se réalise et aucun revenu sinon. Cette condition serait suffisante pour rendre possibles tous les transferts de pouvoir d'achats désirables d'un état à l'autre. Dans ce cas, le nombre de marchés nécessaires passe de $l.S$ à $l+S$, en comptant ceux des valeurs mobilières.

Il s'agit certes d'une situation extrême. Dans la réalité, les titres ne collent pas aussi parfaitement aux états contingents. Toutefois, les titres réels, les actions plus que les obligations, sont assez liés aux états contingents de l'économie. Pensons par exemple à la solvabilité de l'émetteur, forcément dépendante de l'état s. Arrow estime que les titres réels peuvent être considérés comme des formes concaténées de ces valeurs élémentaires imaginaires.

RADNER ET LE MODELE D'ARROW-DEBREU EN ECONOMIE SEQUENTIELLE

L'article « Existence of Equilibrium of Plans, Prices and Price Expectations in a Sequence of Markets » (1972) de Roy Radner expose un modèle appliquant la suggestion d'Arrow. A chaque

période /état, les agents qui désirent transférer du pouvoir d'achat achètent ou vendent des titres. Le cadre le plus adéquat est celui d'une *économie séquentielle*. A chaque période, certains marchés contingents font défaut mais cette lacune est compensée par l'existence de marchés boursiers.

Malgré l'économie séquentielle, on reste loin de l'équilibre temporaire hicksien, car la conception des anticipations diffère fondamentalement. Les anticipations des agents hicksiens pouvaient être démenties par les faits. Sous la forme d'une économie séquentielle, Radner élabore un modèle qui a les propriétés de fond de l'économie d'Arrow-Debreu. Notamment, les agents ont la faculté de prévision parfaite.

Concernant les anticipations, Radner écrit : « We can represent a trader's expectations as a function that indicates what the prices will be at a given date in each elementary event at that date. This includes, in particular, the representation of future prices as random variables, if we admit that the uncertainty of the traders about future events can be scaled in terms of subjective probabilities. I shall say that traders have *common* expectations if they associate the same (future) prices to the same events. This does not necessarily imply that they agree on joint probability distribution »[1]. L'hypothèse des anticipations communes suit logiquement celle des anticipations parfaites.

Radner se place comme observateur à la première période. Les périodes ultérieures appartiennent donc au futur et leurs prix n'existent encore que dans la tête des agents, comme anticipations. Mais l'équilibre général de Radner implique l'équilibre, non seulement des marchés présents mais aussi entre les anticipations. « An equilibrium of plans, prices, and price expectations is a set of prices on the first market, a set of common price expectations for the future and a consistent set of individual plans, one for each trader, such that, given the current prices and price expectations, each individual trader's plan is optimal for him, subject to an appropriate sequence of budget constraints »[2]. Les prix d'équilibre qui s'établiront dans le futur confirmeront les anticipations présentes. Radner impose donc la condition que les prix anticipés annulent la demande excédentaire des échanges qui se produiront dans le futur.

Nous n'allons pas développer le modèle mais simplement l'esquisser. Une notion importante est la *paire date-événement*, c'est-à-dire une situation à un moment donné, caractérisée par la période *t* et l'état contingent *s*. Radner se réfère à l'arbre des événements de Debreu (graphique 6.8) et réinsiste sur le nombre FINI de périodes et d'états. A chaque date-événement *m*, les agents passent des contrats entre eux, prévoyant la livraison d'une quantité d'un bien *h* à la date-événement *n*, qui a lieu au même moment que *m* (contrat courant) ou qui lui est postérieure (contrat futur)[3] et pour autant qu'existe un marché de *n* en *m*. Il y a un prix par marché correspondant à une combinaison de *h*, *m* et *n*. Le prix, exprimé en unités de compte, est payé lors de la conclusion du contrat.

Les consommateurs échafaudent des *plans de consommation* $\mathbf{x_i}$, les producteurs des *plans de production* $\mathbf{y_j}$, mais tous deux établissent également des *plans de contrats* $\mathbf{z_i}$ *et* $\mathbf{z_j}$, qui reprennent les achats et ventes qu'ils projettent. Un plan de contrat se caractérise par les quantités achetées ou vendues des différentes combinaisons de *m*, *n* et *h***.** A partir d'une situation *m*, l'ensemble des plans de contrat possibles et autorisés est noté Z_m. Radner impose une limite supérieure à

[1] Radner [291] p. 289
[2] Radner [291] pp. 289-290
[3] Il y a donc la date-événement *m* pour la décision et la date-événement *n* pour la livraison-paiement. Il faut bien sûr que *n* soit compatible avec les événements antérieurs (passage possible le long des branches de l'arbre des événements).

ces livraisons : $z_{hmn} \leq L$ où L est arbitraire[1]. C'est à l'intérieur de cet ensemble, que le producteur ou le consommateur doivent optimiser leur choix en matière de contrats. Les plans de production des producteurs devront se concilier avec leurs plans de contrats, en ce sens qu'il doit y avoir égalité entre la production qu'ils projettent en *n* et la totalité des contrats qui prévoient une livraison en *n*. De même, le plan de consommation de l'individu *i* devra être compatible avec son plan d'achats, compte tenu de son allocation initiale en biens.

Les consommateurs établissent également des *plans de portefeuille*, un plan à chaque *m*. Sur le marché financier se traitent les actions des entreprises *j*. Celles-ci donnent droit à une part du bénéfice de *j* en chaque *n* postérieur à *m.* Si f_{itj} est la part (comprise entre 0 et 1) de l'entreprise *j* que détient le consommateur *i* au temps *t*, la valeur du portefeuille de *i* en *m* est le produit vectoriel $\mathbf{v_m}.\mathbf{f_{im}}$, où $\mathbf{v_m}$ est le vecteur de prix des titres en *m* et $\mathbf{f_{im}}$ est le vecteur des f_{imj} (j = 1…J). Quant au revenu que *i* tire de son portefeuille au temps *n*, il vaut $\sum_j \mathbf{f_{inj}}.\mathbf{z_{nj}}.\mathbf{p_n}$.

Formellement, l'équilibre de Radner se caractérise par :

1. un ensemble de prix (non négatifs et pas tous nuls) pour chaque marché de bien (paramètres *h*, *m* et *n*) et pour chaque titre (paramètres *j* et *m*), répondant à une règle de normalisation.
2. pour chaque producteur un plan de production $\mathbf{y_j}$ et un plan de contrats $\mathbf{z_j}$ qui maximisent ensemble sa fonction d'utilité individuelle. On ne peut parler directement de maximisation du profit, car celui-ci est incertain : il faut donc passer par une fonction d'utilité permettant de départager des profits comportant des risques différents.
3. pour chaque consommateur, un plan de consommation $\mathbf{x_i}$, un plan de contrats $\mathbf{z_i}$ et un plan de portefeuille en actions. Ces plans maximisent ensemble l'utilité rapportée par $\mathbf{x_i}$, compte tenu des contraintes budgétaires. Celles-ci tiennent compte des revenus des actions et des achats et ventes sur le marché des titres, ainsi que de l'allocation initiale en titres.
4. la consistance de tous les *plans de contrat* pour tous les *m*. Cette consistance correspond à l'annulation de la demande excédentaire projetée.
   $\Sigma_k z_{km} = 0$ pour tout marché *m*, où *k* est l'addition des *i* et des *j*.
5. la répartition des titres des entreprises entre les consommateurs, de telle sorte que le capital de toutes les entreprises est entièrement souscrit.

Radner démontre mathématiquement qu'un tel équilibre existe aux conditions habituelles dans ce genre de démonstration : non satiété, convexité (et fermeture) des ensembles de production et de consommation faisables, continuité et de concavité des fonctions d'utilité (donc aversion pour le risque). Il faut toutefois ajouter une condition supplémentaire concernant le marché des actions : « However, an equilibrium can fail to exist if some provision is not made for the elimination of unprofitable enterprises »[2]. Les pertes des entreprises et leurs conséquences sur la situation financière de leurs actionnaires sont un élément perturbateur.

L'émergence de l'équilibre sur les marchés spot n'est possible que parce que l'offre et la demande s'y confrontent réellement. Le commissaire-priseur peut déterminer les variations de prix qui devraient annuler ou réduire la demande excédentaire. Est-il prévu dans l'économie de Radner que les agents doivent soumettre leur *plan de contrat* au commissaire-priseur ? Si ce n'est pas le cas, l'équilibre, certes possible, n'a aucune chance d'advenir. Si c'est le cas, c'est un pas de plus dans l'abstraction.

---

[1] Il écrit : « An upper bound on allowable contracts is natural; for example, a contract to deliver a quantity vastly greater than the total supply of the commodity would not be credible" ([291] p. 291).

[2] Radner [291] p. 289

Il est dommage que le modèle ignore toute relation entre les profits espérés et la valeur des actions.

L'apparition d'une fonction d'utilité au niveau des firmes, mêlant profit et risque, doit être saluée. La prise au sérieux des entreprises est trop rare dans l'économie néoclassique.

Equilibre de Radner et anticipations rationnelles

Dans un autre article, paru en 1979, Radner étudie la relation entre l'information, les décisions prises et l'équilibre. La thèse de l'article, basée sur la distinction entre l'information exogène et l'information endogène, est celle-ci : « When traders come to a market with different information about the items to be traded, the resulting market prices may reveal to some traders something about the information available to other traders (…) a rational expectations equilibrium reveals to all traders the information possessed by all the traders taken together »[1].

Considérons d'abord l'information disponible comme un ensemble de signaux exogènes. Appelons *s*, l'ensemble complet des signaux disponibles dans une économie. Chaque agent (indicé *i*) a accès à un sous-ensemble s_i de cette quantité de signaux. Sur base de ce sous-ensemble de signaux, il construit dans son esprit un système de *probabilités subjectives* quant aux rendements des actifs. Il s'agit de *probabilités conditionnelles* (probabilités de Bayes) : quelle est la probabilité de tel prix ou de tel rendement, SACHANT les différents signaux qui composent s_i. Soit π_i l'ensemble des distributions de probabilité (une par titre) formés subjectivement par l'agent *i* sur base de s_i.

Le vecteur π_i influence les offres et demandes, car les acteurs individuels tentent de maximiser l'*utilité espérée* de leurs choix. L'équilibre se définit toujours par un vecteur de prix **p** qui annule la demande excédentaire. Il suffit donc que change l'information dont dispose un agent, pour que le vecteur des prix d'équilibre **p** soit altéré.

A chaque contenu de l'information disponible *s*, on peut associer un vecteur **p** d'équilibre. La question importante est de savoir si la correspondance entre les valeurs de *s* et celles de **p** est biunivoque. Si elle l'est, il y a moyen, par une réflexion adéquate, de parcourir le chemin en sens inverse et d'inférer le signal sous-jacent à partir de la connaissance du vecteur **p**. Radner qualifie un vecteur de prix satisfaisant cette condition, de *révélateur*.

Radner distingue trois cas :

1- tous les agents ont accès sans limite à la totalité de l'information disponible *s*. L'équilibre est appelé *Full communication equilibrium* (FCE). Un FCE fait correspondre un équilibre (demande excédentaire nulle) à chaque valeur de *s*.
2- les agents se répartissent en deux groupes : les bien-informés et les sous-informés. Un équilibre s'établira, différent du précédent, qui sera dépendant de la manière dont l'information est répartie (les s_i).
3- les agents non informés ont la qualité d'être *sophistiqués*, c'-à-d qu'ils « realize that there is a regular relationship between the total information signal and the market price »[2]. Ils adaptent leur demande en exploitant ce surplus d'information endogène qui s'ajoute à s_i. On obtient alors ce que Radner appelle un *rational expectations equilibrium* (REE). Ici, deux situations sont possibles :
   - soit l'équilibre est *révélateur* : dans ce cas, le vecteur de prix **p** du REE correspondra exactement à celui du FCE.

[1] Radner [292] pp. 655-656
[2] Radner [292] p. 660

- Soit l'équilibre n'est pas révélateur. Alors, « initially uninformed traders cannot infer the state of the environment from the market price »[1]. Dans ce cas, l'équilibre sera le même que celui qui prévalait dans la deuxième hypothèse.

Radner démontre mathématiquement que le cas où il existe un équilibre d'anticipation rationnelle révélateur est de loin celui qui a le plus de chances de se produire. Mais cette démonstration repose sur une hypothèse forte : l'ensemble des signaux d'informations *s* doit être FINI. « One cannot expect in general that equilibrium prices will be revealing if the signal sets are 'too large'. For example, if the set of signals were Euclidian spaces, and the equilibrium prices were smooth functions of the joint signal, then the equilibrium prices could not be revealing if the dimension of the joint signal exceeded the number of commodities »[2]. Un système de mille prix ne peut informer sur un ensemble de dix mille paramètres qui les déterminent.

L'ECONOMIE DES MARCHES INCOMPLETS

Les *marchés incomplets* font l'objet d'une vaste littérature. Une structure de marchés est dite incomplète si l'absence de certains marchés rend inaccessibles certaines allocations de biens futurs contingents aux individus. Ceci implique que la bourse n'a pas réussi à jouer le rôle palliatif conçu par Arrow, pour l'une de ces deux raisons : soit les titres ne sont pas suffisamment nombreux, soit trop de titres différents procurent des rendements qui ne sont pas linéairement indépendants. Les titres ne couvrent pas l'intégralité des états contingents.

Le premier modèle est celui de Peter **Diamond** (1967). Les marchés d'actions ne permettent de compenser l'absence de marchés futurs contingents que partiellement. Les firmes choisissent leur plan de production de façon à maximiser la valeur des actions et les ménages constituent leur portefeuille dans l'optique de maximiser l'utilité de leur consommation. C'est un modèle simplifié à un bien où l'économie n'est pas séquentielle. Le résultat intéressant de ce modèle est que l'équilibre est un *optimum pareto-contraint*, c'est-à-dire qu'il domine suivant le critère de Pareto toutes les autres allocations accessibles, mais qu'il est moins optimal que s'il n'y avait pas de situations inaccessibles, c'est-à-dire si la structure de marchés était complète.

Dans son article « On the Optimality of Equilibrium when the Market Structure is Incomplete » (1975), Oliver **Hart** arrive à des conclusions moins optimistes que celles de Diamond et Radner. Il cible tout particulièrement l'équilibre de Radner. Pour démontrer que l'équilibre n'existe pas toujours ou qu'il n'est pas nécessairement optimal, le modèle présenté par Hart est calqué grosso modo sur celui de Radner. Il apporte toutefois les simplifications suivantes. On est en échange pur : il n'y a pas de production ; sans firmes, les titres ne sont pas des actions mais fonctionnent un peu comme des billets de loterie dont le gain dépend de l'état contingent *s*. A côté de ce gain, les ménages reçoivent à chaque période *t* une allocation $\mathbf{w_{it}(s)}$, un ensemble de biens à échanger contre d'autres biens ou des titres. Il y a trois périodes : des titres sont négociés en t1 et t2 mais pas en t3 car transférer du revenu vers le futur n'a plus de sens la dernière période. Le vecteur $\mathbf{z}_i$ donne la quantité des titres indicés (1,..f..F) détenus par l'agent *i*. Il n'y a pas de *plans de contrats* au sens que Radner attribuait à ce terme.

Comme chez Radner, les agents formulent des anticipations parfaites et donc communes.

Dans ces conditions, l'équilibre est un ensemble de plan de consommation $\mathbf{x_i}$ et de plans de portefeuille $\mathbf{z_i}$ (i = 1…I) ainsi qu'un système de prix (**p**,**v**), tels que :

[1] Radner [292] p. 663
[2] Radner [292] p. 677

- Les contraintes budgétaires propres à chaque période sont respectées
- Pour chaque ménage, l'utilité de la consommation $\mathbf{x_i}$ est maximisée
- A toutes les dates-événements, sur tous les marchés de biens, l'offre (le total des $\mathbf{w_i}$) égale la demande (le total des $\mathbf{x_i}$).
- A toutes les dates-événements, l'offre et la demande de chaque titre s'équilibrent.

Un tel équilibre, existe-t-il nécessairement ? Hart construit un exemple chiffré d'une économie simplifiée « that shows that even under all the standard convexity and continuity assumptions, an equilibrium may fail to exist »[1], tant avec des prix d'équilibre linéairement dépendants qu'indépendants. Comment Radner avait-il échappé à cette conclusion ? Grâce à son hypothèse limitant les quantités échangées dans les contrats suivant la règle $z_{hmn} \leq L$. Si z_{hmn} est trop élevé, cela crée une discontinuité dans la contrainte budgétaire et donc dans la fonction de demande, qui rend l'équilibre impossible. Hart écrit : « The main objection to Radner's approach is that the upper bound L is not determined endogenously, but instead is fixed a priori »[2]. Le côté arbitraire de L dérange Hart puisque, comme il le montre, l'équilibre dépend de la valeur choisie. Dans son modèle, il préfère ne pas fixer de limite quitte à accepter l'absence possible d'un équilibre.

Nous avions vu au sous-chapitre 6.3.2 que la *substituabilité brute* est une condition suffisante de l'unicité de l'équilibre d'Arrow-Debreu. Hart précise que ce n'est plus le cas en situation de *marchés incomplets*.

Hart se demande ensuite si lorsque l'équilibre existe, il est optimal. Pour simplifier, Hart limite la comparaison à des allocations qui peuvent être atteintes comme équilibres concurrentiels. Il donne des exemples chiffrés d'économies simplifiées comportant deux équilibres ; calculant l'utilité des (deux) agents, il montre que chaque fois, un équilibre pareto-domine l'autre, ce qui implique forcément qu'un équilibre n'est pas nécessairement optimal. Ces exemples « show that if markets are incomplete and there are at least two goods or at least three dates, one equilibrium may be pareto-dominated by another equilibrium"[3]. Quand un équilibre révèle des vecteurs de prix linéairement indépendants dans l'état s_1 et dans l'état s_2 et l'autre des vecteurs linéairement dépendants, c'est évidemment le premier qui domine. Vu que les dividendes des titres sont payés sous forme d'un vecteur de biens, la dépendance des prix se répercute sur les rendements des titres.

Hart se demande encore "what happens if new securities become available at date 1 or date 2". Intuitivement, on penserait que l'utilité ne pourrait que s'en trouver améliorée. Pourtant, Hart construit des exemples d'économies simplifiées où l'ouverture de marchés supplémentaires fait empirer la situation des (deux) consommateurs. Pour s'assurer que l'ouverture de nouveaux marchés améliore la situation, il faut rendre la structure complète.

[1] Hart [135] p. 427
[2] Hart [135] p. 431
[3] Hart [135] p. 435

6.5- UNE PLACE POUR LA MONNAIE DANS L'EQUILIBRE GENERAL ?

6.5.1. La filière Walras – Hicks – Lange – Patinkin

Nous avons vu comment Fisher et Wicksell ont « dichotomisé » l'économie en un secteur réel où les équations walrassiennes déterminent les prix relatifs et un secteur monétaire où l'équation quantitative détermine les prix absolus. On ne retrouvait pas cette séparation nette entre la monnaie et les marchandises chez Walras et chez Hicks. Mais ces deux auteurs n'ont pas suffisamment clarifié la relation entre la monnaie et l'économie réelle. Voyons maintenant deux contributions qui tentent d'avancer sur cette voie. L'ouvrage de Patinkin qui sera examiné ci-après porte le sous-titre évocateur « Integrating monetary and value theory ».

LANGE : LA LOI DE WALRAS ET LA LOI DE SAY

Lange lance la théorie néowalrassienne de la monnaie en 1942 avec son article « Say's Law : a Restatement and Criticism ».

Soit une économie à *n* biens : n-1 marchandises et le n^{ième} bien : la monnaie. Comme Walras et Hicks avant lui, Lange intègre la monnaie dans l'équilibre général en tant qu'un bien parmi les autres. Contrairement à Hicks, il ignore les titres.

Pour chaque bien (indice *i* variant de 1 à *n*), on a une équation d'équilibre entre l'offre et la demande (fonctions du système de prix) :

$$D_i(p_1,p_2,\ldots p_n) = S_i(p_1,p_2,\ldots p_n) \qquad (6.32)$$

Comme seules n-1 équations sont indépendantes, quelques manipulations algébriques suffisent à Lange pour démontrer l'égalité suivante :

$$\sum_{i=1}^{n} p_i.D_i = \sum_{i=1}^{n} p_i.S_i \qquad (6.33)$$

Walras l'avait déjà formulée dans le sillage de l'équation (3.12). Lange baptise donc l'équation (6.33) *loi de Walras*[1]. Elle est réputée valable en toutes circonstances, qu'il y ait ou non équilibre sur tous les marchés. Cette loi est la représentation de l'interdépendance des marchés.

Vu cette interdépendance, l'équilibre entre l'offre globale et la demande globale sur les n-1 marchés de marchandises n'est possible que si le marché monétaire est également équilibré. L'équilibre monétaire correspond à l'égalité $D_n = S_n$. Appelons ΔL la demande excédentaire de monnaie ($\Delta L = D_n - S_n$), pour le stock de monnaie existant considéré comme une donnée exogène. Ecrivons :

$$\sum_{i=1}^{n-1} p_i.D_i = \sum_{i=1}^{n-1} p_i.S_i \quad \Leftrightarrow \quad \Delta L = 0 \qquad (6.34)$$

Ce qui signifie : les marchés des marchandises ne seront en équilibre (globalement) que si les agents sont satisfaits de leur encaisse et ne sont donc pas demandeurs d'une modification de celle-ci.

L'économie politique orthodoxe a fait sienne la loi des débouchés exprimée par Say, dont il découle qu'un excédent d'offre sur le marché d'un bien ou facteur est toujours compensé par

[1] Elle est indépendante du fait que le n^{ième} bien est la monnaie.

une demande excédentaire sur le marché d'un autre bien ou facteur. La surproduction générale de marchandises est impossible. Lange exprime cette théorie par l'équation :

$$\sum_{i=1}^{n-1} p_i . D_i \equiv \sum_{i=1}^{n-1} p_i . S_i \qquad (6.35)$$

Ici, il ne s'agit plus d'une condition d'équilibre comme dans l'équation (6.34), mais d'un postulat, ce qu'indique le signe d'égalité particulier. Lange appelle l'équation (6.35) *loi de Say*. Il s'agit d'une loi de Walras réduite aux n-1 biens non monétaire, lorsque le $n^{ème}$ bien est la monnaie.

De cette loi, on peut déduire que les achats de marchandises sont toujours financés par les ventes de marchandises et jamais par des variations de l'encaisse. ΔL est toujours nul. La monnaie ne sert pas de réserve de valeur.

Les marchandises incluent les produits et services ainsi que les facteurs de production. Même dans un monde régi par l'identité de Say, on pourrait avoir une demande excédentaire de produits et services compensée par une offre excédentaire de facteurs de production. Vu l'identité de Say, le déséquilibre ne peut être que croisé. Le mécanisme normal des prix ramènera toutefois l'équilibre. La demande excédentaire des produits élèvera leur prix pendant que l'offre excédentaire de facteurs réduira le leur. Dans un monde non soumis à la loi de Say, une offre excédentaire générale des produits et des facteurs est possible, moyennant une demande excédentaire de monnaie ($\Delta L > 0$). Dans ce cas, la solution pour revenir à l'équilibre passe par le renoncement du public à vouloir augmenter ses encaisses.

Si la loi de Say prévaut, il y a n-2 équations indépendantes pour déterminer les n-2 prix relatifs des marchandises. Pas de problème de ce côté. Par contre, l'équation monétaire ($D_n = S_n$) ne permet pas de déterminer les prix absolus (exprimés en unités monétaires), car il n'y a pas d'échange net entre les marchandises et la monnaie. C'est cette indétermination des prix monétaires, causée par l'identité de Say, qui a poussé l'économie traditionnelle à introduire l'équation quantitative, pour rétablir la détermination. Réécrivons l'équation quantitative cambridgienne (4.26) avec le revenu *R* remplacé par le produit des transactions réelles *T* par le niveau général des prix *P* :

$$M = k.P.T \qquad (6.36)$$

Si on assimile le membre de gauche (l'offre de monnaie) à S_n et le membre de droite (la demande de monnaie) à D_n, en vertu du résultat $Dn \equiv Sn$ impliqué par (6.35), on obtient l'identité :

$$M \equiv k. \sum_{i=1}^{n-1} S_i \, . p_i \qquad (6.37)$$

Lange conteste que la détermination soit rétablie. La loi de Say implique que l'équation (6.37) soit inconditionnellement satisfaite. Si cette loi prévaut, l'équation quantitative sera donc valable pour toute valeur de *P* moyennant l'ajustement de *k* ; les valeurs de *k* et *P* sont indéterminées. La conclusion de l'article est donc : « Say's law precludes any monetary theory »[1].

[1] Lange [208] p. 66

Lange démontre également que la loi de Say implique le *postulat d'homogénéité*[1]. D'après celui-ci, les fonctions d'offre et de demande des marchandises telles qu'elles apparaissent dans l'équation (6.32) sont homogènes de degré zéro par rapport à l'ensemble de prix (p_1,p_2,…p_{n-1}). Autrement dit, la multiplication conjointe de l'ensemble de ces prix par quelque coefficient que ce soit, laisse intactes l'offre et la demande de ces marchandises. Celles-ci sont insensibles au niveau des prix absolus ; seuls comptent les prix relatifs. Ce postulat est la pierre angulaire de la dichotomie entre les secteurs réel et monétaire. Lange plaide pour l'abandon de la loi de Say[2] et de ses appendices, le postulat d'homogénéité et la dichotomie ainsi que pour la détermination directe des prix monétaires dans le système d'équations général.

La conception monétaire de Lange sera critiquée par **Modigliani** dans un article publié en 1944 sur lequel nous reviendrons au sous-chapitre 7.1.2. Modigliani y défend la dichotomie. Il admet la conclusion quant à l'indétermination du niveau général des prix lorsque prévaut la loi de Say. « But the traditional theory of money is not based on Say's law »[3]. La neutralité monétaire orthodoxe repose sur le *postulat d'homogénéité* que Modigliani défend car il découle de l'hypothèse de rationalité des agents. Le postulat de Say implique la neutralité monétaire, mais celle-ci peut se passer de lui. Modigliani restaure ainsi la détermination du niveau des prix *P*.

PATINKIN ET L'EFFET D'ENCAISSE REELLE

L'ouvrage de Patinkin « Money, Interest and Prices » (1956) est rapidement devenu un grand classique de la littérature sur la monnaie. La première partie, une espèce de microéconomie de la monnaie, sera exposée ici ; la seconde, purement macroéconomique, sera examinée au sous-chapitre 7.5.1.

Patinkin s'engage sur le chemin tracé par Hicks (cf. chapitre 6.2). On s'intéresse à l'équilibre général d'une économie à n+1 marchés : n-1 marchandises, la monnaie et les prêts. Chaque semaine, un nouvel équilibre est recherché sur cet ensemble de marchés. Mais Patinkin ajoute une simplification : l'échange pur. Comme il n'y a pas de production et que les marchandises sont supposées périssables (donc, non transférables vers une période ultérieure), chaque semaine, chaque agent reçoit une allocation de marchandises ; en outre, la première semaine mais la première semaine uniquement, il se voient également allouer une somme de monnaie. Donc, pour analyser les variations de l'offre de monnaie, Patinkin pratique la statique comparative.

Dans ce marché global, la monnaie est un bien parmi d'autres, ayant donc son marché spécifique avec son offre et sa demande. Mais au fond, pourquoi la monnaie est-elle demandée ? Patinkin avance une explication qui a été largement adoptée. Selon lui, la monnaie est demandée à cause de la *non-synchronisation des recettes et des dépenses* ; nous désirons posséder un stock de monnaie parce que le hasard peut amener une succession de dépenses dont le total dépasse les recettes de la même période, ce qui nécessite la constitution d'une réserve de pouvoir d'achat à titre de précaution. Certes, les titres pourraient jouer ce rôle en lieu et place de la monnaie et ils auraient l'avantage de procurer un revenu ; si une période voyait les dépenses dépasser les recettes, l'agent paierait avec des titres les marchandises qu'il ne pourrait pas payer avec d'autres marchandises. Patinkin balaie à l'avance ce contre-argument : les achats-ventes de titres incessants qui s'ensuivraient seraient trop coûteux en temps et en ressources.

---

[1] Cf. supra (sous-chapitre 5.1.4) : Leontief et Ohlin.

[2] La loi de Walras est acceptée ; la loi de Say est rejetée. Nous verrons plus loin que cette bienveillance envers la loi de Walras n'est pas partagée par tous les économistes.

[3] Modigliani [259] p. 70.

Patinkin se rend compte que sa justification de l'usage de la monnaie s'intègre mal dans son modèle d'inspiration néo-walrassienne. Comme nous l'avons vu, le tâtonnement walrassien n'a pas besoin de la monnaie. Patinkin assume cette contradiction, sans lui apporter une véritable solution.

Procédons par ordre : dans un premier temps, centrons-nous sur la relation entre les marchandises et la monnaie et ignorons les obligations qui intégreront l'analyse par après. Et avant de nous attacher à l'équilibre des marchés, considérons avec Patinkin le comportement individuel en la matière.

La demande de telle marchandise par un agent dépend de trois facteurs :
- le système de prix relatifs
- le revenu réel : la valeur REELLE de l'allocation de biens que l'agent reçoit par hypothèse le lundi matin
- l'encaisse REELLE avec laquelle l'agent commence la semaine, c'est-à-dire le pouvoir d'achat de celle-ci sur les marchandises.

« This dependence on real balances is the crucial element of the following analysis ». Nous arrivons directement au concept central de l'ouvrage : l'*effet d'encaisse réelle*. « To say that an individual adjusts his money balances as to maintain a desired relationship between them and his planned expenditures on commodities is the same as to say that he adjusts these expenditures so as to maintain a desired relationship between them and his money balances »[1]. Si l'encaisse réelle est supérieure au niveau désiré, l'agent accroîtra ses dépenses ; il les réduira dans le cas inverse.

L'important, c'est que toute variation du niveau général des prix (*P*) affecte l'encaisse réelle ; c'est par ce biais qu'une diminution de *P* stimulera la dépense et qu'une augmentation de *P* la freinera.

Enchaînons avec un deuxième concept fondamental de l'ouvrage : l'*illusion monétaire*. Un individu est sujet à l'illusion monétaire si sa demande de telle marchandise dépend d'autres facteurs que sa richesse réelle (comprenant l'allocation de marchandises et l'encaisse) et les prix relatifs, par exemple, si elle est influencée par les prix absolus. Lorsque l'agent est libre d'illusion monétaire, une variation EQUIPROPORTIONNELLE de son encaisse et de tous les prix monétaires ne le pousse pas à modifier sa demande des marchandises. Patinkin suppose que l'agent rationnel est libre d'illusion monétaire.

Tout achat de marchandise doit être financé par une vente ou par une réduction de l'encaisse[2] ; telle est la contrainte budgétaire. La demande excédentaire de monnaie vaut nécessairement l'offre excédentaire de marchandises (en valeur). Cela ne signifie pas que la monnaie est l'élément passif et les marchandises l'élément actif ; toutes ces demandes se déterminent simultanément. Pour ajuster son encaisse, l'agent doit nécessairement moduler la demande excédentaire de marchandises en adaptant ses achats et ventes.

Nous pouvons maintenant passer de l'étude du comportement individuel à l'équilibre des marchés. Pour tous les biens y compris la monnaie, on obtient les demandes de marché par l'agrégation des demandes individuelles. Patinkin démontre l'existence de l'équilibre général. Il

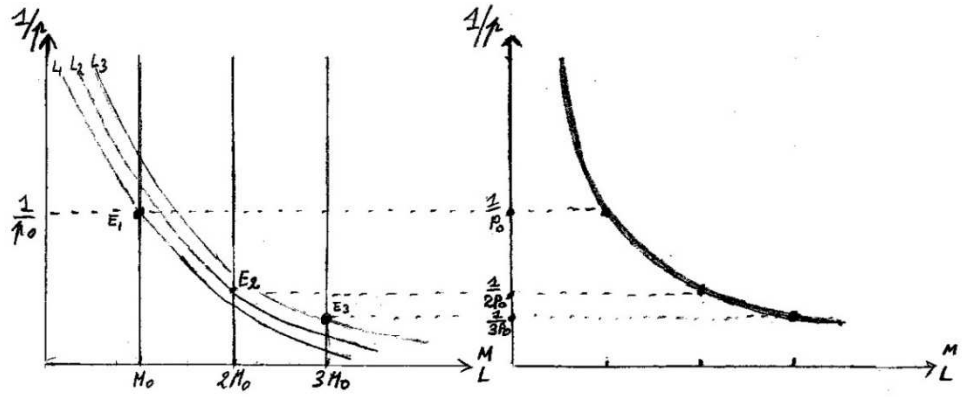
[1] Patinkin [276] p. 18
[2] La masse de monnaie en circulation est fixe, mais l'échange peut modifier sa répartition entre les agents.

faut déterminer les prix MONETAIRES de n – 1 marchandises[1], ce que permettent les n-1 équations indépendantes (de type 6.34) des *n* marchés[2]. La monnaie est traitée indistinctement des autres biens. Ce procédé permet d'éviter l'indétermination des prix monétaires. Imaginons une situation où règne l'équilibre des prix relatifs mais où les prix monétaires sont inférieurs aux prix d'équilibre. Les encaisses seront trop élevées, ce qui produira une demande excédentaire de marchandises et donc un gonflement de leur prix : l'équilibre des prix monétaires sera rétabli.

Le graphique 6.10 explique la relation, essentielle en macroéconomie, entre la masse monétaire en circulation et le niveau des prix des marchandises.

Patinkin montre comment l'équilibre entre l'offre et la demande de monnaie aboutit à l'hyperbole équilatère de Wicksell (cf. sous-chapitre 4.3.2) qui caractérise la relation enttre la monnaie et sa valeur 1/P qui est l'inverse du niveau des prix. L'astuce est que la hausse de l'offre de monnaie élève la demande de tous les biens normaux, y compris celle de la monnaie. Lorsque l'offre passe de M_0 à $2M_0$, la demande passe de L_1 à L_2 et le nouvel équilibre E_2 est à la croisée de deux nouvelles courbes. L'hyperbole équilatère est la courbe enveloppe de tous ces équilibres. Le prix varie de façon inversement proportionnelle à la quantité de monnaie. Le produit de l'offre de monnaie par sa valeur unitaire reste donc constant.

Graphique 6.10: demande de monnaie et équilibre monétaire



Les courbes de la figure de gauche rappellent le graphique 4.22 des Cambridgiens.

Il est temps d'introduire les prêts dans le modèle, qui conserve les propriétés déjà mentionnées. Par souci de simplification, Patinkin n'envisage que les obligations. Cela signifie qu'il n'y a pas de prêt sans émission d'obligations. L'hypothèse de l'échange pur restant d'application, la question du financement du capital des entreprises ne se pose pas.

Passons brièvement en revue les mécanismes du modèle de Patinkin :

- Selon une logique fisherienne, les agents peuvent, durant telle semaine, dépenser plus que leur revenu en empruntant à ceux qui dépensent moins que leur revenu. Il n'est pas prévu que les agents accumulent indéfiniment des dettes ou des créances. Pour chaque agent, l'endettement d'une semaine implique donc l'épargne d'une autre semaine et inversement.
- Au début de la première période, outre la monnaie et les marchandises, les agents héritent également d'un portefeuille d'obligations initial (négatif pour les débiteurs). Comme ce sont les agents qui décident de prêter ou emprunter en vue d'optimiser leur situation, l'encours total des obligations peut varier, contrairement à celui de la monnaie.

[1] Le prix monétaire de la monnaie est évidemment l'unité et n'est donc pas une inconnue.
[2] Patinkin est conscient du caractère vieux jeu de ce mode de démonstration pré-Cowles.

- Par rapport à Hicks, Patinkin simplifie en gommant l'incertitude et en attribuant à ses agents des anticipations où les prix futurs restent invariables

L'agent doit opérer une répartition temporelle de sa consommation. Compte tenu de ses goûts et de la *forme de son flux de revenu*, il transférera du pouvoir d'achat du futur vers le présent ou du présent vers le futur. Pour transférer du pouvoir d'achat vers une semaine ultérieure, l'agent a le choix entre la souscription d'obligations qui procurent de l'intérêt et la détention de monnaie qui le protège contre l'illiquidité. Il y a donc un choix à optimiser. L'une comme l'autre impliquent une réduction de la consommation présente.

Quelle est l'influence du taux d'intérêt sur la demande de marchandises ? Patinkin met en évidence un double effet. D'une part, il suppose qu'une hausse réduira la demande de marchandises actuelles par rapport à celles marchandises futures. « In other words, we shall assume that savings are positively related to the rate of interest »[1]. C'est la position classique bien connue. Mais en même temps, un taux d'intérêt plus élevé rendra les obligations plus attrayantes par rapport à la monnaie. C'est la thèse keynésienne. Ces deux effets jouent dans l'équilibre général de Patinkin.

Dans le système d'équilibre général, le taux d'intérêt est un prix relatif comme les autres. Par rapport à l'équilibre sans les prêts, nous avons ajouté une inconnue (l'intérêt) et une équation indépendante (celle du marché des obligations). L'existence de l'équilibre reste donc assurée. Le taux d'intérêt pourra être positif ou nul, mais pas négatif, car cela laisserait en permanence une demande d'emprunt insatisfaite.

Patinkin re-pose la question de l'impact d'une variation de la masse monétaire, plus précisément ici, en statique comparative, d'une variation de l'encaisse initiale. Supposons un doublement de l'encaisse initiale de chaque individu, alors que l'économie était en équilibre. L'augmentation générale du niveau des prix va enrichir les débiteurs et appauvrir les créditeurs, ce qui poussera les premiers à moins s'endetter et les seconds à moins prêter. L'offre et la demande d'obligations chuteront toutes deux ; le taux d'intérêt haussera ou baissera selon que ce soit l'une ou l'autre de ces forces opposées qui l'emporte. Si les obligations faisaient l'objet d'une réévaluation, de telle façon que leur encours double de valeur au même titre que la masse monétaire, et pour autant que les agents sont libres d'*illusion monétaire*, alors les prix relatifs seraient stables, le niveau général des prix doublerait exactement et la taux d'intérêt resterait constant. Patinkin semble attacher une certaine considération à ce cas, qui confirme la traditionnelle *neutralité de la monnaie*. Pourtant l'indexation de la valeur des obligations est un cas très rare.

Comme nous l'avons vu, Cassel avait analysé l'équilibre général en se centrant sur les seules demandes excédentaires, alors que Walras avait effectué la même démarche à partir d'une théorie de l'utilité. Jusqu'à présent, l'analyse de Patinkin s'est faite à la Cassellienne. Mais l'auteur tient à la répéter, cette fois à partir de la notion d'utilité, pour aboutir aux mêmes conclusions. Il ne nous est pas possible d'approfondir cette matière ; contentons-nous de l'une ou l'autre remarques.

L'utilité marginale de la monnaie a fait couler beaucoup d'encre ; certains auteurs estiment que l'idée même est contradictoire ; d'autres assimilent l'utilité marginale de la monnaie à celle des marchandises achetées. Selon Patinkin, la monnaie a une utilité intrinsèque fondée sur sa liquidité, ce qui la distingue des obligations, qui n'ont pas d'utilité en sus du pouvoir d'achat

[1] Patinkin [276] p. 66

qu'elles représentent ; par contre, l'intérêt qu'elles rendent au créditeur augmente son pouvoir d'achat futur et donc l'utilité totale que rapportera la consommation future.

Comment l'individu maximisera-t-il son utilité ? Les paramètres sur lesquels il peut jouer sont essentiellement l'encaisse et la détention d'obligations à la fin des différentes semaines. Suivant la règle d'optimisation bien connue, l'agent égalisera l'utilité du dernier dollar dépensé dans les différents usages.

- Si l'agent bénéficiait d'un dollar supplémentaire à la période *t*, celui-ci devrait lui rapporter une utilité identique, qu'il soit consommé ou thésaurisé au cours de cette période. Or l'utilité en cas de thésaurisation n'est autre que la somme de l'utilité procurée par le sentiment de liquidité éprouvé en *t* et de l'utilité procurée par la consommation de ce dollar en t+1.
- Si l'agent bénéficiait d'un dollar supplémentaire à la période *t*, celui-ci devrait lui rapporter une utilité identique, qu'il soit consommé ou placé en obligations au cours de cette période. Or l'utilité en cas de placement en obligations n'est autre que l'utilité de la consommation de $ (1+r) en t+1, où *r* est le taux d'intérêt.

Dans ce chapitre, Patinkin ne s'intéresse pas seulement au motif de liquidité qui fait détenir une encaisse de transaction mais également au motif de précaution qui fait détenir une encaisse de spéculation pour se prémunir contre l'incertitude quant à la valeur future des obligations. Il arrive à une conclusion importante : « the uncertainty generated by his bond operations makes the risk-adverting individual less willing than under certainty to modify his given income stream by means of borrowing and lending »[1]. En conséquence, non seulement la valeur actualisée du revenu mais également ce que Fisher appelait la *forme du revenu* influence la demande des marchandises ; ceci est en contradiction avec les principes de l'économie fishérienne.

Laissons là l'analyse de l'utilité et examinons maintenant le chapitre qui a rendu Patinkin célèbre. Il y critique la **dichotomie** qui caractérise l'économie traditionnelle tant classique que néoclassique. Patinkin dirige sa critique principalement vers une version plus moderne de cette dichotomie, basée sur le *postulat d'homogénéité*, qu'on retrouve notamment chez Modigliani (cf. l'article cité ci-dessus). Cette vision découpe l'économie en un secteur réel et un secteur monétaire. Dans le second, la demande de monnaie dépend des prix monétaires alors que dans le premier, les demandes des marchandises dépendent des prix RELATIFS ; c'est le *postulat d'homogénéité*. Patinkin l'explique par un souci légitime d'écarter l'illusion monétaire, qui se serait égaré.

Patinkin écrit : « …much of the attractiveness of this dichotomy lay in the belief that it was a necessary condition for the validity of the quantity theory of money. … but its basic assumption is even a denial of the quantity theory itself! »[2]. Patinkin démontre qu'une théorie économique associant la théorie quantitative et cette dichotomie arrive à des conclusions contradictoires. Plus précisément, elle implique que le niveau général des prix soit à la fois déterminé et indéterminé.

- Ecrivons l'équation de l'équilibre monétaire (annulation de la demande excédentaire de monnaie) à partir de l'équation quantitative cambridgienne.

$$k.P.T - M = 0 \qquad (6.40/6.38)$$

[1] Patinkin [276] p. 132
[2] Patinkin [276] p. 175

Supposons que cet équilibre est perturbé par un accroissement de *P*, toutes autres choses restant constantes. La demande k.P.T sera supérieure à l'offre *M*. Cette demande excédentaire suscitera une augmentation de la valeur de la monnaie, autrement dit une diminution de *P*, qui reviendra à son niveau d'équilibre. L'équilibre exprimé par l'équation (6.38) est donc une base solide pour la DETERMINATION de *P*.

- Vue sous un autre angle, cette augmentation perturbatrice de *P* (qui laisse intacts les prix relatifs) n'affecte pas les demandes excédentaires des marchandises, vu le postulat d'homogénéité. Par la loi de Walras, le marché monétaire reste également en équilibre. Le nouvel état avec *P* plus élevé est donc également un équilibre. Cette conclusion est valable quelle que soit la valeur que prend *P*. Cette variable est donc INDETERMINEE.

Le postulat d'homogénéité est à blâmer[1]. Son grand tort est d'être incompatible avec l'effet d'encaisse réelle, la clé de voûte du système patinkinien, dans lequel la variation du niveau général des prix *P* affecte la demande de marchandises lorsque l'encaisse ne change pas.

Les « Patinkin controversies »

L'œuvre de Patinkin suscita tout un débat : le modèle classique est-il effectivement contradictoire ; l'effet d'encaisse réelle, est-il la réponse adéquate ? Un article très remarqué de cette controverse est « Monetary and Value Theory : A Critique of Lange et Patinkin », publié par Archibald et Lipsey en 1958.

Le leitmotiv de Patinkin est la relation étroite entre le marché de la monnaie et les marchés des marchandises. Les agents désirent des encaisses et ils ne peuvent les ajuster qu'en adaptant la consommation des marchandises. Mais alors, si les agents sont satisfaits de leur encaisse, le lien se relâche. Patinkin, n'aurait-il pas été mieux inspiré d'étudier la quantité de monnaie qui satisfait l'agent au point qu'il s'abstient de thésauriser au déthésauriser ? Telle est la thèse d'Archibald et Lipsey. L'équilibre de Patinkin n'est donc pas un « full equilibrium », puisqu'il lui manque la caractéristique de se répéter à l'identique de semaine en semaine. Il n'est qu'un jalon sur la route vers l'équilibre complet. D'après eux, en équilibre complet, les agents, nécessairement satisfaits de leur encaisse et de leur portefeuille, les maintiennent constants, ce qui suppose que leur dépense correspond exactement à leur revenu. La consommation est donc indépendante du niveau général des prix (puisqu'elle égale nécessairement le revenu). L'effet d'encaisse réelle ne peut donc jouer le rôle que lui assigne Patinkin. Ils écrivent : « Since in full equilibrium, consumption is equal to income, a change in real balances can only change real consumption during the process of adjustment. The real balance effect is a transitory phenomenon, which is operative in some disequilibrium situations »[2]. L'effet d'encaisse réelle n'influence donc pas le point d'équilibre mais explique seulement comment une économie cherche un nouvel équilibre.

D'après Archibald et Lipsey, si on ne considère que des équilibres COMPLETS, la demande d'encaisse REELLE ne dépend que des goûts et du revenu ; elle est indépendante de l'encaisse que les agents détiennent initialement. En conséquence, la demande de monnaie NOMINALE augmente proportionnellement à *P* ; contrairement à l'opinion de Patinkin, elle est correctement décrite par une hyperbole équilatère.

A l'aide d'un exemple chiffré, Archibald et Lipsey montrent que le modèle classique, incluant le postulat d'homogénéité, n'est pas inconséquent. Dans leur exemple, le niveau général des

[1] C'était déjà la conclusion de Lange. Plus encore que ce dernier, Patinkin associe la loi de Say et le postulat d'homogénéité comme deux versions d'une même loi. Il entre en totale contradiction avec Modigliani.
[2] Archibald & Lipsey [10] p. 9

prix dépend à la fois de l'équation quantitative et de la demande excédentaire de marchandises, et pourtant le problème a une solution qui respecte toutes les contraintes. Cette solution peut même être un équilibre complet à condition de postuler l'identité de Say (qui garantit que les agents dépensent exactement leur revenu).

Partant de cet équilibre complet, Archibald et Lipsey le mettent délibérément en déséquilibre en supposant un doublement de *P*, toutes autres choses restant égales. Leur modèle révèle alors ce paradoxe que la demande excédentaire de monnaie $D_n - S_n$ (le complément de la demande excédentaire de marchandises) est nulle alors qu'elle est positive si on la déduit de l'équation (6.38). Ceci ressemble à la critique de Patinkin contre le modèle classique, mais seulement en apparence. Ici, la contradiction est propre au déséquilibre. La lacune principale du modèle classique est l'absence de mécanisme pour corriger un déséquilibre. La solution selon Archibald et Lipsey consiste en l'abandon du *postulat d'homogénéité* et l'introduction de l'*effet d'encaisse réel*. En l'occurrence, les remèdes de Patinkin, mais pour des raisons différentes...

La source de l'erreur se trouve chez Lange lorsqu'il identifie la demande d'encaisse-stock (ΔL) avec le flux d'encaisse $D_n - S_n$ (équation 6.34). Or les deux peuvent diverger. Si on privilégie la demande de monnaie-stock (ΔL) plutôt que le flux $D_n - S_n$, alors la loi de Walras ne tient pas : un individu peut très bien avoir un ΔL positif, associé à un ensemble de demandes excédentaires nulles pour les marchandises. Plus précisément, la *loi de Walras* ne tient qu'à l'équilibre, quand $\Delta L = 0$. Le système classique est dichotomique : l'équilibre monétaire et celui des marchandises sont deux circuits fermés juxtaposés : il est normal que dans ce cas, la *loi de Walras*, basée sur le principe d'interdépendance, soit hors de propos. Parodiant Lange, ils écrivent : « The classical dichotomy precludes Walras' law »[1].

6.5.2. Nouvelles ouvertures

HAHN CONTRE LA MONNAIE « INESSENTIELLE »

L'œuvre de Patinkin fut généralement saluée comme une avancée, mais la majorité des économistes demeuraient insatisfaits. C'est Hahn qui exprime le mieux cette insatisfaction.

La monnaie a comme caractéristique que si son prix est nul, sa demande excédentaire ne peut être positive. Pourquoi demanderait-on de la monnaie si son pouvoir d'achat sur les marchandises était nul ? Considérons un modèle d'équilibre général du type de celui de Patinkin. Le bien *n* est la monnaie. Sur base de la caractéristique de la monnaie stipulée ci-dessus, Hahn démontre mathématiquement que pour un tel modèle[2], on peut toujours trouver une solution où $p_n = 0$. « We therefore reach the rather displeasing conclusion that the Patinkin model always contains a "non-monetary" solution »[3]. Certes, Patinkin imposa à son modèle la propriété de n'être défini que pour des valeurs de p_n qui sont positives. Pour que le modèle conserve une solution d'équilibre avec cette contrainte supplémentaire, Patinkin dut lui ajouter l'hypothèse que la demande excédentaire de monnaie est toujours positive si $p_n > 0$. L'argument invoqué était que l'incertitude quant au timing des encaissements et des décaissements rend nécessaire l'encaisse de précaution.

---

[1] Archibald & Lipsey [10] p. 16

[2] La démonstration repose sur les hypothèses suivantes : le postulat d'homogénéité, la loi de Walras, la rareté des biens et la continuite de la fonction de demande excédentaire X(P), bornée par en-dessous. Rappelons que Patinkin rejetait le postulat d'homogénéité.

[3] Hahn [127] p. 150

Cet argument ne convainc pas Hahn. Lorsqu'ils tombent « out of cash », les agents du modèle économique pourraient acheter à crédit, comme cela se pratique régulièrement dans la vie courante. La théorie de Patinkin n'offre pas de raison excluant la généralisation de ce procédé. En conclusion, « while Patinkin has rendered signal services he has failed to provide a model which can serve as an adequate foundation for a monetary theory »[1].

Quelles caractéristiques devrait avoir une théorie monétaire digne de ce nom, selon Hahn?

1. Elle doit se démarquer du monde d'Arrow-Debreu avec son processus de *tâtonnement* « en une fois » et sa myriade de marchés futurs, qui ne laissent pas à la monnaie la possibilité de jouer un rôle significatif. Il faut recourir à des modèles d'*économie séquentielle*.
2. Elle doit reconnaître l'incertitude quant aux prix futurs.
3. L'allocation initiale des agents doit être évaluée en termes monétaires, ce que les modèles d'équilibre général font très rarement. Si le niveau général des prix varie, toutes autres choses restant égales, la richesse effective de l'agent change, ce qui peut avoir des effets dévastateurs. Un tel facteur ne peut être mis en lumière que si l'allocation initiale est également valorisée en monnaie.
4. Elle « requires a precise statement of the methods of transactions open to an individual with their attendant costs »[2]. Le sens de cette remarque apparaîtra dans la sous-section consacrée aux *coûts de transaction.*

Prenons une économie séquentielle et monétaire ; supposons que l'équilibre se caractérise par des prix relatifs et des quantités consommées et produites parfaitement identiques à ce qu'ils seraient dans le modèle d'Arrow-Debreu (tâtonnement unique à l'origine portant sur toutes les dates sans intervention de la monnaie). Pour qualifier ce cas, Hahn parle d'une « *monnaie inessentielle* ». Hahn décrit cette situation ainsi : « There is nothing we can say about the equilibrium of an economy with money which we cannot also say about the equilibrium of a non-monetary economy »[3]. De tels modèles sont malheureusement légion.

Malgré son zèle pour promouvoir la recherche de théories où la monnaie est essentielle, Hahn s'en tient à l'« *axiom of reals* », qu'il définit ainsi : « The objectives of agents that determine their actions and plans do not depend on any nominal magnitudes. Agents care only about "real" things… leisure and effort. We know this as the axiom of the absence of money illusion, which it seems impossible to abandon in any sensible sense ». Davidson trouve cette position de Hahn contradictoire avec ce qu'il dit par ailleurs : « It should have been obvious to Hahn, but apparently is not, that his acceptance of one of the fundamental axioms of neoclassical economic theory is the reason that there is no room for money in orthodox theory »[4].

CLOWER ET LA CONTRAINTE FINANCIÈRE

Dans son article « A Reconsideration of the Microfoundations of Monetary Theory » (1967), Clower pose le même constat que Hahn : « Modern attempts to erect a general theory of money and prices on Walrasian foundations have produced a model of economic phenomena that is suspiciously reminiscent of the Classical theory of a barter economy. My purpose in this paper is to show that the conception of a money economy implicit in these constructions is

[1] Hahn [127] p. 153
[2] Hahn [127] p. 153
[3] Hahn [127] p. 160
[4] Davidson [167] p. 152

empirically and analytically vacuous and to propose an alternative microfoundation for the pure theory of a money economy »[1].

Clower estime que la théorie classique de la monnaie (par exemple Patinkin) s'appuie sur la double équation suivante, relative à l'optimum de l'individu *j* :

$$\text{Max } U_j(d_1,\ldots d_i,\ldots d_n, M_j'/P) \qquad (6.39)$$
$$\text{Sous contrainte : } \Sigma_i^n\, p_i.X_{ij} + M'_j - M^\circ_j = 0 \qquad (6.40)$$

Les d_i sont les quantités demandées des n marchandises. M°_j est l'encaisse initiale et M'_j est l'encaisse désirée[2]. Les X_{ij} sont les demandes excédentaires des marchandises et les p_i sont leurs prix, dont P est l'index composite. L'équation (6.40) est la très classique contrainte budgétaire. Il est sous-entendu que ces variables se rapportent à une période de référence.

Ce qui dérange Clower, c'est que ce modèle ne reconnaît aucune spécificité à la monnaie. Il pourrait très bien s'accommoder d'une solution où de nombreuses marchandises sont troquées directement les une contre les autres, avec la monnaie qui n'interviendrait que dans une minorité d'échanges (voire aucun). Dans l'économie réelle, la quasi-totalité des transactions impliquent l'échange d'une marchandise contre de la monnaie. L'économie monétaire se caractérise par l'usage généralisé de la monnaie. Parodiant Jean-Baptiste Say, Clower écrit : « Money buys goods and goods buy money, but goods do not buy goods »[3].

Pour intégrer ces éléments dans la théorie monétaire, Clower propose d'assortir le système d'équations d'une *contrainte financière*, imposant l'usage de la monnaie. On aurait :

$$\text{Max } U_j(d_1,\ldots d_i,\ldots d_n, m_j'/P, m_j/P) \qquad (6.41)$$
$$\text{Sous contrainte 1 : } \Sigma_i^n\, p_i.X_{ij} + m'_j - M^\circ_j = 0 \quad \text{pour } X_{ij} \geq 0 \qquad (6.42)$$
$$\text{Sous contrainte 2 : } \Sigma_i^n\, p_i.X_{ij} + m_j = 0 \quad \text{pour } X_{ij} < 0 \qquad (6.43)$$

L'encaisse désirée comporte maintenant deux parts :

- m'_j est l'encaisse de PRÉCAUTION : "total initial cash balances less prospective (gross) depletions of cash balances for currently scheduled purchases of goods".
- m_j est l'encaisse de TRANSACTION : "to replace cash currently scheduled for disbursement from initial holdings of money balances"

L'équation (6.42), appelée *contrainte de dépense*, qui se rapporte aux demandes excédentaires positives (les achats), « asserts that all (net) purchase offers must be backed by a readiness to supply money in exchange ». L'équation (6.43), appelée *contrainte de revenu*, qui se rapporte aux demandes excédentaires négatives (ventes), « asserts that all (net) sale offers involve a demand for just one other commodity, namely money, in exchange »[4]. L'addition de m_j et de m'_j permettra à l'agent de repartir avec une encaisse initiale adéquate au début de la période suivante.

Il conclut : « Supposing that the decision problem has a solution, however, we arrive at individual demand and excess demand functions that differ in crucial respect from those of

[1] Clower [57] p. 131
[2] Dans cette façon d'écrire, la monnaie n'est évidemment pas le énième bien, mais le $(n+1)^{ème}$. Les titres sont ignorés.
[3] Clower [57] p. 135
[4] Clower [57] p. 136 (les définitions des deux encaisses et des deux contraintes).

established theory ». Notamment, l'offre de marchandises ne crée pas automatiquement la demande d'autres marchandises : le passage par la monnaie pourrait gripper la loi de Say.

La suggestion de Clower fut fraîchement accueillie, car elle donne l'impression que l'usage de la monnaie nécessite le respect d'une obligation. Si l'économie monétaire est comme le troc avec une contrainte en plus, on pourrait même en déduire la proposition absurde qu'elle est moins efficace que le troc[1]. Plutôt que d'imposer l'usage de la monnaie, il vaut mieux rechercher les raisons qui assurent à l'économie monétaire un surplus d'efficacité. C'est à cette tâche que s'attèleront de nombreux économistes, qui feront appel à des explications variées. On verra ainsi fleurir une prose abondante expliquant de façon savante les avantages de l'économie monétaire sur le troc, qui ne semblent évidents que pour le profane.

Examinons trois de ces contributions qui explorent chacune une piste différente : la coïncidence des besoins, les coûts de transaction et l'information imparfaite.

STARR ET LA COÏNCIDENCE DES BESOINS

Carl Menger, l'un des pères de l'économie néoclassique, est le premier à avoir exploré cette piste, dans un article publié en 1892 et resté célèbre, qui s'intitulait « On the origins of money ». Les économistes qui furent inspirés par cet article sont parfois appelés « néo-mengeriens ». Ross Starr est l'un d'eux. Voyons d'abord l'article de Menger.

Menger estime que l'essence de la monnaie est restée un mystère pour l'économie politique. Il conteste l'explication la plus courante, qui impute l'existence de la monnaie à une convention générale ou une volonté de l'autorité publique, ce que l'histoire ne corrobore pas. La question réellement importante est : certains biens ont accédé au statut de moyen d'échange ; quelle est la genèse de ce phénomène ?

Menger s'attache d'abord à l'échange primitif, le troc, qui limite étroitement le champ des échanges possibles. Pensons au détenteur d'un bien indivisible de grande valeur qui voudrait se procurer une variété de biens de faible ou moyenne valeur. Ou bien cet exemple devenu classique : l'agent *A* voudrait obtenir le bien de l'agent *B* ; l'agent *B* voudrait obtenir le bien de l'agent *C* ; l'agent *C* voudrait obtenir le bien de l'agent *A*. Le troc (bilatéral) ne procure pas aux agents de possibilité d'échange qui leur permette de satisfaire leur besoin. Le troc n'est efficace que lorsqu'il y a coïncidence réciproque des besoins des personnes qui traitent bilatéralement sur le marché.

Menger met en avant une propriété permettant de caractériser les différents biens, qu'il appelle la *vendabilité* ; il la définit ainsi : « A commodity is more or less *saleable* according as we are able, with more or less prospect of success, to dispose of it at prices corresponding to the general economic situation… ». Il y a tout un éventail de vendabilités allant des biens « hardly marketable » comme des « astronomical instruments, anatomical preparations, Sanskrit writings » jusqu'aux métaux précieux, dont la vendabilité est presque infinie. Menger énumère une longue liste de facteurs qui influencent la vendabilité : divisibilité, durabilité, quantité disponible et étendue géographique de cette disponibilité, coûts du transport, de l'entretien et de la conservation … La plupart des marchandises ne peuvent être revendues qu'à perte et n'ont qu'une vendabilité limitée. La monnaie s'explique tout simplement par l'accession à cet état, des biens les plus vendables, à la suite d'un processus quasiment naturel et spontané.

[1] Clower a été raillé par plusieurs auteurs à cause de cette déduction qu'il n'a pas tirée mais qui aurait pu l'être. Cette critique est évidemment ridicule ; le sens du mot « contrainte » n'est pas le même dans un modèle mathématique et dans la société.

Quelques individus ont compris les premiers, le bénéfice qu'ils retireraient s'ils fréquentaient les marchés, nantis de biens très vendables. Voyant l'avantage de ce procédé, les autres l'ont imité. Lorsqu'un bien très vendable devient moyen d'échange, sa vendabilité s'accroît encore. C'est ainsi que certains biens se sont imposés comme monnaie. La sanction légale a suivi le processus spontané.

Dans l'article « The Structure of Exchange in Barter and Monetary Economies », Ross **Starr** démontre mathématiquement la supériorité d'une économie monétaire.

Pour qu'une transaction de troc soit possible entre les agents *A* et *B*, il faut que l'agent A ait besoin du bien détenu par *B* et que *B* désire le bien offert par *A*. Dans une économie monétaire, la monnaie est le bien qu'on accepte sans en avoir besoin.

Pour apprécier les performances respectives des deux types d'économie, Starr se donne l'outil suivant. Soit la matrice **A** qui représente la totalité des échanges ayant lieu. Il y a *n* biens et *m* individus. Notre matrice a *n* colonnes et m^2 lignes. Il y a une ligne par paire d'agents. Les lignes sont les vecteurs $\mathbf{a_{ij}}$ composés de *n* éléments correspondant aux quantités de chaque bien que l'agent *j* procure à l'agent *i* (négatifs si l'agent *j* reçoit). On doit donc avoir $\mathbf{a_{ij}} = -\mathbf{a_{ji}}$. Chaque agent *i* a également un vecteur de demande excédentaire $\mathbf{x_i}$ qui est fonction du système de prix donné par le vecteur **p** ; ces deux derniers vecteurs ont donc *n* éléments.

Dans cette économie plus réaliste, l'équilibre postule plus que la simple égalité de l'offre et de la demande des *n* biens. Il faut que les transactions individuelles soient justifiées et équilibrées. En troc, cette condition prend la forme de la coïncidence des besoins, mais elle doit également être satisfaite dans une économie monétaire. Pratiquement, dans le modèle, la satisfaction de cette condition implique :

1- ***la cohérence des prix*** : pour chaque ligne de **A**, on a l'égalité : $\mathbf{a_{ij}}.\mathbf{p} = 0$. Chaque agent donne autant qu'il reçoit, non seulement globalement mais vis-à-vis de chaque interlocuteur. « What price consistency requires is that all goods acquired must be paid for by sending goods of equal value from the trader acquiring the purchased goods to the trader supplying them »[1].
2- ***la réduction de la demande excédentaire*** : la relation entre $\mathbf{a_{ij}}$, $\mathbf{x_i}$ et $\mathbf{x_j}$ est telle que si pour le bien *k*, on a $a_{ij}^{\ k}$ non nul (le $k^{ième}$ élément du vecteur $\mathbf{a_{ij}}$), la transaction $a_{ij}^{\ k}$ réduit la demande ou l'offre excédentaires des deux agents concernés pour le bien k (réduction de $x_i^{\ k}$ et de $x_j^{\ k}$).
3- ***la satisfaction de la demande excédentaire*** : consolidons les vecteurs $\mathbf{a_{ij}}$ en un vecteur $\mathbf{a_i}$ dont les éléments sont les quantités totales des *n* biens, échangées par l'agent *i* avec les autres agents pris comme un ensemble. Il y a au moins coïncidence partielle entre $\mathbf{a_i}$ et $\mathbf{x_i}$. Cette propriété suppose que chaque agent *i* voit sa demande excédentaire de certains biens s'annuler à la suite de ses transactions. Cette propriété complète la deuxième car la satisfaction de la demande excédentaire d'un bien peut nécessiter des transactions avec plusieurs vendeurs.

Concernant le troc, Starr démontre ce théorème : si **p*** est le vecteur de prix qui égalise l'offre et la demande sur tous les marchés, alors on pourra toujours lui associer une matrice **A** qui satisfait deux des trois conditions ci-dessus mais rarement une matrice qui satisfait les trois simultanément.

[1] Starr [347] p. 294. Au passage, Starr remarque que « Payment for goods purchased seems a concept almost absent from general equilibrium theory ».

Voilà ce dont est capable l'économie de troc. Passons maintenant à l'économie monétaire. La matrice **A** comporte maintenant une colonne en plus, car un $(n+1)^{ème}$ bien est ajouté : la monnaie. Le prix de la monnaie est arbitrairement fixé à l'unité. Le vecteur de prix d'équilibre **p**** comporte les éléments de **p*** plus le prix unitaire de la monnaie. Il n'est donc pas affecté par l'offre et la demande de monnaie mais seulement par les facteurs réels, suivant en cela le principe classique de la dichotomie.

La condition de la réduction de la demande excédentaire (propriété 2) sera considérée comme satisfaite par l'économie monétaire pour autant que sa partie réelle (la matrice **A** sans sa dernière colonne) la satisfait. « Unlike other goods, money will be accepted in exchange whether it is desired or not »[1].

Starr démontre le théorème suivant : si le vecteur **p**** égalise l'offre et la demande des *n* biens dans l'économie monétaire, on pourra TOUJOURS lui associer une matrice **A** qui a les TROIS propriétés ci-dessus. L'économie monétaire ouvre donc des possibilités d'échange supérieures à l'économie de troc. Starr démontre également qu'à toute économie de troc satisfaisant les deuxième et troisième propriétés à l'équilibre, on peut associer une économie monétaire identique (à l'exception de la $n+1^{ème}$ colonne de **A** et du $n+1^{ème}$ élément de **p**) qui répond aux trois conditions de la coïncidence des besoins.

NIEHANS ET LES COUTS DE TRANSACTION

Considérons l'équilibre général d'une économie à *n* biens, mais envisageons-la à un haut degré de généralité, où nous ne savons pas au départ si les membres de cette économie pratiqueront le troc ou élèveront un des biens au statut de monnaie ; notre but est précisément de déterminer les conditions poussant à l'émergence de l'un ou l'autre de ces deux régimes d'échange. L'attention doit donc se porter, non sur les *n* marchés, mais sur ce que Niehans nomme les *postes d'échange*, consacrés au troc de deux marchandises spécifiées (par exemple les noix contre les pommes, les noix contre les oranges…). Le nombre de ces postes s'élève à (n/2)(n-1).

« It is an essential feature of the model that exchange causes costs (…) It will be assumed for simplicity that all transaction costs are expressed in terms of "work", measuring the time and trouble in making transactions »[2]. Concernant les coûts de transaction, Niehans pose certaines hypothèses simplificatrices : notamment, ils diffèrent d'un bien à l'autre, d'un individu à l'autre…Il y en a des fixes (par rapport à la quantité échangée) et des variables, mais Niehans exclut les fixes, car ils sont trop difficiles à modéliser.

Dans un premier temps, Niehans s'en tient à une analyse statique ; seules interviennent des livraisons au temps présent. Il faut déterminer les prix sur les (n/2)(n-1) postes. L'égalité entre l'offre et la demande doit s'établir sur chacun d'eux. Niehans montre que le nombre d'inconnues de ce problème égale le nombre de ses équations indépendantes entre elles ; la solution d'équilibre existe donc.

A noter qu'à l'équilibre, certains biens circuleront pour satisfaire directement un besoin de l'acheteur alors que d'autres circuleront comme moyen d'échange, c'est-à-dire que les agents les acquerront à la seule fin de les échanger par la suite. Le choix des biens utilisés comme moyens d'échange, ainsi que leur nombre, sont indéterminés a priori et variable selon la personne. A noter qu'à ce stade, Niehans n'inclut pas encore de biens sans utilité directe ; les

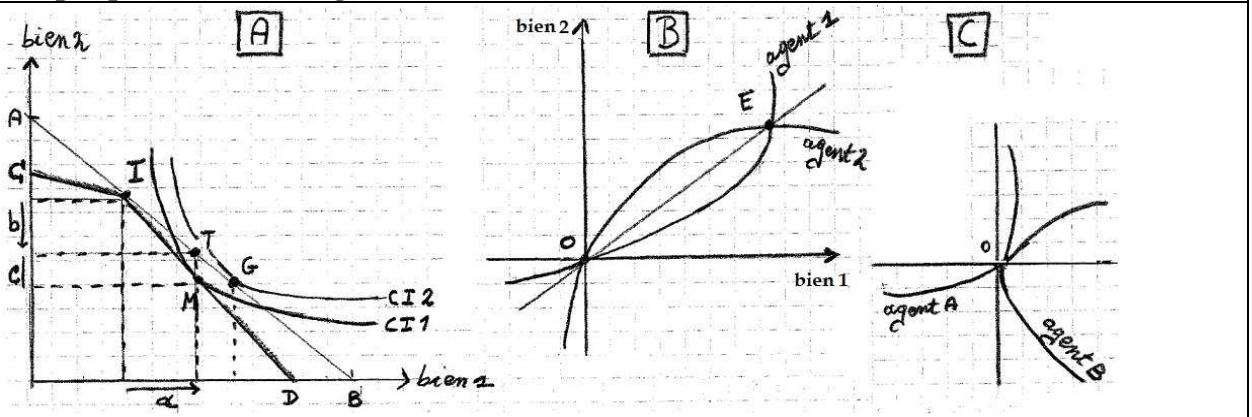
[1] Starr [347] p. 298
[2] Niehans [267] p. 774

biens utilisés comme moyen d'échange sont tous susceptibles d'utilité directe et ne sont donc pas prédestinés à ce statut.

Dans la solution d'équilibre, nombre de postes feront sans doute l'objet d'un échange nul. Niehans montre que « the no-trade solution is more likely, the higher the transaction cost rate »[1]. Le graphique 6.11 illustre un cas simplifié à deux agents et deux biens. La figure *A* montre l'optimisation de la situation de l'agent un, suivant la méthode habituelle de la tangence de la droite de budget avec une courbe indifférence. Supposons que le point *I* indique son allocation initiale, autrement dit son point de non-échange. Contrairement à ce qu'il en est dans le graphique standard, la droite de prix AB ne coïncide pas avec la droite de budget (capacité d'achat effective) qui est ici la ligne brisée CID. En effet, plus on échange et on s'éloigne donc du point *I*, plus les coûts de transaction variables écartent la droite de budget de la droite de prix. L'optimum se situe au point *M* ; pour obtenir une quantité *a* du bien un, notre agent offre une quantité *b* du bien deux et paye des frais de transaction s'élevant à *c*[2]. La courbe d'indifférence qui passe par *G* montre que s'il n'y avait pas de coût de transaction, l'agent serait incité à échanger une quantité supérieure.

Sur la figure B, emboitons les champs d'indifférence des deux agents de telle façon que les points de non échange des deux agents coïncident avec l'origine. L'ordonnée indique au-dessus de zéro les achats du bien deux par l'agent deux et les ventes par l'agent un et en-dessous de zéro l'inverse. L'abscisse indique à droite de zéro les achats du bien 1 par l'agent un et les ventes par l'agent deux et à gauche de zéro l'inverse. La pente des droites passant par l'origine indique le rapport de prix. Sont dessinées les courbes d'échange des deux agents qui renseignent les achats et ventes qu'ils souhaitent aux différents rapports de prix. Le point de non-échange (l'origine) en fait naturellement partie. Si les deux courbes se coupent également en un autre point, celui-ci (*E* sur la figure B) sera l'équilibre (avec p_E comme prix d'équilibre). L'équilibre de non-échange est celui où les courbes d'offre ne se coupent qu'à l'origine. Le cas a plus de chances de se produire quand les coûts de transaction sont élevés. La brisure de la droite de budget au point *I* sur la figure A implique une brisure des courbes d'échange au point *O* sur la figure B. Si les coûts de transaction sont élevés, les deux courbes d'offre seront plus saillantes à l'origine. Dans ce cas, elles auront moins de chances de se couper en un autre point (cf. figure C).

Graphique 6.11: l'échange avec coûts de transaction



Sur la figure A, si T s'était trouvé au nord-ouest de I, le sens des transactions a et b eût été inversé ; il y aurait eu achat du bien 1 et vente du bien 2.

[1] Niehans [267] p. 778

[2] Si les frais de transaction sont payés avec le bien 2.

La façon différenciée dont les coûts de transaction affectent les différents biens sera donc déterminante quant à leur place dans l'échange. On constate notamment :

- plus l'ensemble des coûts de transaction sont élevés, plus réduits seront les échanges.
- si tous les coûts de transaction sont nuls, on est dans le cas d'Arrow-Debreu, avec un équilibre général qui sera normalement un grand troc.
- Si un bien se distingue de tous les autres par ses coûts de transaction quasiment nuls, il s'imposera comme moyen d'échange exclusif et les postes d'échange dont il est absent seront désertés.

Le bien du troisième cas accède au statut de monnaie ; l'économie devient monétaire. On reste au stade de la *monnaie marchandise* ; il ne s'agit pas encore de la *monnaie fiduciaire*. Cette apparition de la monnaie est neutre, car le flux des biens sera identique à celui qui caractériserait l'équilibre général sans coûts de transaction.

Niehans élargit ensuite son modèle pour y introduire des propriétés caractéristiques d'une économie monétaire moderne. D'abord, il devient intertemporel. On admet que les biens ne sont pas tous consommés immédiatement mais qu'ils sont transférables à la période suivante, moyennant un coût de stockage. Dans ce cas, on pourra voir émerger comme réserves de valeur, des biens incapables de rendre un service réel mais dont l'utilité découle de leur potentiel comme moyen d'échange. Si l'un d'eux a de faibles coûts de stockage et des coûts de transaction nuls, il accédera au statut de monnaie fiduciaire. Cette monnaie restera neutre, mais Niehans insiste sur le fait que la neutralité est imputable à ces deux propriétés : le bien n'a pas d'utilité intrinsèque et son coût de transaction est nul.

BRUNNER ET MELTZER ET LE COUT D'ACQUISITION DE L'INFORMATION

Niehans, pas plus que Hahn qui écrivit sur le même sujet la même année, ne crurent nécessaire de décrire plus précisément en quoi consistent ces *coûts de transaction*. Dans l'article « The Uses of Money » qu'ils publient en 1971, Brunner et Meltzer distinguent deux composantes :

1. le coût d'exécution des transactions
2. le coût de l'acquisition d'information « about market opportunities » : quels partenaires offrent ou demandent tel bien ; quelle est la qualité offerte ou demandée ; quel prix est-il raisonnable de payer ou d'exiger ? On conçoit qu'en l'absence de monnaie, l'information à obtenir devient très ample et très coûteuse.

C'est sur ce deuxième type de coûts que nos auteurs concentrent leur attention : leur conclusion est : « For individuals, money is a substitute for investment in information and labor allocated to search »[1]. Leur modèle rompt radicalement avec l'hypothèse d'information parfaite (donc libre et gratuite) présente dans les modèles traditionnels d'équilibre général.

Voyons ce modèle plus en détail. Le contexte est celui d'une économie non monétaire. De façon novatrice, le modèle se départit de ce marché global où en une seule opération, avec l'aide du commissaire-priseur, tous les intervenants vendent, achètent et obtiennent le panier optimal pour le restant de la semaine hicksienne. Ici, les échanges ne sont pas tous simultanés et l'indice *t* des variables du modèle augmente d'une unité à chaque échange. Considérons un agent doté initialement d'une allocation sous la forme d'un panier de biens. Dans les modèles traditionnels, on a l'égalité $R_0 = C_0 + X_0$ qui devient $R_1 = C_0 + Y_1$, où *R* est le panier disponible pour la consommation, *C* est la part de ce panier consommée directement, *X* est le prélèvement dans ce panier pour offrir une contrepartie aux achats *Y*. Ces variables sont toutes des vecteurs à *n* biens.

[1] Brunner & Meltzer [42] p.799

Un aspect essentiel du modèle est que les agents ne connaissent que partiellement la masse d'information nécessaire concernant les qualités des biens, les caractéristiques des autres agents, les rapports de prix habituels, la disponibilité des biens… De leur information dépendra l'optimalité de leurs échanges. Ils peuvent investir un montant I_t pour l'améliorer. C'est là un premier moyen d'action que Brunner et Meltzer confèrent à leurs agents, qui est absent de la plupart des modèles. Et ils en attribuent un deuxième : l'échange indirect : on acquiert des biens, non pour les consommer mais parce qu'on pense qu'ils seront plus faciles à écouler pour obtenir les biens convoités.

La fonction d'utilité à maximiser et la contrainte budgétaire sont respectivement :

$$U = U(\varepsilon, v) \qquad (6.44)$$

$$R_0 = C_0 + X_0 + I_0 + S_0 \qquad (6.45)$$

L'utilité augmente avec la qualité du panier ε que l'agent espère constituer et diminue avec l'incertitude *v* quant au contenu de ce panier. Comme dans les autres modèles, le consommateur affecte une part de son budget à *C* et *X*, mais il investit également *I* pour améliorer son information ; enfin, une dernière part est absorbée par les frais de transaction proprement dits *S*. L'équation (6.45) se répète avec des indices temporels croissants à chaque étape dans le processus d'échange.

Les ratios d'échange ont la forme d'une matrice *E* de taille *n*x*n* ; les éléments e_{ij} sont la quantité du bien *j* que je peux obtenir en cédant une unité du bien *i*. Cette matrice est importante pour valoriser mon offre potentielle *X*. Cette matrice de l'ensemble des rapports d'échange est elle-même une variable aléatoire, plus précisément une variable conditionnelle à l'information connue notée *P*. Pour chaque *P*, il existe une matrice E* la plus avantageuses. La densité de probabilité de la variable *E* influence directement les paramètres ε et *v* de la fonction d'utilité.

Les auteurs formalisent également le processus d'échange. La séquence des transactions est donnée par une suite de matrices $T_0,\ldots,T_\theta$ (où θ n'est pas déterminé a priori). Les éléments de ces matrices renseignent la transformation de X_t en Y_{t+1}. Ces transactions ont un coût *S* où S = S(T).

L'agent doit optimiser *S* et *I*. Tous deux obèrent *C* et *X*, ce qui réduit l'utilité. Mais *I* permet d'augmenter ε et de diminuer *v*. La dépense *S* élargit la gamme des transactions possibles et contribue donc également à accroître ε. Leur utilité marginale à tous deux est décroissante ; il y a donc une valeur optimale pour *I* et pour *S*. Mais les budgets affectés à *I* et *S* ne sont pas tout ; le choix des biens échangés importe également, car la variabilité de *I* et *S* d'un bien à l'autre est considérable.

Ici intervient un postulat des auteurs : lorsqu'un bien est plus couramment échangé, ses propriétés sont de plus en plus connues au sein du groupe social. Les membres du groupe ont donc un intérêt commun à lui réserver une place importante dans l'échange pour réduire les coûts et l'incertitude. Ce postulat implique une certaine convergence des comportements. Ainsi, « individuals' optimizing behavior leads to the social choice of a small number of medium of exchange assets »[1]. C'est ainsi que des biens peu coûteux en frais de transaction et dont les qualités sont bien connues accèdent au statut de monnaie. Les métaux précieux conviennent parfaitement : la représentation de ces métaux par du *papier monnaie* ne modifie pas fondamentalement la nature de la monnaie, sauf si une mauvaise gestion aboutit à une forte

[1] Brunner & Meltzer [42] p. 793.

inflation, auquel cas les avantages de la monnaie comme arme contre l'incertitude perdent leur efficacité.

La monnaie n'est pas une réponse à la non-synchronisation entre les dépenses et les recettes. En fait, toutes deux sont la conséquence d'une même cause qui est l'efficacité supérieure de l'échange indirect.

Brunner et Meltzer concluent : "The use of money remained puzzling as long as the theory of exchange was restricted to the case of perfect certainty, a world in which information about market prices and the qualities of goods and services is obtained at zero cost"[1].

CONCLUSION

Le lecteur aura remarqué l'absence du commissaire-priseur dans les trois modèles que nous venons d'étudier. Mais en quelque sorte, c'est la monnaie qui reprend le rôle du commissaire. Le déséquilibre et le chômage restent donc impossibles.

[1] Brunner & Meltzer [42] p. 804.

TITRE VII- LA MACROECONOMIE APRES LA THEORIE GENERALE

A la suite de la révolution keynésienne, le débat a évidemment fait rage et il n'est pas encore clos. Le sera-t-il jamais ? Par la même occasion, le centre de gravité de l'économie politique s'est déplacé de la micro à la macroéconomie.

Les premières réactions suivirent immédiatement la publication de la Théorie générale. Leurs auteurs étaient des économistes de l'ancienne école néoclassique ou proches d'elle, principalement Pigou et les wickselliens (Ohlin et Robertson). En moyenne, la jeune génération était plutôt favorable à la révolution keynésienne.

Après la deuxième guerre mondiale, le théâtre du débat changea ; de nouvelles écoles se formèrent. Les opposants ne pouvaient plus attaquer la pensée keynésienne en faisant fi de ses apports. Quant à ses défenseurs, ils occupèrent tout un éventail d'attitudes à l'égard de la pensée néoclassique.

Dans le titre VII, nous examinerons le principales écoles et tendances qui prirent part aux disputes. Nous y évoquerons également quelques controverses opposant les écoles.

7.1- LA SYNTHESE NEOCLASSIQUE-KEYNESIENNE

INTRODUCTION

Ce sont diverses écoles qui incarnèrent le keynésianisme. Notamment, il trouva un soutien important à Cambridge, le lieu de sa naissance. Mais ce fut un autre courant, parfois appelé les *néo-keynésiens*, qui fut le porte-drapeau le plus remarqué du keynésianisme dans les débats macroéconomiques. Contrairement aux Cambridgiens, qui concevaient la théorie keynésienne en opposition avec la théorie néoclassique, les néo-keynésiens étaient attachés aux bases néoclassiques et tentèrent la synthèse entre les deux conceptions. Comment intégrer Keynes dans le corpus général de la théorie néoclassique sans récuser les bases de celle-ci ?

L'école de la synthèse domina la pensée économique dans les décennies 1940, 1950 et au début des sixties-. On y trouve notamment pas mal de nos parétiens du chapitre 4.2 : John Hicks, Paul Samuelson, Abba Lerner. On y trouve également Robert Solow, qui fut partie prenante à la controverse du capital comme Samuelson. Ainsi que Franco Modigliani, James Tobin et de nombreux autres tant aux Etats-Unis qu'en Grande Bretagne.

Le souci majeur de ces néoclassiques est l'explication du chômage involontaire dans un système économique conservant les principes néoclassiques. La solution de facilité consiste à pointer la rigidité du salaire nominal ; Keynes y fit certes allusion ; mais un non-keynésien peut également adhérer à cette explication : si le keynésianisme se limite à ce truisme, la révolution keynésienne se ramène à une tempête dans un verre d'eau. Les néo-keynésiens, avancent-ils d'autres arguments ? La connivence avec la Théorie générale, comme nous le verrons, est très variable d'une individualité à l'autre.

7.1.1. Hicks et le modèle IS-LM

LE MODELE IS-LM

Dès l'année qui suit la parution de la Théorie Générale (donc en 1937), John Richard Hicks, publie « Mister Keynes and the Classics ; a Suggested Interpretation », un court article dont l'impact sera énorme. Y apparaît pour la première fois le très fameux *modèle IS-LM*. Pendant près d'un demi-siècle, ce modèle dominera la macroéconomie, où il sera servi à toutes les sauces. Il a pour objet de produire une version synthétique, simplifiée et pédagogique de la théorie keynésienne, qui permet aussi de la comparer plus facilement à d'autres théories. L'interprétation hicksienne, qui se dégage à la lecture de cet article, est que la théorie de Keynes n'est pas aussi différente de la théorie « classique »[1] que ne le laisse croire la *Théorie Générale*.

Le modèle IS-LM et ses équations (7.1)

	a- Théorie classique	b- Théorie de Keynes	c- Généralisation Keynes
1-Monnaie	M = k.Y	M = L(Y,i)	M = L(Y,i)
2-Investissement	I = I(i)	I = I(i)	I = I(Y,i)
3-Epargne	I = S = S(Y,i)	I = S = S(Y)	I = S = S(Y,i)

Le modèle est centré sur l'analyse à court terme : le volume du capital est fixe. En outre, le salaire nominal est considéré comme donné.

[1] au sens keynésien du terme « classique ».

Voyons d'abord comment Hicks présente le modèle classique. Il le fait tenir dans les trois équations de la première colonne du tableau (7.1). L'équation [a.1] n'est autre que la théorie quantitative dans sa version cambridgienne. L'équation [a.2] indique que l'investissement est une fonction décroissante du taux d'intérêt. L'équation [a.3] égalise l'investissement avec l'épargne et fait dépendre celle-ci du taux d'intérêt et du revenu (de façon croissante).

Ces trois équations permettent la détermination des trois inconnues : *Y, I*, et *i*. Le capital étant considéré constant par une hypothèse simplificatrice (également utilisée par Keynes), le travail est le seul facteur variable. Connaissant *Y* et *I*, on peut donc en déduire le volume de l'emploi.

C'est donc principalement par des variations de *M* et de *k* (l'indice de la demande de monnaie par rapport au revenu) que les « classiques » peuvent expliquer les variations importantes du revenu nominal au cours du cycle économique. Hicks, tout en reconnaissant un fond de vérité à ces explications, les juge insuffisantes.

Hicks traite ensuite de façon analogue la théorie keynésienne (deuxième colonne). La principale nouveauté, l'équation [b.1], n'est qu'une réécriture simplifiée de l'équation (5.3) de Keynes. La préférence pour la liquidité est l'innovation de Keynes que Hicks préfère. Son cœur parétien ne pouvait qu'être séduit par une théorie qui place la monnaie au centre d'un choix optimisateur, comme les autres biens.

Venons-en maintenant au fameux diagramme IS-LM reproduit au graphique 7.1. Chaque point du cadran est une combinaison entre le taux d'intérêt et le revenu. La courbe LM[1] correspond à l'équation [b1] et donne les combinaisons qui assurent l'équilibre du marché monétaire. Considérons l'équation (5.3) de Keynes. Si *M* est fixé, lorsque *i* baisse, l'encaisse spéculative s'élève, laissant moins de monnaie pour satisfaire la demande de transaction, ce qui n'est compatible qu'avec un *Y* plus bas. Les variables *i* et *Y* évoluent donc dans le même sens, ce qui revient à dire que la courbe LM est croissante. Une augmentation de *M* se marquerait par un glissement de la courbe LM vers la droite, car pour un taux d'intérêt donné, le surplus de monnaie disponible pour les transactions nécessite une augmentation du revenu

La courbe IS découle des équations [b.2] et [b.3]. Comme l'investissement est fonction décroissante de l'intérêt et l'épargne est fonction croissante du revenu, le maintien de l'égalité entre *I* et *S* (qui est la condition d'équilibre du secteur réel de l'économie) postule une relation inverse entre *i* et *Y* : la courbe IS est donc descendante.

L'équilibre global de l'économie correspond évidemment à l'intersection entre les deux courbes, la monétaire et la réelle. Comme l'explique Hicks, les courbes IS et LM déterminent le revenu et l'intérêt simultanément, comme les courbes d'offre et de demande déterminent la quantité et le prix simultanément.

La politique budgétaire de l'Etat qui soutient la demande globale, augmente le MEC ; pour maintenir l'équilibre, l'accroissement de *I* qui s'ensuit nécessite celui de *S* et donc celui de *Y*, ce qui déplace la courbe IS vers la droite. Cette conclusion keynésienne, où la politique budgétaire augmente effectivement le revenu, s'oppose à la vision classique où elle aboutissait juste à élever le taux d'intérêt. Mais la figure 7.1-B montre que si on passe de IS1 à IS2, la hausse du revenu s'accompagne d'une hausse de l'intérêt. Comment comprendre que ce facteur ne se retrouve pas dans la *Théorie Générale* ?

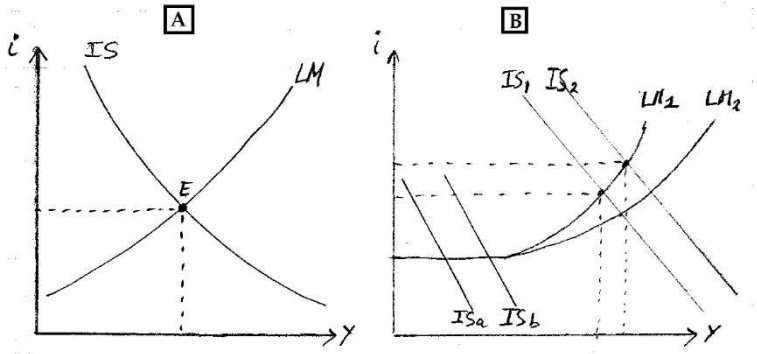
Hicks l'explique par la forme particulière de la courbe LM. Celle-ci découle de deux facteurs :

[1] Dans l'article, elle apparaît sous l'appellation « LL ». Elle sera rebaptisée par Hansen (1949).

- Il y a un minimum en dessous duquel le taux d'intérêt ne descendra pas, même si on accroît indéfiniment *M* ; autrement dit, les courbes LM correspondant aux différentes valeurs de *M* ont le même plancher. Ce phénomène est évoqué dans la Théorie Générale. Devenu un concept essentiel, en partie grâce au présent article de Hicks, il est appelé *trappe des liquidités*[1]. Hicks le considère comme "the most important thing in Mr. Keynes' book"[2].
- Il y a un revenu maximum pouvant être « financé » par une valeur donnée de *M*.

La courbe LM est donc horizontale dans sa partie gauche et croissante ensuite, pour devenir finalement verticale. La hausse de la MEC déplace la courbe IS vers la droite. Si ce déplacement a lieu alors que les deux courbes se coupent dans la partie horizontale de LM (passage de IS_a à IS_b), le schéma keynésien se vérifie pleinement : il en résulte exclusivement une hausse du revenu. Par contre, à droite de cette zone, le déplacement de IS (passage de IS_1 à IS_2) produira une hausse à la fois de *Y* et de *i*. Plus on va vers la droite, plus la théorie classique se montre valide, d'où il ressort que le modèle économique de Keynes convient particulièrement à une économie en dépression[3].

Graphique 7.1 : le modèle IS-LM



Outre la politique budgétaire, la forme de la courbe *LM* influence également la politique monétaire. Supposons qu'on passe de LM_1 à LM_2, sur la figure 7.1-B. La politique monétaire ne sera pas efficace si la courbe IS coupe LM dans sa partie horizontale, car cette partie horizontale est commune aux différentes courbes LM.

Après les théories classiques et keynésiennes, Hicks propose un troisième modèle, qu'il appelle « modèle keynésien généralisé » qui mélange tant bien que mal les deux théories rivales : la théorie classique et la théorie keynésienne deviennent toutes deux des cas particuliers du modèle généralisé. Cette fois, les trois variables, *L*, *I* et *S* dépendent toutes, à la fois de *Y* et de *i*, ce qui, selon Hicks, procure au modèle une certaine « élégance mathématique ». La réintégration de *i* dans l'équation de l'épargne n'est pas étonnante de la part d'un néoclassique, mais elle contredit diamétralement la vision de Keynes. Quant à l'intégration de *Y* dans l'équation de l'investissement, Hicks la justifie ainsi : un revenu plus élevé élève la demande globale, ce qui doit stimuler l'investissement. Comment réagit la courbe IS lorsque *I* et *S* sont tous deux fonctions du revenu *Y* ?

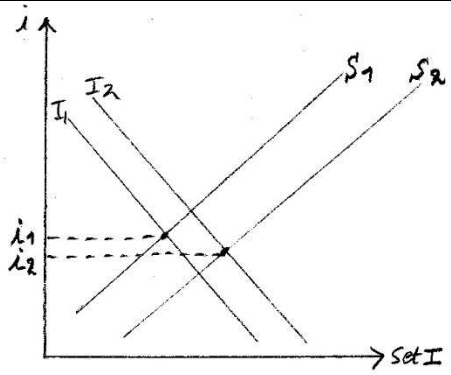
---

[1] Keynes écrit à son sujet : « Peut-être ce cas limite prendra-t-il une importance pratique dans l'avenir, mais nous n'en connaissons pas d'exemples dans le passé. A vrai dire, étant donné la répugnance de la plupart des autorités monétaires à opérer hardiment sur les créances à long terme, on n'a pas souvent l'occasion d'en faire l'expérience » ([186] p. 215). Les termes « trappe des liquidités » apparaissent pour la première fois dans l'article « Mister Keynes and the Rate of Interest » (1940) où Robertson critique la théorie keynésienne.

[2] Hicks [152] p. 154.

[3] Ce qui est une évidence, y compris pour Keynes lui-même.

Graphique 7.2 : l'influence du revenu sur l'épargne et l'investissement.



Hicks ressuscite ici le fameux diagramme classique du graphique 4.23, auquel Keynes s'opposa.

Si le revenu (non visible sur la figure 7.2) augmente de Y1 à Y2, l'épargne passera de S1 à S2 et l'investissement de I1 à I2. Si $\Delta S > \Delta I$, le taux d'intérêt baissera ; dans le cas contraire, il augmentera. Une courbe IS descendante postule donc que l'effet de ΔY sur *S* soit plus fort que son effet sur *I* (ce qui est assurément le cas aux yeux de Keynes). Si les deux effets s'équivalent, la courbe IS sera une droite horizontale, dont l'altitude correspondra au *taux d'intérêt naturel* de Wicksell.

Y a-t-il une place pour le sous-emploi dans le modèle IS-LM ? Le principal argument est évidemment la *trappe des liquidités*, qui se caractérise par une demande de monnaie parfaitement élastique par rapport au taux d'intérêt. Elle peut empêcher le taux d'intérêt de descendre au niveau qui serait nécessaire pour atteindre le plein emploi.

Telles semblent être les principales armes des néo-keynésiens pour défendre l'existence d'un équilibre de sous-emploi contre leurs adversaires.

Les Cambridgiens (principalement J. Robinson et Pasinetti) ont reproché à Hicks, la simultanéité walrassienne entre les équilibres monétaires et réels, qui est une hérésie par rapport à la pensée keynésienne. La Théorie générale présentait les choses sous la forme d'un enchaînement où l'équilibre monétaire était en situation d'antériorité : de lui dépendait le taux d'intérêt, qui confronté au MEI déterminait le niveau d'investissement. Il en résultait le revenu et donc le niveau de l'emploi.

Pour une génération d'économistes, le *modèle* IS-LM ETAIT le keynésianisme. Le lecteur conviendra que le caractère novateur de la *Théorie générale* est bien émoussé dans le digest hicksien, dont l'existence arrangera pourtant Keynes soucieux d'assurer une acceptabilité à sa théorie. Toutefois, au crépuscule de sa carrière, Hicks prendra ses distances avec son modèle et se rapprochera des thèses cambridgiennes plus radicales. Il reconnaîtra notamment la validité de l'argument contre la simultanéité.

LA STABILITE DE L'EQUILIBRE TEMPORAIRE

Revenons un instant à l'ouvrage « Value and Capital » (1939), qui réunit dans un chapitre macroéconomique terminal les deux concepts fondamentaux étudiés au chapitre 6.2 : la question posée est celle de la **stabilité** de **l'équilibre temporaire**. Hicks réussit le tour de force de proposer une deuxième théorie macroéconomique deux ans après la première, non contradictoire avec elle.

Comme nous l'avons vu au chapitre 6.2, la stabilité nécessite que toute variation d'un prix entraîne une substitution importante des offres et demandes sur les différents marchés de façon à ramener celles du marché déséquilibré à leur niveau d'équilibre ; la flexibilité des prix s'avère donc essentielle. Dans un équilibre temporaire, cette flexibilité doit jouer entre les prix présents et les prix futurs. Mais les prix des biens futurs sont forcément des prix anticipés. La façon dont sont formées les anticipations détermine donc la stabilité de l'ensemble. Pour en rendre compte, Hicks introduit le concept d'*élasticité des anticipations*[1]. Une élasticité nulle signifie que les prix anticipés sont indépendants des prix actuels et que les variations actuelles sont donc perçues comme temporaires. Si l'élasticité vaut un, il est prévu que la hausse actuelle maintiendra exactement ses effets ; entre zéro et un, ces effets seraient atténués ; plus grande que un, cette hausse est extrapolée.

Si l'élasticité des anticipations est nulle, une hausse des prix actuels aurait pour effet de rendre les achats futurs relativement plus intéressants, ce qui entraînerait un report des achats. Cet important effet de substitution temporel engendrera l'offre excédentaire de biens présents nécessaire à la stabilité. Par contre, si l'élasticité égale l'unité, la hausse est considérée comme permanente et aucune substitution dans le temps ne se produira. A partir de l'unité, l'élasticité des anticipations rend l'équilibre temporaire instable. La moindre perturbation peut faire déraper l'économie hors du chemin de l'équilibre.

Hicks considère l'élasticité unitaire comme l'hypothèse la plus plausible. Pourtant, l'observation de l'économie réelle montre que l'instabilité existe, certes bien présente, n'est pas générale. Des facteurs stabilisateurs compensent donc partiellement l'influence déséquilibrante des anticipations. Principalement les variations du taux d'intérêt. Si la hausse des prix actuels s'accompagne d'une hausse du taux d'intérêt, il redevient avantageux de reporter les achats même si les prix se maintiennent. Une telle hausse de l'intérêt est probable car la prévision d'une hausse des prix continue stimulera la demande de monnaie, sauf si l'offre de monnaie est freinée par l'autorité monétaire.

Malheureusement, cet effet stabilisateur n'est pas symétrique. La *trappe des liquidités* entrave la baisse du taux d'intérêt à court terme et plus encore sa transmission au taux d'intérêt à long terme. En cas de dépression, (donc d'offre excédentaire des biens présents), la conjonction de l'élasticité des anticipations trop élevée avec la trappe des liquidités rend inefficace une baisse des prix présents. Du fait de la grande instabilité du système, il sera très difficile de remédier à l'offre excédentaire des biens présents. En conséquence, comme l'avait déjà remarqué Keynes, la politique monétaire est efficace pour lutter contre l'inflation, mais pas contre le chômage.

Contre une instabilité caractérisée par la chute des prix alors que l'élasticité des anticipations est unitaire, la rigidité des salaires nominaux joue paradoxalement le rôle de rempart ; si le rempart cède, la spirale déflationniste empêchera la baisse du salaire de résorber le chômage lié à la rigidité.

7.1.2. Modigliani, la rigidité des salaires et la préférence pour la liquidité

Dans son article « Liquidity Preference and the Theory of Interest and Money » publié en 1944, que nous avons déjà évoqué (cf. sous-chapitre 6.5.1), Franco Modigliani canonne la théorie keynésienne. Bizarrement, l'auteur est souvent compté parmi les économistes de la synthèse.

[1] Il la définit ainsi : « l'élasticité des anticipations (d'un individu donné), concernant le prix d'un bien X, est définie comme le rapport entre l'accroissement relatif des prix attendus de X et l'accroissement relatif du prix actuel » ([154] p.192).

Le modèle IS-LM hicksien sert de canevas à son analyse, mais Modigliani l'élargit, notamment en y intégrant la production et le marché du travail. A partir des trois équations du modèle IS-LM (7.1) et de quatre équations qu'il leur joint, Modigliani construit le système d'équations suivant :

$M = k.Y$	(7.2-A)	ou	$M = L(Y,i)$ (7.2-A')
$I = I(Y,i)$	(7.2-B)		
$I = S \; et \; S = S(Y,i)$	(7.2-C)		
$Y \equiv P.X$	(7.2-D)		
$X = X(N)$	(7.2-E)		
$W = X'(N).P$	(7.2-F)		
$W = f(N).P$	(7.2-G)	ou	$W = w_0$, si $N < N_0$ (7.2-G')

Que disent les équations ajoutées ? Le revenu nominal *Y* est par définition la production *X* multipliée par le niveau des prix *P*. La production est fonction du niveau d'emploi *N*. La demande de travail égalise le salaire réel *W/P* avec la productivité marginale du travail X' = dX/dN (7.2-F). L'offre de travail est une fonction du salaire réel (7.2-G), mais dans la version keynésienne, le salaire nominal est donné de façon exogène si le plein emploi n'est pas atteint (7.2-G').

Ce système de 8 équations à 8 inconnues permet de déterminer un équilibre, plus précisément deux équilibres : le keynésien et le classique, car l'équation monétaire (7.1-A/A') et l'équation du salaire nominal (7.2-G/G') ont deux versions.

Avec les équations keynésiennes, les courbes IS et LM peuvent se couper à gauche du revenu de plein emploi, alors que le plein emploi est assuré avec les équations classiques. Le sous-emploi keynésien, provient-il plutôt de la théorie de la préférence pour la liquidité ou de l'hypothèse de la rigidité du salaire nominal ? Pour le savoir, Modigliani va « croiser » les deux systèmes d'équations. Il arrive aux conclusions suivantes :

- Un système comportant l'équation quantitative (7.2-A) à la place de la préférence pour la liquidité (7.2-A') mais avec le salaire nominal rigide (7.2-F') aboutit à la possibilité du sous-emploi.
- Un système comportant la préférence pour la liquidité (7.2-A') combinée avec le salaire flexible (7.2-F) aboutit au plein emploi, sauf dans le cas particulier de la trappe des liquidités.

La rigidité du salaire nominal est donc la clé du système keynésien : « …except in a limiting case (NDLR : la trappe des liquidités), this result is due entirely to the assumption of 'rigid wages' and not to the Keynesian liquidity preference »[1].

La dichotomie joue un rôle important dans sa démonstration et celle-ci s'appuie sur le controversé *postulat d'homogénéité*. Modigliani montre notamment que sans la rigidité du salaire, les équations 7.2-B à 7.2-G suffisent à déterminer le taux d'intérêt sans le concours de l'équation monétaire (7.2-A'). Un des effets de la rigidité du salaire est précisément de rendre les grandeurs réelles dépendantes de la masse monétaire *M*, alors qu'avec des salaires flexibles, *M* ne détermine que le niveau des prix *P*, que ce soit avec (7.2-A) ou (7.2-A'). Avec des

[1] Modigliani [259] p. 57

salaires flexibles, même la demande de monnaie de spéculation n'influence pas le taux d'intérêt.

Quel sera l'effet d'une baisse de l'efficacité marginale de l'investissement ? Voici l'enchaînement : le taux d'intérêt s'adapte à la baisse ; la demande d'encaisse de spéculation augmente ; l'encaisse de transaction s'en trouve réduite ; le revenu nominal s'adapte à la baisse. Pour la suite, il faut distinguer : si le salaire nominal est flexible, la baisse de *Y* prend la forme d'une réduction de *P* et *W*. S'il est rigide, c'est le revenu réel *X* qui souffrira et avec lui l'emploi ; il y aura également une baisse de *i* qui restera toutefois au dessus de son niveau assurant le plein emploi et de *P*, qui fera hausser le salaire réel W/P.

Mais Modigliani ne commente pas seulement les théories existantes : il propose sa propre explication : « It is true that a reduced level of employment and a reduced level of investment go together, but this is not in general the result of causal relationship. It is true instead that the low level of investment and employment are both the effect of the same cause, namely (…) that money wages are too high relative to the quantity of money that explains why it is unprofitable to expand employment to the 'full employment' level"[1].

Modigliani clôture son article en apportant sa réponse à la question largement débattue depuis la Théorie Générale : le taux d'intérêt se détermine-t-il dans la sphère monétaire ou la sphère réelle ? Il est assez fier de répondre : les deux. A court terme, à l'intérieur de la semaine hicksienne, le taux d'intérêt équilibre la demande de monnaie et la demande de titres. Mais à long terme, il équilibre l'investissement ex ante et l'épargne ex ante. D'ailleurs, Modigliani ne considère pas que la courbe IS représente l'économie réelle et la courbe LM le secteur monétaire. Selon lui, la courbe LM représente l'équilibre monétaire à court terme ; la courbe IS est nécessaire pour déterminer l'équilibre monétaire à long terme, celui qui, à conditions inchangées, ne varie pas de semaine en semaine.

7.1.3. Samuelson : variables autonomes et variables induites

En 1948, Samuelson publie « The Simple Mathematics of Income Determination », un exposé didactique de la théorie keynésienne du revenu. L'équation (5.1) de Keynes, $Y = C + I$, sert de point de départ. Au cœur du raisonnement, il y a la distinction entre variable *induite* et variable *autonome* :

- La consommation *C* est une donnée *induite* en ce sens qu'elle est fonction de *Y*, ce qu'on note $C = C(Y)$
- L'investissement *I* est une donnée *autonome* c'est-à-dire exogène, se comportant comme une constante, ce qu'on note $I = I°$.

On obtient donc l'égalité :

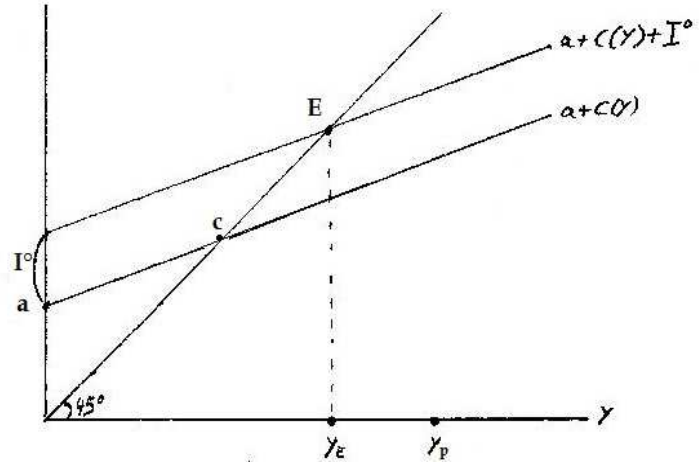
$$Y = C(Y) + I° \qquad (7.3)$$

Cette équation a une solution. Il y a une et une seule valeur de *Y* qui satisfait cette condition d'équilibre. C'est ce que montre le fameux *diagramme en croix*[2].

[1] Modigliani [259] pp. 68-69

[2] Ce diagramme habite littéralement l'article de 1948 mais il n'y est pas dessiné.

Graphique 7.3 : le diagramme en croix des keynésiens



C(Y) est souvent dessinée rectiligne par simplification. Vu la propension marginale à consommer décroissante, il serait plus logique de la dessiner concave.

C est une fonction croissante de *Y*. Vu la constance de I°, les droites *C* et $C+I$ sont parallèles. La droite qui monte suivant un angle de 45° transpose en ordonnée la valeur de *Y* en abscisse. L'intersection entre cette droite et celle de $C+I$ détermine le revenu d'équilibre[1].

Par définition, la variable autonome résulte de facteurs extérieurs à l'équation. Selon Keynes, l'investissement dépend des anticipations des entrepreneurs et du taux d'intérêt. La variable autonome I° est donc sensible à la politique budgétaire et à la politique monétaire. Quel sera l'effet sur *Y* d'une variation donnée de I° ? On obtient la réponse en calculant la différentielle de *Y* par rapport à I° dans l'équation (7.3) :

$$\frac{dY}{dI^\circ} = \frac{1}{1 - C'(Y)} \qquad (7.4)$$

C' n'est rien d'autre que la *propension marginale à consommer*, la variation de la consommation induite par la variation du revenu. La différentielle dY/dI° est évidemment le *multiplicateur d'investissement*. Le lecteur peut vérifier que la formule (7.4) correspond bien à la formule (5.2) de Keynes.

Voilà pour le scénario de base. Pour le rendre plus réaliste où l'adapter à certains cas que l'on désire étudier, il peut s'avérer opportun d'introduire une *consommation autonome* et un *investissement induit*. Si la fonction de consommation est du type *a+C(Y)*, la valeur *a* représente la *consommation autonome*, que peut expliquer par exemple le minimum vital. Le graphique 7.3 comporte une consommation autonome. A gauche du point *c*, le revenu est inférieur à la consommation : il y a désépargne ; à droite de *c*, le revenu est supérieur à la consommation : il y a épargne[2]. L'effet multiplicateur d'une variation de *a* sur le revenu est le même que celui d'une variation de I°.

Selon Samuelson, considérer un *investissement induit* I(Y) peut être légitime si on analyse les variations de l'investissement à court terme. Le terme I'(Y) indique alors la *propension marginale à investir*. Le multiplicateur se détermine toujours par le calcul différentiel mais la

[1] Nous avons placé le revenu de plein emploi Y_P à droite du revenu d'équilibre Y_E, puisque la théorie keynésienne attribue une probabilité élevée à une telle situation.

[2] La désépargne durant la période *t* implique qu'il y eu épargne au cours d'une période antérieure.

formule se complique un peu. Lorsque l'équation du revenu est du type $Y = a + C(Y) + I(Y)$, le multiplicateur de la dépense autonome est :

$$\frac{dY}{da} = \frac{1}{1 - C'(Y) - I'(Y)} \qquad (7.5)$$

Samuelson introduit ensuite l'action du gouvernement dans le modèle. Les équations s'enrichissent des variables *G* (les dépenses publiques à l'exclusion des transferts) et *W* (le montant des recettes fiscales)[1]. Commençons avec l'hypothèse la plus simple où *G* et *W* sont des variables autonomes (indépendantes du revenu). On a donc $G = G°$ et $W = W°$. Sur le diagramme en croix, la fonction C+I+G serait une droite parallèle à C+I, surélevée par rapport à celle-ci. La consommation est maintenant fonction du *revenu disponible* noté *y*, qui égale $Y - W°$. Les équations deviennent :

$$Y = C(y) + I° + G° \qquad (7.6)$$

$$\frac{dY}{dG°} = \frac{dY}{dI°} = \frac{1}{1 - C'(y)} \qquad (7.7)$$

$$\frac{dY}{d(-W°)} = \frac{C'(y)}{1 - C'(y)} = \frac{dY}{dG°} - 1 \qquad (7.8)$$

L'équation (7.7) montre que l'effet bénéfique des dépenses gouvernementales sur le revenu est le même que celui de l'investissement privé. Par contre, l'effet multiplicateur d'une baisse des taxes est inférieur d'une unité aux multiplicateurs de l'investissement et des dépenses gouvernementales. Alors qu'on peut calculer les multiplicateurs de ΔG et de ΔW, il est impossible de déterminer le multiplicateur d'un déficit $\Delta D = \Delta G - \Delta W$ sans savoir comment il se distribue entre ΔG et ΔW.

Tout ceci conduit au fameux et paradoxal *théorème du budget équilibré*[2]. « According to this theorem, a deficit is not at all necessary for an expansionary fiscal policy »[3]. Si *G* et *W* augmentent simultanément d'un montant égal, le revenu s'accroîtra de ce même montant ; autrement dit, le multiplicateur vaudra l'unité. Ce théorème requiert les deux hypothèses que la relation fonctionnelle entre *C* et *y* ne change pas et que l'investissement privé reste constant. Du fait que *Y* augmente de ΔG, on peut déduire que le prélèvement de ΔW laisse le revenu disponible *y* inchangé, ce que Samuelson démontre.

Samuelson abandonne ensuite l'hypothèse irréaliste de la constance de *W*. Au lieu de $W = W°$, on a $W = r + W(Y)$. La variable *r* représente le niveau de pression fiscale indépendamment du revenu ; *W(Y)* est la relation fonctionnelle entre les recettes fiscales et le revenu. Cette optique conduit à une diminution de tous les multiplicateurs de dépense autonome, car la hausse du revenu contient en elle le germe d'un effet contraire (mais pondéré) : la hausse des taxes obère le revenu disponible pour la consommation : $y = Y - r - W(Y)$. Les multiplicateurs deviennent :

---

[1] On pourrait considérer également que *W* est le solde des taxes moins les transferts.

[2] Ce théorème était déjà connu depuis le début de la décennie 1940 et avait déjà été traité par nombre d'auteurs (notamment par Hansen et Haavelmo). Samuelson écrit : « The explanations given for this paradoxical result are numerous » ([323] p. 25). Il en cite quatre, dont celle d'un certain Mr Salant que voici: la chaîne du multiplicateur de ΔG est $1+C'+(C')^2+(C')^3+\ldots$; la chaîne du multiplicateur de ΔW est : $-C'-(C')^2-(C')^3-\ldots$ En effet, au premier stade, une réduction de taxe de 100 ne réduit pas la consommation de 100 mais de *C'*.100.

[3] Samuelson [323] p. 24.

$$\frac{dY}{dG^{\circ}} = \frac{dY}{dI^{\circ}} = \frac{1}{\left[1 - C^{'}(y)\right].[1 - W^{'}(Y)]} \qquad (7.9)$$

$$\frac{dY}{d(-r)} = \frac{C'(y)}{\left[1 - C^{'}(y)\right].[1 - W^{'}(Y)]} \qquad (7.10)$$

Comme on le constate, à investissement constant, le multiplicateur d'une réduction d'imposition reste inférieur à celui d'un accroissement des dépenses publiques.

Samuelson prend encore en considération un dernier élément qui obère le revenu disponible pour la consommation : les profits que les entreprises mettent en réserve en réserve *B*. Faisant abstraction de l'intervention de l'Etat, l'équation (7.3) devient Y = C(Y-B)+I°.

7.1.4. Tobin et la théorie du portefeuille

Au contraire de Hicks qui a simplifié Keynes et de Modigliani qui l'a réduit à un cas particulier de l'économie néoclassique, James Tobin s'est attaché à approfondir l'explication de certains mécanismes keynésiens.

LA PREFERENCE POUR LA LIQUIDITE

Dans la Théorie générale, Keynes reconnaît à la préférence pour la liquidité un rôle essentiel. Le taux d'intérêt a précisément pour objet de la compenser. Comme le montre le graphique 5.2, la demande de monnaie est une fonction décroissante du taux d'intérêt, mais cette relation provient exclusivement de la demande d'encaisse de spéculation. La demande d'encaisse de transaction est insensible au taux d'intérêt.

Dans « The Interest-Elasticity of Transactions Demand for Cash » (1956), Tobin pose la question: "Why not hold transaction balances in assets with higher yield than cash, shifting into cash only at the time an outlay must be made"[1]. Explorant cette possibilité, il estime "plausible that the share of cash in transaction balances will be related inversely to the interest rate on other assets"[2] et il le démontre mathématiquement. La procédure préconisée procure à l'individu l'avantage de percevoir un intérêt mais lui impose le coût tant pécuniaire que non pécuniaire des transferts entre le cash et les obligations. Une répartition optimale de l'encaisse de transaction entre le cash et les actifs à rendement peut être trouvée en combinant au mieux cet avantage et cet inconvénient. Sur base de certaines hypothèses comme la prévisibilité de la dépense et son étalement parfaitement constant sur la période, Tobin détermine la formule de l'encaisse moyenne et du volume moyen d'obligations optimaux. Il s'avère que la part du cash diminue quand le taux d'intérêt augmente.

[1] Tobin [360] p. 241.
[2] Tobin [360] p. 242.

Baumol et l'encaisse de transaction

Quatre ans avant Tobin, William Baumol s'était intéressé au niveau optimal de l'encaisse de transaction dans l'article « The transaction Demand for Cash : An Inventory Theoretic Approach ». Comme le titre l'indique, l'idée est d'appliquer à cette encaisse un traitement comparable à celui de la gestion des stocks dans les entreprises. Comme chez Tobin, l'optimum de l'individu ou de la firme consiste à combiner idéalement la perte d'intérêt et les coûts des transferts entre l'encaisse et les titres. Baumol arrive à une formule où l'encaisse optimale est une fonction linéaire de la racine carrée du volume des transactions pendant la période considérée. Donc, "demand for cash rises less than in proportion with the volume of transactions". Cette constatation conduit à la déduction que "The effect on real income of an injection of cash into the system may have been underestimated"[1]. Certes d'autres effets de l'injection, comme la baisse de l'intérêt et la hausse de prix modéreront l'effet positif. Il faut en outre examiner comment l'addition des comportements individuels peut être assumée par l'économie globale pour déterminer si l'effet bénéfique au niveau de l'individu peut être généralisé au niveau de l'économie.

Tobin réexamine ensuite la relation inverse entre l'encaisse de spéculation et le taux d'intérêt dans la Théorie générale, dont l'explication a été critiquée par beaucoup de commentateurs pour son manque de réalisme. Rappelons-nous : si l'épargnant s'attend à ce que le taux futur dépasse le taux présent, il gardera tout son avoir en liquide et n'achètera pas de titres ; s'il s'attend à une stagnation du taux, il placera tout son avoir en titres pour ne pas garder un euro stérile. Seules les divergences d'appréciation entre les acteurs expliquent la coexistence des titres avec des avoirs monétaires. Dans la réalité, les agents ne pratiquent évidemment pas cette politique du tout ou rien. L'article « Liquidity Preference as Behavior Towards Risk » (1958) de Tobin présente une explication alternative, débarrassée de cet irréalisme ; la monnaie y est demandée en tant qu'actif sans risque.

Voici les données du modèle : pour composer son portefeuille, l'épargnant peut combiner deux sortes d'actifs : la monnaie qui est un actif sans risque et sans return, et les « consolidées », qui sont un titre donnant droit à une rente perpétuelle. Ces consolidées rapportent un intérêt fixe dont le taux est *r*. Les consolidées émises à des périodes différentes n'auront pas nécessairement le même taux d'intérêt. Il s'ensuivra des variations de leur cours boursier. Le cours des anciennes consolidées baisse lorsque le taux d'intérêt augmente et inversement. Le rendement des consolidées vaut donc (r+g) où *g* est le gain en capital (positif ou négatif). Le risque des consolidées est mesuré par l'écart type de *g* (noté σ_g)[2]. Il faut déterminer la part du portefeuille détenue en monnaie, notée A_1 et celle placée en consolidées, notée A_2 qui vaut 1-A_1.

Sélectionner la meilleure combinaison de A_1 etA_2, c'est choisir la combinaison optimale entre le rendement espéré et le risque du portefeuille combiné. L'espérance mathématique de *g* est nulle car, sauf hypothèses particulières, a priori la valeur des consolidées a autant de chances de varier dans un sens que dans l'autre. Comme seule la partie « consolidées » du portefeuille rapporte, les caractéristiques du portefeuille sont les suivantes :

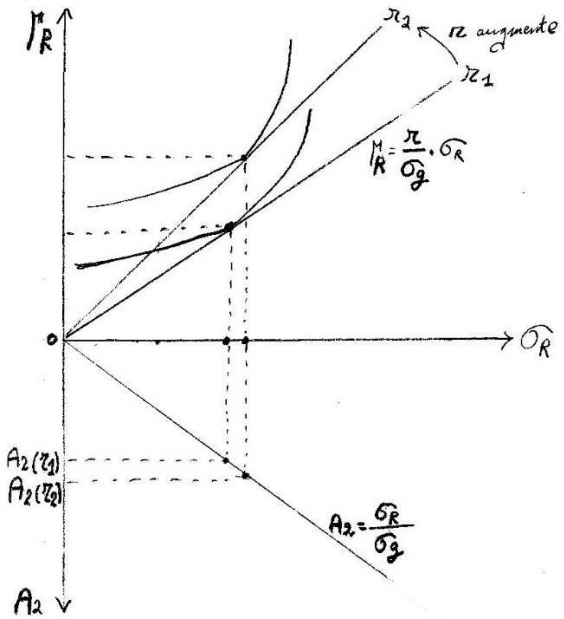
[1] Baumol [26] p. 349

[2] Dans le monde des économistes et des analystes financiers, il est habituel d'estimer le risque d'un placement par l'écart type de son return. L'écart type du return est la moyenne des écarts du return par rapport au return moyen sur la période considérée. Un actif plus risqué verra son return fluctuer plus amplement et l'écart type sera donc plus élevé.

Rendement effectif :	$R = A_2.(r+g)$	(7.11)
Rendement espéré :	$\mu_R = A_2.r$	(7.12)
Risque :	$\sigma_R = A_2.\sigma_g.$	(7.13)

Pour sélectionner la combinaison idéale entre R et σ_R, Tobin fait appel à une méthode connue. On dessine le champ d'indifférence de l'agent entre le rendement et le risque et on cherche le point de tangence avec la droite de prix ; c'est ce qu'indique le graphique 7.4.

Graphique 7.4 : la préférence pour la liquidité comme attitude face au risque.



Vers le bas, A_2 ne diminue pas, mais augmente puisqu'on s'éloigne de l'abscisse.

Il est généralement admis que les agents sont *risk averters*, que pour un return donné, leur satisfaction diminue lorsqu'augmente le risque. Les courbes d'indifférence doivent donc avoir une pente positive[1]. Qu'est-ce qui assure leur convexité ? Il faut pour cela poser des hypothèses concernant la fonction d'utilité qui sous-tend les courbes d'indifférence, notamment l'utilité marginale décroissante du return. Sous ces conditions, « all risk averters are diversifiers: plungers do not exist »[2].

La droite de prix, croissante, donne le « prix » du risque en abscisse. Son équation est donnée par la relation fonctionnelle qu'on peut établir entre μ_R et σ_R. De (7.12) et (7.13), on déduit :

$$\mu_R = (r/\sigma_g).\ \sigma_R. \qquad (7.14)$$

La pente de la droite de prix est donc r/σ_g. L'optimum correspond au point de tangence avec une courbe d'indifférence. De cette combinaison optimale du rendement et du risque du portefeuille, il est facile de déduire les parts de la monnaie et des consolidées, à partir de (7.13) et (7.14), comme le montre la partie inférieure du graphique 7.4.

Si le taux d'intérêt augmente (passage de r_1 à r_2 sur le graphique), la pente de la droite μ_R augmente aussi et celle-ci sera tangente à une autre courbe d'indifférence en un point où le

[1] Par sa forme, cette courbe d'indifférence diffère de celles que nous avons rencontrées jusqu'ici. L'un des éléments du choix, le risque, a une connotation négative ; il doit être compensé, non par un sacrifice de l'élément positif (le rendement) mais par un supplément.

[2] Tobin [361] p. 76

risque est plus élevé ; à ce nouveau point d'équilibre, A_2 sera supérieur. La part détenue en monnaie diminue donc quand le taux d'intérêt augmente ; c'était déjà la conclusion de Keynes ; nous y parvenons par un autre cheminement plus réaliste.

Une réduction du risque perçu (subjectivement) par l'agent fera augmenter la pente de la droite de μ et de la droite de A_2, puisque toutes deux comportent σ_g au dénominateur de leur équation. Il en résultera tout logiquement l'augmentation de A_2.

LE COEFFICIENT « q »

Ce n'est pas le moindre mérite de Keynes que d'avoir été le premier à mettre en avant les décisions des agents quant à l'agencement de leur patrimoine et d'avoir intégré la monnaie dans ce processus. Mais étonnamment, la décision se ramène à un choix entre deux actifs : la monnaie qui ne porte pas intérêt et les obligations qui portent intérêt. Certes, Keynes a consacré plusieurs pages de réflexion à la bourse et au marché des actions, démontrant ainsi l'importance qu'il leur attachait. Toutefois, à l'instant décisif, lorsqu'il s'agit de déterminer le volume des investissements, le marché des actions est incroyablement absent.

Dans des articles parus en 1968 et 1969, Tobin reconstruit la relation entre le secteur réel (les investissements productifs) et le système financier et monétaire en accordant la place centrale aux actions. L'optique de Tobin est franchement innovante. Il considère les différents acteurs : l'autorité publique et monétaire, les banques privées, le public et les entreprises ; il envisage l'actif et le passif du patrimoine de chaque catégorie d'acteurs, ainsi que le type de rationalité qui la guide. Les diverses catégories d'actifs sont offertes et demandées : le capital réel de l'entreprise, les dépôts dans les banques et l'encours de leurs prêts, les réserves libres ou obligatoires à la banque centrale, la dette publique du gouvernement… Chaque actif a son marché où se détermine son prix d'équilibre. Pour tenir compte de cette globalité interdépendante, Tobin recourt à la méthode walrassienne de l'équilibre général. Le principe est bien sûr d'égaliser le nombre des équations avec le nombre des inconnues. Sur chaque marché, l'offre et la demande sont fonction, non seulement du taux d'intérêt considéré mais également de ceux des actifs concurrents[1].

De nombreuses variables interviennent dans les équations[2] ; citons sans être exhaustifs : *W* (le patrimoine net total de la société), *Y* (le revenu national), les différents taux de rendement, *K* (le capital des entreprises), *M* (la masse monétaire), S (la dette publique), *k* (les réserves obligatoires des banques auprès de la banque centrale) … Il faut maintenant insister sur l'une d'elles, qui a fait la réputation du modèle : le fameux *coefficient q*, qui joue un rôle central. Le modèle s'attache au rapport entre le capital réel et le capital financier. Prenons une entreprise. On peut la valoriser de deux façons. Premièrement, en additionnant les valeurs de remplacement ses éléments individuels, telles qu'elles ressortent des marchés de ces biens. C'est la variable *K*. Deuxièmement, en prenant sa valeur boursière *V*. Celle-ci correspond à la somme des revenus futurs escomptés au coût du capital (r_K). Le *coefficient q* est en fait le rapport entre ces deux valeurs. Par définition, on a :

$$q = V/K \text{ ou } \quad V = q.K \qquad (7.15)$$

[1] La demande des différents actifs est également fonction du revenu (plus précisément le rapport entre le revenu et le patrimoine), car de ce facteur dépend le besoin en encaisse de transaction qui influence toutes les autres demandes.

[2] Tobin a présenté cette théorie dans plusieurs articles avec des variantes quant aux équations, aux variables prises en compte, à leur caractère endogène ou exogène.

Si q >1, alors V >K. Cela signifie que racheter une entreprise existante revient plus cher que de réaliser un investissement direct correspondant. Une telle situation favorise donc les nouveaux investissements directs, ceux qui accroissent *K* et donc l'emploi.

De cette équation, on peut en déduire une autre, relative aux rendements. Considérons les deux taux *R* et r_K : *R* est l'*efficacité marginale du capital* ; par définition, elle égalise les revenus futurs actualisés de l'entreprise avec le coût de remplacement de ses actifs sur le marché des biens. Pour sa part, le taux r_K égalise les revenus futurs de l'entreprise avec la valeur boursière. Il en résulte donc l'équation :

$$q = R/r_K \text{ ou } \qquad q.r_K = R \qquad (7.16)$$

Le coefficient *q* est une des inconnues du système d'équations. Sa valeur d'équilibre est évidemment un , puisque si q > 1, les investissements seront encouragés, ce qui réduira *R* jusqu'à ce que l'égalité avec le coût du capital soit atteinte ; *q* sera alors ramené à sa valeur d'équilibre. Les investissements dont il est question ici ne sont pas liés à la croissance économique ; on est dans un modèle stationnaire. Ils sont générés par un déséquilibre et l'incitation à investir disparaît avec le déséquilibre qui l'a causée. Ce qui est intéressant, c'est qu'un tel déséquilibre peut être produit artificiellement, par la politique monétaire.

Supposons que les autorités monétaires accroissent la masse monétaire en circulation. Les agents sont amenés à réajuster leur patrimoine en tentant de substituer d'autres actifs à la monnaie pléthorique, ce qui les rendra moins exigeants quant au rendement de ces actifs. Comme d'autres taux d'intérêt, le coût du capital propre des entreprises, r_K, va baisser et tomber en dessous de la MEC ; autrement dit, *q* s'élèvera au-dessus de un. Les firmes auront alors intérêt à investir car ce nouveau capital rapportera plus qu'il ne coûte. Tobin écrit : « the valuation of investment goods relative to their cost is the prime indicator and proper target of monetary policy »[1]. Chez Keynes, l'expansion monétaire faisait passer le taux d'intérêt sous la MEC. Chez Tobin, elle fait hausser *q*, ce qui a le même effet.

Dans la liste des variables, Tobin n'oublie jamais de citer le taux d'intérêt de la monnaie pure. Certes, il vaut normalement zéro, que ce soit pour raison légale ou par convention sociale. Ceci ne signifie pas son inexistence mais son exogénéité ! Cette caractéristique du modèle tombe à point nommé. Suivant le principe walrassien, les équations relatives à *n* marchés d'actifs interdépendants ne peuvent déterminer que (n-1) taux. Tous les autres taux que celui de la monnaie sont donc susceptibles d'être considérés comme endogènes (seulement SUSCEPTIBLES, car l'économiste reste libre de choisir ses variables endogènes).

Quelle caractéristique particulière de la monnaie, absente chez les autres actifs, lui confère une influence dans la politique d'expansion ? « It is not because asset n° 1 has been called "money" and asset n° 2 "securities". It is not because asset n°1 is a means of payment or has any other intrinsic properties asset n° 2 lacks. It is not that asset n° 1 bears no interest-it may or may not (…) The essential characteristic- the only distinction of money from securities that matters for the results given above- is that the interest rate on money is exogenously fixed by law or convention, while the rate of return on securities is endogenous, market-determined »[2]. Si la monnaie avait un taux endogène, l'effet d'une variation de la masse monétaire se résumerait en grande partie à faire varier ce taux ; l'effet sur les autres taux (notamment r_K) serait amorti et la politique monétaire serait moins efficace.

[1] Brainard & Tobin [40] p 104.
[2] Tobin [362] pp. 25-26

7.2- LE MONETARISME

Alors que la pensée néo-keynésienne devînt dominante dès les années quarante, la résistance vînt de l'Université de Chicago où à partir de 1956, Milton Friedman construisît patiemment une théorie alternative, qu'on appellera le *monétarisme*. Il connut son heure de gloire dans les années soixante-dix, lorsque la présence simultanée d'une crise économique et d'une inflation presque galopante mit à mal certaines théories néo-keynésiennes. Le monétarisme n'est pas un retour à la conception macroéconomique d'avant Keynes ; c'est une conception nouvelle qui se place sur le terrain macroéconomique défini par Keynes, pour le contredire. La contre-révolution néoclassique, qui niera tout l'apport keynésien, viendra plus tard avec les *nouveaux classiques*.

Le monétarisme fit tache d'huile à partir de Chicago et devînt très influent mais il est intimement lié à la personnalité de Milton Friedman, son concepteur. Cette école compte également les économistes Karl Brunner (l'inventeur du terme « monétarisme » en 1968), Allen Meltzer, Philip Cagan, Edmund Phelps, Harry Johnson…

La réputation de Friedman a débordé du cercle des économistes : un public plus large le connaît pour ses positions opposées à l'intervention de l'Etat dans la vie sociale et économique, ce qui le met également en opposition avec les conclusions politiques de la théorie économique keynésienne.

7.2.1. La théorie quantitative remise à l'honneur

LA THEORIE QUANTITATIVE

La grande passion de Friedman est la théorie quantitative et notamment ses deux équations, celle de Fisher (4.23) et celle de Pigou (4.26). Patinkin prétend que *l'effet d'encaisse réelle* sous-tend la théorie quantitative chez tous les auteurs qui la professent, même si eux-mêmes n'en sont pas conscients. Friedman approuve cette mise en évidence de l'effet d'encaisse réelle. Dans l'équation de Fisher, l'augmentation de M entraîne celle de P à cause de la manière dont les agents s'y prennent pour liquider leurs encaisses excédentaires : ils paient plus cher.

Dans l'équation de Fisher, le terme PT comprend toutes les transactions, comptées à leur valeur totale. L'inclusion de la totalité de la circulation monétaire n'est pas nécessairement intéressante. A la version TRANSACTIONS de Fisher, Friedman oppose la version REVENU, formulée ainsi :

$$M.V = P.y \qquad (7.17)$$

La variable y représente le revenu national réel. La valeur de $P.T$ sera normalement supérieure à celle de $P.y$, car :

1- L'équation (7.17) n'inclut que les transactions créatrices de revenu et non celles qui le distribuent ou le font circuler (donations et cadeaux, impôts etc).
2- $P.T$ compte les transactions à leur valeur totale, alors que $P.y$ ne comprend que la valeur ajoutée.

L'EQUATION FRIEDMANIENNE DE LA DEMANDE DE MONNAIE

C'est son article « The Quantity Theory- A Restatement » publié en 1956 qui lança le monétarisme. Le fil conducteur tient dans cette phrase : « The quantity theory is in the first instance a theory of the demand for money ». Friedman remarque que la monnaie est une manière de détenir de la richesse parmi d'autres. La demande individuelle de monnaie dépend de la répartition optimale du patrimoine. Sur base de cette analyse utilitarienne, Friedman

construit sa fameuse équation de la demande de monnaie. On y rencontre donc trois types de facteurs :

1) la richesse totale à partager entre les différentes affectations.
2) les rendements attendus de la monnaie et des placements alternatifs.
3) les goûts et préférences des agents

L'article analyse la demande de monnaie en trois étapes : d'abord, celle d'un ménage représentatif, ensuite celle d'une entreprise représentative, enfin l'agrégation des demandes au niveau de la société.

Commençons par le ménage qui doit répartir sa richesse totale *W*. Friedman la conçoit de façon très abstraite et générale. C'est plutôt une espèce de capitalisation de l'ensemble des revenus de tous types que l'agent s'attend à recevoir dans sa vie. En prenant la formule de la capitalisation perpétuelle, on a donc :

$$W = Y/r \qquad (7.18)$$

Ici, le revenu nominal *Y* n'est pas le revenu COURANT, mais le *revenu PERMANENT*, un concept important défendu par Friedman, sur lequel nous reviendrons au chapitre 9.5. Grosso modo, il s'agit du niveau moyen de revenu que l'agent s'attend à avoir au cours de la période considérée. Similairement, le taux d'intérêt *r* de cette formule est un taux moyen sur une longue période et non le taux courant.

Friedman considère cinq types d'actifs entre lesquels le patrimoine sera partagé : la monnaie *M*, les obligations *B*, les actions *E*, les biens physiques durables *G* et le capital humain, c'est-à-dire la capacité de gagner de l'argent en travaillant *H*. Voyons en quoi consistent les rendements de ces cinq actifs, qu'on retrouvera dans la fonction de demande de monnaie.

1- La **monnaie** paye son return en nature par la facilité et la sécurité qu'elle procure. D'après Friedman, ce service dépend du pouvoir d'achat d'une unité de la monnaie. Le niveau général des prix *P* est donc un facteur déterminant.

2- Les **obligations**. Soit r_b égale le coupon annuel divisé par le cours de bourse. Outre l'intérêt, il faut tenir compte des variations du cours. Friedman simplifie en prenant le cas d'une obligation perpétuelle : le cours vaut donc $1/r_b$. Ses fluctuations ne proviennent que de celles du taux, données par dr_b/dt ; le signe est négatif, car le cours varie en sens contraire du taux. Le rendement total est :

$$r_B - \frac{1}{r_b}.\frac{dr_b}{dt} \qquad (7.19)$$

3- Les **actions**. Le principe est le même que pour les obligations, mais avec un élément en plus, correspondant à l'augmentation des prix moyenne attendue [$1/P.(dP/dt)$]. En effet, contrairement aux obligations, les actions ne sont pas appelées à se déprécier du fait de l'inflation. Le rendement total est donc :

$$r_E - \frac{1}{r_e}.\frac{dr_e}{dt} + \frac{1}{P}.\frac{dP}{dt} \qquad (7.20)$$

4- Le return d'un **bien physique** se limite à la réévaluation dont il peut faire l'objet par rapport à la monnaie :

$$\frac{1}{P}.\frac{dP}{dt} \qquad (7.21)$$

5- Le **capital humain**. N'étant pas substituable avec les autres actifs, sa taille est donnée. Représentons par *w* le rapport du capital non humain sur le capital humain.

Il faut également mentionner les goûts de l'individu comme déterminant de la construction du patrimoine. Friedman regroupe dans la variable *u* l'ensemble des facteurs de ce type.

Il reste à réunir tous ces éléments dans la fonction de demande de monnaie. Moyennant quelques simplifications, on obtient finalement pour le ménage individuel la fonction :

$$M_D = \mathrm{f}(\mathrm{P}, \mathrm{r_b}, \mathrm{r_e}, \frac{1}{P}.\frac{dP}{dt}, w, Y, u) \qquad (7.22)$$

Le premier *P* se justifie par le fait que *u*, r_b, r_e et *(1/P).(dP/dt)* sont des rendements NOMINAUX alors que ce sont les variables REELLES qui gouvernent la demande de monnaie. Pour déterminer l'encaisse désirée, il faut tenir compte du pouvoir d'achat de la monnaie. Dans le même ordre d'idées, la demande de monnaie doit être indépendante de l'unité de mesure monétaire. Avec un dollar valant 10% de moins, la demande de monnaie vaut 10% de dollars de plus. L'équation (7.22) est donc homogène de degré un par rapport aux variables exprimées en unités monétaires *P* et *Y*.

Venons en aux entreprises. Pour elles, la monnaie est un capital à l'actif de leur bilan qui est compensé par des actions ou des dettes au passif. Sur cette base, Friedman considère qu'il n'y a pas de nouvelles variables à ajouter dans l'équation (7.22) : *P*, r_b, r_e, (*1/P).(dP/dt*) représentent bien le coût de la détention de monnaie pour l'entreprise[1]. Il faut juste adapter *u* pour qu'il intègre également les données technologiques et commerciales qui peuvent influencer la demande de monnaie (degré d'intégration verticale par exemple).

Friedman estime que l'équation (7.22) convient à la fois pour la demande de monnaie des agents individuels et pour celle de la société. On peut négliger les problèmes que pose l'agrégation. Il reconnaît toutefois que si les anticipations des agents quant à la hausse des prix diffèrent, l'agrégat dP/dt devient un concept très abstrait.

Nous arrivons maintenant au moment critique du raisonnement friedmanien. L'équation (7.22) va être transformée en équation quantitative. D'abord, l'offre de monnaie (M) remplace la demande (M_D) dans le membre de gauche, puisqu'elles doivent s'égaliser à l'équilibre. Friedman la considère comme fixée de façon autonome[2]. Comme *M* est homogène de degré un par rapport à *Y* et *P*, on peut transformer (7.22) en :

$$M = \mathrm{Y.f}(\mathrm{P/Y}, \mathrm{r_b}, \mathrm{r_e}, \frac{1}{P}.\frac{dP}{dt}, w, Y, u) \qquad (7.23)$$

On en déduit

$$\mathrm{M.V} = \mathrm{Y}. \qquad (7.24)$$

$$o\grave{u} \quad V = \mathrm{V}(\mathrm{Y/P}, \mathrm{r_b}, \mathrm{r_e}, \frac{1}{P}.\frac{dP}{dt}, w, Y, u) \qquad (7.25)$$

L'équation (7.24) n'est rien d'autre que l'équation quantitative (7.17) où le terme *P.y* est remplacé par le revenu nominal *Y*. L'équation (7.25) énumère les facteurs qui influencent la vitesse de circulation selon Friedman. Ils sont manifestement différents de ceux de Fisher. Notamment, les arrangements institutionnels réglant la régularité et la fréquence des paiements

[1] Certes, les entreprises peuvent aussi obtenir leur monnaie en recourant à la création de monnaie par les banques, plutôt qu'en s'adressant aux ménages. Friedman considère cette différence comme non essentielle.

[2] Comme Fisher, Friedman voit dans le caractère autonome de l'offre de monnaie une caractéristique fondamentale de la théorie quantitative. Il écrit « there are important factors affecting the supply of money that do not affect the demand for money » ([99] p. 16).

n'y figurent pas. Friedman pense que ces arrangements dépendent eux-mêmes de la recherche de l'utilité maximale par les agents, qui a guidé l'élaboration de ses équations.

La question essentielle est : quel sera l'effet d'une variation de *M* ? Prenons l'exemple d'une hausse. Il y a deux sous-questions :

1- le revenu nominal, augmentera-t-il ; autrement dit, est-il exclu qu'une variation inverse de *V* compense celle de *M* ?
2- si le revenu nominal *Y* augmente, comment cette augmentation se répartira-t-elle entre le revenu réel et le niveau des prix ?

Concernant la première sous-question, Friedman écrit : « The quantity theorist accepts the empirical hypothesis that the demand for money is highly stable ». La demande de monnaie étant l'inverse de sa vitesse de circulation, la stabilité de l'une implique celle de l'autre. Les variations de *M* ne devraient donc pas être compensées par celles de *V* . Friedman précise toutefois que cette stabilité ne doit pas être comprise comme une constance dans le temps, mais comme une relation stable entre la demande de monnaie et les facteurs qui la déterminent. Cette mise au point, fût-elle judicieuse, a l'inconvénient qu'on ne voit plus très bien si oui ou non une variation du taux d'intérêt affectera la demande de monnaie. Dans un article paru en 1959, Friedman considérera la demande de monnaie comme inélastique par rapport au taux d'intérêt, mais cette idée ne se retrouvera pas dans tous ses écrits et elle semble en contradiction avec l'équation (7.22) ci-dessus.

La deuxième sous-question ne trouvera une réponse étayée que dans des publications ultérieures, mais Friedman en donne un avant-goût : « money does matter (…) any interpretation of short term movements in economic activity is likely to be seriously at fault if it neglects monetary changes »[1]. Cette citation suffit à illustrer le fossé entre Friedman et les théoriciens quantitativistes antérieurs[2].

Friedman compare sa théorie représentée par l'équation (7.23) avec celle de Keynes reflétée par le graphique 5.2. Chez Keynes, l'équilibre du marché monétaire ne détermine que le taux d'intérêt. Chez Friedman, la hausse de *M* entraînera des ajustements de tous les facteurs de l'équation (7.23). Notamment, une augmentation de *M* pourrait stimuler la demande de biens physiques.

LA MONNAIE, LE REVENU NOMINAL, LE REVENU REEL ET LES PRIX

Friedman publie en 1963 son étude monumentale « A Monetary History of the United States 1867-1960 ». Un chapitre de plus de cent pages est consacré à la période 1929-1933, qui a vu la plus grande dépression économique de l'histoire. Ce chapitre commence par égrener quelques chiffres qui donnent la mesure de l'événement. Le produit national des Etats-Unis en prix courants a baissé de plus de moitié ; en prix constants, il a baissé de plus du tiers ; les prix de plus d'un quart. Pour trois travailleurs actifs, il y avait un chômeur. Le stock de monnaie se réduisit d'un tiers. Le nombre de banques commerciales également d'un tiers. La vitesse de circulation de la monnaie déclina de près d'un tiers. Une baisse de *V* est habituelle en dépression et vu la sévérité de celle-ci, on aurait pu s'attendre à une baisse de *V* encore plus sensible, mais elle fut freinée car les nombreuses faillites bancaires modérèrent l'envie du public de détenir des encaisses. Friedman estime que l'attitude des autorités monétaires américaines pendant la grande crise fut inepte. Elles ne s'opposèrent pas activement à la baisse de la masse monétaire et ne soutinrent pas suffisamment les banques en difficulté en leur

[1] Friedman [99] p. 3
[2] Friedman rappelle toutefois que Fisher permettait une influence de *M* sur *T* à court terme.

apportant des liquidités. Sans leur incompétence[1], la crise eût été moins grave. Voilà donc une illustration du pouvoir de la monnaie : mal gérée, elle dérègle toute l'économie. Les statistiques exposées dans l'ouvrage montrent toutefois que la monnaie de base émise par la réserve fédérale augmenta pendant la période incriminée et que la baisse de *M* s'explique par la chute des deux ratios bancaires, le ratio *dépôts / réserve* et le ratio *dépôts / billets de banque*, qui dépendent autant du public et des banques commerciales que de l'autorité monétaire[2].

Venons-en aux conclusions tirées par Friedman concernant la relation entre l'évolution de la masse monétaire *M* et celle des variables *Y, y* et *P* sur l'ensemble de la période étudiée. Le tableau suivant donne le degré de corrélation de chacune des trois variables avec *M*, tel qu'il ressort des recherches empiriques de Friedman :

Variables		Mouvements à court terme	Mouvement séculaire
Y	*revenu nominal*	élevé	élevé
y	*revenu réel*	élevé	bas
P	*niveau des prix*	élevé	élevé

Passons maintenant de la corrélation à la causalité. « One thing is abundantly clear from our narrative. Monetary changes have in fact often been independent, in the sense that they have often not been an immediate or necessary consequence of contemporaneous changes in business conditions »[3]. La corrélation s'explique donc par l'influence de la monnaie sur les autres variables. Friedman reconnaît qu'il est difficile à l'économiste d'isoler les effets des actions des autorités monétaires ; d'autres événements économiques interagissent, auxquels les effets observés pourraient également être attribués. Pour étayer sa vision de la causalité, Friedman s'appuie sur une argumentation empirique plutôt que théorique. A ce titre, Friedman reconnaît l'existence d'influences en sens inverse ; l'un des trois déterminants de la masse monétaire, le *ratio dépôts / réserves*[4] se montre sensible au cycle des affaires.

Le tableau que dépeint Friedman pour la période 1929-1933 ne semble pas soutenir ses conclusions ; est-ce dû à la particularité de cette période ?

7.2.2. La politique monétaire et budgétaire

Les propositions de Friedman ont varié dans le temps mais elles s'appuyaient toujours sur les mêmes principes. Le premier d'entre eux est l'opposition à toute politique contracyclique. Le niveau des taxes, celui des dépenses publiques et la masse monétaire doivent connaître une croissance stable et indépendante de la phase du cycle dans laquelle se trouve l'économie. Même si on admet l'avantage d'une politique qui renforce la demande globale lorsqu'elle est faible et qui la bride lorsqu'elle surchauffe, il reste le problème de la synchronisation des effets de ces politiques avec l'état de la demande globale. Or l'Etat maîtrise les effets de ses politiques budgétaire et monétaire trop imparfaitement. Friedman énumère trois types de délais, dont l'addition rend improbable que les effets se fassent sentir au moment opportun ; il les énonce ainsi[5] :

1- the lag between the need for action and the recognition of this need

---

[1] Cette incompétence est une simple constatation empirique que Friedman ne s'explique pas.
[2] Friedman & Schwartz [108] p. 333.
[3] Friedman & Schwartz [108] p. 686
[4] Les deux autres sont : la *monnaie de base* émise par la banque centrale et le ratio *dépôts / billets de banque*.
[5] Friedman [97] p. 382.

2- the lag between recognition of the need for action and the taking of action
3- the lag between action and its effects

Friedman disqualifie totalement la politique budgétaire. Qu'en est-il de la politique monétaire ? Il conteste l'efficacité de la proposition keynésienne bien connue : la création de monnaie fait baisser le taux d'intérêt, ce qui relance l'investissement. Certes, dans un premier temps, la recette marche. Mais ce succès apparent n'est que le début du processus, pas la fin. L'accroissement des dépenses va doper la demande de monnaie, la demande de prêts et les prix, ce qui réduira même peut-être les encaisses réelles. Après la tendance à la baisse, le taux d'intérêt connaîtra une phase de pression à la hausse. Et il finira par revenir au niveau qui serait le sien si la politique monétaire n'avait pas rêvé de l'en écarter. Friedman compte grosso modo une année pour que le taux d'intérêt commence à augmenter après l'action de l'autorité monétaire et une deuxième année pour qu'il revienne au taux naturel. Et comme le taux d'intérêt MONETAIRE inclut une prime pour se prémunir contre l'inflation anticipée (cf. Fisher), ce taux se trouvera même plus élevé, car les agents anticiperont vraisemblablement la poursuite de la hausse des prix. Friedman conclut : « As an empirical matter, low interest rates are a sign that monetary policy has been tight- in the sense that the quantity of money has grown slowly ; high interest rates area sign that the monetary policy has been easy- in the sense that the quantity of money has grown rapidly »[1].

Les objectifs raisonnables pour la politique monétaire sont[2]:
1- "prevent money itself from being a major source of economic disturbance
2- "provide a stable background for the economy
3- "offsetting major disturbances in the economic system arising from other sources. Les perturbations d'importance secondaire ne doivent pas être combattues, car la réaction ne viendra jamais au bon moment avec la force juste.

Concrètement, les objectifs de la politique monétaire doivent être formulés sur base d'une variable qui est effectivement contrôlable par les pouvoirs publics. Il faut donc exclure des variables comme le taux d'intérêt, le taux de chômage… Même le taux d'inflation n'est pas suffisamment contrôlable pour servir d'objectif à la politique monétaire[3]. L'objectif doit porter sur les agrégats monétaires eux-mêmes. Dans le cas des Etats-Unis pendant les années soixante, Friedman préconisait une croissance de 3 à 5%. Mais comme il le précise lui-même, c'est moins le taux qui importe que sa stabilité et le fait qu'il soit connu à l'avance par le public.

En conclusion, nous avions constaté chez Keynes le paradoxe qu'après avoir fait de la monnaie le centre de l'économie, il pensait que la politique monétaire avait peu d'effet. Le paradoxe continue avec les monétaristes, car ceux-ci reprochent aux keynésiens de sous-estimer l'influence de la politique monétaire mais ils en déconseillent l'emploi vu son caractère potentiellement perturbateur.

7.2.3. « The Opimal Quantity of Money »

Tel est le titre d'un article assez étonnant écrit par Friedman en 1969. Il y défend une autre règle que celle des 3 à 5% mentionnée ci-avant, en se basant sur une argumentation quasi-microéconomique, comme Patinkin l'avait fait (cf. 6.5.1), mais avec des conclusions différentes. Il intitule d'ailleurs sa conclusion « A Final Schizophrenic Note ». La règle de 3 à

[1] Friedman [101] p. 7.
[2] Friedman [101] pp. 12-14.
[3] Lors de l'instauration de la monnaie unique européenne, l'objectif imparti à l'autorité monétaire était une inflation annuelle de 2%.

5% n'est pas répudiée mais ramené à une règle à court terme facile à mettre en place. C'est par pragmatisme que Friedman la recommande ; mais la réflexion théorique- il faut bien le dire : fort abstraite- le conduit à une autre conclusion. Voyons-en la substance.

L'individu représentatif[1] répartit son revenu entre la consommation et les variations de son patrimoine mobilier qui comporte divers types d'actif, principalement la monnaie, les obligations et les actions. Pour ce faire, il doit tenir compte des facteurs suivants :

- L'IRD (internal rate of discount) qui est le taux d'impatience fisherien entre consommation présente et consommation future. L'IRD d'origine, sans prêt ni emprunt, est noté IRD(0).
- Les rendements des actifs, c'-à-d zéro pour la monnaie, r_e pour les actions, r_b pour les obligations.
- Le taux d'inflation anticipé μ qui déprécie la monnaie et les obligations mais pas les actions.
- Les avantages non pécuniaires marginaux des actifs[2] : $MNPS_e$ (actions), $MNPS_b$ (obligations) et $MNPS_m$ (monnaie). Au sous-chapitre 6.5.2, il a été abondamment question des services de la monnaie. La nouveauté est que "just as the holding of money balances yields non-pecuniary returns in the form of a feeling of security and pride of possession, so also the possession of a bond may yield similar non-pecuniary returns"[3]. Friedman formule deux hypothèses les concernant :
    - $MNPS_m$ et $MNPS_b$ sont chacun une fonction décroissante des deux variables m_i et v_i ; m_i est le ratio de l'encaisse sur le revenu et v_i est le ratio de la valeur du portefeuille d'obligations sur le revenu.
    - $MNPS_m > MNPS_b$.

    Ces deux hypothèses me paraissent discutables mais elles sont indispensables au raisonnement friedmanien.

Selon un principe bien connu, l'agent optimise son bien-être en consommant, thésaurisant, prêtant et empruntant de façon à égaliser l'avantage marginal de chaque option. Le capital par actions étant ignoré, l'équation d'équilibre individuel est :

$$MNPS_m - MNPS_b = IRD(0) - MNPS_b + \mu = r_b \qquad (7.26)$$

Friedman montre qu'au niveau social agrégé, partant d'un équilibre avec des prix stables, si on crée une inflation positive par un accroissement de *M* constant et prévisible, il en résultera une perte de bien-être du fait de la baisse de la valeur réelle de l'encaisse[4]. Inversement, une baisse de *M* suscitant un μ négatif accroît l'encaisse réelle et donc le bien-être[5]. Dans ce cas, le bien-être augmente jusqu'à un maximum au-delà duquel il redescend, car l'encaisse très élevée entraîne des coûts. Cet optimum social sera atteint pour $r_b = 0$; en ce point, le taux d'inflation négatif compense exactement le taux d'intérêt réel, qui reste donc positif. En bon néoclassique, Friedman considère que le taux d'intérêt réel est insensible à la politique monétaire qui ne peut gouverner que le taux nominal ; en générant une inflation négative, elle peut réduire le taux d'intérêt nominal qui égale le taux d'intérêt réel exogène plus l'inflation anticipée. « Our final

[1] Friedman tient également compte des entreprises ; je les ignore à titre de simplification.
[2] « Marginal Non Pecuniary Services ».
[3] Friedman [102] p. 24. La citation ne mentionne pas les actions. Friedman les fait intervenir pour mémoire ; il les élimine ensuite du raisonnement à titre de simplification.
[4] Le niveau des prix *P* augmentera plus que la masse monétaire parce que les agents tenteront de détenir moins d'actifs sans risque pour fuir l'effet de la hausse des prix : se heurtant à l'impossibilité de réaliser leur désir de diminuer leur encaisse (*M* est exogène), il ne leur restera qu'à pousser encore les prix à la hausse.
[5] Tant que $r_b > 0$, $MNPS_m$ l'est également. L'accroissement de l'encaisse réelle crée donc du bien-être.

rule for the optimum quantity of money is that it will be attained by a rate of price deflation that makes the nominal rate of interest equal to zero »[1].

En situation stationnaire, la politique de l'autorité monétaire doit viser à faire baisser *M* au taux de déflation souhaité. Si l'économie est en croissance, *M* devra varier à un rythme valant le taux de croissance moins le taux de déflation souhaité.

Cette règle rééquilibre le patrimoine en faveur de l'encaisse monétaire dont l'attrait souffre plus de l'inflation que celui des autres actifs, car elle ne rapporte pas d'intérêt dont le taux peut être ajusté à la hausse des prix. Commentant l'article de Friedman, Sanches le résume ainsi : "When the nominal interest rate is positive, households and firms hold too little money"[2]. Friedman remarque lui-même que le même objectif peut être atteint par une voie alternative : l'attribution d'un intérêt sur les avoir monétaires.

Selon Sanchez, les principales critiques qui furent adressées à cette théorie friedmanienne sont :
1- La perte de bien-être due aux encaisses trop restreintes est faible.
2- La politique monétaire vise d'autres objectifs plus importants que le coût d'opportunité de la détention de monnaie.

7.2.4. Comment Friedman perçoit sa controverse avec les keynésiens

Dans l'article « A Theoretical Framework for Monetary Analysis » (1970), Friedman fait le point sur sa controverse avec les keynésiens. Pour apprécier l'effet d'un élargissement de l'offre de monnaie, il part de l'équation quantitative cambridgienne :

$$M = k.P.y \qquad (7.27)$$

La théorie néoclassique traditionnelle, suivant l'exemple de Marshall, considère qu'à court terme, l'équilibre s'ajuste à un choc quelconque par les prix (variable *P*) et non par les quantités. Keynes, qui, d'après Friedman, ne s'intéressait qu'au court terme, inverse la relation. Il privilégie l'ajustement par les quantités, à tel point que Friedman considère l'hypothèse de prix rigides comme sous-jacente dans la théorie keynésienne. Dans l'équation (7.27), une variation de *M* laissera donc *P* intact ; qui de *y* ou *k* supportera l'ajustement ? De deux choses, l'une :

- ou bien la demande de monnaie spéculative ne suffit pas à éponger la monnaie nouvellement émise et le taux d'intérêt baissera, stimulant le revenu réel *y*. L'hypothèse des prix rigides est nécessaire, car la variation du taux d'intérêt suppose que les agents se comportent comme si la variation de la masse monétaire NOMINALE équivalait à une variation de la masse monétaire REELLE.
- soit, prévaut la trappe des liquidités ; alors la demande de monnaie, représentée par *k*, absorbe la totalité de l'ajustement et le taux d'intérêt reste constant et avec lui le revenu réel.

Friedman considère que sa théorie distribue l'ajustement sur les trois variables *k, y* et *P*. Délibérément, Friedman est réticent à être plus précis quant à l'importance des variables dans l'ajustement, mettant de nouveau en exergue le REVENU NOMINAL (*y.P*). Il conclut toutefois : « I regard the description of our position as "money is all that matters for changes in NOMINAL

[1] Friedman [102] p. 34.
[2] Sanches [325] p. 10

income and for SHORT-RUN changes in real income" as an exaggeration but one that gives the right flavor of our conclusions »[1].

Dans le même article, il présente son fameux « simple common model ». L'idée sous-jacente, toute paradoxale, est la proximité entre les théories néoclassique et keynésienne ; les divergences sont plus économétriques que théoriques et peuvent être réduites au poids à attribuer à quelques variables. Un jour, les études empiriques mettront d'accord les économistes des deux bords.

L'économie peut être synthétisée par un modèle à sept équations et sept inconnues : *C* (la consommation), *I* (l'investissement), *Y* (le revenu nominal), *r* (le taux d'intérêt), *P* (le niveau des prix), M^D (la demande de monnaie) et M^S (l'offre de monnaie). Keynésiens et monétaristes sont d'accord sur six des sept équations. Voici ce tronc commun :

$$C/P = f(Y/P,r) \qquad (7.28\text{-A})$$
$$I/P = g(r) \qquad (7.28\text{-B})$$
$$Y/P = C/P + I/P \qquad (7.28\text{-C})$$
$$M^D = P.l(Y/P,r) \qquad (7.28\text{-D})$$
$$M^S = h(r) \qquad (7.28\text{-E})$$
$$M^D = M^S \qquad (7.28\text{-F})$$

Le lecteur est maintenant habitué à rencontrer les relations fonctionnelles et les égalités (7.28-A à 7.27-F) ; la seule qui surprend est la cinquième, qui fait dépendre l'offre de monnaie du taux d'intérêt ; Friedman l'intègre « to be consistent with the literature ». Mais il précise qu'elle peut être remplacée par une équation exprimant l'exogénéité.

La divergence porte sur ce que Friedman appelle la « missing equation » :

$$\underline{\text{keynésiens}} : \quad P = P_0 \qquad (7.29\text{-A})$$
$$\underline{\text{néoclassiques}} : \quad y = y_0 \qquad (7.29\text{-B})$$

L'équation (7.29-A) exprime la rigidité des prix à laquelle il a été fait allusion ci-dessus. L'équation (7.29-B) souligne la dichotomie néoclassique : le revenu réel est déterminé de façon exogène par la théorie de la valeur, c'est-à-dire que le système d'équations walrassien régit le secteur réel de l'économie.

Friedman attribue au modèle néoclassique et au modèle keynésien une faiblesse commune : les hypothèses 7.28-A et 7.28-B les dispensent d'élaborer un mécanisme expliquant comment la variation de *Y* se distribue entre *y* et *P*. Friedman a ici le tort de sous-estimer la théorie de Keynes. Rappelons-nous le huitième mécanisme au sous-chapitre 5.1.1. Keynes y analyse les facteurs répartissant l'influence d'une variation de *M* sur *y* (ou *T*) et *P* ; ce mécanisme de la théorie générale dément l'hypothèse de la rigidité des prix que Friedman impute à Keynes.

Dans l'idée de l'analyse économétrique, Friedman explore la problématique du mécanisme de répartition entre *y* et *P*. Il imagine ainsi un système de deux équations non indépendantes expliquant dP/dt et dy/dt par différentes variables, qui pourrait être testé empiriquement.

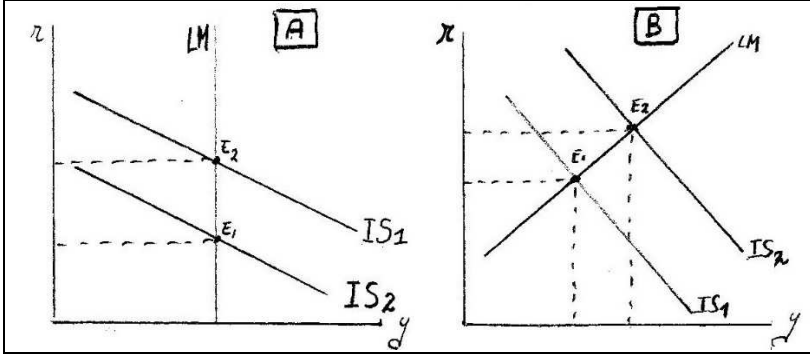
Dans un article-réponse publié en 1972, **Tobin** s'étonne de la vision de la controverse exposée par Friedman. Selon lui, les « missing equations » (7.29) ne reflètent ni la théorie monétariste, ni la théorie keynésienne. « I had thought that both monetarists and neo-Keynesians agreed that short-run variations of money income (*P.y* or *M.V*), however caused, were generally divided

[1] Friedman [103] p. 217

between changes in output and changes in price »[1]. Nous avons déjà remarqué que le point de vue de Friedman n'est pas constant en cette matière.

Tobin s'attendait à trouver en exergue l'opposition quant à la courbe LM. Une opposition telle que reflétée dans le graphique 7.5. A gauche, la courbe LM monétariste, qui est verticale, car la demande de monnaie est réputée insensible aux variations du taux d'intérêt[2]. A droite, la courbe LM élastique des keynésiens.

Graphique 7.5 : l'équilibre IS-LM chez les monétaristes et les keynésiens



Toujours selon Tobin, de ces courbes LM différentes découlent des conclusions divergentes pour la politique économique. Sur la figure de gauche, une élévation de la courbe IS par une politique budgétaire ad hoc ne permet pas d'accroître le revenu réel *y* et n'a d'autre effet que l'accroissement du taux d'intérêt. Pour accroître *y*, il faudrait pousser la courbe LM vers la droite en augmentant M/P, ce qui n'est pas simple dans l'optique monétariste, car l'augmentation de *M* peut entraîner celle de *P*. Chez les keynésiens, la hausse du taux d'intérêt générée par l'accroissement des dépenses publiques libère des encaisses de spéculation qui deviennent des encaisses de transactions pouvant financer la demande supplémentaire sur le marché des biens.

Comme on le voit, monétaristes et keynésiens ne parvinrent même pas à se mettre d'accord pour définir leurs points de convergence et de divergence.

7.2.5. La contre-révolution monétariste : acte deux

L'année 1968 marque une étape importante dans la carrière de Friedman. Pour la première fois apparaît dans une publication son nouveau cheval de bataille, le *taux de chômage naturel*. En fait, Friedman propose une nouvelle manière d'appréhender comment une variation du revenu nominal se partage entre le revenu réel et le niveau des prix, un sujet resté assez obscur dans ses écrits antérieurs. Treize ans après la remise à l'honneur de la théorie quantitative, Friedman parvient une nouvelle fois à emporter l'adhésion d'une frange importante de la profession économique.

Le nouvel assaut de Friedman ne peut être compris que dans son contexte : le débat sur la courbe de Phillips, que nous examinerons au sous chapitre 7.3.3. Les keynésiens avaient intégré la courbe de Phillips dans leur arsenal pour fonder théoriquement l'idée d'un possible arbitrage entre l'inflation et le chômage dans le chef de la politique économique. L'attaque de Friedman contre cette théorie lança le débat qui écorna sérieusement le crédit de l'école néo-keynésienne.

[1] Tobin [363] p. 854
[2] L'avis de Friedman sur cette question ne fut pas constant, comme nous l'avons vu.

Ne déflorons pas ici l'exposé consacré à ce débat, mais résumons en deux mots le nouveau leitmotiv de Friedman. Toute tentative de la politique monétaire ou budgétaire de réduire le chômage en dessous de son taux naturel, au prix d'une poussée inflationniste, peut obtenir des résultats à court terme, mais ceux-ci seront suivis par le retour au taux naturel. En voici la raison : les agents déterminent leurs actions sur base des grandeurs REELLES anticipées (prix relatifs, taux d'intérêt et salaire réels…). L'inflation n'affecte donc pas les comportements lorsqu'elle est anticipée correctement. Par contre, si elle prend les agents par surprise, ils seront dupés et la prendront pour une variation des grandeurs réelles. Les salariés au chômage répondront à cet appel, croyant que le salaire réel a augmenté. Lorsqu'ils se rendront compte de leur méprise, les agents ajusteront leur offre vers le bas. Le chômage reviendra au taux naturel. Par contre, l'inflation est irréversible.

7.3- QUELQUES GRANDS DEBATS

7.3.1. Baisse des prix et effet de richesse

Peu après la parution de la Théorie Générale éclate la première riposte : l'*effet de richesse* apparaît chez Pigou (1941, 1943, 1947). Ce fut une arme majeure contre la doctrine keynésienne. On l'appelle également *effet Pigou*.

L'idée de base est celle-ci : « the more accumulated wealth people have, the less keen they are to get more »[1]. Cela revient à dire que la propension marginale à consommer est une fonction croissante du rapport des possessions sur le revenu. C'est évidement la valeur REELLE du patrimoine et non sa valeur NOMINALE qui importe. Les variations du niveau général des prix influencent donc le phénomène.

PIGOU INTRODUIT L'EFFET DE RICHESSE

Voyons comment Pigou présente la question dans son article « The Classical Stationary State » (1943). L'article s'attaque à la thèse keynésienne selon laquelle un équilibre stationnaire de plein emploi peut s'avérer inaccessible, peu importe la baisse du salaire.

A cette fin, Pigou construit un modèle ad hoc d'équilibre stationnaire. Il commence par résumer la doctrine classique en cinq équations :

$$S\,(K,Y,i) = 0 \qquad (7.30\text{-A})$$
$$I\,(K,i) = 0 \qquad (7.30\text{-B})$$
$$Y = Y^* \qquad (7.30\text{-C})$$
$$\rho = i \qquad (7.30\text{-D})$$
$$\rho = \rho\,(K,Y) \qquad (7.30\text{-E})$$

La première équation explique l'offre réelle de capital ; elle croît avec le revenu réel *Y* et avec le taux d'intérêt *i* ; elle est aussi négativement liée au capital réel existant *K*. La deuxième équation explique la demande réelle de capitaux, négativement corrélée à *K* (productivité marginale décroissante) et au taux d'intérêt. En régime stationnaire, l'investissement et l'épargne nets sont nuls. La troisième équation impose que le revenu réel soit à son niveau de plein emploi Y*, l'hypothèse classique habituelle. La quatrième équation énonce le principe fisherien que « the rate of interest must be equal to the representative man's rate of discounting future satisfactions »[2]. Ce qui assure sa non négativité, puisque ρ est toujours positif. La cinquième équation détermine ρ.

Avec cinq équations pour les quatre inconnues (Y,K,i,ρ), le système semble surdéterminé. Il ne l'est pas, car la première équation n'est pas indépendante de la quatrième et de la cinquième. Le système a donc une solution.

La quatrième équation représente l'hypothèse très ancrée dans la théorie classique que « investment (i.e. saving) is made solely for the sake of the income that is expected to yield »[3]. Mais Pigou reconnaît qu'elle n'est pas réaliste. « People save (i.e. invest) partly from other

[1] Pigou [284] pp. 344-345
[2] Pigou [284] p. 345
[3] Pigou [284] p. 346

motives, the desire for possession, conformity to tradition or custom and so on »[1].En conséquence, les épargnants peuvent se satisfaire d'un taux d'intérêt inférieur à ρ. Dans ce cas, l'équation (7.30-A) pourrait n'être satisfaite que pour un taux d'intérêt inférieur à zéro. Or le taux d'intérêt ne peut descendre sous zéro. Pigou remplace les équations 7.30-D et E par :

$$i > 0 \qquad (7.31)$$

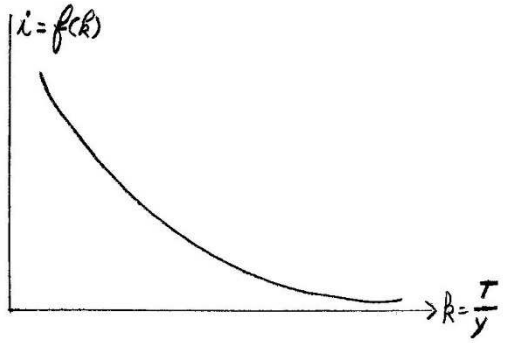
Le risque existe alors que prévale un équilibre de sous-emploi, parce que le taux d'intérêt ne peut pas tomber au niveau assurant le plein emploi, une situation qui donnerait raison aux keynésiens face aux néoclassiques. Pigou veut démontrer que ce cas est impossible.

Thèse n° 1 : ***S'il existe, l'équilibre ne peut être de sous-emploi***. Certes, avec l'équation (7.31) remplaçant 7.30-D et E, le système peut ne pas comporter de solution. Mais alors, le problème ne découle pas de la fausseté de (7.30-C) mais de l'incompatibilité de (7.31) avec le reste du système. Supposons un état stationnaire de sous-emploi avec un taux d'intérêt à peine supérieur à zéro. Pour démontrer la validité de (7.30-C), Pigou montre qu'une réduction suffisante du salaire réel peut toujours mener au plein emploi. Il dénonce l'opinion souvent entendue selon laquelle la diminution du salaire nominal entraînerait la réduction, non du salaire réel mais du niveau des prix. Dans ce scénario, le salaire réel resterait au dessus de son niveau d'équilibre et le revenu nominal chuterait.

L'argumentation de Pigou pour exclure l'éventualité d'une baisse du revenu nominal est assez tortueuse. Il invoque l'existence d'une relation fonctionnelle entre ledit revenu nominal et le taux d'intérêt ; le premier ne peut diminuer que moyennant la baisse du second. Or, par hypothèse, le taux d'intérêt est au minimum admissible ; il ne peut donc pas baisser.

Mais pourquoi le revenu nominal est-il rivé au taux d'intérêt ? Suivant la règle habituelle d'optimisation, l'intérêt des titres et la détention de monnaie doivent procurer une satisfaction marginale égale. L'équation quantitative cambridgienne nous dit que le coefficient *k* vaut l'encaisse REELLE (que Pigou note par la lettre *T*) divisée par le revenu réel *Y*. Il représente la tendance à thésauriser. Le graphique 7.6 présente son utilité marginale f(k) sur laquelle le taux d'intérêt doit se régler. Il apparaît clairement que le taux d'intérêt doit évoluer dans le même sens que le revenu réel *Y*. Mais pour un néoclassique, si la masse monétaire est constante, la proportion entre le revenu nominal et le revenu réel l'est aussi ; le taux d'intérêt et le revenu nominal fluctuent donc de concert. En conséquence, les prix ne diminuant pas, la baisse du salaire nominal stimule l'emploi.

Graphique 7.6 : l'utilité des encaisses et le taux d'intérêt



Thèse n° 2 : ***Il existe un équilibre de plein emploi***. Il ne s'agit plus de montrer que l'équation (7.30-C) ne peut pas être cause de déséquilibre mais qu'avec cette équation, le déséquilibre est

[1] Pigou [284] p. 346

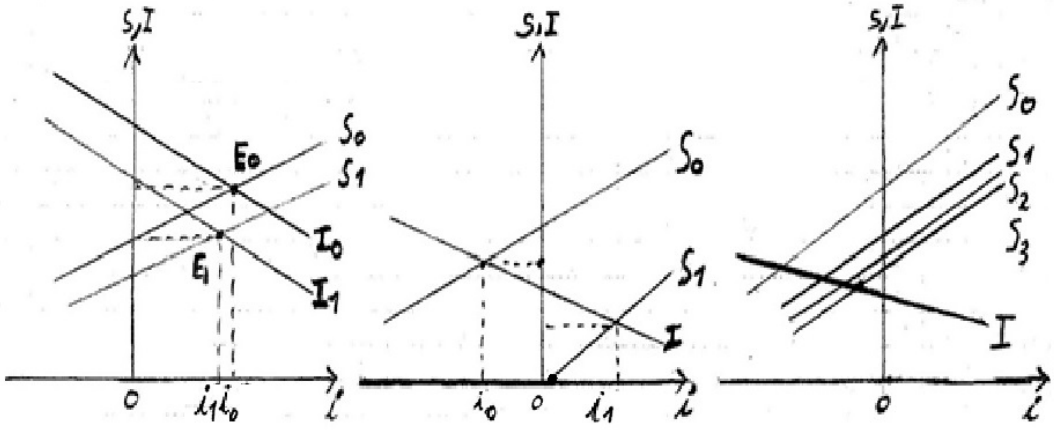
impossible. Pigou suppose une situation où *I* et *S* sont poussés à la limite supérieure, avec forcément *i* très bas. Le plein emploi règne mais l'économie est en déséquilibre car à ce taux d'intérêt, le désir d'épargner reste élevé pour les raisons exposées ci-avant, alors que l'incitation à investir a atteint la nullité. Le retour à l'équilibre entre *I* et *S* implique la réduction de *Y*, ce qui provoque du chômage.

Le salut viendra de la baisse du salaire nominal qui permettra de maintenir l'emploi et suscitera la baisse des prix et du revenu NOMINAL. Mais la spirale infernale entre la baisse des salaires nominaux et celles des prix sera évitée grâce à l'*effet de richesse*. La baisse des prix accroît la valeur des encaisses réelles qui constituent une part du patrimoine. Mieux nantis, les agents n'auront plus besoin d'un taux d'intérêt négatif pour les dissuader d'épargner. Avec plus de consommation, l'équilibre entre *I* et *S* redevient compatible avec le plein emploi.

CRITIQUE DE L'EFFET PIGOU SELON PATINKIN

En 1948, Patinkin critique l'état stationnaire de Pigou dans son article « Price Flexibility and Full Employment ». Pour le mettre en perspective, il remonte au fameux graphique 4.23, base de la théorie néoclassique de l'épargne, de l'investissement et du taux d'intérêt. La figure de gauche du graphique 7.7 montre cet équilibre en inversant l'abscisse et l'ordonnée. On voit les courbes S_0 et I_0 correspondant à un revenu réel Y_0 déterminer un taux d'intérêt i_0. Il y a une paire de courbes S et I pour chaque niveau du revenu ; si le revenu avait valu Y_1 (moins élevé), on aurait eu les courbes S_1 et I_1 avec le taux i_1. L'existence de motifs d'épargne autres que l'intérêt amène un taux d'intérêt négatif sur la figure centrale ; mais le taux devient positif grâce à l'effet de richesse, qui fait prévaloir la nouvelle courbe S_1, où l'épargne s'annule pour un taux d'intérêt juste positif.

Graphique 7.7 : l'épargne, l'investissement et l'effet de richesse



Bien que tous les actifs sont concernés par l'effet de richesse, Patinkin comprend que Pigou se soit surtout préoccupé des encaisses monétaires, car le mécanisme pigovien peut se satisfaire de cette seule sorte de patrimoine.

Patinkin critique assez sévèrement le raisonnement pigovien. Pas tellement l'*effet de richesse* en tant que tel, dont l'*effet d'encaisse réelle*, qui lui est cher, peut être considéré comme un cas particulier, mais l'usage qu'en fait Pigou. Il émet principalement trois reproches :

Reproche n°1 : « The role of real wages in Pigou's system is very ambiguous »[1]. Il est tout à fait clair que l'*effet de richesse* peut fonctionner avec un salaire réel décroissant, constant ou même croissant dans le cas où le niveau des prix baisserait plus que le salaire nominal. Une telle idée est totalement en rupture par rapport à la macroéconomie classique, pour laquelle seule la baisse du salaire réel pouvait accroître l'emploi.

Patinkin estime que l'ambiguïté aurait été réduite si Pigou avait déterminé le niveau des prix dans son système, moyennant une équation et une inconnue supplémentaire

Reproche n°2 : Patinkin souscrit au SENS de la relation pigovienne entre le niveau des prix et l'épargne. Mais l'INTENSITE de cette relation lui semble largement exagérée par Pigou. Il faut envisager l'hypothèse où, dans le cours de sa progression, la baisse des prix épuise son potentiel. « The savings function moves less and less to the right, until eventually it might respond only infinitesimally, no matter how much prices fall »[2].Cette situation est illustrée par la figure de droite du graphique 7.7. La courbe S_3 est la position limite que l'épargne peut atteindre dans sa chute. Patinkin démontre mathématiquement la possibilité d'une telle situation. En fait, cette situation adviendra, si l'agent a atteint un niveau de consommation suffisamment élevé pour que l'utilité marginale de cette consommation tombe en dessous de l'utilité qu'il retire de la thésaurisation. Dans ce cas, l'accroissement des encaisses réelles donnerait lieu à thésaurisation et ne serait pas consommé.

Reproche n°3 : Pigou prend comme hypothèse que l'autorité monétaire maintient le stock monétaire constant pour permettre la baisse des prix et des salaires nominaux. Ce qui ne signifie pas qu'il recommande cette politique ; il s'agit d'une hypothèse de travail. Selon Patinkin, l'autorité qui appliquerait cette politique ferait fausse route. Le risque est celui d'une spirale déflationniste. La baisse des prix peut induire l'anticipation de baisses ultérieures. La consommation serait postposée, la demande globale faiblirait et le chômage s'accroîtrait au lieu de se réduire. La spirale déflationniste serait accélérée par deux phénomènes :
- l'incertitude favoriserait la tendance à thésauriser.
- L'accroissement du fardeau de la dette supporté par les débiteurs causerait des faillites

Une politique économique comptant sur l'effet Pigou pour faire reculer le chômage pourrait se révéler désastreuse. L'expérience montre que cette crainte est fondée : pendant les années trente, la baisse des prix fut sensible au point d'accroître considérablement les encaisses réelles, mais le revenu réel chuta.

MONNAIE INTERNE ET MONNAIE EXTERNE

En 1944, **Kalecki** adressa une autre critique à l'*effet de richesse*. Le stock monétaire peut être décomposé en deux parts :
- la monnaie EXTERNE dont la contrepartie est le stock d'or, les devises ou les créances sur les pouvoirs publics.
- La monnaie INTERNE, dont la contrepartie se compose de titres privés et créances sur le secteur privé[3].

[1] Patinkin [277] p. 186
[2] Patinkin [277] p. 188
[3] La terminologie *monnaie interne* et *monnaie externe* est de Gurley & Shaw (1960).

Selon Kalecki, un accroissement en termes réels de la monnaie interne appauvrit les débiteurs du système bancaire en même temps que s'accroît la richesse des détenteurs d'encaisses. Seule la monnaie externe peut produire un effet de richesse.

Cet argument fut assez largement accepté, au point que dans un article de 1947, Pigou lui-même admit cette restriction. L'acceptation ne fut pourtant pas unanime. Par exemple, Pesek & Saving (1967) estiment que la source interne ou externe de la monnaie importe peu ; quelle qu'en soit l'origine, la monnaie est un bien désiré parce qu'elle est source d'utilité ; à ce titre, elle mérite d'être considérée intégralement comme une richesse. Ils critiquent vertement les économistes comme Gurley & Shaw qui considèrent la monnaie principalement comme une dette.

L'argument de Kalecki peut être étendu à d'autres actifs que la monnaie. Les titres privés ne sont pas une richesse nette dans la mesure où l'existence d'une créance implique celle d'une dette. Concernant l'effet des baisses de prix, Patinkin note : « to every stimulated creditor corresponds a discouraged debtor »[1]. Le problème est extrêmement complexe, car l'effet d'une baisse des prix sur les débiteurs et sur les créditeurs n'est pas nécessairement symétrique[2].

Le capital physique et les biens durables ne voient pas leur valeur réelle augmenter lorsque les prix diminuent, car leur valeur nominale suit normalement l'évolution des prix. En fin de compte, il ne reste que deux actifs sujets à l'effet de richesse : la monnaie externe et la dette publique[3]. Comme nous le verrons dans la section suivante, la possibilité d'un effet de richesse sur la dette publique sera également mise en doute.

Dans son article « Wealth, Saving and the Rate of Interest » paru en 1951, **Metzler** aborde la question sous un tout autre angle. Plutôt que de discuter la réalité de l'effet de richesse, Metzler préfère poser la question : quelles sont les implications pour l'économie politique si on en accepte l'existence ? D'après lui, elles sont considérables. L'effet de richesse représente une rupture par rapport à la tradition classique et néoclassique. « Thus, in salvaging one feature of classical economics- the automatic tendency of the system to approach a state of full employment, Pigou and Haberler have destroyed another feature, namely, the real theory of interest rate »[4].

- La théorie classique du taux d'intérêt est une théorie REELLE, car ce sont des facteurs réels, l'épargne et l'investissement, qui le déterminent.
- La théorie keynésienne du taux d'intérêt est une théorie MONETAIRE, car celui-ci dépend de l'offre et la demande de monnaie.
- Si on l'assortit de l'*effet de richesse*, la théorie classique reste réelle pour les variations de la *monnaie externe* mais devient monétaire pour les variations de la *monnaie interne* : comme chez Keynes, l'*open market policy* est capable de diminuer durablement le taux d'intérêt.

[1] Patinkin [277] p. 182

[2] Si les créanciers sont des ménages et les débiteurs sont des entreprises, l'effet est d'autant plus difficile à analyser que les motivations des comportements diffèrent complètement. A mon avis, l'accroissement de la dette réelle des entreprises n'entraîne pas d'effet direct sur la consommation et l'épargne. Toutefois, l'insolvabilité peut dérégler considérablement l'économie.

[3] En supposant que l' « appauvrissement » de l'Etat n'affecte pas négativement la consommation.

[4] Metzler [257] p. 95

Partons d'un achat de titres par la banque centrale et voyons-en les conséquences chronologiquement. Il y aura d'abord une première phase keynésienne où les liquidités apportées aux agents poussent à la demande de titres et donc font baisser le taux d'intérêt. Mais une deuxième phase suit presque immédiatement qui annule les effets de la précédente. Le niveau des prix hausse sous l'effet de la demande globale dopée par les liquidités ajoutées. Face à ces prix qui haussent, les entrepreneurs ont besoin de plus de capitaux pour continuer à fonctionner comme avant. La demande de capitaux rejoint l'offre qui l'avait précédée dans sa hausse et le taux d'intérêt revient à sont niveau initial.

L'open market policy transfère des titres des détenteurs privés vers la banque centrale et affecte donc le patrimoine mobilier des épargnants en tant que groupe social. Ceux-ci ont cédé une part de leur portefeuille mais les titres qu'ils conservent gagnent de la valeur du fait de la baisse du taux d'intérêt. Inversement, les épargnants ont obtenu des liquidités en échange des titres mais l'inflation grignote la valeur de ces avoirs monétaires.

Metzler suppose que la dévalorisation des encaisses est plus sévère que les gains en capital sur les titres. Les ménages s'appauvrissent donc. Le processus décrit par Metzler entre dans sa troisième phase et c'est ici qu'intervient l'*effet de richesse*. La propension à épargner des ménages appauvris augmente. La hausse de l'épargne qui s'ensuit pousse le taux d'intérêt à la baisse. Contrairement à la baisse constatée dans la première phase, celle-ci est durable ; le nouvel équilibre est stable.

Si la création monétaire avait porté sur la monnaie EXTERNE, le processus se serait arrêté à la deuxième phase. Toutes les grandeurs nominales auraient augmenté dans la même proportion que les prix ; la richesse réelle n'aurait donc pas varié, pas plus que sa répartition entre les titres et les encaisses, et l'effet de richesse n'aurait pas joué.

Metzler tire cette conclusion intéressante : « The power of the banking authorities to alter the equilibrium interest rate is attributable not to their influence upon the nominal quantity of money but to their influence upon the quantity and value of privately held securities »[1].

7.3.2. Les déficits publics

LE FARDEAU DE LA DETTE

Comme on le sait, la théorie keynésienne préconise qu'en cas de récession, les dépenses publiques soutiennent la demande globale. Mais quelle est la meilleure façon de les financer ? Trois méthodes sont envisageables :
- les impôts
- les obligations d'Etat
- la création de monnaie par le Trésor ou la banque centrale

Le budget de l'Etat est en déficit si ses dépenses sont supérieures aux recettes de l'impôt. Un déficit budgétaire suppose que la taxation soit accompagnée par au moins une des deux autres formes de financement. Tout un débat anime depuis longtemps le monde des économistes quant aux mérites respectifs des trois modes de financement ; la polémique porte tout particulièrement sur les vices et les vertus de la dette non monétaire.

Les keynésiens lui vouent une certaine sympathie (très nuancée toutefois). Pour Keynes, le revenu privé qui n'est ni consommé ni investi est un revenu perdu. L'émission d'obligations par

[1] Metzler [257] p. 107

l'Etat pour financer un surcroît de dépenses offre la possibilité d'augmenter le revenu social à partir d'un revenu non consommé. Il a en outre l'avantage de ne pas obérer le revenu disponible pour la consommation.

Dans un article publié en 1943, **Lerner** prône ce qu'il appelle la *finance fonctionnelle* : « The central idea is that government fiscal policy, its spending and taxing, its borrowing and repayment of loans, its issue of new money and its withdrawal of money, shall all be undertaken with an eye only to the results of these actions on the economy and not to any established traditional doctrine about what is sound or unsound (…) In brief, Functional finance rejects completely the traditional doctrines of "sound finance" and the trying to balance the budget over a solar year or any arbitrary period »[1]. La plupart des auteurs keynésiens sont plus modérés que Lerner et ne défendent pas la subordination complète des finances publiques à la politique économique.

Selon Lerner, l'application de la finance fonctionnelle repose sur deux lois:
1- maintenir les dépenses publiques au niveau qui égalise l'offre et la demande globales, au niveau des prix existant. Dépenser plus générerait l'inflation, dépenser moins causerait le chômage ; dépenser juste évite l'un et l'autre[2] ;
2- la taxation ne doit être pratiquée que si conjointement à l'attribution de moyens supplémentaires à l'Etat, il est désirable de réduire ceux des particuliers pour lutter contre l'inflation. La dette non monétaire n'est à conseiller que s'il est souhaitable de substituer les obligations à la monnaie dans le portefeuille des ménages. Dans tous les autres cas, la création monétaire s'impose comme le moyen de financement le moins perturbant.

Même si la finance fonctionnelle ne vise pas l'équilibre budgétaire sur quelque période que ce soit, Lerner estime qu'à long terme, les forces du marché établiront un tel équilibre. Son argumentation en appelle à l'*effet de richesse* : la dette publique causée par un déficit accroît le patrimoine mobilier des ménages et stimule donc leur consommation. Le soutien de la demande globale par le déficit public devient alors moins nécessaire. En finance fonctionnelle, l'Etat commence à dépenser moins quand les ménages commencent à dépenser plus, et inversement. Cette politique ne peut donc pas conduire à des situations extrêmes comme l'accumulation infinie de la dette publique. Lerner reconnaît que menée de façon incompétente, la finance fonctionnelle n'aboutit pas à cet équilibre à long terme.

L'efficacité de la politique budgétaire résulte d'un double effet : celuit de la dépense publique et celui du moyen de financement ; la dépense est à la fois stimulante et inflationniste. Selon Lerner, la dette non monétaire est déflationniste. Elle réduit la liquidité des ménages et fait donc hausser le taux d'intérêt. Lorsqu'il s'agit de relancer l'économie, cet outil doit donc céder la place à la création monétaire. La crainte de l'inflation vaut à celle-ci une mauvaise réputation, ce que déplore Lerner : la finance fonctionnelle ne recourt à des outils inflationnistes que tant que sévit la déflation et elle les abandonne dès que le vent tourne.

Dans un autre article, Lerner estime devoir plaider la cause de la dette publique. Sa réputation, excessivement mauvaise tant dans l'opinion que parmi les économistes, est totalement injustifiée. Lerner critique deux arguments couramment opposés à la dette publique :
1- ***Le cas de la dette publique serait analogue à celui de la dette privée ; la même réticence à s'endetter devrait donc prévaloir***. Lerner admet que cette analogie est valable pour la dette extérieure, celle que l'Etat contracte envers des créanciers étrangers. Mais elle est invalide dans le cas de la dette intérieure, parce que « we owe it to ourselves ».

[1] Lerner [222] p. 39 et p. 41
[2] Est donc sous-jacente l'hypothèse que l'inflation et chômage sont deux dangers exclusifs l'un de l'autre.

2- « A variant of the false analogy is the declaration that ***national debt puts an unfair burden on our children's*** »[1]. Les générations futures, à qui il reviendra de rembourser la dette émise aujourd'hui, la rembourseront … à elles-mêmes. Lerner reconnaît que les poches qui donnent et celles qui reçoivent ne seront pas nécessairement les mêmes, mais la redistribution qui en découlera le laisse étonnamment serein[2].

En 1958, **Buchanan** publie « The Public Principles of Public Debt ». C'est l'anti-Lerner. Oui, l'analogie entre la dette privée et la dette publique est valide aussi bien pour les emprunts à l'intérieur que pour la dette extérieure. Oui, le fardeau de la dette est rejeté sur les générations futures.

Un argument pour nier le report du fardeau prétend que la dépense publique impliquerait le sacrifice CONCOMITANT d'une dépense privée. Buchanan le conteste ; il n'y a pas de sacrifice présent. Les épargnants souscrivent aux emprunts d'Etat, parce qu'ils y trouvent leur avantage ; ils n'y sont pas forcés. Appelons t_0 la période à laquelle l'emprunt public est émis et t_n le moment de son remboursement. Il est simpliste de mettre aux prises la génération *0* et la génération *n* considérées comme deux blocs monolithiques. Buchanan distingue TROIS intervenants :

- l'acquéreur d'obligations d'Etat : il touchera les intérêts et le remboursement et n'a pas de raison d'être mécontent.
- le contribuable de la période t_0 : ce n'est pas lui qui paie la dépense financée par l'emprunt.
- le contribuable de la période t_n : il bénéficie de l'équipement collectif financé par la dette ; il en supporte l'intérêt et le remboursement. Des trois, il est le seul à supporter un fardeau. Certes, ce fardeau aura idéalement une compensation, mais il suffit d'imaginer le cas d'un gaspillage pour apprécier plus clairement qui supporte le fardeau.

[1] Lerner [229] p. 243

[2] A mon avis, le raisonnement de Lerner est parfaitement logique et il disqualifie totalement l'affirmation du fardeau de la dette transmis comme héritage. Par contre, les générations futures héritent bien d'un fardeau : le problème social que représente l'opposition d'intérêt entre les contribuables détenteurs de fonds d'Etat et les contribuables qui en sont dépourvus.

Ricardo et la dette publique

Ricardo défend la thèse de la *neutralité de la dette*, selon laquelle il est indifférent que les dépenses publiques soient financées par l'emprunt ou par l'impôt. Supposons que l'Etat décide l'annulation pure et simple de sa dette ; il commettrait une injustice vis-à-vis des détenteurs de ses obligations. Mais du point de vue économique, cela ne changerait rien. Il y aurait juste une redistribution des fortunes au désavantage des détenteurs d'obligations publiques et à l'avantage du reste de la société.

L'autre argument de Ricardo en faveur de la neutralité de la dette est la solidarité intergénérationnelle. Parlant des taxes futures nécessaires au remboursement des emprunts présents, il écrit : « but if he leaves his fortune to his son, and leaves it charged with this perpetual tax, where is the difference whether he leaves him 20,000 £ with the tax or 19,000 £ without it »[1]. Il est postulé que les citoyens s'abstiennent volontairement de consommer les provisions pour les taxes futures, tout comme ils renoncent forcément à consommer les taxes présentes.

Quelques pages plus loin, Ricardo prend pourtant fait et cause pour la thèse opposée, sans se rendre compte de la contradiction. Il devient hostile à la dette publique. Il lui reproche de donner l'illusion à la nation d'être plus riche qu'elle n'est en réalité, ce qui freine l'esprit d'épargne. Cette argumentation est connue sous l'appellation de l'*illusion de la dette.*

Buchanan conteste frontalement la neutralité de la dette[2]. L'idée qui sous-tend la neutralité ricardienne est que les agents déduisent de l'évaluation de leur patrimoine, la valeur actuelle des paiements futurs. Buchanan rétorque que les agents qui n'ont pas de patrimoine ne réagissent certainement pas ainsi ; ils ne partagent donc pas le fardeau de la génération future.

Buchanan analyse le cas d'une économie en sous-emploi ; il est possible que les dépenses publiques y font travailler des facteurs inoccupés. Dans ce cas, leur coût social est nul. Il n'est pas raisonnable de vouloir les financer par l'emprunt public. Buchanan préconise la création monétaire, car elle évite aux contribuables de supporter les intérêts à payer aux prêteurs.

D'une façon générale, Buchanan affiche un réel scepticisme envers la dette publique. Il termine sa réflexion par la question « When should government borrow ? ». Les cas où le financement par l'emprunt public lui agrée sont rares. Ce mode de financement s'avère le plus convaincant dans le cas des investissements qui s'autofinancent, par exemple une autoroute à péage, la distribution d'eau ou d'électricité… Les revenus tirés par l'Etat lui permettent alors d'assurer le remboursement de la dette et le paiement des intérêts sans le recours à l'imposition.

L'EFFET D'EVICTION

Après le débat sur les principes, la controverse se déplaça vers la question : l'emprunt public, peut-il coexister avec l'investissement privé sans lui porter ombrage ? Sur ce plan, l'argument contre la dette publique (non monétaire) est appelé l'*effet d'éviction* (crowding out). Le mécanisme est le suivant : la concurrence sur le marché de l'épargne entre les émetteurs de titres privés et les émetteurs publics élève le taux d'intérêt, ce qui freine l'investissement privé.

L'un des moments forts de cette controverse est la publication de l'article « Does Fiscal Policy Matter ? » (1973) par les auteurs keynésiens **Blinder et Solow**. L'article expose le problème à l'aide du graphique IS-LM (graphique 7.8). L'équilibre initial est donné par le point E_0.

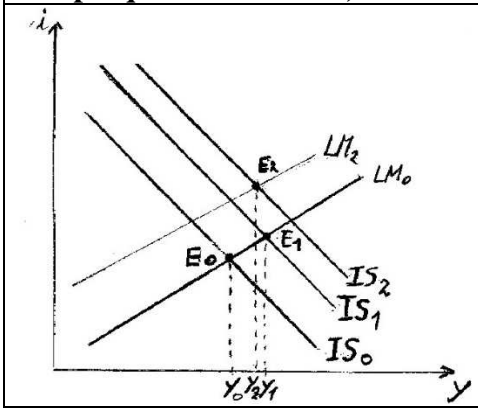
[1] Ricardo [297] p. 539. L'homme laisse à son fils deux héritages : sa fortune et des taxes futures. Que le père ou le fils paie les taxes, la somme des deux héritages reste identique.

[2] Buchanan ne retient de Ricardo que sa première opinion : la *neutralité de la dette.*

L'accroissement de la dépense publique financée par l'émission d'obligations se traduit par le passage de IS_0 à IS_1 plus à droite. On passe à l'équilibre E_1, caractérisé à la fois par un revenu plus élevé ($Y_1 > Y_0$) et un taux d'intérêt plus élevé. Mais E_1 n'est pas un équilibre définitif. L'accroissement du patrimoine mobilier des agents nantis d'obligations d'Etat a lui-même une double répercussion sur l'équilibre IS-LM :

- d'une part, les agents, plus riches, consomment plus, ce qui pousse IS encore plus vers la droite (IS_2). C'est le fameux *effet de richesse*, dont il fut question à la section précédente. La plupart des auteurs keynésiens en ont accepté le principe, bien qu'ils n'adhèrent pas aux conclusions de Pigou.
- d'autre part, la demande de monnaie augmente, car des agents plus nantis désirent une encaisse supérieure. La courbe LM est tirée vers la gauche (LM_2). L'équilibre est déplacé au point E_2, avec un taux d'intérêt supérieur à E_1 et un revenu inférieur.

Graphique 7.8 : Blinder, Solow et l'effet d'éviction



L'accroissement du taux d'intérêt en deux phases révèle l'effet d'éviction. Résumant l'état du débat, Blinder et Solow estiment que le mécanisme de l'éviction n'est « disputed by almost no one » ; les keynésiens ne contestent pas son existence, mais ils considèrent que le *multiplicateur* des dépenses publiques reste positif malgré tout. Blinder et Solow concèdent qu'il est certainement « lower than the naive Keynesian formula » (équation 5.2). Selon eux, la controverse n'est pas tant théorique qu'empirique : quelle est l'ampleur de la hausse du taux d'intérêt et de sa répercussion sur l'investissement privé ?

Comme le remarquent les auteurs, face à la dualité contradictoire des effets sur le revenu, les monétaristes cherchent également à déterminer si la balance penche du côté de l'accroissement ou de la réduction de *Y* ; ils la verraient bien pencher dans le sens d'un multiplicateur négatif. Friedman soutient que la différence essentielle entre les keynésiens et leurs adversaires concerne l'intérêt accordé aux effets de patrimoine comme celui qui poussait ci-avant la courbe LM vers la gauche. Ils auraient généralement tendance à focaliser leur attention sur le « premier round » des événements, négligeant les effets consécutifs en sens inverse.

Par contre, les deux camps s'accordent à considérer le déficit financé par la création monétaire comme nécessairement expansionniste. Non seulement la courbe IS mais également la courbe LM se déplaceraient vers la droite.

Pour démontrer que le déficit budgétaire financé par des emprunts a une forte probabilité d'être expansionniste, Blinder et Solow construisent un modèle basé sur le modèle IS-LM. Des variables et des équations sont ajoutées au système de Hicks pour intégrer les dépenses publiques, leurs trois modes de financement (taxes, emprunt, création monétaire), les effets de patrimoine et de richesse, l'intérêt sur la dette publique…

L'une des équations, une espèce de contrainte budgétaire de l'Etat, dit que le déficit (dépenses *G* moins taxes *T*) égale la création monétaire dM/dt plus la variation de l'encours des obligations dB/dt. Il s'agit donc d'une équation différentielle assurant le caractère dynamique de l'ensemble.

C'est la « long-run steady state solution » qui intéresse Blinder et Solow ; elle implique que dM/dt = dB/dt = 0, ce qui signifie un budget en équilibre. Dans ces conditions, ils obtiennent le multiplicateur à long terme suivant :

$$\frac{dY}{dG} = \frac{1 + \left(1 - T'\right).\frac{dB}{dG}}{T'} \qquad (7.32)$$

où *Y* est le revenu réel
G est la dépense publique
T' est la propension marginale à taxer le revenu
B est le nombre de titres de l'emprunt public[1] (tous identiques dans le modèle)

Ce multiplicateur est positif ; en outre, il apporte une surprise : il est plus élevé quand dG/dB est positif que lorsque cette expression est nulle : « not only is deficit spending financed by bonds expansionary in the long run, it is more expansionary than the same spending financed by the creation of new money »[2]. Ce « paradoxical result » s'explique par le fait que l'intérêt de la dette non monétaire ajoute du déficit au déficit ; la valeur totale des obligations s'en trouve accrue et l'effet de richesse dispose donc d'une assiette plus large. »[3]. Selon Blinder et Solow, l'intervalle de valeurs de ces paramètres où la stabilité est assurée semble correspondre aux caractéristiques de l'économie américaine. La stabilité paraît l'hypothèse la plus plausible, avec dans son sillage l'efficacité de la politique budgétaire sous financement par l'emprunt.

Dans un article publié en 1976, **Buiter et Tobin** remarquent avec regret que le modèle de Blinder et Solow, comme la plupart des modèles qui étudient la dette non monétaire, considèrent celle-ci comme un substitut proche du capital privé plutôt que comme un substitut de la monnaie. Si l'emprunt non monétaire était un substitut de la monnaie, l'émission d'obligations d'Etat n'augmenterait pas la demande de monnaie, mais la diminuerait. Sur le graphique 7.8-A, le passage de LM_0 à LM_2 correspondrait à un glissement de la courbe vers la droite et non vers la gauche. Le déplacement de la courbe LM conforterait celui de la courbe IS vers un revenu plus élevé et un taux d'intérêt plus bas.

BARRO ET L'EQUIVALENCE RICARDIENNE

La réplique à Blinder and Solow vint dans un article publié par **Barro** en 1974 : « Are Government Bonds Net Wealth ? ». L'argument consiste à contester que les obligations d'Etat puissent générer un effet de richesse ; la cause en est que les épargnants décomptent les taxes futures de leur patrimoine, même si ce sont leurs héritiers qui en seront redevables. Déjà rencontrée chez Ricardo (cf. supra), cette idée fut exprimée maintes fois, y compris par des économistes keynésiens. Barro cite cette phrase de Tobin (1952) : « How is it possible that society merely by the device of incurring a debt to itself can deceive itself into believing that it is wealthier ? Do not the additional taxes which are necessary to carry the interest charges

[1] Comme beaucoup d'auteurs, Blinder et Solow intègrent dans leur modèle, l'obligation la plus facile à traiter mathématiquement : une rente perpétuelle de $1 par an.
[2] Blinder & Solow [37] p. 327
[3] Blinder & Solow [37] p. 334

reduce the value of other components of private wealth? »[1]. Barro n'est donc pas l'initiateur de cette théorie mais il est le premier à avoir construit un modèle dont elle est la conclusion, ce qui contribua largement à sa réputation. Voyons plus en détail ce modèle, qui appartient à la famille des *modèles à générations imbriquées*[2].

La bienveillance entre les générations joue un rôle fondamental dans le modèle de Barro. L'héritage est instillé dans l'imbrication entre les générations ; à chaque période, la vieille génération bénéficie d'un supplément de ressources venant de la génération disparue et prend à cœur d'épargner pour léguer un héritage à la génération qui sera âgée la période suivante.

L'économie est stationnaire : ses paramètres contextuels se répètent de période en période. Les hypothèses simplificatrices qui entourent l'émission de la dette publique sont peu réalistes. En période 1, des obligations dont le terme est d'une seule période, sont émises pour un montant *B*. Le modèle choisit d'ignorer les dépenses publiques ; l'impôt finance un transfert à la génération 1 (les vieux de la période 1). Un intérêt (r.B) est payé durant cette même période 1, financé par une taxe particulière frappant les jeunes. A la période 2, l'emprunt est remboursé aux détenteurs d'obligations ; ce remboursement est financé, non par un nouvel emprunt, mais par une taxe particulière frappant les mêmes contribuables entretemps devenus vieux[3].

Les épargnants répartissent leur patrimoine entre le capital privé *K* et la dette publique *B* qu'ils considèrent comme de parfaits substituts ; tous deux rapportent le taux d'intérêt *r*, constant d'une période à l'autre et parfaitement anticipé. Le patrimoine de la génération *i* est noté A^y_i quand cette génération est jeune et A^o_i quand elle est âgée.

Barro adopte le point de vue traditionnel de l'optimisation. L'individu représentatif de la génération *i* tente de maximiser son utilité intertemporelle, dont on peut écrire l'équation :

$$U_i = U_i(c^y_i, c^o_i, U^*_{i+1}) \qquad (7.33)$$

L'utilité est fonction de sa consommation pendant sa jeunesse (c^y_i), de sa consommation pendant sa vieillesse (c^o_i) et de l'utilité optimale de la génération future (U^*_{i+1}), qui traduit sa générosité envers un membre de cette génération. En réalité, l'utilité de la génération qui hérite dépend non de l'héritage brut, mais de ce que Barro appelle l'*héritage net,* les taxes futures étant considérées comme un héritage négatif à déduire. La formule de l'héritage net laissé par la génération 1 à la génération 2 est :

$$(1\text{-}r).A^o_1 - B \qquad (7.34)$$

A partir de l'équation (7.33) et des contraintes budgétaires (non exposées ici), Barro démontre que le problème d'optimisation de la génération 1, ses ressources étant déterminées, peut être ramené à la sélection optimale de l'héritage net, avec la contrainte que l'héritage brut (A^o_1) est non négatif. Cette sélection étant faite, supposons une variation marginale de la dette publique *B*. Elle suscitera la variation de l'épargne A^o_1 juste nécessaire pour maintenir constante l'expression (7.34). Cette correction s'obtient sans variation de la consommation c^o_1 puisque la dette, par hypothèse, finance des transferts à la génération 1. Barro montre que la

[1] Barro [25] p. 1096.

[2] Le premier modèle à générations imbriquées est l'œuvre de Samuelson (1958). A chaque période, l'ancienne génération âgée disparaît, l'ancienne génération jeune devient âgée et apparaît une nouvelle génération jeune.

[3] Comme le constate choi : "There is a more serious mistake committed by Barro. His government gives generation 1 a transfer payment, but he offers no explanation as to where his government gets the money from" ([48] p. 6).

consommation des générations suivantes reste également inchangée. La consommation n'est donc pas stimulée ; il n'y a pas d'*effet de richesse*.

Analysons maintenant l'équilibre du marché des fonds prêtables à la période 1. Considérons la situation avant l'émission des fonds d'Etat : la demande de capital *K* égalise sa productivité marginale avec l'intérêt *r* et celui-ci équilibre la demande de capital *K* avec l'offre qui correspond aux actifs détenus. Intégrons maintenant l'émission d'obligations *B*. On obtient l'équation (7.35), dont le membre de gauche donne la demande de fonds prêtables et le membre de droite en donne l'offre.

$$K(r) + B = A^{o}_{1} + A^{y}_{2} \qquad (7.35)$$

Le membre de gauche augmente de *B* précisément ; Barro montre que c'est également le cas du membre de droite : A^{o}_{1} augmente de B/(1-r) pour maintenir l'héritage net et A^{y}_{2} diminue de B.r/(1-r) à cause de la taxe finançant l'intérêt. L'équilibre entre l'offre et la demande de fonds prêtables est maintenu, sans que le taux d'intérêt doive s'ajuster. La neutralité de la dette exclut donc l'effet d'éviction.

Les mêmes conclusions valent si le modèle s'attache non plus à la dette publique mais à un système de retraites publiques par répartition (financé par l'impôt).

La démonstration s'appuie à la fois sur la volonté et la capacité des membres d'une génération d'éviter à ses héritiers le coût fiscal du remboursement de la dette publique. Il faut donc que les conditions socio-économiques le permettent. Barro examine notamment l'effet de l'impôt sur les successions, l'effet du risque qui pourrait être lié à ce placement.

Au départ, l'effet de richesse était une arme anti-keynésienne. Les keynésiens tentaient d'en limiter la portée. Blinder et Solow l'ont récupéré pour la diriger contre les classiques. C'est maintenant au tour de ceux-ci de la contrer.

Le lecteur ne s'étonnera pas que Buchanan fut l'un des premiers critiques du modèle de Barro. Son article « Barro on the Ricardian Equivalence Theorem » (1976) est à l'origine de l'appellation *équivalence ricardienne* ; c'est sous cette étiquette que le modèle de Barro est devenu l'une des théories les plus célèbres de la macroéconomie moderne. Buchanan reproche à Barro :

- la manière artificielle dont les finances publiques sont introduites dans le modèle
- d'avoir regardé la dette publique par le petit bout de la lorgnette, se focalisant sur cette question d'héritage, somme toute secondaire par rapport à la baisse de la consommation ou de l'investissement privés nécessaires selon lui pour réunir les fonds prêtés à l'Etat. Buchanan croit fermement en l'effet d'éviction.

7.3.3. La courbe de Phillips et le taux de chômage naturel

ORIGINE ET DEVELOPPEMENT DE LA COURBE DE PHILLIPS

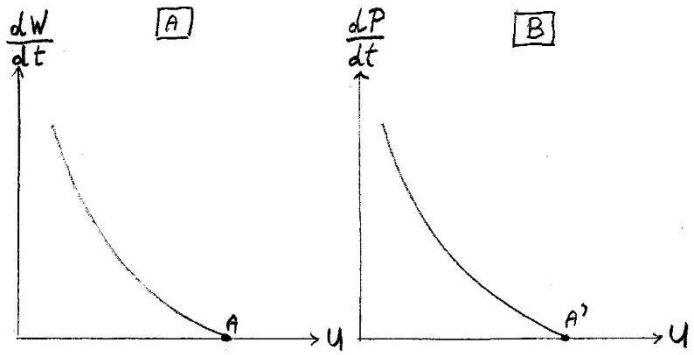
Le graphique 7.9-B nous montre une courbe où l'on voit le taux d'inflation augmenter quand le chômage décroît et vice versa, ce qui suggère que la politique économique du gouvernement jouirait d'une possibilité d'arbitrage entre ces deux maux. La question de savoir si un tel arbitrage existe bel et bien donna lieu à l'une des batailles les plus longues et les plus acharnées de la macroéconomie moderne.

Paradoxalement, à l'origine de cette controverse, on trouve une inoffensive étude statistique sans prétention théorique : l'article « The Relation between Unemployment and the Rate of

Change of Money Wage Rates in the United Kingdom, 1861-1957 », publié en 1958 par Alban Phillips. L'étude se contente d'établir la régression entre les variations du salaire nominal et le taux de chômage durant la période concernée. La conclusion tient dans ces deux constatations :

- Les données empiriques confortent l'hypothèse que le taux de chômage et les variations de celui-ci expliquent, pour une bonne part, les variations du salaire nominal, à l'exception des années où le prix des biens importés a augmenté sensiblement.
- En supposant une hausse annuelle de la productivité de 2% par an, le niveau de la demande globale qui garantit la stabilité des prix des produits implique un taux de chômage légèrement inférieur à 2,5%. Et pour stabiliser le salaire nominal, le taux de chômage devrait être d'environ 5,5%[1].

Graphique 7.9 : la courbe de Phillips



Trouvant cette étude intéressante mais insuffisamment rigoureuse, l'économiste keynésien Richard Lipsey (1960) reprend le travail de régression en l'affinant et l'assortit d'une réflexion théorique expliquant la relation constatée. L'équation de la corrélation devient :

$$dW/dt = a + b.(1/U) + c.(1/U^2) + d.(dU/dt) + e.(dP/dt) \qquad (7.36)$$

- dW/dt est la variation annuelle du salaire nominal
- a, b, c, d et e sont les coefficients dont la régression a déterminé les valeurs
- U est le taux de chômage et dU/dt est la variation annuelle du taux de chômage
- dP/dt est le taux d'inflation

Pour la période 1862-1923, Lipsey obtient une excellente corrélation (inverse) entre dW/dt et *U*. Cette corrélation se dégrade dans la période 1919-1957, alors que la corrélation entre dW/dt et dP/dt se renforce.

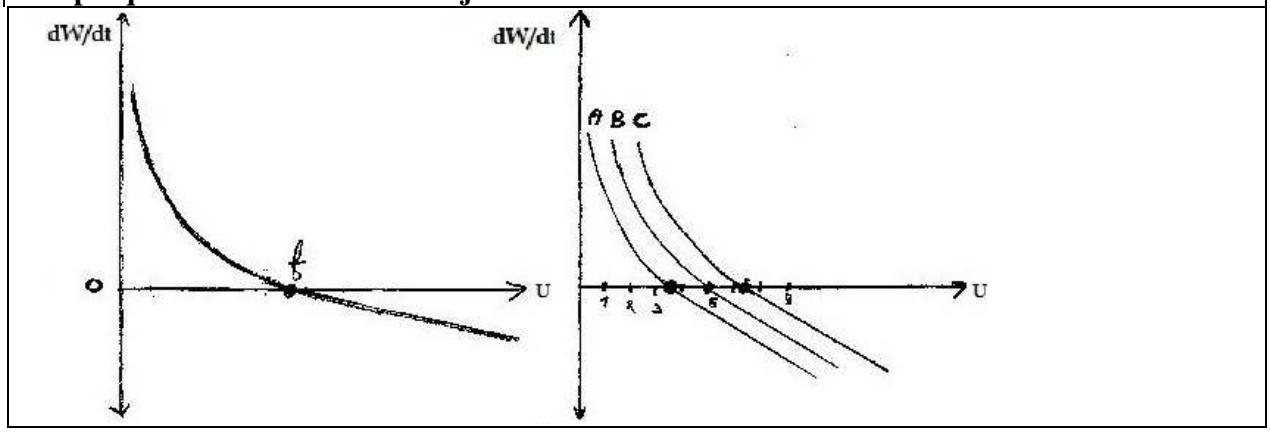
Phillips pensait que la courbe 7.9-A, qu'il avait mise en évidence, reflétait simplement la manifestation de la loi de l'offre et la demande sur le marché du travail. Mais si l'offre de travail est supérieure à la demande, son prix ne devrait normalement pas augmenter ; or la courbe de Phillips révèle une inflation salariale positive pour toute une gamme de taux de chômage ; il faut pouvoir le justifier.

A cette fin, Lipsey désagrège le marché du travail selon les secteurs d'activité et les spécialisations demandées. Pour chaque segment de marché, on peut tracer une courbe *d'ajustement* (graphique 7.10-A), montrant la variation du salaire induite par la demande ou l'offre excédentaires. Au point *f*, l'offre et la demande de travail s'équilibrent ; le chômage se

[1] Ces pourcentages sont évidemment datés ; ce qui nous intéresse, c'est le raisonnement qui est derrière ces chiffres.

résume donc au *chômage frictionnel*[1]. A gauche de *f*, il y a demande excédentaire ; à droite, il y a offre excédentaire. Le salaire nominal est donc constant en *f*, croissant à gauche, décroissant à droite. Ici, Lipsey pose une hypothèse importante : à gauche de *f*, la courbe est convexe, car réduire le chômage frictionnel nécessite une hausse salariale qui va croissant. A droite de *f*, la baisse du salaire entraînée par l'offre excédentaire de travail est linéaire.

Graphique 7.10 : les courbes d'ajustement du salaire à la demande excédentaire



1- Pourquoi une inflation salariale positive tout au long de la courbe de Phillips ?
Même lorsque l'économie globale est en proie au chômage, certains segments connaissent une demande de travail excédentaire. Comme les courbes d'ajustement sont convexes à gauche et linéaires à droite, le salaire augmente plus sur les marchés où la demande est excédentaire qu'il ne baisse là où l'offre est excédentaire. Imaginons une économie I où le marché du travail est segmenté en deux secteurs décrits par les courbes d'ajustement *A* et *C* du graphique 7.10. Une économie II, comparable à la première, a ses deux segments décrits par une unique courbe *B* située à mi-chemin (horizontalement) entre *A* et *C*. Le lecteur peut vérifier que la moyenne des courbes *A* et *B* donne une courbe surélevée par rapport à la courbe *B*. La variété des situations spécifiques est donc un facteur de hausse des salaires.

Comme la variation du salaire annule à terme l'offre ou la demande excédentaires qui l'ont générée, le maintien de la courbe de Phillips dans le temps implique que de nouveaux déséquilibres de ce type prennent constamment le relais de ceux qui se résolvent, ce qui n'est pas une hypothèse déraisonnable.

2- Pourquoi la courbe de Phillips, est-elle décroissante ?
La moyenne des courbes sectorielles décroissantes doit elle-même être décroissante.

3- Pourquoi la courbe est-elle convexe par rapport à l'origine ?
Si sur un intervalle de l'abscisse, une seule des nombreuses courbes d'ajustement segmentaires est convexe, la moyenne de ces courbes sera également convexe. On peut imaginer que la portion la plus à droite de la courbe de Phillips soit rectiligne si toutes les courbes segmentaires le sont sur cet intervalle.

Dans leur article « Analytical Aspects of Anti-Inflation Policy », Samuelson et Solow rassemblent les données empiriques de la relation entre le chômage et la variation du salaire nominal aux Etats-Unis ; ils calculent la régression, dont ils comparent le résultat avec celui de Phillips. La corrélation entre dW/dt et *U* reste bonne, mais à la condition d'appliquer la

[1] Il y a égalité entre les demandes d'emploi et les postes de travail vacants.

régression successivement sur plusieurs intervalles de temps plus courts. Sur l'ensemble de la période, la relation entre dW/dt et U ne s'avère pas très stable. Nos auteurs constatent aussi que le taux de chômage associé aux différentes valeurs de dW/dt semble plus élevé qu'au Royaume Uni ; ils l'expliquent par la moindre perfection du marché du travail américain, que l'étendue du pays rend plus hétérogène : des poches de demande excédentaire de travail s'y développent plus facilement ; le salaire nominal ne devient stable qu'avec un taux de chômage plus élevé.

Tout à la fin de l'article, les auteurs « translate the Phillips' diagram showing the American pattern of wage increase against degree of unemployment into a relation diagram showing the different levels of unemployment that would be "needed" for each degree of price level change »[1]. Voici enfin dessinée la fameuse *courbe de Phillips* du graphique 7.9-B qui animera tant les débats. Pour passer de l'une à l'autre, il suffit de retrancher l'accroissement de la productivité de la variation du salaire nominal.

L'intersection de cette courbe avec l'abscisse (point *A'*) donne le taux de chômage correspondant à la stabilité des prix, estimé à 5,5% pour les Etats-Unis. L'objectif d'un chômage limité à seulement 3% devrait être « payé » par une inflation de 4 à 5%. « … much price rise would seem to be the necessary cost of high employment and production in the years immediately ahead »[2]. Voici posée la fameuse théorie de l'arbitrage (trade off) entre l'inflation et le chômage.

LA CRITIQUE DE PHELPS ET FRIEDMAN

Les premières flèches furent décochées par Edmund **Phelps**, dans son article « Phillips Curves, Expectations of Inflation and Optimal Unemployment Over Time » (1967). Il y discute la politique optimale de contrôle de la demande globale par l'Etat. A cette fin, il construit une fonction d'utilité sociale dépendant des deux variables que sont le revenu national et l'inflation. L'utilité sociale marginale :

- augmente avec le revenu mais de façon décroissante parce que l'utilité marginale de la consommation décroît, alors que la désutilité marginale du travail ne décroît pas.
- diminue avec l'inflation, car en faisant augmenter le taux d'intérêt nominal, celle-ci suscite une fuite devant la liquidité, qui est un facteur de bien être.

La fonction *U* donne l'utilité instantanée au temps *t* en tant qu'elle est dépendante des combinaisons du revenu national *y* et de l'inflation π. Mais ce n'est pas l'utilité instantanée qui importe, c'est l'utilité intertemporelle ; on l'obtient par l'intégrale de la fonction *U* définie sur l'intervalle de temps allant de maintenant (t = 0) à l'infini. Selon que oui ou non, la société déprécie l'utilité instantanée future par rapport à celle du présent, il faudra maximiser la première ou la seconde de ces deux équations :

$$W = \int_0^\infty e^{-\lambda t}.U(y,\pi)\ dt \quad \text{avec } \lambda \geq 0 \qquad (7.37\text{-A})$$

$$W = \int_0^\infty U(y,\pi)\ dt \qquad (7.37\text{-B})$$ [3]

La courbe de Phillips joue un rôle essentiel dans le modèle, car elle associe les différentes combinaisons entre l'inflation et le revenu national (lié inversement au chômage) qui s'offrent à l'arbitrage de l'Etat à chaque instant *t*.

---

[1] Samuelson & Solow [324] pp. 191-192

[2] Samuelson & Solow [324] p. 192

[3] Lorsque le futur n'est pas escompté, la méthode est un peu plus compliquée que cette formule, car il faut éviter d'obtenir une utilité intertemporelle infinie. Je n'entre pas dans ces détails.

Phelps introduit un élément qui vient perturber le mécanisme keynésien du trade off et, par là, troubler le résultat de la politique économique : les anticipations des agents relatives à l'inflation. « It is reasonable to suppose that the participants in product and labour markets will learn to expect inflation (and the concomitant money wage trend) and that, as a consequence of their rational anticipatory behaviour, the Phillips curve will gradually shift upward (in uniform vertical displacement) by the full amount of the newly expected and previously actual rate of inflation »[1]. Phelps insiste sur le fait que la montée de la courbe sur le graphique se fait « one to one », par rapport à l'inflation anticipée. Tant que dure l'inflation et qu'elle est correctement anticipée, l'élévation de la courbe de Phillips perdure. Le taux d'inflation associé à un taux de chômage donné est donc appelé à s'élever.

Phelps démontre que les comportements rationnels aboutissent à l'existence d'un taux de chômage d'équilibre u* qui est le seul stable à long terme. Si on s'en écarte (dans le sens d'un abaissement), l'anticipation de l'inflation élève la courbe de Phillips de telle façon qu'on revient à u* : « the only steady state Phillips curve is a vertical line intersecting the horizontal axis at u* »[2]. Le *trade off* n'est que temporaire. Une politique inflationniste permet un accroissement transitoire de l'emploi et du revenu dans le PRESENT, au prix d'une inflation et de taux d'intérêt nominaux toujours plus élevés persistant dans le FUTUR. Ce type de politique ne sera donc préférée (équations 7.37) que si le taux d'escompte du futur est assez élevé.

Dans son modèle, Phelps doit supposer un mode d'anticipation de l'inflation. Comme l'indique la fin de l'avant-dernière citation ci-dessus, il considère que l'inflation réelle d'aujourd'hui est l'inflation anticipée pour demain. « The expected inflation rate adjusts gradually to the actual inflation rate »[3]. L'inflation anticipée varie proportionnellement à la différence entre le taux d'inflation réel et le taux d'inflation anticipé. Tel est le mécanisme de correction automatique des anticipations.

Les mêmes idées seront formulées par **Friedman** à de multiples reprises à partir de 1968 quand parut « The Role of Monetary Policy ». Voici comment Friedman décrit l'enchaînement des événements faisant suite à une action de relance par l'Etat et qui conduisent à l'échec final.

Partons d'une situation stable où aucun changement n'est anticipé. De façon inattendue, la demande globale connaît une stimulation. L'augmentation du prix qui lui est associée, fait croire aux entrepreneurs que leur produit est plus prisé, ce qui les incite à produire plus et donc à embaucher ; ils attirent la main d'œuvre supplémentaire en augmentant le salaire nominal proposé. Les travailleurs (actifs ou à la recherche d'un emploi), prenant l'augmentation du salaire nominal pour une augmentation du salaire réel, répondent à cette demande supplémentaire et travaillent donc plus[4]. Mais cette situation est transitoire. Après quelques temps, les entrepreneurs et les travailleurs se rendent compte de leur méprise : les prix relatifs n'ont pas varié, pas plus que le salaire réel, qui a plutôt baissé, car les prix auront probablement plus augmenté que les salaires. Le salaire réel ne couvre plus la désutilité du travail. Les agents adaptent leur offre en conséquence. En effet, ce sont les grandeurs réelles qui les intéressent et non les grandeurs nominales ; ne pas l'avoir compris est l'erreur la plus importante des partisans de la courbe de Phillips.

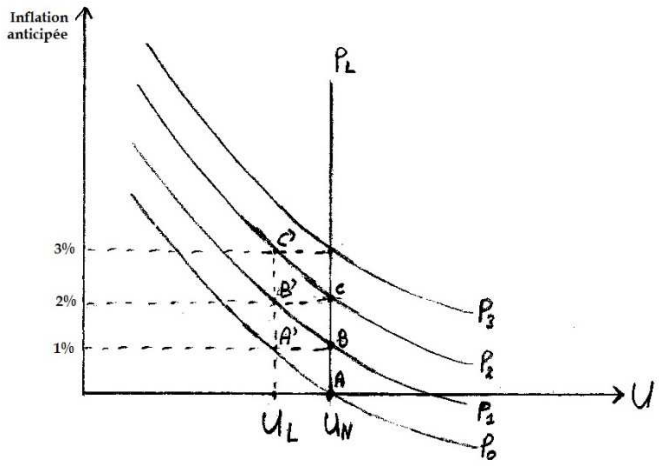
---

[1] Phelps [281] p. 255
[2] Phelps [281] p. 256
[3] Phelps [281] p. 254
[4] Selon Friedman, le *salaire de réserve* (cf. sous-chapitre 4.3.4) est un salaire réel et non un salaire nominal.

Après l'embellie temporaire, l'emploi et le revenu reviennent à leur niveau initial. Pour empêcher ce retour, l'Etat doit produire une inflation supérieure à celle qui est anticipée de façon à leurrer à nouveau les agents qui sont maintenant sur leur garde.

C'est ce qu'indique le graphique 7.11. Il y a une courbe de Phillips à court terme par taux d'inflation anticipé. On part du point A sur la courbe P_0, correspondant à une inflation anticipée nulle. La relance de la demande globale fait glisser l'économie le long de la courbe vers la gauche. Lorsqu'une inflation de 1% est anticipée, l'économie revient au taux de chômage naturel sur la courbe P_1 (point B). Une nouvelle relance de la demande globale porte l'économie au point B' et ainsi de suite. On assiste à une *accélération de l'inflation*. Le zigzag ramène toujours au *taux de chômage naturel* U_N, base de la droite verticale qui constitue la courbe de Phillips à long terme (P_L).

Graphique 7.11 : la courbe de Phillips à court et à long terme



Voyons de plus près *le taux de chômage naturel* brièvement évoqué au sous-chapitre 7.2.5 et qu'on peut rapprocher du taux u* de Phelps. Selon son auteur, il s'agit du taux résultant de l'équilibre général walrassien, aménagé pour tenir compte des imperfections des marchés du travail et des produits (coût de la mobilité, information imparfaite…). Friedman reconnaît que la science économique ne dispose pas encore du moyen de l'estimer avec précision.

Trois aphorismes de Friedman résument bien sa théorie :
- « There is always a temporary trade-off between inflation and unemployment; there is no permanent trade off » [1].
- « A rising rate of inflation may reduce unemployment, a high rate will not ».
- « Only surprise matters » [2].

A partir de 1974, le monde connut une période de *stagflation*, c'est-à-dire la conjonction de l'inflation et du sous-emploi. Considérant cette situation, Friedman estime qu'elle est caractérisée par une courbe de Phillips croissante au lieu de la droite verticale qui est la norme selon lui. Comment explique-t-il cette anomalie ? Le système des prix, moteur de l'équilibre économique, est brouillé par l'inflation, conduisant à des situations infra optimales, notamment le chômage.

[1] Friedman [101] p. 11
[2] Friedman [106] p. 456

LA DEFENSE DE TOBIN

Dans un article publié en 1972, Tobin défend la courbe de Phillips face aux critiques de Phelps et Friedman. Leurs arguments- constate-t-il- impliquent que le chômage aux alentours du taux naturel soit volontaire. Selon ces auteurs, l'accroissement du volume de travail correspondant aux déplacement A-A', B-B' et C-C' sur le graphique 7.10 implique une perte de bien-être, vers laquelle les agents ne tendent que sous l'effet d'une illusion. Tobin estime au contraire qu'à un taux de chômage de 5 à 6% (marge où devrait se situer le taux naturel), l'économie est parfaitement capable créer des emplois dont la productivité marginale atteint le seuil qui les rend plus attractifs que le loisir. Les chômeurs de Friedman sont des gens qui s'abstiennent de saisir les emplois proposés aujourd'hui, mus par l'espoir que les salaires proposés demain seront plus élevés. Tobin doute que le chômage soit une position favorable pour dénicher les emplois de qualité ; un travailleur actif peut changer d'emploi sans trop de difficulté.

Dans l'optique keynésienne, est du *chômage involontaire* celui qui peut être épongé par l'accroissement de la demande globale. Les néoclassiques considèrent la rigidité à la baisse du salaire nominal comme une preuve que le chômage est volontaire. Tobin rappelle la théorie de Keynes selon laquelle il existe un motif rationnel expliquant que les travailleurs acceptent une baisse du salaire réel par l'inflation alors qu'ils rejettent une baisse du salaire nominal : la crainte d'une déstructuration de l'échelle des salaires (cf. sous-chapitre 5.1.1). L'argument contre la courbe de Phillips selon lequel les agents ne se préoccupent que du salaire réel n'est donc pas acceptable. La courbe de Phillips est, selon Tobin, la digne héritière des chapitres de la *Théorie Générale* consacrés à l'emploi et au salaire.

Tobin critique également le concept du *taux de chômage naturel*. Friedman le fondait sur l'équilibre général walrassien adapté pour intégrer les imperfections du marché. Que savent concrètement les économistes sur l'existence d'un tel équilibre, que savent-ils sur son optimalité ?

La théorie de l'*accélération de l'inflation* est excessivement alarmiste. Le discours dominant chez les économistes identifie l'inflation à une catastrophe majeure. Tobin estime au contraire qu'elle est un mal bien léger en comparaison avec le chômage. Le manque de liquidité mis en avant par Phelps paraît maigre pour soutenir la vision apocalyptique de l'inflation. Il est exact que l'inflation lèse les détenteurs d'actifs à revenu fixe. Pour résoudre ce problème, Tobin préconise que le gouvernement émette ses bons de trésor avec une clause d'indexation de leur valeur sur la hausse des prix.

*

Les quelques articles résumés ci-dessus ne constituent que le début de la controverse, dont la suite sera évoquée dans les sous-chapitres consacrés aux *nouveaux classiques* et aux *nouveaux keynésiens*. Les monétaristes et leurs héritiers feront de la courbe de Phillips et de sa critique le centre de gravité de la macroéconomie. Avant d'explorer ce sujet plus avant, arrêtons-nous quelques instants sur la problématique de l'inflation.

L'INFLATION VUE PAR LES KEYNESIENS

Contrairement aux monétaristes, les keynésiens ont plutôt pour vocation de combattre le chômage que l'inflation. Ils ont toutefois consacré quelques réflexions à cet objet. Le premier à s'y être intéressé n'est autre que **Keynes** lui-même. En 1940, il lui consacra un ouvrage de

circonstance intitulé « How to Pay for the War ». On y trouve le premier exposé de ce qui sera plus tard appelé l'*écart inflationniste*[1].

Keynes constate que le contexte est en quelque sorte l'opposé de celui qui prévalait lors de la crise des années trente, dont on sait qu'elle fut le terreau de la *Théorie générale*. « We have been accustomed to a level of production which has been below capacity. In such circumstances, if we have more to spend, more will be produced and there will be more to buy… But in war time, the size of the cake is fixed. If we work harder, we can fight better. But we must not consume more »[2].

L'effort de guerre pousse l'économie vers le plein emploi et même plus (travail des femmes, heures supplémentaires…), augmentant largement les revenus individuels disponibles pour la consommation. Mais il détourne parallèlement une part importante de la production vers un canal qui ne lui permet pas de rencontrer cette demande accrue.

Sans réaction du gouvernement, le public dépensera plus d'argent mais ne consommera pas plus de biens, « for prices will rise just enough for the money we spend to be used up by the increased cost of what there is to buy »[3]. L'inflation est alors inévitable, mais elle heurte le sens de la justice de Keynes, car elle avantage ceux qui vivent de la vente des biens par rapport aux catégories qui perçoivent un revenu fixe.

Comme réponse politique à cette menace, Keynes préconise l'EPARGNE FORCEE. Une partie du revenu serait obligatoirement déposée sur des comptes bloqués (portant intérêt) que les citoyens ouvriraient dans une institution financière de leur choix. Les sommes déposées financeraient l'effort de guerre ; elles seraient débloquées à la fin de la guerre, idéalement lors de la première récession qui suivrait la guerre, servant alors à la relance de la consommation nécessaire à la sortie de crise.

Ce mécanisme permettrait aux salariés de figurer parmi les créanciers de l'Etat, de garder le fruit de leur travail et du consommer après la guerre. L'inflation non contrariée aurait de toute façon imposé une épargne globale équivalente, puisque l'épargne vaut le revenu moins la consommation et que ces deux grandeurs sont déterminées par ailleurs ; la différence est que, dans ce dernier cas, seule la classe capitaliste aurait épargné, grâce à ses bénéfices accrus. Les syndicats commettraient donc une lourde erreur en exigeant le paiement total et immédiat des salaires. Comme avec le code de la route, le public peut bénéficier des contraintes qu'il s'impose.

Keynes pointe les insuffisances des solutions alternatives :
- le financement de l'effort de guerre par les **taxes** réduirait l'épargne volontaire, alors que celle-ci constitue un complément indispensable ;
- le **contrôle des prix** produirait des pénuries et l'ampleur des subsides qu'il implique mettrait à mal le budget de l'Etat ;
- l'**épargne volontaire** va dans le bon sens et doit être encouragée, mais seule l'épargne obligatoire garantit que la classe salariée ne sera pas mise hors jeu.

Selon Keynes, c'est l'excès de demande globale qui crée l'inflation. En 1949, dans « The Inflationary Process: Some Theoretical Aspects », **Lerner** répète cette conception.

[1] Cet écart fait référence au surplus de la demande globale par rapport à l'offre globale, lorsque l'outil de production tourne à sa capacité maximum.
[2] Keynes [185] p. 4
[3] Keynes [185] p. 5

Il écrit : « This is the definition of inflation I wish to use- an excess of demand over supply »[1]. L'inflation est le contraire de la déflation (caractérisée notamment par le chômage). Dans cette optique, la hausse des prix est la CONSEQUENCE de l'inflation, au même titre qu'une baisse de prix peut résulter de la déflation. Cet effet sur les prix n'est même pas inévitable, puisque des forces politiques ou économiques peuvent le contrecarrer [2].

A la question « why is it bad ? », Lerner répond que le mal ne réside pas dans la hausse des prix en soi, mais dans les perturbations résultant des erreurs d'anticipation. Les choix des agents ne leur apportent dès lors pas le résultat escompté. Lerner pose le même constat que Friedman et, comme lui, déplore l'injustice subie par certaines catégories comme les créanciers. « Inflation is bad because it sabotages one of the most valuable instruments of modern society- the price mechanism »[3].

Pour vaincre l'inflation, il faut réduire la demande excédentaire par les impôts, par la modération des dépenses publiques et par l'encouragement à l'épargne et aux importations[4]. Lerner déconseille les mesures de restriction monétaire. La dépense *MV* peut très bien augmenter à cause de la vitesse de circulation *V*[5]. En outre, la réduction de *M* pourrait accroître la préférence pour la liquidité au point de faire entrer l'économie dans une phase de déflation. Lerner pense aussi que moyennant un effort pédagogique, le public est capable de comprendre qu'en période d'inflation, les augmentations NOMINALES ne se concrétisent pas en augmentations REELLES.

Cette thèse de l'*inflation tirée par la demande* n'est pas la seule qu'avancèrent les keynésiens. Il y eut également une explication concurrente : l'*inflation poussée par les coûts*, qui repose sur l'entraînement mutuel cumulatif entre les revenus (principalement salariaux) et les prix. L'opposition entre les deux explications n'est pas aussi simple qu'il y paraît. La spirale salaires-prix intervient également dans l'*inflation tirée par la demande* mais pas en tant que cause première. C'est ainsi que Lerner écrit dans l'article précité: « the cumulative spiralling rise in prices and incomes is merely a result of the attempt to overcome the lag in expectations »[6]. Parallèlement, la demande globale en excès joue un rôle de transmission dans le processus de l'*inflation poussée par les coûts*.

Deux ans plus tard, dans « Fighting Inflation », Lerner passe la frontière et développe résolument la deuxième explication. Chaque individu, quelle que soit sa catégorie sociale, contribue involontairement à l'inflation : quand il voit les prix et les autres revenus augmenter, il souhaite avant tout éviter d'être le dindon de la farce. Lerner compare la situation inflationniste avec celle d'une panique générale. L'*inflation par les coûts* a comme terreau un haut niveau d'emploi. « The very high level of employment raises the bargaining power of labor to a point where wage increases are easily obtained and passed on in higher prices of the product »[7]. Lerner plaide pour une réorganisation du marché du travail, qui permettrait au salaire de varier sur un intervalle donné lorsque le rapport entre l'offre et la demande le justifie,

---

[1] Lerner [224] p. 194
[2] On ne peut pas dire que cette conception terminologique se soit imposée.
[3] Lerner [224] p. 196.
[4] La décision de taxer doit toutefois prendre en considération le risque de décourager la production.
[5] Les keynésiens ne partagent pas la vision monétariste de la constance de *V*.
[6] Lerner [224] p. 195.
[7] Lerner [225] p. 195.

tout en maintenant le niveau général des salaires solidement arrimé au niveau général de productivité.

Un exposé notable de l'*inflation poussée par les coûts* fut fourni par **Kaldor** en 1959. Selon lui, les économistes s'accordent sur l'existence d'un lien entre l'inflation et l'augmentation des salaires nominaux. Sans cette interaction, une augmentation des prix en une fois non seulement rétablirait l'équilibre du marché, mais effacerait les causes de la persistance du déséquilibre. Mais, note-t-il, la part des salaires dans le revenu national n'est pas affectée par cette augmentation des salaires nominaux : « when money wages rise at a faster rate, money profits will also rise at a faster rate »[1]. L'unanimité des économistes disparaît lorsqu'il est question de la cause de l'augmentation des salaires :

- les uns l'attribuent à la demande des employeurs sur le marché du travail, qui se disputent la main d'œuvre disponible trop rare (demand-pull).
- D'autres la font résulter de la pression des syndicats lors des négociations collectives (cost-push).

Kaldor dit croire à la deuxième cause. Selon lui, le facteur essentiel est « the bargaining strength of labour ». Or celle-ci fluctue avec le cycle économique et atteint son maximum pendant les périodes de prospérité. La raison en est que les employeurs sont beaucoup plus pénalisés par les grèves en période d'intense activité que lorsque celle-ci est faible ; les travailleurs font donc triompher leurs revendications plus facilement lorsque règne le plein emploi. Cette raison est plus puissante qu'une autre pourtant plus souvent citée : qu'en basse conjoncture, les travailleurs se disputent les emplois disponibles et sont donc plus enclins aux concessions.

Selon Kaldor, l'augmentation des prix n'est que le produit INDIRECT des hausses salariales, par le biais de l'accroissement de la demande de produits et services par les salariés. En effet, s'il suffisait aux entreprises de répercuter les hausses salariales sur leur prix de vente, de façon quasi-automatique, elles n'auraient pas de raisons de s'opposer à ces hausses.

La bonne conjoncture économique fait gonfler les profits. Ce sont les profits accrus qui jouent le rôle déclencheur des revendications salariales, bien plus que la hausse du coût de la vie. Les salariés revendiquent de participer au partage du revenu nouvellement créé. En retour, les hausses salariales, en accroissant la demande de biens, favorisent la profitabilité des entreprises. « It is therefore more correct to speak of a profit-wage spiral than of a price-wage spiral »[2].

La distinction entre les processus inflationnistes *cost-push* et *demand-pull* est au centre de l'article de **Samuelson et Solow** examiné ci-avant. Les auteurs mettent en garde contre certaines conclusions simplistes.

1- le cost-push se caractériserait par une hausse des salaires nominaux plus élevée que celle de la productivité alors que le demand-pull se distinguerait par un accroissement de la dépense supérieur à celui de la production réelle. En réalité, les deux phénomènes sont présents dans chacun des deux processus.

2- Une augmentation des salaires antérieure à celle des prix révèlerait un processus cost-push. Dans le cas contraire, on aurait affaire à un processus demand-pull. De nombreux arguments peuvent être opposés à cette assertion :
   - Déterminer lequel est antérieur suppose que soit fixé un début de la période d'observation ; or celui-ci est nécessairement arbitraire.
   - Avec les anticipations, l'effet peut précéder la cause

[1] Kaldor [175] p. 190
[2] Kaldor [175] p. 194

- Supposons que l'offre globale et la demande globale s'équilibrent mais qu'existent de fortes variations sectorielles. La cause de l'inflation ne résidera pas dans les secteurs à forte demande qui augmentent leur prix mais dans les secteurs à faible demande qui ne diminuent pas le leur.
- En microéconomie, l'augmentation des coûts entraîne une baisse de la production. Empiriquement, les périodes d'inflation ne coïncident de loin pas avec une telle baisse. Peut-être est-il erroné de considérer l'offre et la demande comme des grandeurs indépendantes.

7.4- LES NOUVEAUX CLASSIQUES

Le débat sur la courbe de Phillips vit progressivement une nouvelle école prendre l'ascendant. Bien que les critiques de Friedman contre la courbe de Phillips furent un succès, la vieille garde monétariste fut supplantée par les *Nouveaux Classiques*. La nouvelle école reprit les thèmes monétaristes en les articulant autour d'une nouvelle hypothèse : les *anticipations rationnelles*. Les principaux économistes de la nouvelle école sont John Muth (le précurseur), Robert Lucas (le leader), Thomas Sargent, Neil Wallace, Robert Barro (déjà connu du lecteur), Edward Prescott et Finn Kydland.

Dans la vision monétariste, les fluctuations monétaires pouvaient affecter l'économie réelle à court terme. La nouvelle école représente un retour à la thèse classique de la neutralité monétaire. Certes, elle reconnaît un effet réel aux variations monétaires non anticipées, de même que Friedman et Phelps. Mais, précisément, en prêtant aux agents des anticipations rationnelles, les nouveaux classiques rendent ce scénario moins probable (mais pas impossible). Comme l'écrivent Sargent et Wallace: « By the virtue of the assumption that expectations are rational, there is no feedback rule that the authority can employ and expect to be able systematically to fool the public. This means that the authority cannot expect to exploit the Phillips curve even for one period »[1]. Même à court terme, il n'y a plus d'arbitrage entre le chômage et l'inflation. La vision s'est donc radicalisée. Décrivant la nouvelle école, Holland écrit : « It also contrasts with standard monetarist analysis, in which money is neutral in the long run, but has expansionary short-run effects »[2].

A cette différence près, la nouvelle théorie est très proche du monétarisme, dont elle s'est largement inspirée. Elles ont notamment en commun :

- « The assumption of equilibrium. In other words, supply is assumed to equal demand in all markets at all times. This is a departure from traditional 'Keynesian' economics »[3]. Lorsque cette règle s'applique au marché du travail, cela signifie l'impossibilité du chômage involontaire.
- le taux de chômage naturel, dont il est vain de vouloir s'écarter.
- la préférence pour une politique monétaire sans prétention anticyclique, particulièrement une « règle de x % ».

7.4.1. Les anticipations rationnelles

Commençons par les origines : l'article de John Muth, « Rational Expectations and the Theory of Price Movements » (1961). Muth estime qu'un bon modèle ne peut pas prêter aux agents un type d'anticipation différent de ses propres équations. Les modèles qui concluent à des effets XYZ alors qu'ils prêtent aux agents un mode d'élaboration de leurs anticipations les faisant prévoir WXY se méprennent. Les agents ne peuvent répéter systématiquement la même erreur. « Expectations, since they are informed predictions about future events, are essentially the same as the relevant economic theory »[4].

Les anticipations rationnelles s'appuient sur cette double justification :

- A la longue, les agents apprennent intuitivement comment fonctionne le système économique ; leurs prévisions sont fondées sur cette compréhension, qui n'a pas

[1] Sargent & Wallace [327] p. 178
[2] Holland [158] p. 5
[3] Holland [158] p. 5
[4] Muth [265] p. 316

(nécessairement) un caractère scientifique, mais qui est néanmoins aussi fiable que les modèles de l'économie politique.
- Les agents exploitent toute l'information disponible. « Information is scarce and the economic system generally does not waste it ».

La rationalité des anticipations est cohérente avec les deux grands postulats de l'économie politique : les agents sont rationnels et ils agissent selon leur propre intérêt. Muth insiste sur le fait que l'hypothèse des anticipations rationnelles n'implique pas que les prévisions des agents individuels sont parfaites ni qu'elles sont toutes identiques. Les individus peuvent se tromper mais en règle générale, le marché ne se trompe pas.

En calcul des probabilités, on appelle *distribution* une fonction qui attribue une densité de probabilité à chaque valeur que peut prendre une variable. Les agents capables d'anticipation rationnelle estiment correctement la distribution des variables telles que les rendements, les prix etc. L'anticipation rationnelle se caractérise par ceci qu'à information disponible égale, la probabilité SUBJECTIVE[1] se règle sur la probabilité OBJECTIVE. La probabilité objective n'est pas une caractéristique de la réalité économique elle-même mais elle traduit les prédictions de la théorie économique.

Remarquons qu'il s'agit ici de *probabilités conditionnelles*, c'est-à-dire de probabilités « sachant que … ». Prenons la probabilité que le dividende de telle action vaille 100 ; il est évident qu'elle ne sera pas identique selon qu'on connaît ou qu'ignore telle information relative à l'entreprise. Les probabilités conçues lors des anticipations sont conditionnelles par rapport à l'information disponible.

Muth expose trois cas qu'il développe algébriquement. Ses exemples, de nature plutôt microéconomique, n'ont pas la généralité de l'exposition par Sargent and Wallace (1976)[2], que je résume ici. Un agent forme à la période *t-1*, sur base de l'information disponible, une anticipation sur la valeur que prendra une variable *p* (qu'on peut supposer être le prix sur un marché ou le niveau général des prix) à la période *t* :

$$_{t-1}p^{*}_{t} = \mathbf{E}_{t-1}p_t + u_t \qquad (7.38\text{-A})$$
$$\mathbf{E}_{t-1}u_t = 0 \qquad (7.38\text{-B})$$

L'astérisque accolé à la variable *p* marque une anticipation. L'indice préfixé (t-1) précise le moment de la formation tandis que l'indice suffixé (t) indique la période couverte par l'anticipation. La lettre **E** désigne une espérance mathématique sur base de l'information disponible à la période en indice suffixé. L'équation (7.38-A) établit que l'anticipation de *p* vaudra l'espérance mathématique de sa probabilité objective plus un terme représentatif des erreurs de prévision, « allowing for what may be very large random deviations from rationality ». Les erreurs de prévision sont purement aléatoires, c'est-à-dire non corrélées entre elles. Leur espérance mathématique est donc nulle, comme l'indique la deuxième équation. On peut en conclure qu'en moyenne, les anticipations sont correctes, parce que la moyenne des erreurs individuelles est nulle. L'hypothèse de la non-corrélation des erreurs est essentielle pour soutenir ce résultat.

Lors de son énoncé par Muth, l'idée des anticipations rationnelles était novatrice. La conception la plus répandue parmi les économistes jusqu'au milieu des années soixante-dix était celle

[1] Le terme « probabilité subjective » est assez intuitif. Ce phénomène sera analysé plus en détail au chapitre 9.4 consacré à l'incertitude.
[2] Sargent & Wallace [327] p.180

d'anticipations qui extrapolent le passé et le présent, soit naïvement, soit en lui appliquant une correction formée a priori. Pensons au modèle de Phelps : à chaque période, on y voit l'anticipation s'adapter aux observations courantes ; elle a donc toujours une période de retard sur la réalité. L'anticipation rationnelle est basée sur l'information présente et passée, mais elle cherchera à y déceler les éléments prémonitoires de ruptures possibles.

De ce point de vue, l'analyse de Muth comporte un élément équivoque si pas contradictoire. Bien dans la tradition néoclassique, il voit dans les prix eux-mêmes, en ce compris les prix passés, la source d'information principale des anticipations. Dans un de ses exemples, le producteur décide de la quantité offerte en anticipant le prix qui égalise l'offre et la demande. Mais l'offre dépend partiellement d'un élément aléatoire (par exemple, la météo). A titre de variante, Muth abandonne l'hypothèse que $\boldsymbol{E}_{t-1}u_t = 0$ et conçoit u_t comme "a linear combination of the past history of normally and independently distributed random variables e_i with zero mean and variance σ^2", ce qu'exprime l'équation :

$$u_t = \sum_{i=0}^{\infty} w_i.e_{t-i} \qquad (7.39)$$

La non-corrélation est reportée sur la variable e_i. Pour anticiper correctement les prix, les agents n'ont comme information que les prix passés entre lesquels, au-delà de leur caractère aléatoire, il existe une relation fonctionnelle qu'ils doivent découvrir. Il s'agit de déterminer les V_j de l'équation :

$$p^*_t = \sum_{j=1}^{\infty} V_j.p_{t-j} \qquad (7.40)$$

Il existe une relation objective entre le V_j et les w_i. Muth démontre qu'ayant déterminé les V_j, les agents formeront la même anticipation p^*_t que dans le cas où $\mathbf{E}_{t-1}u_t = 0$. Ce développement me paraît ramener l'anticipation rationnelle à une anticipation adaptive sophistiquée. Elle sous-entend que le public serait incapable de prévoir p_t si la relation future entre p_t et e_i devait emprunter un chemin différent de ce qu'elle était était dans le passé. Ces anticipations ne sont alors que modérément rationnelles. Ou bien e_i représente la totalité du savoir possible, y compris celui qui était inutile dans le passé, mais alors ce développement n'est qu'une tautologie inutile.

7.4.2. Le modèle d'information imparfaite de Lucas

L'article « Expectations and the Neutrality of Money » (1972) de Lucas est généralement considéré comme le joyau de la littérature des nouveaux classiques. L'auteur y expose le *modèle d'information imparfaite*. Ce modèle n'exclut pas l'existence empirique de relations du type « courbe de Phillips », mais il les rend totalement aléatoires au pont de ne fournir aucune prise à une politique d'arbitrage entre le chômage et l'inflation.

Lucas part du très walrassien principe selon lequel les producteurs décident de la quantité offerte sur base du seul système de prix relatifs. Les prix relatifs fluctuent en fonction des mouvements de l'offre et de la demande sur les divers marchés. Mais en macroéconomie, ces variations se mélangent avec celles du pouvoir d'achat de la monnaie, occasionnées par les modifications du stock monétaire. « Information on the current state of these real and monetary disturbances is transmitted to agents only through prices in the market where each agent happens to be. In the particular framework presented below, prices convey this information only imperfectly, forcing agents to hedge on whether a particular price movement results from a relative demand shift or a nominal (monetary) one »[1]. Mal interprétée, une variation du stock monétaire peut donc susciter des actions qui ont des effets REELS. La cause n'est plus

[1] Lucas [237] p. 103

l'irrationalité des anticipations (notamment par l'*illusion monétaire*) mais l'imperfection de l'information.

Le modèle a pour objet une économie simplifiée et arrangée. Le salariat en est absent. On a affaire à des producteurs indépendants qui décident de produire plus ou moins selon que le prix de leur produit est plus ou moins élevé. C'est donc la même variable n qui exprime la quantité de travail fournie et la quantité produite. Il n'y a qu'un seul bien, mais il est échangé (contre la monnaie) sur deux marchés différents. Le commissaire-priseur dirige les demandeurs du produit vers l'un des deux marchés de façon à égaliser la quantité demandée sur chacun d'eux (si la demande totale vaut a, la demande sur chaque marché vaut $a/2$). Par contre les offreurs sont affectés à l'un des deux marchés de façon aléatoire (les proportions respectives de l'offre sur l'un et l'autre marchés sont q et $1-q$). Par ce stratagème, Lucas simule les variations des prix relatifs de l'économie réelle.

Le modèle est à *générations imbriquées* avec une population qui se reproduit à l'identique. Les jeunes produisent. Ils vendent aux vieux la part du produit qu'ils ne consomment pas et thésaurisent la recette. Celle-ci sera entièrement consommée lors de leur retraite (deuxième et dernière période de leur vie).

Lors de chaque période, l'Etat distribue gracieusement un supplément de monnaie à tous les membres de la vieille génération, de telle façon que l'encaisse de chacun est multipliée par un coefficient discrétionnaire x non corrélé d'une période à l'autre. Ainsi est simulée la politique monétaire. Les producteurs ne connaissent ni q ni x qu'ils doivent deviner, mais sur base de l'information disponible relative au stock monétaire m et au niveau des prix p présents, ils sont capables de formuler les distributions probabilistes[1].

Contrairement aux individus âgés qui doivent juste consommer leur encaisse, les jeunes font face à un problème d'optimisation : combien thésauriser pour maximiser leur utilité intertemporelle qui va croissant avec les consommations présente et future et qui diminue avec la quantité de travail. Une substitution intertemporelle de consommation se joue en fonction des variations du rapport entre les prix futur et présent p'/p. Prenons le marché à offre q comme représentatif. A partir de l'optimisation susmentionnée et de l'égalisation de l'offre avec la demande, Lucas détermine l'équation (compliquée) du prix d'équilibre. Il montre que cette équation a une solution, que celle-ci est unique et qu'elle a une caractéristique intéressante : le prix d'équilibre p est fonction continue, différentiable et positive $m.\varphi(x/q)$. Si on avait $p = \varphi(x)$, le producteur saurait qu'il ne doit pas réagir aux variations de prix ; si on avait $p = \varphi(q)$, il saurait comment adapter sa production à une variation de p. Mais on comprend qu'avec $p = \varphi(x/q)$, les variations de p le laissent perplexe et peuvent l'induire en erreur.

Une version plus simple du modèle d'information imparfaite est exposée dans un autre article de Lucas « Econometric Policy Evaluation : A Critique » (1976), que nous allons résumer. L'article a pour objet de contester les études empiriques qui confirment l'existence de la courbe de Phillips et de l'arbitrage entre l'inflation et le chômage. Ces études ont le tort de garder la même équation de régression, qu'elles analysent les effets d'une politique ou d'une autre totalement différente. Or des politiques divergentes suscitent des anticipations différentes ; les modèles devraient donc différencier les équations selon la politique étudiée.

Les caractéristiques du présent modèle sont notamment :
- il n'est plus à générations imbriquées

[1] La démonstration de Lucas s'appuie sur des restrictions imposées à ces distributions.

- la création de monnaie n'y intervient plus, si ce n'est de façon sous-jacente ; l'attention est centrée sur le lien direct entre les prix et la quantité offerte (courbe de Phillips)
- il y a un bien unique échangé sur N marchés (i = 1,…,N).

Soit y_{it} la quantité offerte sur la marché *i* pendant la période *t*. Ou plutôt son logarithme[1]. Comme le montre l'équation (7.41), elle résulte de l'addition de deux composantes :

$$y_{it} = y^{p}_{it} + y^{c}_{it} \qquad (7.41)$$

- une composante PERMANENTE : y^{p}_{it}. Déterminée par les tendances fondamentales de l'économie, elle est insensible aux changements de prix.
- une composante CYCLIQUE : y^{c}_{it}. Au contraire de la précédente, celle-ci « varies with perceived changes in the relative price of goods in *i* »[2], comme l'indique l'équation :

$$y^{c}_{it} = \beta.(p_{it} - p^{e}_{it}) \qquad (7.42)$$

Le terme p_{it} est le prix sur le marché *i* à la période *t* et p^{e}_{it} est le niveau général des prix tel qu'il est estimé par les participants au marché *i*. La différence $(p_{it} - p^{e}_{it})$ renseigne évidemment sur les fluctuations des prix relatifs estimés[3].

Ainsi que l'indique l'équation (7.43), le prix de marché p_{it}, ou plutôt ses variations, est également décomposable en :
- une composante, p_t, qui varie dans le temps mais est commune à tous les marchés
- l'autre composante, z_{it}, au contraire, reflète les variations des prix relatifs dans le temps.

$$p_{it} = p_t + z_{it} \qquad (7.43)$$

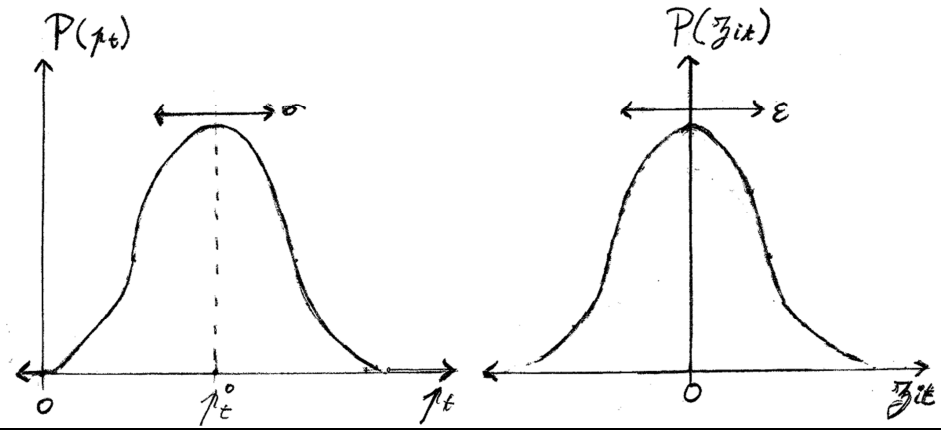
L'équation (7.43) réunit les deux causes de variation du prix de *i*. Comme nous l'avons vu précédemment, les producteurs ne peuvent pas les observer séparément, ce qui leur impose des exercices divinatoires. Heureusement, ils en connaissent la distribution probabiliste, que nous montre le graphique 7.12. Pour ces deux variables, la loi normale (Gauss-Laplace) est d'application. La composante p_t a pour espérance mathématique la valeur $p°_t$, avec un écart-type de σ. La composante z_{it} n'est corrélée ni avec p_t, ni avec les z_t des autres marchés ; sa moyenne est nulle et son écart-type vaut ε.

[1] Dans ce modèle, toutes les variables comportant la lettre *y* ont trait aux quantités offertes. Ces *y* sont toujours le logarithme de la quantité et non la quantité elle-même. Les variables avec la lettre *p* ont trait aux prix et sont également des logarithmes. Par souci de simplifier l'exposé, je ne répéterai pas chaque fois « le logarithme de… ». L'usage des logarithmes a pour conséquence que les multiplications deviennent des additions, puisque $\log x.y = \log x + \log y$.

[2] Lucas [238] p. 36.

[3] La différence entre ces deux logarithmes est représentative d'un rapport de prix.

Graphique 7.12 : distribution probabiliste du niveau général des prix et des prix relatifs



Le niveau général des prix est la moyenne des prix sur les différents marchés. Lucas montre que c'est le cas de p_t, pour autant que *N* soit suffisamment élevé.

p^e_{it} est l'anticipation de p_t que forment les producteurs du marché *i*. Vu l'hypothèse des anticipations rationnelles, p^e_{it} égale l'espérance mathématique de la probabilité objective, ce qu'exprime l'équation (7.43) :

$$p^e_{it} = \mathbf{E}\{p_t \mid p_{it}, I_{t-1}\} = (1-\theta).p_{it} + \theta.p°_t \qquad \text{où } \theta = \varepsilon^2/(\sigma^2+\varepsilon^2) \quad (7.44)$$

Les probabilités intervenant dans le modèle sont des probabilités conditionnelles, connaissant l'information disponible la période précédente notée I_{t-1} et les prix observés p_{it}.

En substituant (7.44) dans l'équation (7.42), on obtient comme explication des variations cycliques de l'offre des producteurs en *i* :

$$y^c_{it} = \theta.\beta.(p_{it} - p°_t) \qquad (7.45)$$

Changeons de point de vue ; quittons le marché *i* et plaçons-nous au niveau macroéconomique, ce qui nous permet de remplacer p_{it} par p_t, pour autant que le nombre de marchés *N* soit suffisamment élevé. Par analogie, avec (7.41), l'offre globale de l'économie est décrite par l'équation :

$$y_t = y^p_t + \theta.\beta.(p_t - p°_t) \qquad (7.46)$$

L'offre globale, et avec elle l'emploi, suivent le chemin y^p_t qui est indépendant des variations du niveau général des prix ; ils n'en divergent que lorsque le niveau général des prix (p_t) s'écarte des anticipations. « These deviations occur because agents are obliged to infer current general price movements on the basis of incomplete information »[1].

7.4.3. Vision de la politique économique

TAUX D'INTERET REEL ET POLITIQUE MONETAIRE

Dans l'esprit keynésien, la politique d'expansion de l'offre monétaire cherche à abaisser le taux d'intérêt. A travers le taux nominal, c'est évidemment le taux d'intérêt REEL qui est visé. D'après les nouveaux classiques, cet objectif est inaccessible, toujours pour la même raison : anticipant l'inflation qui résultera de l'expansion monétaire, les détenteurs de capitaux adapteront leur exigence de rendement nominal de façon à contrecarrer le rabotage du revenu par l'inflation.

[1] Lucas [238] p. 37

Sargent consacre à cette problématique l'article « Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment » (1973). Cette analyse veut moderniser la théorie fisherienne de la relation entre les taux d'intérêt réels et nominaux. La théorie monétariste traditionnelle s'appuie sur des hypothèses comme la fonction IS horizontale, la fonction LM verticale ou la courbe de Phillips verticale pour nier l'efficacité de la politique d'expansion monétaire. Sargent montre qu'avec l'hypothèse des anticipations rationnelles, la démonstration est plus simple et moins contraignante.

Résumée en une phrase, la conclusion est celle-ci : « the real rate of interest is independent of the systematic, or foreseen, part of the money supply, which therefore can influence the nominal rate only through effects on expected inflation"[1].

Le modèle comporte trois équations qui décrivent la structure macroéconomique :

$$y_t = k_t + \alpha.(p_t - {}_{t-1}p^*_t) + u_1 \qquad (7.47\text{-A})$$

$$y_t = k_t - \beta.[r_t - ({}_tp^*_{t+1} - p_t)] + u_2 \qquad (7.47\text{-B})$$

$$m_t - p_t = y_t - \gamma r_t + u_3 \qquad (7.47\text{-C})$$

Les variables *m*, *p*, *k* et *y* désignent les logarithmes naturels respectivement de la masse monétaire, du niveau des prix, de la capacité productive normale et du revenu réel. Le taux d'intérêt nominal est *r*. L'astérisque indique une anticipation. Les trois *u* sont des variables aléatoires, de distribution normale, non corrélées entre elles. α, β et γ sont des coefficients positifs.

- L'équation (7.47-A), courbe d'offre globale, dit que le revenu réel dévie de sa trajectoire normale, lorsque l'évolution des prix a démenti les anticipations
- L'équation (7.47-B), courbe de demande globale, dit que le revenu réel descend sous sa trajectoire normale, lorsque le taux d'intérêt réel augmente.
- L'équation (7.47-D) dit que la demande d'encaisses réelles croît avec le revenu réel et diminue avec le taux d'intérêt réel.

A cela s'ajoutent deux équations. Celle qui affirme la rationalité des anticipations par l'égalité entre ${}_tp^*_{t+1}$ et $\mathbf{E}_t p_{t+1}$ (cf. supra). Et celle qui explique le comportement des autorités monétaires ; les variations de *m* sont fonction de deux types de facteurs : d'une part de ce que Sargent appelles des « feedback rules » qui sont des combinaisons linéaires des valeurs passées et présentes de divers paramètres tels *k* et les *u* ; d'autre part d'une composante imprévisible ε_t censée représenter le pouvoir discrétionnaire, de distribution normale et de moyenne nulle. Les deux équations sont indirectement liées dans la mesure où tous ces paramètres des décisions monétaires et leur effet sont connus et compris par le public, à l'exception évidemment de ε. L'information sur laquelle est basée l'anticipation ${}_tp^*_{t+1}$ comporte les valeurs présentes et passées de *m*, *k* et des *u*.

Sargent obtient ainsi un modèle cohérent dont il relève ces caractéristiques essentielles : "First, a natural rate of output exists in the sense that the deviation of output from its normal level is statistically independent of the systematic parts of monetary and fiscal policies (…) Second, the real rate of interest is independent of the systematic parts of the money supply"[2]. Par contre, "the only part of money supply at *t* that affects the real rate is the random component ε_t"[3]

[1] Sargent [326] p. 431.
[2] Sargent [326] pp. 442-443
[3] Sargent [326] p. 444

SARGENT ET WALLACE : LA POLITIQUE DISCRETIONNAIRE IMPUISSANTE

En cette matière, les nouveaux classiques adhèrent à la thèse friedmanienne : la meilleure politique monétaire consiste en l'adoption d'un taux de croissance de la masse monétaire stable et indépendant des vicissitudes économiques. Sargent & Wallace défendent cette optique dans leur article « Rational Expectations and the Theory of Economic Policy » (1976) Selon eux, plus aucun économiste ne soutient l'idée que la politique monétaire doit être laissée à la discrétion de l'autorité. La question n'est « faut-il une règle ? » mais bien « quel type de règle est le plus souhaitable ? ». Les keynésiens préconisent une *règle de feed back*, faisant varier la masse monétaire en fonction de l'évolution du revenu national. Les monétaristes lui opposent la règle des x%.

Supposons que l'autorité monétaire puisse contrôler le taux de croissance du revenu national *y* en manipulant le taux de croissance de la masse monétaire *m*, comme l'indique l'équation :

$$y_t = \alpha + \lambda.y_{t-1} + \beta.m_t + u_t \qquad (7.48)$$

Les coefficients α, β et λ sont des paramètres qu'une régression pourrait déterminer. L'intervention de y_{t-1} est nécessaire car le contrôle doit idéalement s'opérer de façon stable dans la durée. La variable u_t est aléatoire, sériellement indépendante, a une espérance nulle et est uniformément distribuée. Le but est de minimiser la variance de *y* autour d'un taux-cible y*. On aurait comme règle de feedback l'équation (7.49):

$$m_t = g_0 + g_1.y_{t-1} \qquad (7.49)$$

Si l'on pose que g_1 est nul, l'équation (7.49) correspond à la règle monétariste. Les auteurs calculent que pour égaliser la moyenne semi-stationnaire E(y) avec y* et minimiser var(y), la politique monétaire doit s'appuyer sur $g_0 = (y^*-\alpha)/\beta$ et $g_1 = -\lambda/\beta$. En substituant ces valeurs dans (7.49) et cette dernière dans (7.48), on arrive à la conclusion que la règle de feedback keynésienne atteint mieux l'objectif que la règle qui annule g_1.

Nos auteurs introduisent ensuite deux hypothèses habituelles des nouveaux classiques :

- l'hypothèse monétariste que seules les variations non anticipées de la masse monétaire influencent y_t,. L'équation (7.48) est modifiée de telle façon que dans le membre de droite, m_t est remplacé par $(m_t - \mathbf{E}_{t-1}m_t)$[1].
- Les anticipations rationnelles : « the public knows the monetary authority's feedback rule and takes this into account in forming expectations »[2]. L'espérance mathématique de la probabilité subjective (probabilité conditionnelle par rapport à l'information disponible en t-1) correspond aux paramètres réels, comme l'indique l'équation :

$$\mathbf{E}_{t-1}m_t = g_0 + g_1.y_{t-1} \qquad (7.50)$$

Sargent & Wallace démontrent qu'avec ces nouvelles hypothèses, l'équation (7.48) peut être réécrite d'une façon qui rend le revenu national y_t totalement indépendant de g_0 et de g_1. Supposons que la politique monétaire décide de modifier g_0 et g_1 ; les agents, anticipant ces changements, modifieront leur comportement, ce qui affectera α et β. "In our hypothetical model, the resulting differences in α and λ just offset the differences in g_0 and g_1, leaving the behavior of *y* identical as a result"[3]. La politique monétaire se révèle donc incapable d'affecter le revenu national dans le sens souhaité.

---

[1] $\mathbf{E}_{t-1}m_t$ indique l'espérance mathématique de la variable m_t, compte tenu de l'information disponible en t-1.
[2] Sargent & Wallace [327] p. 173
[3] Sargent & Wallace [327] p. 174

On touche ici à l'essentiel de la conception politique des nouveaux classiques : ce que Holland appelle la « policy ineffectiveness proposition ». Toute politique visant à stimuler ou stabiliser le revenu réel et l'emploi est vouée à l'échec aussi bien à court qu'à long terme, si ce n'est par l'effet du hasard. La critique contre la courbe de Phillips est l'exemple le plus notoire. Mais la conclusion est la même pour toute action politique. L'équivalence ricardienne posée par Barro va dans le même sens : une génération d'épargnants anticipe les taxes que ses héritiers devront acquitter pour rembourser la dette publique, ce qui les amène à réduire leur consommation présente pour compenser les effets de ce remboursement ; l'emprunt public ne peut donc pas financer une politique de relance.

Alors, qu'est-ce qui motive l'acharnement des nouveaux classiques contre la politique monétaire, à laquelle ils ne reconnaissent pas d'effets réels ? Comme l'avait relevé Lucas, les agents sont confrontés à la difficile tâche de distinguer les variations des prix relatifs de celles du niveau général des prix. Si la politique monétaire rend la masse monétaire plus instable, les variations du niveau général des prix deviendront prépondérantes objectivement et subjectivement. Les agents auront tendance à surestimer la part des variations du niveau général des prix par rapport aux variations des prix relatifs. En conséquence, la firme représentative adaptera plus mollement sa production aux variations des prix, même lorsque ceux-ci résultent de chocs réels. « This means that the price system is less effective as a mechanism for allocating resources »[1]. Cet argument, ne met-il pas involontairement en lumière une contradiction inhérente à cette théorie : si les anticipations rationnelles traitent une information trop imparfaite, les anticipations rationnelles sont des anticipations imparfaites.

KYDLAND & PRESCOTT : POLITIQUE OPTIMALE OU COHERENTE

Dans une série d'interactions mutuelles, les actions politiques affectent les conditions économiques, mais également les anticipations et les comportements des agents qui à leur tour modifient les conditions, ce qui induit de nouvelles actions politiques et ainsi de suite. Cette série pourra être convergente ou divergente. Dans ce dernier cas, les actions politiques tendent à générer de l'instabilité économique alors qu'elles ont une vocation stabilisatrice. Même dans le cas où la série converge, le résultat sera sous-optimal et moins performant que ne l'aurait été une règle, annoncée a priori et respectée tout au long de la période. Telle est la thèse fondamentale de l'article « Rules Rather than Discretion : the Inconsistency of Optimal Plans » que publient Finn Kydland et Edward Prescott en 1977.

Pour la politique économique, c'est la quadrature du cercle. Imaginons une politique qui veut être cohérente dans le temps en appliquant à chaque période les mesures optimales compte tenu des paramètres économiques de la période sans une vue au-delà de cette période. Une telle politique est vouée à long terme à être à la fois sous-optimale et inconstante. La raison en est que les agents rationnels comprennent le problème des autorités et anticipent qu'elles adapteront les paramètres de leur politique à ceux, modifiés, de l'économie. Le pouvoir politique aura donc effectué son exercice d'optimisation sur base de données incomplètes. Si elle était rationnelle, l'autorité anticiperait la modification des comportements des agents dans le calcul des paramètres de sa politique[2]. Mais souvent, elle espère pouvoir imposer sa politique à un public qui resterait inerte, ce qui l'amène à devoir toujours réadapter son action aux conditions changeantes et à ne jamais tenir parole de ses engagements antérieurs. Un problème

[1] Holland [158] p. 10

[2] Certes, les comportements au temps t-1 précèdent la décision politique au temps *t*, qu'ils orientent. Mais indirectement, la décision politique au temps *t* influence ces comportements, car elle avait été anticipée au temps t-1 et les comportements résultent de cette anticipation. Cette projection se reproduit à chaque période et doit être pris en compte par une autorité rationnelle.

de crédibilité se fait jour ; le public adapte son comportement non à l'action annoncée mais à celle probable. Ce chemin se caractérise par sa sous-optimalité. La démarche (erronée) reste cohérente, mais une incohérence temporelle se développe dans le sens que le pouvoir doit recalculer les paramètres de la pseudo-optimisation à chaque période.

Les auteurs donnent l'exemple des constructions résidentielles en zone inondable, que la collectivité aurait intérêt à éviter. En cas d'inondation, l'Etat peut indemniser les citoyens sinistrés et engager des travaux de protection ou laisser les sinistrés face à leur seule responsabilité. Si les citoyens doutent de la détermination de l'Etat, ils construiront, convaincus que l'Etat n'osera pas les abandonner à leur sort.

La solution, c'est que l'autorité comprenne que sa crédibilité est un paramètre plus important que les éléments techniques eux-mêmes. Si l'autorité s'est organisée avec des rigidités institutionnelles qui l'empêchent de fléchir, le public fera confiance en ses déclarations, grâce à quoi l'autorité gardera un contrôle des anticipations et des comportements. Le gouvernement, soucieux du bien-être des citoyens, est incité à augmenter la circulation monétaire pour faire baisser le chômage. Mais voulant leur bien, il fera empirer leur situation, car il stimulera l'inflation sans empêcher le maintien du chômage à son taux naturel. Par contre, si la politique monétaire est aux mains d'un organisme solidement indépendant, moins susceptible d'empathie avec les citoyens, la situation de chacun sera améliorée. En l'absence d'institutions ad hoc, le risque de dérapage est très important.

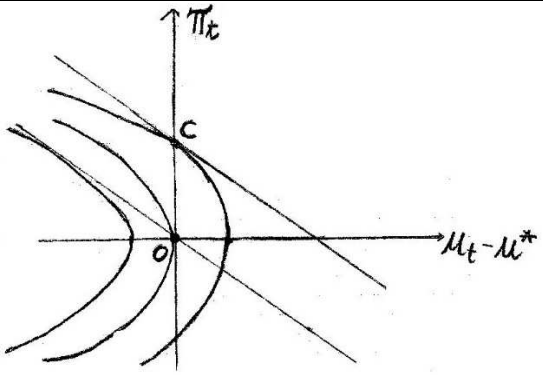
Les auteurs modélisent la problématique politique ainsi. Soient $x_1 \ldots x_T$, la séquence des situations économiques aux périodes comprises entre 1 et *T*. Soient $\pi_1 \ldots \pi_T$ les décisions politiques sur ce même intervalle de temps. Le but de ces décisions est de maximiser une fonction de bien-être social $S(\pi_1 \ldots \pi_T, x_1 \ldots x_T)$. Considérons les décisions prises au temps *t* ($1 < t < T$). Les auteurs les définissent comme COHERENTES si $\forall t$, le gouvernement veille à maximiser *S* en tenant pour acquit $\pi_1 \ldots \pi_{t-1}, x_1 \ldots x_{t-1}$. Ils les définissent comme OPTIMALES si $\forall t$, il veille à maximiser *S* considérant $\pi_1 \ldots \pi_t, x_1 \ldots x_{t-1}$.

Prenons l'exemple de l'arbitrage entre l'inflation et le chômage. Le gouvernement joue sur le taux d'inflation π_t pour réduire le chômage u_t. La fonction de bien-être social $S(\pi_t, u_t)$ diminue évidemment avec l'augmentation tant de l'inflation que du chômage qui sont deux maux. Les coefficients de cette fonction pondèrent la répulsion ressentie par la société envers l'un et l'autre ; les auteurs ne précisent pas cette fonction, si ce n'est que le taux d'inflation préféré est 0%. Cette politique se heurte à la contrainte absolue que représente le *taux de chômage naturel* (la courbe de Phillips verticale) :

$$u_t = u^* - \lambda.(\pi_t - \pi^e_t) \qquad (7.51)$$

où u* est le taux de chômage naturel et π^e_t est le taux d'inflation anticipé ; vu l'hypothèse des anticipations rationnelles, ce taux égale l'espérance mathématique $\mathbf{E}\pi_t$.

Graphique 7.13 : arbitrage optimal entre l'inflation et le chômage



Dans l'exemple dessiné, les courbes d'indifférence sont symétriques par rapport à l'abscisse, laissant supposer que l'inflation négative dérange autant que l'inflation positive.

Plus l'inflation s'écarte du taux optimum nul, plus le taux de chômage doit être bas, ce qu'exprime la forme des courbes d'indifférence. La contrainte (7.51) s'exprime par les droites parallèles de pente $-1/\lambda$ qui coupent l'ordonnée à un niveau représentant l'inflation anticipée. Le résultat de toute politique se trouvera toujours sur l'axe des ordonnées qui correspond à la courbe de Phillips verticale. Supposons une politique COHERENTE entraînant une anticipation d'inflation OC. Le point C en est la représentation. Une politique OPTIMALE parvient à convaincre le public que l'inflation restera nulle ; elle est représentée par l'origine. On constate qu'elle se situe sur une courbe d'indifférence plus à l'ouest, donc plus satisfaisante que le point C.

Le retentissement de cet article fut considérable. Un double retentissement :

- D'une part, théorique : "It has inspired a large cross-disciplinary literature at the intersection between economics and political science (…) According to Rogoff's analysis, if an independent central bank is managed by a 'conservative' central banker- i.e. an agent who is more inflation averse than citizens in general- then better welfare outcome can be achieved"[1].
- D'autre part, pratique : "Indeed the reforms of central banks undertaken in many countries as of the early 1990s have their roots in the research initiated by Kydland and Prescott"[2]. Il est certain que les statuts de la BCE, censés en assurer l'indépendance, s'intègrent dans cette évolution.

Laissons la conclusion aux auteurs : "We find that a discretionary policy for which policymakers select the best action, given the current situation, will not typically result in the social objective function being maximized. Rather, by relying on some policy rules, economic performance can be improved"[3].

[1] [357] p. 3 et p. 10.
[2] [357] p. 3.
[3] Kydland & Prescott [204] p. 473.

7.5- LA MACROECONOMIE DU DESEQUILIBRE

L'école que nous abordons maintenant a établi son camp de base en terre walrassienne ; mais précisément, un camp de base a pour vocation qu'on s'en écarte[1]. L'école du déséquilibre s'est forgé de concepts hétérodoxes à partir des matériaux walrassiens dans le but de soutenir la thèse keynésienne.

Keynes considérait que le sous-emploi pouvait correspondre à une situation d'équilibre, de par l'absence d'incitant à augmenter l'emploi et la production. Dans l'optique walrassienne, l'équilibre suppose l'égalité entre l'offre et la demande sur tous les marchés, y compris celui du travail. Par définition, le chômage y est donc un déséquilibre. L'école du déséquilibre s'intéresse à des situations qui sont un déséquilibre au sens walrassien mais pas nécessairement au sens keynésien.

L'école du déséquilibre est montée en puissance dans les années soixante, a atteint son apogée dans les années soixante-dix et a presque complètement disparu du paysage de l'économie politique dès la fin des années quatre-vingts. Elle mérite toutefois notre attention, car on ne comprend vraiment l'équilibre qu'après avoir appréhendé le déséquilibre.

7.5.1. Patinkin, le précurseur

Nous avons résumé la première partie de l'ouvrage « Money, Interest and Price » au sous-chapitre 6.5.1. Après des chapitres restés très orthodoxes malgré l'apparente dénonciation de la dichotomie, arrive enfin à la fin de l'ouvrage le chapitre où le déséquilibre fait sa première apparition.

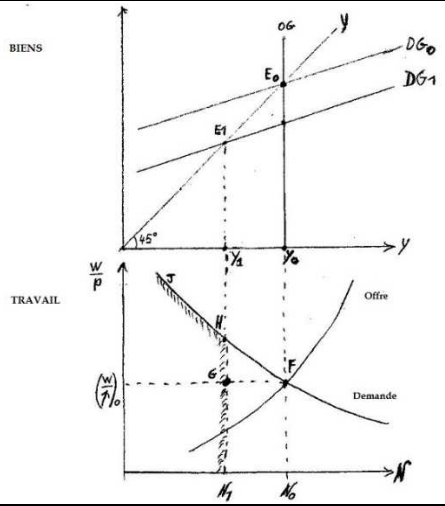
Patinkin s'intéresse particulièrement à la relation entre le marché du travail et celui des marchandises, dont rend compte le graphique 7.14.

Lorsque l'économie est dans sa position d'équilibre, le plein emploi règne sur le marché du travail. Selon Patinkin, le plein emploi implique que les travailleurs soient sur leur courbe d'offre. Cette condition est nécessairement satisfaite lorsque le marché est en équilibre d'équilibre. N'oublions pas que dans l'optique walrassienne, cette courbe est le résultat d'un processus d'optimisation individuelle ; elle ne peut donc pas comporter de positions involontaires. Le chômage involontaire suppose que les travailleurs vendent une quantité de travail située à gauche de leur courbe d'offre.

Le point E_0 correspond à l'équilibre général. Il se situe à l'intersection de TROIS courbes : la demande globale DG_0, la droite *Y*, qui indique la quantité effectivement produite, et l'offre globale OG_0. Comme il y a équilibre sur tous les marchés, y compris celui du travail, l'offre globale est une droite verticale au niveau de production Y_0, résultant du plein emploi sur le marché du travail. L'équilibre sur ce marché est représenté par le point *F*, déterminant un salaire réel de $(w/p)_0$ et un emploi de N_0.

[1] Cette métaphore est reprise chez de Vroey

Graphique 7.14: chômage involontaire et déséquilibre suivant Patinkin



Comment introduire le chômage involontaire dans le modèle ? Supposons qu'une baisse de la consommation ou de l'investissement fasse régresser la demande globale de DG_0 à DG_1. Normalement, un mécanisme autocorrecteur devrait jouer : le prix des marchandises devrait baisser tout comme le taux d'intérêt à la suite de l'augmentation de la demande de titres ; l'investissement serait relancé. On assisterait alors juste à un gonflement temporaire des stocks, sans autre inconvénient. Mais supposons que l'élasticité de l'investissement par rapport au taux d'intérêt et celle de la consommation par rapport aux prix soient insuffisante. Le salaire n'ayant pas varié, l'offre globale de marchandises reste représentée par la droite OG. Mais au point E_1, les entreprises, contraintes de ne pas produire plus qu'Y_1, sont hors de cette courbe. Elles adaptent l'emploi à leurs débouchés, le réduisant de N_0 à N_1. La courbe de demande de travail est comme impuissante ; une courbe de comportement d'embauche J-H-N_1 prend le relais. Le point *G* du graphique indique la nouvelle situation du marché du travail, en déséquilibre. Tout comme les entrepreneurs, les travailleurs sont donc hors de leur courbe d'offre.

Evidemment, l'économie ne demeurera pas bloquée à ce stade. Les offres excédentaires sur le marché des biens et sur celui du travail exercent conjointement une pression à la baisse du salaire et du niveau général des prix. La demande globale sera stimulée. Que la baisse du salaire nominal soit supérieure, égale ou inférieure à celle du niveau des prix, dans les trois cas, la demande globale finira par revenir à sa position initiale. L'économie retrouvera l'équilibre E_0 avec le plein emploi. Ce n'est que si les prix ou les salaires sont rigides que l'accroissement de la demande globale s'arrêtera dans la zone comprise entre DG_1 et DG_0 et que le chômage ne sera pas entièrement résorbé. La rigidité des salaires ou des prix n'est pas la cause première du chômage, mais elle entraîne sa persistance.

Patinkin distingue deux interprétations possibles du chômage keynésien dans le cadre walrassien :

- Soit on est dans un cas de non-existence de l'équilibre général. Aucun vecteur de prix n'assure l'équilibre simultané sur tous les marchés ;
- soit l'équilibre existe mais n'est pas stable ; le choc qui écarte l'économie de son équilibre la maintient durablement dans le déséquilibre.

Patinkin n'adhère ni à l'une ni à l'autre de ces interprétations qu'il juge fort radicales. Son explication fait simplement appel au facteur temps. Certes, le retour à l'équilibre peut être long, intolérablement long pour ceux qui souffrent du déséquilibre.

7.5.2. L'hypothèse de décision duale

CLOWER INTRODUIT L'HYPOTHESE DE DECISION DUALE

L'article « The Keynesian Counter Revolution : A Theoretical Appraisal » de Robert Clower paru en 1965 met la théorie du déséquilibre sur les rails. Le titre fait allusion à la trahison de la révolution keynésienne par la synthèse keynésienne-néoclassique.

Selon Clower, la théorie orthodoxe explique adéquatement les transactions en situation d'EQUILIBRE. Par contre, face au DESEQUILIBRE, elle est lacunaire. Il me semblerait plus logique de prétendre qu'elle fait montre de lacunes qui ne se révèlent qu'en cas de déséquilibre.

Clower assimile la *loi de Walras* à l'aphorisme de Keynes parodiant Say : « l'offre crée sa propre demande ». Il lui semble évident que la pensée keynésienne est incompatible avec la *loi de Walras*, sans quoi elle n'aurait rien apporté de neuf. Beaucoup d'économistes se sont éloignés de Keynes, car ils n'ont pas osé rompre avec cette loi. Clower, libre de cette inhibition, s'en débarrasse allègrement.

Clower relève que la théorie orthodoxe fait de l'offre et la demande, donc de la demande excédentaire, des fonctions de la seule variable indépendante que sont les prix. Cela lui semble bien étroit. Les quantités échangées, principalement des facteurs de production, influencent le revenu et donc méritent également d'entrer comme variable indépendante dans ces fonctions. Selon Clower, Keynes se référait tacitement à une conception des échanges plus large qui inclut des signaux relatifs aux quantités parmi les variables indépendantes des fonctions de demande excédentaires, en plus du système de prix ; pensons par exemple à sa fonction de consommation $C = C(Y)$. Sont notamment une source importante de signaux à intégrer dans les fonctions de demande excédentaire, les fluctuations des stocks, qui gonflent lorsqu'il y a des invendus.

Bien plus que des fonctions de demande excédentaires elles-mêmes, le problème provient d'une hypothèse implicite qui leur est accolée : « established preference analysis tacitly presupposes that selling, buying and saving plans are carried out simultaneously »[1]. En réponse au prix crié par le commissaire-priseur, la fonction de demande excédentaire synthétise les choix combien gagner, comment le gagner, combien épargner et combien dépenser et comment le dépenser. Clower conclut : « we must surely question the universal relevance of the unified decision hypothesis »[2] ; il propose de lui substituer ce qu'il appelle l'*hypothèse de décision duale*.

Un peu de terminologie avant de continuer. Clower appelle
- *offre notionnelle* et *demande notionnelle*, les fonctions d'offre et de demande bien connue de la théorie orthodoxe, telles qu'elles émanent du processus d'optimisation.
- *demande contrainte*, la fonction de demande dont l'agent devra se contenter si sa contrainte budgétaire subit une révision à la baisse comme suite à un rationnement exercé sur l'offre de ses services.

Selon Clower, les agents décident de leur demande excédentaire en deux temps :
1- Ils définissent leur offre et leur demande notionnelles et testent si le marché leur permet de les satisfaire.
2- Si le *revenu courant effectif* reste en deçà du *revenu notionnel*, parce que l'offre de ses services se trouve rationnée, notre agent se résout à un « second round of decision making », révisant ses demandes vers le bas, le résultat étant la *demande contrainte*.

[1] Clower [58] pp. 47-48
[2] Clower [58] p. 48

Sur le marché, les seuls signaux émis par l'agent qui sont effectivement perçus par les autres agents sont l'offre notionnelle et la demande contrainte. Supposons un agent dont l'offre notionnelle de travail et la demande notionnelle de biens valent 120. Mais confronté à une limite de 100 pour son offre, il doit restreindre sa demande à 100. Une demande notionnelle s'élevant à 20 n'est tout simplement pas signalée aux autres agents. Signalée, elle aurait pu servir de contrepartie à une hausse de la demande de facteurs et donc de ses services. La demande notionnelle, qui permettrait le plein emploi, est impuissante, parce que non signalée.

L'équation (7.52) expose la *loi de Walras*. L'indice *i* concerne les *m* biens de consommation et l'indice *j* les *n* facteurs de production. Les variables d_i, d_j, s_i et s_j sont respectivement leurs demandes et leurs offres notionnelles, fonctions exclusivement du vecteur des prix **p**.

$$\Sigma_i^m \, p_i.[d_i(\mathbf{p}) - s_i(\mathbf{p})] + \Sigma_j^n \, p_j.[d_j(\mathbf{p}) - s_j(\mathbf{p})] = 0 \qquad (7.52)$$

$$\Sigma_i^m \, p_i.[d_i^*(\mathbf{p},\mathbf{y}) - s(\mathbf{p})] + \Sigma p_j.[d_j(\mathbf{p}) - s_j(\mathbf{p})] \leq 0 \qquad (7.53)$$

Clower préconise son remplacement par l'équation (7.53). La demande notionnelle des biens de consommation en disparaît pour être remplacée par la demande contrainte, notée d*, qui dépend à la fois du vecteur des prix et des quantités *y*. La nouveauté essentielle est la substitution d'une inégalité à l'égalité walrassienne. L'équation (7.52) révèle cette caractéristique bien connue de l'économie walrassienne qu'une offre excédentaire de facteurs (chômage) doit être compensée par une demande excédentaire de produits. L'équation (7.53) rend possible une offre excédentaire conjointe des produits et des facteurs, ce qui est plus conforme à la réalité économique.

Selon Clower, l'*hypothèse de décision duale* et l'équation (7.53) sont présentes implicitement chez Keynes. Ce qui faire dire à Clower que la théorie classique doit être comprise comme un cas particulier de la théorie keynésienne. Il apparaît en effet que l'équation (7.52) est la forme que revêt l'équation (7.53), plus générale, lorsque prévaut le plein emploi. Clower remarque en outre que l'hypothèse de décision duale est compatible avec la flexibilité des prix.

On constate que la monnaie et les titres sont absents du modèle clowerien. C'est une faiblesse, car en cas de baisse de leur revenu, les agents pourraient réduire l'acquisition de ces actifs plutôt que de réduire leur consommation.

LEIJONHUFVUD SOUTIENT LE KEYNESIANISME CLOWERIEN

Trois ans après la publication de l'article de Clower paraît un livre qui reste considéré comme un des grands classiques de la littérature économique : « On Keynesian Economics and the Economics of Keynes » par Axel Leijonhufvud. Le but de ce livre est de montrer l'écart énorme qui sépare la *synthèse néoclassique-keynésienne* de la *Théorie Générale*, synthèse à laquelle il oppose l'interprétation clowerienne en termes de déséquilibre. Il écrit: « one must conclude, I believe, that Keynes' theory, although obscurely expressed and doubtlessly not at all clear even in his own mind, was still in substance that to which Clower has recently given a precise statement »[1].

Cet ouvrage sera étudié plus en détail au chapitre 9.7. Voyons maintenant quelques remarques que l'auteur exprime dans le chapitre consacré au déséquilibre.

[1] Leijonhufvud [215] p. 102

Rappelons-nous les trois périodes de Marshall pour l'égalisation de l'offre et la demande sur un marché (cf . sous-chapitre 3.2.3). La *période de marché* correspond à un ajustement INSTANTANE du prix ; les quantités ne s'adaptent qu'à la deuxième ou la troisième périodes. Selon Leijonhufvud, le caractère révolutionnaire de la pensée keynésienne vient justement de l'inversion de cet ordre[1]. La vitesse d'ajustement des prix cesse d'être infinie. Leijonhufvud reproche aux classiques d'« identifying imperfect flexibility of prices with the existence of monopolistic practice of one sort or another »[2], alors qu'il s'agit d'un mode de fonctionnement normal de notre système économique.

La vitesse d'ajustement infinie des prix, implicite chez les classiques, implique l'information parfaite, c'est-à-dire complète et sans coût, des intervenants sur le marché. Par exemple, pour décider que produire et en quelle quantité, les offreurs doivent connaître exactement les besoins des consommateurs. Selon Hayek, le génie de l'économie de marché est justement de simplifier ce problème : toute l'information nécessaire est concentrée dans le système de prix ; connaissant uniquement les prix des biens qu'il est susceptible d'acheter ou de vendre, l'entrepreneur est à même de prendre sa décision. Leijonhufvud remarque que le système de prix que l'offreur devrait connaître est le prix d'équilibre. Cette théorie doit donc exclure les *false tradings*, c'est-à-dire les transactions conclues à d'autres prix que le prix d'équilibre, dont l'existence brouille l'information. Le *tâtonnement* walrassien interdit explicitement ces transactions. Mais précisément, le tâtonnement n'est pas en vigueur dans l'économie réelle. L'absence du commissaire priseur a encore une autre conséquence dévastatrice : elle oblige les participants au marché à être des *price makers*. Chaque offreur devrait donc connaître les prix demandés par ses différents concurrents. Le problème de l'information atteint ainsi une complexité extrême.

Le temps nécessité par la production est un autre frein à l'ajustement parfait et instantané des prix. Vu ce délai, la demande d'aujourd'hui est satisfaite par l'offre décidée hier et avant-hier. Il faut donc espérer que les anticipations des offreurs étaient adéquates.

Leijonhufvud soutient l'idée clowerienne que chez Keynes, l'offre excédentaire sur le marché du travail n'est compensée par aucune demande excédentaire sur un autre marché. Quoiqu'en dise une interprétation courante, le marché keynésien de la monnaie n'est pas en demande excédentaire lorsque prévalent la surproduction et le chômage (pensons aux équations de Lange : cf. sous-chapitre 6.5.1). Une telle situation s'accompagnerait d'une hausse du taux d'intérêt, ce qui est peu fréquent en sous-emploi keynésien.

7.5.3. Solow et Stiglitz : emploi et salaire

L'article « Output, Employment, and Wages in the Short Run » (1968) pèse les atouts respectifs des théories cambridgienne et néoclassique de la répartition des revenus ; il nous intéresse par la méthode utilisée dans laquelle intervient le déséquilibre. Les acteurs principaux de l'intrigue sont ces quatre variables : le salaire réel, l'emploi, l'offre globale et la demande globale, comme dans le chapitre treize de Patinkin.

Le modèle permet aux salaires nominaux et aux prix de varier dans les deux sens ; simplement, leurs fluctuations sont incapables d'équilibrer INSTANTANEMENT les marchés du travail et des

[1] Dans un article ultérieur, Leijonhufvud considérera toujours que cette inversion est indispensable, mais n'en créditera plus Keynes, resté fort attaché à ses racines marshalliennes.
[2] Leijonhufvud [215] p. 67

produits. Si les termes *rigidité des salaires et des prix* ont un sens légitime, selon les auteurs, c'est ce sens là.

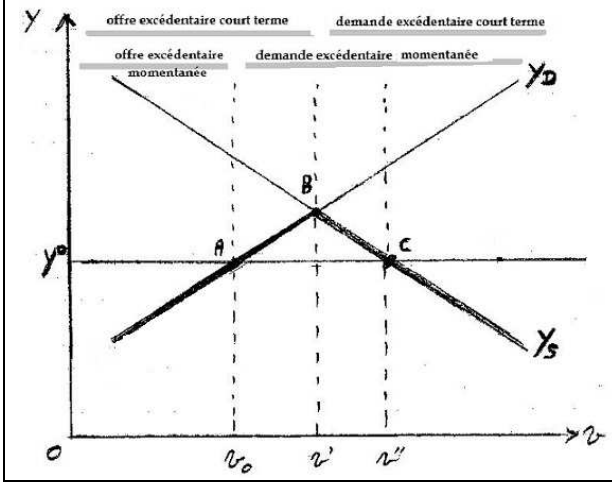
Les auteurs montrent d'abord comment l'offre et la demande globales sont des fonctions du salaire réel. L'OFFRE GLOBALE Y_s est assimilée à la production. Comme le modèle raisonne à court terme, l'emploi *N* est le seul facteur variable. On a donc $Y_s = F(N)$ qui est croissante. Mais le salaire réel *v* correspond à la productivité marginale du travail. On a $v = F'(N)$ décroissante, donc $N = f(v)$ décroissante, donc $Y_s = F(f(v))$ décroissante. L'offre globale est dès lors une fonction décroissante du salaire réel *v*.

Parallèlement, la DEMANDE GLOBALE Y_d est montrée être une fonction croissante du salaire réel. La hausse du salaire réel entraîne celle de la consommation qui est un composant essentiel de la demande globale, car la propension à consommer des salariés est supérieure à celle de la classe vivant du profit.

Le modèle distingue le COURT TERME et le terme MOMENTANE . Au cours de ce dernier, non seulement le capital est fixe mais également l'emploi ; seuls sont alors susceptibles de varier, le prix et le salaire nominal. Partons d'un équilibre momentané : l'emploi est fixé au niveau N° déterminé historiquement, correspondant à une offre globale momentanée Y°, parfaitement inélastique. Les firmes porteront logiquement leur production à un niveau correspondant au minimum de ces deux valeurs que sont Y° et Y_d, noté min $(Y°,Y_d)$. Ces deux courbes se coupent pour un salaire v_0, qui assure donc l'équilibre momentané sur le marché des biens (point A sur la figure 7.15). Pour les autres valeurs du salaire réel, le déséquilibre prévaut : *offre excédentaire momentanée* si $v < v_0$, *demande excédentaire momentanée* si $v > v_0$[1].

Si l'on passe du momentané au court terme, les entreprises ne maintiendront pas l'emploi à N° mais l'ajusteront en fonction d'un objectif cible que les auteurs identifient au minimum de Y_s et Y_d. Cette façon de privilégier le *côté court* de l'offre et la demande plutôt que le point d'égalité, déjà opéré ci-avant entre Y° et Y_d, est typique des modèles de déséquilibre.

Graphique 7.15 : offre excédentaire et demande excédentaire chez Solow et Stiglitz



Sur le graphique 7.15, la ligne brisée en pointe, en caractère gras, représente le minimum de Y_s et Y_d. Les entreprises veulent migrer de la droite horizontale Y° vers cette courbe. Cela signifie

[1] Ici, les auteurs appellent « offre excédentaire, » ce que d'autres qualifieraient de demande excédentaire négative.

que si le salaire réel est compris entre v_0 et v", elles augmenteront la production et pour ce faire embaucheront du personnel. Par contre, si $v < v_0$ ou $v > v"$, l'offre actuelle Y° est au-dessus de l'objectif et elles réduiront la production et l'emploi. A partir d'une offre momentanément excédentaire, les entreprises réduisent l'emploi ; à partir d'une demande momentanément excédentaire, elles l'augmentent, sauf dans le cas où le salaire est tellement élevé (supérieur à v") qu'il rend la production non rentable au-delà de Y_s.

Les auteurs avouent leur gêne quant à la situation paradoxale de l'offre excédentaire à court terme : le prix y excède le coût marginal puisque la production est inférieure au niveau qui égalise le coût marginal et le prix, niveau correspondant à la courbe d'offre globale Y_s. « Each producer could increase his profits by selling more at the going price and, as a perfect competitor, he ought to try to do so, and he ought to succeed. But all producers together can sell no more than Y_d for the going real wage »[1]. Ils sont réticents à la suggestion d'Arrow de considérer qu'une industrie en situation d'offre excédentaire est nécessairement en concurrence imparfaite ; comme Patinkin, ils se contentent de considérer l'insuffisance de la demande globale comme une « force majeure ».

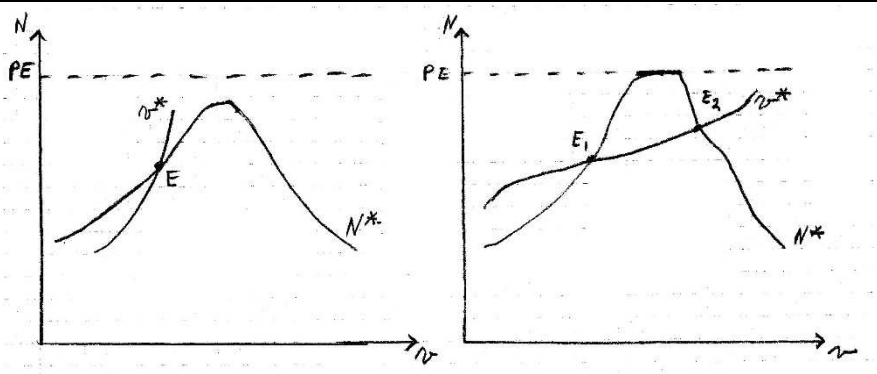
Voyons maintenant l'équilibre du système. Dans le graphique 7.16, Solow et Stiglitz considèrent l'espace (N,v) et y tracent deux courbes : la courbe N* où il n'y a pas de pression à la hausse ou à la baisse de l'emploi et la courbe v* où c'est la variable *v* qui est au repos. La courbe N* est déduite de la courbe en pointe comportant les points A, B et C (graphique 7.15), par le changement d'ordonnée $N = f(Y)$; naturellement, la forme est un peu similaire, moyennant la réserve que la courbe N* est écrêtée si et seulement si elle atteint le plein emploi (PE). Aux points situés au dessus de la courbe N*, l'emploi subit une pression à la baisse ; en dessous, une pression à la hausse.

La relation entre *N* et *v* est soumise à deux influences opposées : d'une part, un niveau d'emploi élevé va de pair avec le salaire réel élevé ; d'autre part, la hausse du salaire réel est source d'inflation et si cette tendance est forte, elle peut pousser le salaire réel à la baisse. La courbe v*, résultante, peut donc être croissante ou décroissante, mais les auteurs estiment la croissance plus probable. A la gauche de v*, le salaire réel est poussé à la hausse, à sa droite, il est poussé à la baisse.

L'équilibre du système procède évidemment de l'intersection entre les courbes v* et N*, soit le point *E*. Il s'agit d'un équilibre à court terme. On voit clairement qu'en fonction de la forme de ces deux courbes, l'intersection peut être unique ou multiple. Dans ce dernier cas, certains équilibres seront stables et d'autres instables. « Except in the very special case that the intersection occurs right at the cusp of the (N*) curve, the commodity market is not cleared at "equilibrium" »[2]. Comme le montre le graphique 7.15, à gauche de la pointe, le marché des biens révèle une offre excédentaire et à droite une demande excédentaire. Dans le premier cas, le chômage est plus que probable ; dans le second, le chômage est possible sauf s'il empêche le salaire nominal d'augmenter au même rythme que l'inflation.

[1] Solow & Stiglitz [343] p. 542
[2] Solow & Stiglitz [343] p. 550

Graphique 7.16 : l'équilibre de l'emploi et du salaire réel



Deux exemples parmi d'autres d'intersection entre les courbes N et v* : unique à gauche, double à droite.*

Dans la Théorie Générale, Keynes restait fidèle à l'idée traditionnelle que la hausse de l'emploi ne peut s'obtenir qu'au détriment du salaire réel. Mais les données empiriques ne confirment pas ce point de vue. Solow et Stiglitz se félicitent que leur modèle est compatible avec cette relation inverse mais ne l'implique pas.

7.5.4. L'école française du déséquilibre

La théorie du déséquilibre stimula des vocations en France : J-M. Grandmont, G. Laroque, Y. Younès, J-P. Benassy, E. Malinvaud ainsi que J. Drèze en Belgique. Question vocabulaire, le terme *déséquilibre* s'efface progressivement au profit des *équilibres non walrassiens*. Cette appellation s'explique par le fait que les modèles renoncent à deux canons walrassiens importants :

- l'équilibre n'implique pas que les marchés soient *soldés*[1]
- les agents sont sensibles aux signaux quantitatifs pour fixer offre et demande

Comme chez Walras, l'équilibre reste une caractéristique du système en ce sens qu'il représente la compatibilité des plans des agents mais ici les plans sont contraints.

BENASSY, LE RATIONNEMENT ET L'EQUILIBRE « FIXPRICE »

Benassy produisit de nombreux modèles d'équilibre non walrassien Son chapitre « Non-Walrasian Equilibria » du « Handbook of Monetary Economics » en donne une synthèse.

Par hypothèse, certains marchés ne sont pas soldés. Soit l'offre y dépasse la demande, soit la demande y dépasse l'offre. L'une des deux est donc ***rationnée***. Malgré cela, les TRANSACTIONS sur un marché s'équilibrent toujours : les achats égalent nécessairement les ventes. Dans les équations, l'ACHAT et la VENTE apparaissent donc comme des variables distinctes de l'OFFRE et la DEMANDE.

Le *rationnement*[2] de l'offre ou de la demande peut prendre diverses formes, ce que Benassy appelle des *schémas de rationnement*. Il en existe toute une variété : répartition uniforme, queues (les premiers arrivés sont servis), répartition proportionnelle à l'offre ou la demande annoncées… Chaque marché a son organisation particulière. Benassy exclut de son analyse l'échange forcé, les marchés inefficients où des acheteurs et des vendeurs sont rationnés

[1] Les auteurs anglo-saxons utilisent l'expression « to clear the market » pour désigner l'égalisation de l'offre et la demande sur un marché. Je la traduis par « solder le marché ».

[2] Le RATIONNEMENT est le concept central de l'économie du déséquilibre. Son acception y est évidemment plus large que l'acception commune. Par exemple, le chômage est un rationnement de l'offre de travail.

simultanément ainsi que la manipulation par laquelle un agent parvient à atténuer le rationnement qu'il subit en déclarant une offre ou une demande plus élevée.

On appelle *côté court* du marché celle de l'offre ou de la demande qui est inférieure à l'autre au prix fixé ; le *côté long* est celle qui dépasse l'autre. La *règle du côté court* stipule que les échanges se montent à la quantité correspondant au côté court ; les agents du côté court réalisent donc leurs plans sans contrainte à l'opposé de ceux du côté long qui souhaitent acheter ou vendre plus au prix en vigueur. Cette règle prévaut nécessairement dans les conditions postulées ci-avant.

Le modèle comporte *n* individus indicés *i* et *l* marchés indicés *h*. Sur chaque marché, l'individu *i* a une demande excédentaire $z_{ih} = d_{ih} - s_{ih}$ (excès positif ou négatif de sa demande sur son offre) et il passe des transactions pour des quantités $z^*_{ih} = d^*_{ih} - s^*_{ih}$ (excès de ses achats sur ses ventes).

Pour définir sa demande, l'agent s'appuie sur le système de prix *p* et de salaires *w*. Mais il est également sensible à certains signaux quantitatifs. Compte tenu du schéma de rationnement en vigueur, les offres et demandes des autres agents créent une limite à la capacité d'acheter ou de vendre de l'agent *i*. On a la contrainte perçue $\overset{\circ}{d}_{ih}$ qui plafonne les achats possibles de l'agent *i* sur le marché *h* et la contrainte perçue $\overset{\circ}{s}_{ih}$ qui y plafonne ses ventes. Le côté court s'écrit $z^*_{ih} = \min(z_{ih}, \overset{\circ}{d}_{ih})$ si *i* est demandeur et $z^*_{ih} = \max(z_{ih}, -\overset{\circ}{s}_{ih})$[1] s'il est un offreur. Les contraintes quantitatives sont une fonction de l'ensemble des demandes excédentaires des autres agents sur ce marché ; on a : $d^\circ_{ih} = G_i^d(z_{-ih})$ et $s^\circ_{ih} = G_i^s(z_{-ih})$, où *-i* représente l'ensemble des autres agents.

Après avoir précisé les ingrédients du modèle, Benassy analyse l'*équilibre à prix fixe* (« fixprice »), où le prix est imposé comme donnée exogène. Il appelle *demande effective* de l'agent *i* le vecteur de demande adapté pour tenir compte des contraintes $\overset{\circ}{d}_{ih}$ et $\overset{\circ}{s}_{ih}$, le résultat de la « deuxième décision » clowerienne en quelque sorte. Elle s'obtient comme la solution à la maximisation de l'utilité moyennant les contraintes du budget et du rationnement :

$$\max U_i(\mathbf{e_i} + \mathbf{z_i}, m_i) \quad \text{st} \quad (7.54)$$
$$\mathbf{pz_i} + m_i = m_i^\circ$$
$$\overset{\circ}{s}_{ik} \leq z_{ik} \leq \overset{\circ}{d}_{ik} \quad \text{où } k \neq h$$

L'argument de la fonction d'utilité comporte à la fois l'allocation après échange et l'encaisse finale. La contrainte budgétaire est bien connue, faisant intervenir l'encaisse avant et après l'échange. La contrainte de rationnement impose que l'agent, lorsqu'il offre ou demande sur le marché *h*, tienne compte des limites rencontrées sur les AUTRES marchés et qu'il procède ainsi sur chaque marché. La contrainte rencontrée sur le marché *h* n'influence donc pas la demande de l'agent sur ce marché, alors qu'elle influence sa demande sur tous les autres marchés. Benassy montre que sans cette spécification, le problème aurait une infinité de solutions.

Benassy démontre qu'un *équilibre à prix fixe*s défini par les comportements et les contraintes explicités ci-dessus existe et qu'il a même de fortes chances d'être unique. La qualification d'EQUILIBRE se justifie vu la possibilité simultanée pour tous les agents d'atteindre la situation qui maximise leur utilité moyennant les contraintes.

Le commissaire-priseur est évidemment absent, car sinon il n'y aurait pas de déséquilibre. Par ricochet, il faut renoncer à l'hypothèse des agents *preneurs de prix*. Benassy arrange alors son

[1] Dans le cas de l'offre, les valeurs de s_{ih} et s°_{ih} sont négatives.

modèle pour introduire les *faiseurs de prix* : « At this stage, the theory is still in need of a description of decentralized price making by agents internal to the system (…) We shall deal here with a particular organization of the pricing process where agents on one side of the market (most often the suppliers) quote price and agents on the other side act as price takers »[1]. Le *price maker* agit comme monopoleur sur son marché. Cette situation est fort proche de la concurrence monopolistique. L'offreur faiseur de prix fait face à une demande déclinante[2]. Ignorons le cas peu réaliste où l'offreur connaît objectivement sa demande. Il se construit alors une *demande subjective* $S_{ih}(p_i,\theta_i)$ où θ_i représente les paramètres de cette demande, que l'agent infère des signaux prix-quantité qu'il a observés empiriquement jusqu'à présent. Les contraintes d°_{ih} et s°_{ih} refont leur apparition, mais il y en a autant qu'il y a de prix possibles. Benassy imagine l'établissement de l'équilibre en deux étapes : le faiseur de prix choisit le prix qui maximise son utilité ; ensuite, la demande excédentaire *z*, la quantité échangée z* et les contraintes d° et s° se règlent selon la méthode *fixprice*. Les conditions d'existence de ce type d'équilibres sont examinées dans des articles antérieurs.

La compatibilité des plans contraints des agents est assurée, mais l'*équilibre à prix fixes* ne sera pas *Pareto-optimal*, sauf si par hasard il correspond à l'équilibre walrassien. Souvent, il sera même en-deçà d'un optimum dégradé intégrant les contraintes de rationnement ; notamment lorsque de nombreux marchés sont en suroffre, comme c'est le cas en dépression keynésienne. Potentiellement, il existe un espace pour des transactions qui amélioreraient le bien-être des parties, mais sans la double coïncidence des besoins, ces transactions ne se réaliseront pas. Il faut que les travailleurs inemployés soient productifs pour certaines firmes et que les biens surproduits soient utiles à certains consommateurs.

MALINVAUD REEXAMINE LA THEORIE DU CHOMAGE

Le livre de Malinvaud « Réexamen de la théorie du chômage », paru en 1977, est un des ouvrages phares de la macroéconomie du déséquilibre.

Malinvaud adhère à la *méthode à prix fixe* et prend soin de l'expliquer et de la justifier au début de son ouvrage. L'*ajustement par la quantité* est le corollaire de la rigidité des prix et des salaires nominaux. Lorsque l'offre ou la demande diminuent, l'agent, ne pouvant adapter son prix, reporte l'ajustement sur une variable quantitative : la production, le temps de travail, l'emploi… Malinvaud cite encore d'autres variables : les délais de livraison, les files d'attente, les stocks… Ces dernières, moins critiques, jouent le rôle de tampons. L'ajustement par les prix n'est pas nié, mais « à ce niveau, l'adaptation des prix reste très lente par rapport à ce qu'exigerait une égalisation permanente des offres et des demandes »[3].

Cette rigidité n'est pas simplement une anomalie rendant l'équilibre irréalisable. Elle est une caractéristique inévitable de la vie économique réelle qu'il convient d'intégrer dans les modèles. Elle correspond à un comportement rationnel des agents qui tentent de réduire l'incertitude du futur. Tant pour les employeurs que pour les employés, il peut être avantageux de créer des liens durables stipulant des conditions salariales indépendantes des fluctuations de l'offre et la demande de travail au jour le jour[4]. Le désir de ne pas vendre à perte est également un frein à la flexibilité des prix. Un peu agacé par les critiques exigeant toujours des

[1] Benassy [30] p. 131.
[2] Dans ce schéma, le *price maker* considère les prix sur les autres marchés comme une donnée fixe.
[3] Malinvaud [245] p. 11
[4] Dans une économie purement walrassienne où l'offre et la demande s'égalisent EN PERMANENCE, les contrats de travail auraient une durée unitaire, une journée ou à peine plus. Les inconvénients pratiques nuiraient à la productivité des entreprises et à l'utilité des ménages.

justifications à la rigidité des prix, Malinvaud remarque malicieusement qu'il est beaucoup plus difficile de justifier la flexibilité walrassienne, dépendante de l'hypothèse irréaliste du commissaire-priseur.

L'un des apports marquants de l'ouvrage de Malinvaud est l'établissement d'une typologie des équilibres avec rationnement. Il distingue trois types d'équilibre :

		Marché des biens	
		Offreurs rationnés	Demandeurs rationnés
Marché du travail	Offreurs rationnés	Chômage keynésien	Chômage classique
	Demandeurs rationnés		Inflation contenue

En situation de **chômage keynésien**, il y a interdépendance entre le rationnement des entreprises vendant sur le marché des biens et le rationnement des particuliers vendant leur travail. Une réduction du salaire ne peut résoudre le problème parce que la baisse du salaire réduirait le revenu des travailleurs et donc la demande globale et aggraverait le rationnement des vendeurs sur le marché des biens, qui est la cause profonde du chômage. Le concept de « **chômage classique** » nous ramène à la loi de Walras : celle-ci implique qu'un déséquilibre sur le marché du travail soit compensé par un déséquilibre inverse sur les autres marchés. L'**inflation contenue** se caractérise par une demande excédentaire sur tous les marchés. La pression inflationniste existe, mais l'hypothèse de rigidité des prix maintient les prix à un niveau trop bas. Quant au quatrième cas, l'association d'une offre excédentaire de biens et une demande excédentaire de travail, il n'a aucune chance de survenir.

Malinvaud échafaude un modèle économique qui envisage ces trois types d'équilibre avec rationnement. A partir des équations caractéristiques de chacun d'eux, il évalue le résultat d'une part de la POLITIQUE BUDGETAIRE et d'autre part d'une POLITIQUE DE REVENU tentant d'influencer les prix et les salaires nominaux. La POLITIQUE MONETAIRE ne peut être évaluée, car à des fins simplificatrices, le modèle ne prête aux agents qu'un type d'actif, la monnaie ; il ignore donc l'effet du taux d'intérêt sur la répartition des portefeuilles. Contentons-nous de survoler les conclusions.

Effet des dépenses publiques :

- **Chômage keynésien** : elles font augmenter le revenu et baisser le chômage, comme l'affirme la théorie keynésienne traditionnelle.
- **Chômage classique** : elles aggravent le rationnement sur les marchés des biens. De ce fait, les actifs réduiront leur offre de travail. En conséquence, les dépenses publiques réduisent le chômage sans augmentation de la production.
- **Inflation contenue** : à l'inverse du cas précédent, une réduction des dépenses publiques rend possible l'accroissement de la consommation, qui stimulera l'offre de travail, donc l'emploi et réduira ainsi l'excès de demande sur les deux marchés.

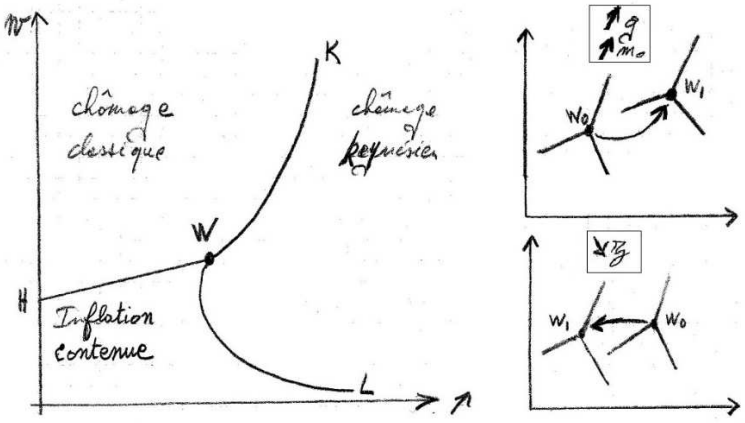
Politique des prix et des salaires

- La lutte contre le **chômage classique** demande que le rapport salaire/prix diminue alors que la solution du **chômage keynésien** implique une baisse des prix plus rapide que celle des salaires nominaux. Ces deux maux nécessitent donc des politiques inverses. Pendant des décennies, les économistes se sont divisés sur la question de la baisse des salaires en tant que solution au chômage. Les uns insistaient sur l'incitation à l'embauche, les autres sur l'affaiblissement de la demande globale. En réalité, note Malinvaud, la conclusion dépend du type de chômage à combattre.
- **Inflation contenue** : la libération des prix et salaires de leur corset de rigidité peut aider à tempérer la demande excédentaire. Il en résultera de l'inflation. L'utilité des

consommateurs pourrait s'en trouver réduite, ce qui explique le soutien populaire aux politiques de contrôle des prix.

Malinvaud établit une cartographie des trois types d'équilibre en localisant leur domaine dans le plan prix (p)-salaires (w). C'est l'objet du graphique 7.17. Le point *W*, au centre, représente la combinaison prix-salaire qui assure l'équilibre walrassien. Le chômage classique prévaut lorsque le salaire est trop élevé par rapport au prix, l'inflation contenue lorsque les prix et salaires sont tous les deux trop bas par rapport à l'encaisse détenue par les agents et le chômage keynésien lorsque les prix sont trop élevés. Les segments WH, WK et WL sont déduits des équations du modèle. Le long de WH, la demande excédentaire vaut exactement zéro sur le marché du travail ; le long du segment WK, la demande excédentaire de biens est nulle.

Graphique 7.17 : les domaines des trois types d'équilibre non walrassien



En fait, *p* et *w* ne sont pas les seules variables exogènes du modèle ; il y encore l'encaisse initiale (m_0), le besoin technique de main d'œuvre par unité de produit (z) et la dépense « autonome » (g) englobant l'investissement et les dépenses publiques. La partie droite du graphique 7.17 montre comment l'ensemble de l'espace prix-salaire se repositionne quand les autres variables exogènes se modifient. Si l'investissement ou la dépense publique augmentent, l'équilibre et les frontières se déplacent vers le nord-ouest. Prenons l'exemple d'une économie qui est à l'équilibre walrassien et supposons la demande autonome augmente alors que les niveaux des prix et des salaires restent inchangés. Cette économie entre alors dans une situation d'inflation contenue.

Une opinion répandue parmi les économistes est que le modèle keynésien décrit correctement la réalité à court terme lorsque la rigidité des prix et salaires fait son œuvre mais qu'à long terme, les prix finissent toujours par s'ajuster et que la théorie classique est alors plus pertinente. Malinvaud conteste cette opinion. Il met en garde contre la tendance à surestimer la faculté d'ajustement des prix à long terme. En outre, constate-t-il, le chômage keynésien se présente plus fréquemment dans la réalité économique que le chômage classique ; le chômage s'accompagne le plus souvent d'un excès d'offre sur le marché des biens. Parmi les raisons de cette prédominance, il y a les fluctuations du taux d'investissement entre un pic en haute conjoncture et un creux en basse conjoncture. Si la constellation prix salaires reste constante, l'économie fait alors un mouvement de va-et-vient entre l'inflation contenue lorsque *g* est élevé et le chômage keynésien lorsque *g* est faible. Le chômage classique est absent de ce mouvement.

7.5.5. Conclusions quant à l'école du déséquilibre

LES CRITIQUES

La théorie du déséquilibre est assise sur un contradiction qui lui fut souvent reprochée : travailler à la walrassienne, mais incorporer des éléments incompatibles avec la théorie walrassienne. Les auteurs n'ont pas tous assumé ce paradoxe de façon cohérente. Patinkin s'inscrit dans l'orthodoxie walrassienne impliquant le commissaire-priseur dans les chapitres qui précèdent, puis, sans explication, en efface l'existence dans le chapitre consacré au déséquilibre. Clower réfute explicitement la simultanéité walrassienne de toutes les décisions, mais il n'explique pas comment fonctionne institutionnellement son système de marché.

Contre Malinvaud qui considérait comme une évidence empirique que la rigidité des prix met de nombreux marchés en déséquilibre, Lucas rétorque que l'équilibre et le déséquilibre sont des caractéristiques, non pas des marchés mais des représentations qu'en fait la science économique. Quel observateur, demande-t-il, peut constater que tel marché réel est ou non en équilibre ? Comme nous le verrons lorsque nous aborderons le cycle conjoncturel, Lucas semble attaché à l'idée que tous les marchés y compris le marché du travail restent généralement à l'équilibre.

LES SUITES

L'école du déséquilibre a décliné dès le milieu des années soixante-dix, à l'exception de sa partie française. L'immense majorité des économistes est restée attachée au dogme walrassien. Les tenants du déséquilibre ne renièrent pas leur œuvre mais cessèrent de s'investir dans la construction de ce paradigme alternatif. Mentionnons toutefois trois contributions de Clower et Leijonhufvud qui font œuvre de testament :

- En 1973, Leijonhufvud consacre un article aux *échecs de coordination*. Dans l'optique orthodoxe, l'équilibre walrassien est un modèle de coordination. Leijonhufvud en relève les ratages, qu'il appelle *coordination failures* ou *effective demand failures*. Ce concept est à rapprocher de l'idée clowerienne déjà rencontrée, selon laquelle la demande notionnelle est insuffisamment signalée au marché. « This failure of the markets to transmit messages about desired transactions from one side to the other is what we mean by the phrase *effective demand failure* »[1].

- Clower et Leijonhufvud considèrent le commissaire-priseur comme un élément à éliminer de la théorie économique. Mais par quoi le remplacer ? Les modèles économiques, disent-ils, doivent avoir l'honnêteté de décrire la technologie des échanges qui les sous-tend. Dans un article coécrit en 1975, ils recherchent cette alternative. Leur solution admet l'échange aux *prix erronés* et rend possibles les contrats bilatéraux entre acheteurs et vendeurs moyennant un coût de transaction. Mais les auteurs ne peuvent se passer de l'intervention d'un superviseur de marché. Inutile de dire que leur proposition fut superbement ignorée.

- Depuis le début des années soixante-dix, les macroéconomistes néoclassiques sont de plus en plus préoccupés par les *microfondations*, c'est-à-dire que les comportements macroéconomiques prêtés aux agents doivent être justifiés par une analyse des décisions individuelles en termes de choix rationnel. Renonçant au canevas walrassien, Leijonhufvud oriente sa recherche de microfondations plutôt du côté de Marshall. Dans un article paru en 1974, il montre que Keynes ne peut être compris qu'à partir de prémisse marshalliennes. La

[1] Leijonhufvud [216] p. 41

microéconomie marshallienne se révèle aussi plus adéquate à l'analyse des échecs de coordination, mais elle doit être combinée avec des éléments walrassiens : par exemple, la demande effective est marshallienne mais la demande notionnelle est walrassienne.

7.6- LES POST KEYNESIENS

Les économistes keynésiens du présent chapitre sont considérés comme les fondamentalistes de ce courant de pensée. Ils dénoncent d'ailleurs sans relâche les fausses interprétations de la pensée keynésienne qu'on retrouve chez tant d'économistes.

Il n'y a pas à proprement parler de théorie post-keynésienne. Il s'agit simplement de la pensée de Keynes. Le seul écart est la thèse de l'offre de monnaie endogène, qu'ils partagent unanimement. Nous avons vu que dans la Théorie générale, Keynes juxtapose deux exposés sur la monnaie :

- les chapitres treize et quinze mettent en scène un marché de la monnaie où le taux d'intérêt s'ajuste pour équilibrer une offre de monnaie exogène avec une demande de monnaie guidée par la préférence pour la liquidité.
- Le chapitre dix-sept, plus subtil, jette les bases d'une espèce d'équilibre général entre les marchés des différents actifs. Ce chapitre est compatible aussi bien avec une offre de monnaie endogène qu'avec une offre exogène. Il inspirera nombre d'économistes post-keynésiens.

Parmi les post-keynésiens, on peut distinguer deux sous-groupes :

- les Cambridgiens, comme J. Robinson, L Pasinetti, N. Kaldor, généralement adeptes d'une microéconomie ricardienne et déjà connus du lecteur.
- Les post-keynésiens américains, principalement Paul Davidson, Hyman Minsky, Sydney Weintraub.

Paul **Davidson** est présent sur le front des débats macroéconomiques depuis le début des années soixante ; sa bibliographie est impressionnante. Voyons quelques thèmes qu'on retrouve de façon répétée dans son œuvre.

D'abord, l'insistance sur l'incertitude. Et pas n'importe quelle incertitude : celle qu'il appelle *non ergodique*. Ainsi qualifie-t-on un processus dont la connaissance du comportement passé n'offre pas le pouvoir statistique de prévoir le comportement futur. Keynes la concevait de la sorte. Entre les prémisses d'une action économique et ses conclusions s'écoule un laps de temps pouvant être long, qui nous conduit, selon les termes de Davidson, « from an irrevocable past to an uncertain future »[1].

Sans l'incertitude, la monnaie est incompréhensible. Avant Keynes, les économistes l'avaient greffée sur des modèles où sa présence n'avait pas de raison d'être. La monnaie ne tombe pas comme une manne céleste. Elle est profondément ancrée dans tous les mécanismes économiques. Ce n'est pas la monnaie qu'il faut analyser mais l'économie monétaire. Celle-ci se caractérise principalement par le fait que les transactions économiques font l'objet de contrats (achat-vente, prêt, contrat de travail) dont la monnaie sert d'unité de compte et que cette monnaie est acceptée comme contrepartie valable dans l'échange : elle a le pouvoir de libérer le débiteur de ses obligations. Cette institution permet à chaque partie de connaître le poids des droits et obligations futurs qui naissent de ses contrats. La monnaie réduit l'incertitude des contractants. « In the absence of money contracts, it is unlikely that entrepreneurs, facing a statistically unpredictable and unknowable future, would undertake large scale long-lived, complex production processes »[2]. L'Etat joue un rôle important dans

[1] Davidson [66] p. 562
[2] Davidson [67] p. 154

l'économie monétaire, car le droit et la puissance publique offrent leur appui pour forcer les parties à respecter leurs obligations.

Davidson souligne le statut particulier du salaire nominal. Dans le système walrassien, le niveau absolu des prix est indéterminé, car le marché du travail est traité comme les autres marchés. Or le salaire n'est pas le fruit du tâtonnement walrassien, il est l'objet de négociations collectives et de règlementations ; ces négociations portent sur le salaire NOMINAL. Un modèle économique réaliste traiterait le salaire nominal comme une donnée exogène. Grâce à lui, le niveau général des prix devient déterminé, car tous les autres prix s'ajustent en fonction de lui.

Dans l'article « Money as Cause and Effect », Davidson et Weintraub proposent un modèle conçu dans cette optique. L'économie y subit un double choc exogène : l'autorité politique crée de la monnaie additionnelle ΔM et, parallèlement, une négociation salariale impose une hausse du salaire nominal *w*. La question, posée en termes fishériens, est celle-ci : comment cette hausse de *M* se partagera-t-elle entre la variation de la quantité produite et celle du niveau des prix ?

Les variables du modèle sont l'emploi *N*, la production en volume *Q*, le niveau des prix P, la productivité moyenne A = Q/N et le revenu nominal Y = QP.

La firme fixe le prix du produit *i* selon un procédé de mark up semblable à celui de Kalecki, que les auteurs agrègent ensuite. On a :

$$P_i = k_i w/A_i \quad \rightarrow \quad P = kw/A \qquad (7.55)$$

Les auteurs écrivent : "A variety of studies reveals *k* to be remarkably constant in Gross Business Product data in which the concept of the market price level has bearing, although there may be a secular trend…"[1].

La demande du produit *i* (7.56) est fonction de l'ensemble des prix et du revenu *Y*.

$$D_i = D(P^1,\ldots P^n, Y) \qquad (7.56)$$
$$Y = PQ\ ;\ D_i = D(P^1,\ldots P^n, kwN) \qquad (7.57)$$
$$M^D{}_t = M^D(P_t,Q_t,r_t,B,P_e,r_e) \qquad (7.58)$$

La demande de monnaie au temps *t*, $M^D{}_t$, dépend du niveau de prix, du volume de la production et du taux d'intérêt en *t*, du montant des obligations *B* ainsi que du niveau des prix et du taux d'intérêt futurs anticipés ; B,P_e,r_e sont considérées comme exogènes.

L'accroissement autonome de M^S par les autorités monétaires via l'open market policy modifie le prix des obligations, donc le taux d'intérêt. De par (7.58), il en résulte un ajustement de M^D. Parallèlement, il se produit une Δw accompagnée ou non d'une évolution de la productivité ΔA. Vu la formule (7.55), le niveau des prix *P* se modifie.

L'accroissement du coût *kwN* entraîne normalement un accroissement correspondant du capital circulant, donc du besoin de financement des entreprises et de leurs emprunts aux banques. L'accroissement monétaire, à peine existant, est mis à contribution, drainé. Selon que ΔM^S soit supérieur, égal ou inférieur au besoin suscité par l'augmentation du besoin de liquidités (ΔM^D), le taux d'intérêt, diminuera, sera constant ou augmentera. Les auteurs en concluent : "Hence, changes in the money-wage numeraire relative to productivity require equal changes in the money supply, if money is to remain neutral with respect to output. Prices however will be

[1] Davidson & Weintraub [69] p. 1125

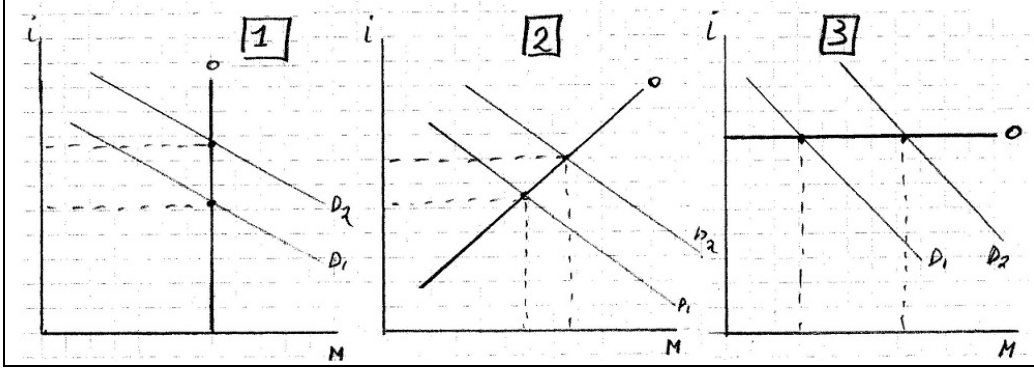
higher by the excess of average money-wage change over the productivity change"[1]. Seule la partie de l'élargissement monétaire qui excède Q.ΔP, en plus d'un éventuel accroissement des encaisses de précaution et de spéculation, est susceptible de pousser le taux d'intérêt à la baisse et donc de stimuler la production. L'effet de Δw est dépendant de l'écart Δw-ΔA ; celui-ci est influencé par les rendements d'échelle lorsque Q varie à la suite de ΔM^S.

En vertu de l'équation (7.55), « so long as the money wages are determined around a bargaining table, monetary policy can have only limited control over the price level »[2].

Les auteurs pensent qu'une hausse du salaire nominal sera presque toujours suivie d'une augmentation de la masse monétaire, ce qui nous amène à la problématique de la monnaie endogène. Dans l'article « Endogenous Money, the Production Process and Inflation Analysis », Davidson décortique « the dispute whether changes in (bank) money supplies play a causal part in influencing economic activity or whether variations are an effect of economic activity »[3], un débat qui remonte très loin, puisque déjà présent dans la controverse entre le *Currency principle* et le *Banking principle*. Selon Davidson, ce débat recouvre deux sous-questions distinctes :

1- Quel est le degré d'élasticité de l'offre de monnaie par rapport au taux d'intérêt ?
Sur la figure 1 du graphique 7.18, l'offre de monnaie est VERTICALE, comme elle l'est sur le graphique 5.2 de Keynes. Les monétaristes et les nouveaux classiques sont également *verticalistes*. La quantité de monnaie offerte est déterminée par l'autorité monétaire indépendamment du taux d'intérêt. Une augmentation de la demande de monnaie de D_1 à D_2 n'affecte pas la masse monétaire en circulation mais seulement le taux d'intérêt.

Graphique 7.18 : l'élasticité de l'offre de monnaie



Sur la figure 2, l'offre de monnaie croît avec le taux d'intérêt, peut-être parce que les banques sont plus enclines à prêter lorsque le taux est plus rémunérateur. Une hausse de la demande de D_1 à D_2 augmente la masse monétaire. La figure 3 représente l'autre cas extrême. L'autorité n'y détermine que le taux d'intérêt : c'est la demande de monnaie qui détermine son abondance. L'économiste *horizontaliste* le plus connu est Kaldor.

2- L'offre de monnaie, est-elle indépendante de la demande ?
Un déplacement exogène de la courbe de demande de monnaie, entraîne-t-il un déplacement de la courbe d'offre dans son sillage ? Ce critère de distinction et le précédent ne se recouvrent pas exactement : on peut concevoir une offre verticale mobile sur l'axe

[1] Davidson & Weintraub [69] p. 1129
[2] Davidson & Weintraub [69] p. 1131
[3] Davidson [67] p. 155

horizontal comme on peut imaginer une offre oblique ou horizontale rigidement cadenassée. En définitive, l'offre de monnaie ne sera exogène qu'au prix de conditions très restrictives : elle doit être à la fois parfaitement verticale et indépendante de la demande.

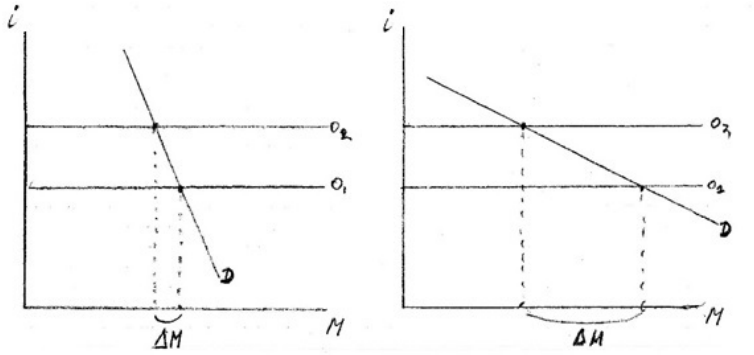
D'après Davidson, l'économie réelle comporte deux mécanismes susceptibles de créer de la monnaie ; tantôt l'un est à l'œuvre, tantôt l'autre.

1- le mécanisme ENDOGENE : un surplus d'activité productive accroît la demande de prêts bancaires. Il est dans l'intérêt des banques d'y répondre favorablement. Si elles peuvent y répondre sans le secours de la banque centrale, l'offre de monnaie est forcément élastique. Si elles doivent recourir à la banque centrale et que celle-ci consent à leur fournir des réserves, il s'ensuivra un déplacement latéral de l'offre de monnaie vers la droite. La responsabilité d'éviter une crise de liquidité doit motiver la banque centrale à intervenir mais la décision dépend de divers paramètres guidant sa politique.

2- Le mécanisme EXOGENE : la banque centrale prend l'initiative d'acheter des actifs aux banques pour les inciter à accorder de nouveaux prêts au secteur non monétaire. Ceci implique un déplacement AUTONOME vers la droite de la courbe d'offre de monnaie.

Restons dans le même domaine, mais changeons d'économiste. **Kaldor** analyse comment l'élasticité de la demande de monnaie influence l'efficacité de la politique monétaire lorsque l'offre est horizontale. Comme le montre le graphique 7.19, si la demande est élastique, une plus faible variation du taux d'intérêt suffit à déclencher un afflux de liquidités que lorsqu'elle n'est que faiblement élastique.

Graphique 7.19 : l'effet de la politique monétaire dans le cas horizontaliste



En conséquence, « the more interest inelastic the demand for money is, the easier it is for the central bank to control the rate of interest and the more difficult it is to control the supply of money »[1].

Friedman considérait que l'évidence empirique d'une bonne corrélation entre la masse monétaire et le revenu national nominal confirme sa thèse que la demande de monnaie est une fonction stable du revenu nominal, peu affectée par d'autres facteurs. Ceci contredit la position keynésienne selon laquelle le taux d'intérêt influence considérablement la demande de monnaie. Kaldor voit en la monnaie endogène une explication alternative à cette observation empirique. La corrélation constatée révèle la séquence causale :

Δ revenu => Δ demande de monnaie => Δ offre de monnaie

[1] Kaldor & Trevithick [176] p. 8

Tout au long de son œuvre, Davidson rappelle les percées de la pensée de Keynes et dénonce les retours en arrière de ses contemporains. Selon lui, la révolution keynésienne rejetait trois postulats néoclassiques :

1- l'*axiome de la substituabilité brute* entre les produits et les actifs non reproductibles.
2- l'*axiom of reals* selon lequel seules les valeurs réelles influencent les agents, à l'exclusion des valeurs nominales
3- la vision d'un monde *ergodique*.

Il reproche aux monétaristes, aux nouveaux classiques et aux néo-keynésiens (qu'il fusionne sous l'appellation « mainstream macroeconomic theory ») de les avoir réintroduits et par là même de réinstaller la loi des débouchés de Say. Comme l'a montré Keynes, la substituabilité entre les actifs non reproductibles et les biens reproductibles a pour conséquence que si les premiers sont très demandés parce que la tendance est à l'épargne, la hausse de leur prix détournera les épargnants vers la reprise de la consommation. Cet axiome offre un parachute à la demande globale, qui n'a pas lieu d'être.

Davidson critique également la *théorie du déséquilibre*, qui fait découler le chômage d'un problème de coordination : les firmes ne reçoivent pas l'information qu'elles pourraient engager profitablement grâce à la *demande effective* qui monterait jusqu'à la *demande notionnelle*. Comme le montre le graphique 7.3, qui reflète bien la pensée de Keynes, en période déflationniste, la demande globale est fondamentalement insuffisante. Au plein emploi, elle est inférieure à l'offre globale. Les entreprises ne sont donc pas motivées d'embaucher les travailleurs jusqu'à ce niveau. La demande globale n'est jamais contrainte par le revenu mais par le fait que celui-ci n'est pas entièrement dépensé. « Spending is constrained only by liquidity and/or timidity considerations »[1].

[1] Davidson [66] pp. 566-567

7.7- LES NOUVEAUX KEYNESIENS

Les *nouveaux classiques* ont pu s'ériger en nouvelle orthodoxie à partir du milieu des années soixante-dix. Mais à partir de la seconde moitié de la décennie suivante, ils doivent partager ce statut avec un outsider : les *nouveaux keynésiens*. Aujourd'hui encore, ce duopole équilibré domine la macroéconomie. Les principaux représentants sont Joseph Stiglitz, Georges Akerlof et son épouse Janet Yellen, Stanley Fischer, Olivier Blanchard, Lawrence Summers, Gregory Mankiw et David Romer.

La nouvelle école tente de réhabiliter les principes keynésiens en les immunisant contre la critique des *nouveaux classiques*. Ceux-ci avaient re-walrassisé la macroéconomie en la connectant directement à la microéconomie walrassienne. Les *nouveaux keynésiens* reprochent aux anciens keynésiens d'avoir présumé la rigidité des salaires nominaux sans l'avoir expliquée, faisant ainsi le jeu de leurs adversaires. Mankiw écrit: « One of the reasons for the resurgence of the equilibrium approach to macroeconomics has been the absence of a theoretical underpinning for this Keynesian price stickiness »[1]. Les nouveaux keynésiens cherchent dans la microéconomie les éléments qui fondent les rigidités[2]. Egalement, ils en déplaceront le centre de gravité des salaires vers les prix des produits.

Romer considère que la synthèse néoclassique (en ce compris l'école du déséquilibre) est assise sur une contradiction fondamentale : d'une part, elle permet le déséquilibre de certains marchés ; d'autre part : « the remainder of the neoclassical synthesis was Walrasian**.** Markets for goods and labor were competitive, externalities were absent, and information was perfect (…) In an environment that is so relentlessly competitive, how can one glaring departure from Walrasian behavior persist? »[3].

7.7.1. Prix et salaires : rigidités réelles et rigidités nominales

Les rigidités nominales affectent le salaire nominal et les prix absolus. Les rigidités réelles concernent le salaire réel et les prix relatifs.

FISCHER : LES CONTRATS DE TRAVAIL A LONG TERME

Le plus ancien modèle (1977) s'intéresse à une cause de rigidité nominale des salaires : le fait qu'ils sont conclus pour des termes relativement longs alors que dans le monde walrassien, tous les salariés sont des journaliers, dont les salaires sont établis à la criée. Fischer présente son modèle ainsi: « Because the money stock is changed by the monetary authority more frequently than labor contracts are renegotiated, and, given the assumed form of the labor contracts, monetary policy has the ability to affect the short-run behavior of output, though it has no

[1] Mankiw [250] p. 529

[2] La rigidité des prix sur les marchés des produits fait l'objet d'études de plus en plus nombreuses qui se situent au carrefour de la micro et de la macroéconomies. Des études empiriques la mettent en évidence (Godley et Nordhaus, Rotemberg). De nombreuses explications théoriques sont avancées. Dans son article « On Sticky Prices : Academic Theories Meet the Real World » ([36]), Blinder en recense pas moins de douze et se livre à l'exercice d'effectuer un sondage parmi les dirigeants d'entreprises pour établir entre elles une espèce de classement de popularité. Sans que ce classement préjuge de la qualité réelle des théories, dévoilons le trio de tête :

1- Les firmes attendent que les autres firmes changent leurs prix en premier.
2- Les hausses de prix sont reportées jusqu'à ce que le coût augmente
3- Les firmes préfèrent modifier d'autres éléments comme le délai de livraison, le service ou la qualité du produit.

[3] Romer [306] p. 6

effects on long-run output behavior »[1]. Fischer prend soin d'inclure l'hypothèse des anticipations rationnelles dans son modèle. Ce qui lui permet d'en tirer la conclusion que cette hypothèse n'est pas déterminante dans les résultats des modèles des nouveaux classiques.

Comment justifier que les agents s'accommodent de cet élément de rigidité qui réduit leur bien-être ? « …the *transaction costs* of frequent price setting and wage negotiations must be part of the explanation ».

En fait, le modèle est décliné en deux versions. Dans la première, tous les salaires sont renégociés au début de chaque période. Durant les périodes, l'équilibre économique apporte son lot de variations aléatoires de la production et des prix, que ni les agents privés ni l'autorité monétaire ne sont capables de prévoir. L'autorité monétaire tente de les stabiliser mais se base sur des paramètres qui sont toujours une période en retard et sont donc connus. L'ayant compris, les agents anticipent correctement ses actions. Fischer montre qu'en pareille situation, l'offre globale n'est sujette qu'aux variations aléatoires ; elle est insensible aux variations de l'offre de monnaie.

Dans la deuxième version, les éléments sont identiques à cette exception près que les salaires sont fixés pour deux périodes et que le timing de la renégociation est décalé : une moitié des firmes négocie le salaire en t-1 et l'autre moitié en *t*. Chaque période, les salaires nouvellement renégociés tiennent compte des perturbations en cours, mais les salaires négociés lors de la période précédente, toujours en vigueur, ne peuvent pas se prémunir contre la politique monétaire, ce qui rend celle-ci efficace. La conclusion est: « The effectiveness of monetary policy does not require anyone to be fooled. In the model of section II, with two-period contracts, monetary policy is fully anticipated but because it is based on information which becomes available after the labor contract is made, it can affect output »[2].

Fischer examine également le cas des contrats qui prévoient l'indexation du salaire sur le niveau général des prix. Selon la méthode d'indexation prévue, la politique monétaire exercera ou non une influence.

SHAPIRO ET STIGLITZ : LES SALAIRES D'EFFICIENCE

Le présent modèle amène une rigidité réelle sur le marché du travail. Divers modèles s'attachent au *salaire d'efficience*, un salaire réel que les employeurs estiment de leur intérêt de fixer au dessus du niveau égalisant l'offre et la demande de travail, parce que ce surplus de revenu est un incitant pour les travailleurs à répondre à leurs attentes. Le modèle le plus connu est celui de Shapiro et Stiglitz, présenté dans leur article « Equilibrium Unemployment as a Worker Discipline Device » (1984).

La thèse est celle-ci: « …the inability of employers to costlessly observe workers' on-the-job effort can explain involuntary unemployment as an equilibrium phenomenon. »[3].

Dans la concurrence parfaite walrassienne, il n'y a jamais de chômage involontaire. Les employeurs n'ont donc aucun moyen de pression sur les tire-au-flanc. S'ils les licencient, ces derniers retrouveront immédiatement un nouvel emploi au même barème. L'idée de Shapiro et Stiglitz est donc que les employeurs offrent délibérément un salaire réel supérieur à celui qui

[1] Fischer [90] p. 191
[2] Fischer [90] p. 203
[3] Shapiro & Stiglitz [331] p. 433

solde le marché. Ce salaire rend l'offre supérieure à la demande, ce qui implique le chômage involontaire. Ainsi, la menace du chômage sert à dissuader les employés de carotter[1]. Ce système n'est pas nécessairement plus coûteux qu'un système d'observation permanent de l'effort fourni par chaque travailleur. « The equilibrium unemployment must be sufficiently large that it pays workers to work rather than to take the risk of being caught shirking ».

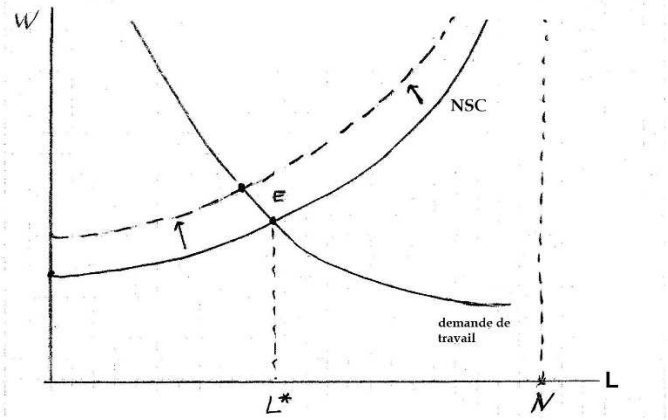
Le modèle proposé simplifie la réalité : le choix du salarié se limite à carotter ou fournir un effort fixe. Les travailleurs sont considérés comme identiques (pas de problème de réputation) et les firmes également (alors que selon le type de travail, le contrôle direct est sans doute plus ou moins aisé).

Les travailleurs et les employeurs veillent chacun à maximiser une fonction d'utilité multipériodes qui tient compte de nombreux paramètres :
- la probabilité de se faire attraper (et licencier) si l'employé carotte (q);
- la probabilité qu'un employé soit licencié pour des raisons ECONOMIQUES (b) ;
- le salaire, qui augmente l'utilité (w) et l'effort, qui la diminue (e);
- l'indemnité de chômage, exogène mais ici à charge de l'employeur (w°) ;
- le taux de chômage (u) ;
- le taux d'escompte de l'utilité future (r) ;
- la probabilité instantanée qu'un chômeur trouve un emploi, qui est l'inverse de la durée moyenne de chômage (a).

Nous n'exposons pas ici les équations qui sont assez complexes. Il est toutefois évident qu'on peut déduire de ces paramètres la grandeur *w**, le « critical wage for non shirking ». Celui-ci augmente avec *e*, *b*, *r*, *w°*, et *a* et diminue avec *q*. La « non shirking curve » (NSC) relie les valeurs de *w** pour les différents niveaux d'emploi. Elle est croissante, car plus on s'approche du plein emploi *N*, plus la dissuasion du carottage a besoin du salaire pour compenser la faiblesse de la menace du chômage (N-L).

L'équilibre du marché procède de l'intersection entre la demande agrégée de travail et la NSC agrégée.

Graphique 7.20 : le marché du travail avec salaire d'efficience



[1] Dans un autre article, Stiglitz distingue quatre motifs qui peuvent pousser les entreprises à payer un *salaire d'efficience* :
- diminuer la rotation du personnel ;
- attirer à elle les meilleurs éléments ;
- inciter les salariés à l'effort (le cas présenté ici) ;
- procurer aux travailleurs la satisfaction morale d'obtenir un salaire digne.

Le graphique 7.20 montre une élévation de la courbe NSC faisant suite à un accroissement de *e*, *b*, *r*, *w*°, *a* ou une diminution de *q*.

Sur le plan du bien-être, l'équilibre *E* est sous-optimal, car trop peu de travailleurs sont employés. En contrepartie, ce système épargne à l'économie un coût de monitoring qui peut être élevé.

IMPACT DE PREMIER ORDRE POUR DES FRICTIONS NOMINALES DE DEUXIEME ORDRE

En 1985 sont publiés deux modèles qui, partant de frictions nominales très légères produisent des variations très sensibles de la production totale. Les deux modèles comportent pas mal de ressemblances, mais ciblent des frictions nominales de natures différentes :

- **Akerlof** et **Yellen** mettent en avant un comportement des agents qui ne serait que *presque rationnel*.
- **Mankiw** s'attache aux *coûts de menu*. Ce terme, qui couvre les coûts réels que doit supporter une entreprise qui désire changer ses prix, fait référence aux menus que doit réimprimer le restaurant qui changerait ses prix.

L'effet de la presque-rationalité

Le modèle envisage une économie dont les firmes sont en position de *price maker*, tant sur le marché de leur produit qui fonctionne en mode concurrence monopolistique que sur le marché du travail où elles appliquent le système du *salaire d'efficience*. Les ventes de la firme représentative dépendent de deux facteurs : la demande globale et le prix relatif (rapport entre le prix de la firme et le niveau moyen des prix). La demande globale dépend du total des encaisses réelles détenues par le public.

Les autorités augmentent la masse monétaire. Face à cette nouvelle donne, les firmes qui maximisent leur profit adaptent immédiatement le prix et le salaire qu'elles fixent. Mais une fraction β des firmes, par inertie, laissent leurs prix inchangés au niveau antérieur qui correspondait à un équilibre à long terme. L'augmentation des encaisses réelles du public et de la baisse de leur prix relatif sont deux facteurs qui poussent la production de ces firmes et leur effectif à la hausse.

L'idée de base est que « inertial wage-price behavior by firms may not, in fact, be very costly; it may be near-rational. (…) Near-rational behavior is behavior that is perhaps suboptimal but that nevertheless imposes very small individual losses on its practitioners relative to the consequences of their first-best policy. Technically, *very small* is defined as being second-order in terms of the policy shocks that create disturbance from a long-run fully maximizing equilibrium »[1].

Mais les conséquences de la rigidité sur le bien-être social sont de *premier ordre*. Si le comportement presque rationnel était commun à toutes les firmes, l'augmentation des encaisses réelles équivaudrait à celle de l'accroissement monétaire et les autres variables réelles se modifieraient dans le même ordre de magnitude.

Akerlof et **Yellen** précisent les éléments clés qui doivent servir de base aux modèles du type de celui qu'ils proposent :

[1] Akerlof & Yellen [2] p. 825

- La nécessité d'avoir des modèles où la monnaie n'est pas neutre. Les rigidités réelles ne suffisent pas pour expliquer que les prix et salaires ne s'ajustent pas immédiatement à un choc nominal de la demande globale.
- Le profit ou l'utilité des agents doit être une fonction différentiable du prix ou du salaire qu'il fixe. Il faut donc exclure la concurrence parfaite, car la fonction de profit y connaît une discontinuité au niveau du prix du marché ; vendre plus cher fait perdre tout débouché et vendre moins cher n'a aucun avantage.

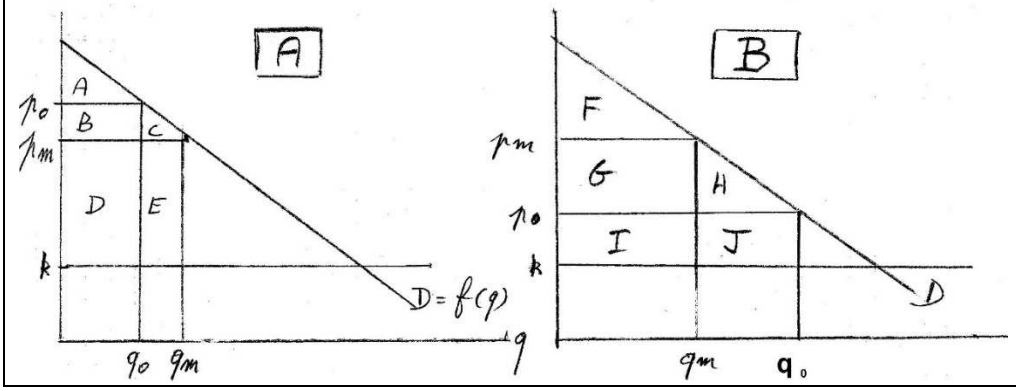
Les coûts de menu

Dans l'introduction de son article "Small Menu Costs and Large Business Cycles: a Macroeconomic Model of Monopoly", **Mankiw** écrit: « This note shows that sticky prices can be both privately efficient and socially inefficient (…) The claim that price adjustment costs are small does not rebut the claim that they are central to understanding economic fluctuations »[1].

L'analyse de Mankiw est conduite essentiellement en termes de bien-être. Le modèle calcule le bien-être social en additionnant les profits des firmes avec le *surplus des consommateurs* (cf. Marshall). La firme est en concurrence imparfaite ; sur sa demande décroissante ; elle fixe son prix et la quantité de façon à maximiser son profit. Le prix (p_0 sur le graphique 7.21) est déterminé en début de période sur base d'une anticipation du niveau général des prix durant la période. Supposons que les fluctuations de la situation monétaire amènent un niveau général des prix qui s'écarte de l'anticipation. De ce fait, le prix p_0 ne maximise plus le profit. Dans les nouvelles conditions, la firme devrait annoncer le prix p_m pour réaliser cet objectif. Problème, cette correction implique un coût d'ajustement *z*.

Mankiw analyse successivement les cas d'une baisse de la demande globale (le niveau des prix diminue) et d'une hausse (le niveau des prix augmente). Voyons d'abord la baisse de la demande globale qu'illustre la figure A du graphique 7.21. La firme, n'ayant pas anticipé cette variation, avait fixé son prix p_0 trop haut en ce sens qu'il est supérieur à p_m. Commençons avec quelques explications à propos de ce graphique. Le coût marginal de la firme est constant et vaut *k*. Il est supposé que la baisse du niveau général du prix affecte dans une même proportion le prix du produit de la firme et les éléments qui constituent son coût. Les situations des courbes *D* et *k* l'une par rapport à l'autre restent donc identiques sur le graphique au temps zéro et en cours de période. Ce qui permet à Mankiw de ne dessiner qu'une seule courbe *D* et une seule courbe *k*, qui sont ici des variables réelles et non nominales.

Graphique 7.21 : l'effet des coûts de menu sur le bien-être



Au prix p_0, le profit vaut B+D et le surplus du consommateur A. Le bien-être se monte à A+B+D. Au prix p_m, le profit vaut D+E, qui est supérieur à B+D d'un écart valant la différence

[1] Mankiw [250] p. 529

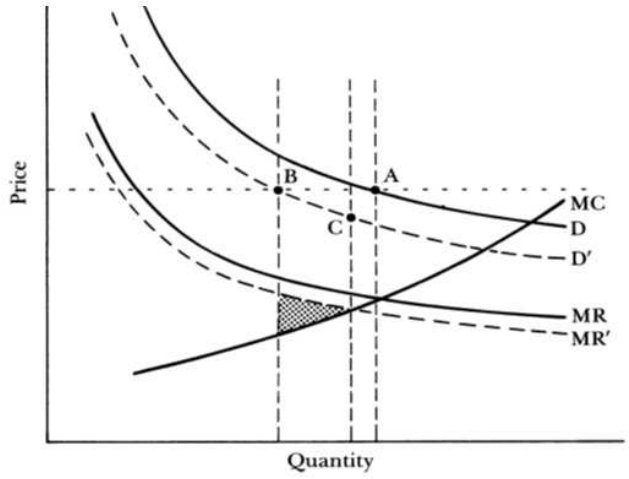
E-B. Le bien-être vaut quant à lui A+B+C+D+E. La firme ne corrigera son prix, le faisant passer de p_0 à p_m, que si E-B > z. On pourrait donc se trouver dans la situation paradoxale et inefficiente où C+E > z > E-B. L'économie est privée d'un accroissement de bien-être sensiblement supérieur à *z*, car le profit de la firme n'augmente pas suffisamment.

Mankiw s'intéresse ensuite à la hausse de la demande globale, le cas où p_0 a été fixé en dessous de p_m tout en étant supérieur à *k*, soit la figure B du graphique 7.21. Le bien-être est affecté de façon fondamentalement différente. En effet, au prix p_0, le bien-être total est plus élevé qu'au prix p_m d'un écart valant H+J. La firme fixera son prix à p_m si G - J > z. Ce résultat surprenant s'explique par la sous-optimalité intrinsèque de la concurrence imparfaite. Le prix optimal, celui qui prévaudrait en concurrence parfaite n'est autre que *k*. Plus le prix effectif s'en écarte vers le haut, moindre sera la quantité produite et donc le bien-être. Descendre le prix de p_0 vers p_m augmente le bien-être ; l'augmenter de p_0 vers p_m le réduit.

Romer et l'incitation à adapter les prix

Dans son article « The New Keynesian Synthesis » (1993), Romer analyse ce qui incite la firme à adapter plutôt le prix ou plutôt que la quantité. Le graphique 7.22 illustre son propos.

Graphique 7.22 : l'incitation à adapter les prix selon Romer



A la suite d'une baisse de la demande globale, la firme représentative fait face à une baisse de sa demande qui passe de D à D'. En conséquence, la recette marginale passe de MR à MR'. Le point A était son équilibre initial découlant de l'intersection de MR avec la courbe de coût marginal MC. La firme peut soit maintenir le prix initial et produire la quantité correspondant au point B, soit ajuster son prix produire la quantité qui maximise son profit dans les nouvelles conditions, l'intersection de MC avec MR' la poussant vers le point C[1].

La surface hachurée est le profit auquel l'entreprise renonce si elle opte pour la position B plutôt que la position C. Sa taille influence fortement la stratégie choisie. Elle s'accroît avec :

- la pente de la courbe de coût marginal, qui dépend à la fois de l'évolution du salaire et de celle de la productivité. Si l'offre de travail est inélastique, la réduction de la production fera baisser sensiblement le salaire réel et donc le coût marginal. Si la productivité marginale du travail varie sensiblement avec l'emploi, le coût marginal sera plus sensiblement affecté.

[1] Romer conserve la courbe MC d'origine pour ne pas compliquer le graphique à l'excès, mais elle aurait également pu reculer.

- La pente de la courbe de recette marginale. Celle-ci dépend de l'élasticité de la demande et de la variation qu'elle subit éventuellement du fait de la baisse de la demande globale.

La firme peut encourir une perte minime en n'ajustant pas le prix, mais si un grand nombre d'entre elles réagissent ainsi, la demande effective baissera et toute l'économie en souffrira. Romer considère cette situation comme une *externalité*.

Cette analyse met en relief que les frictions nominales et la concurrence imparfaite ne suffisent pas à expliquer la baisse de la production. Il faut encore que les déplacements des courbes MR et MC soient tels que la surface hachurée soit faible, ce que Romer appelle des *rigidités réelles*. Il conclut : « In short, a complete model of large real effects of nominal disturbance requires both nominal frictions and real rigidities »[1]. Même si des explications aux comportements rigides existent, une incitation élevée à l'ajustement des prix ne paraît pas insensée. Le côté récurrent, presque incontournable des comportements rigides reste un peu mystérieux. La recherche dans ce domaine doit encore progresser, mais Romer constate avec satisfaction une exploration assez intense des rigidités réelles.

Romer revient également sur l'économie du bien-être, une préoccupation permanente des *nouveaux keynésiens*. Dans l'optique walrassienne traditionnelle, l'économie est toujours à son optimum. Même en récession, s'écarter de cet équilibre réduirait le bien-être. Romer pense au contraire qu'une politique des autorités visant à stabiliser l'économie augmenterait le bien-être. Vraisemblablement, les fluctuations réduisent plus le bien-être en récession qu'elles ne lui sont favorables en bonne conjoncture. Il semble logique que l'utilité marginale combinée résultant de la consommation et de la désutilité du travail soit plus élevée en récession que dans une période de boom. En outre, il est plausible que les fluctuations conjoncturelles réduisent le niveau moyen de la production par rapport à ce qu'il serait dans une économie stabilisée. Romer rappelle « the traditional view that negative aggregate demand shocks mainly reduce output while positive shocks mainly raise prices. If this view proves correct, it could imply significant welfare gains from countercyclical stabilization policy »[2].

BLANCHARD & SUMMERS ET L'HYSTERESE

Remémorons-nous la controverse à propos de la *courbe de Phillips*. En fonction des errements de l'inflation, le taux de chômage REEL fluctue, mais, selon les néoclassiques, il est rappelé tôt ou tard à son taux NATUREL. Se basant sur les données empiriques relative à la situation européenne des années quatre-vingts, Blanchard et Summers constatent une évolution neuve par rapport aux décennies précédentes. Dans leur article "Hysteresis in Unemployment" (1987), ils écrivent : « As actual unemployment has gone up, equilibrium unemployment, as estimated from Phillips curve relations has risen in tandem »[3]. Le chômage a crû très sensiblement sans que l'inflation ne lui emboîte le pas. Les auteurs concluent: « The recent European experience has led to the development of alternative theories of unemployment embodying the idea that the equilibrium unemployment rate depends on the history of the actual unemployment rate. Such theories may be labelled hysteresis theories after the term in the physical sciences referring to situations where equilibrium is path-dependant ».

Pour expliquer la persistance du chômage, ces théories s'appuient sur deux facteurs :

---

[1] Romer [306] p. 11
[2] Romer [306] p. 14
[3] Blanchard & Summers [34] p. 288. *Chômage d'équilibre* peut être compris comme synonyme de *chômage naturel*.

- la distinction entre les *insiders* (travailleurs actifs dans l'entreprise) et les *outsiders* (chômeurs). D'après la théorie classique, lors des négociations salariales ils devraient intervenir indistinctement. En réalité, le poids des insiders est prépondérant ;
- Si les chômeurs de courte durée font encore partie de l'offre de travail dans laquelle les entreprises puisent pour recruter, les chômeurs de longue durée, par contre, semblent exclus, à la fois par le découragement dans leur propre chef et par le préjugé des employeurs à leur égard.

Dans un premier temps, le modèle envisage un cas très simplificateur. Les insiders participent seuls aux négociations salariales et ont priorité à l'embauche ; ils fixeront le salaire de telle façon que l'emploi dans leur firme corresponde à leur effectif. Les chocs non anticipés de la demande globale nominale affectent l'emploi en ajoutant ou retranchant des postes à l'effectif négocié. Le résultat auquel aboutit ce modèle est que chaque période, le taux de chômage d'équilibre égale le taux de chômage réel de la période précédente. D'une période à l'autre, le taux de chômage fluctue aléatoirement, sans faire montre d'une tendance à retourner à une position d'équilibre.

Les hypothèses sont ensuite adaptées dans le sens plus réaliste où le chômeur récemment licencié ne quitte le cercle des insiders que s'il reste chômeur plusieurs périodes et où le nouvel entrant ne devient insider qu'après quelques périodes de travail dans l'entreprise. Dans ce cas, il y a un taux de chômage d'équilibre assez stable, sauf dans le cas où plusieurs périodes consécutivement, la demande globale se modifie dans le même sens ; dans ce cas, il se crée un nouveau taux d'équilibre, d'où le chômage ne sera délogé qu'à la suite d'une nouvelle séquence de changements consécutifs de même signe.

Le modèle est encore affiné pour tenir compte de l'affaiblissement de la position des insiders dans la négociation salariale quand le chômage anticipé est élevé. Plus de chômage signifie un salaire moindre. Les auteurs introduisent ici une variable supplémentaire, la sensibilité du salaire au chômage anticipé. Si cette sensibilité est faible, le cas diffère peu du précédent. Si elle est élevée, le chômage sera moins sujet à la persistance.

Le modèle subit encore un dernier raffinement : la pression à la baisse du salaire du paragraphe précédent n'est le fait que des chômeurs de moins d'un an, ce qui semble assez réaliste, ce qui renforce la persistance du chômage. Les auteurs remarquent : « …an increase in long term unemployment increases equilibrium unemployment for some time »[1].

7.7.2. Stiglitz et l'économie de l'information

En 2001, Joseph Stiglitz reçoit le prix Nobel d'économie. A cette occasion, comme c'est la coutume, le lauréat donne un cours qui fait l'objet d'une publication par la fondation Nobel. Le texte « Information and the Change of Paradigm in Economics » est extraordinaire a bien des égards. On peut presque parler d'un écrit révolutionnaire. Stiglitz y retrace sa carrière d'économistes et notamment le malaise qu'il ressentit dès le début du fait de l'écart énorme entre le monde réel plein de contradictions et la science économique empreinte d'équilibre et d'optimum. Le ton est donné dès les premières pages : à propos de la fameuse *main invisible* d'Adam Smith, fétiche de beaucoup d'économistes, Stiglitz ironise : « … the reason that the hand may be invisible is that it is simply not there- or at least that if it is there, it is palsied »[2].

[1] Blanchard & Summers [34] p. 294
[2] Stiglitz [350] p. 473

Dans la conclusion, Stiglitz appelle les chercheurs académiques à se préoccuper de la justice sociale autant que de la prospérité économique.

Stiglitz critique vertement l'économie politique mainstream, celle dont l'exposé occupe la majorité des pages du présent ouvrage. L'article énumère une liste de déficiences tellement longue qu'à la fin, il doit poser la question : « One might ask, how can we explain the persistence of the paradigm for so long ? » Le fait qu'une minorité de marchés, les marchés organisés, se comportent plus ou moins selon la théorie ne peut l'expliquer que partiellement. « But one cannot ignore the possibility that the survival of the paradigm was partly because the belief in that paradigm, and the policy prescriptions, has served certain interests »[1].

Beaucoup des défauts de la science économique mainstream peuvent être rapportés directement ou indirectement à une déficience générique : « For more than a hundred years, formal modeling in economics has focused on models in which information was perfect »[2]. L'*information asymétrique* signifie que l'une des parties d'une transaction est mieux informée que l'autre[3]. Dans la réalité économique, l'information est généralement insuffisante ou asymétrique. L'erreur à ne pas commettre serait de penser qu'il s'agit là d'une question secondaire. Le cours Nobel dresse une impressionnante liste d'articles que Stiglitz a écrits ou coécrits, qui démontrent, chacun pour un problème particulier, qu'en *information imparfaite*, le comportement des agents s'écarte du modèle standard au point d'affecter l'équilibre du marché[4].

Stiglitz dit tenter depuis quinze ans de mettre sur pied un nouveau modèle macroéconomique basé sur l'information imparfaite avec en son centre une nouvelle théorie microéconomique de la firme : cette théorie ferait la part belle aux éléments financiers (solvabilité), s'intéresserait à la gouvernance des entreprises et à leur structure organisationnelle, des éléments négligés par l'économie néoclassique traditionnelle. Mais il s'avère que l'information imparfaite est très difficile à modéliser.

Passons en revue non exhaustive quelques effets dévastateurs de l'information imparfaite pour la théorie économique :

- La théorie des *salaires d'efficience* s'appuie elle-même sur l'information asymétrique, car l'employeur ignore le niveau d'effort effectivement fourni par le salarié.
- Avec l'information imparfaite, l'hypothèse de la convexité des préférences ne tient plus, car l'utilité de l'information n'est pas une fonction concave de sa quantité. La loi des rendements décroissants se révèle avoir un champ d'action moindre que ce qu'on estimait.
- Lorsque le prix demandé « sélectionne » le type de demandeur ou lorsqu'il sert d'incitant à la qualité du service offert, il n'égalise pas nécessairement l'offre et la demande.
- L'unicité du prix d'équilibre ne résiste pas non plus. Les marchés réels sont caractérisés par des *distributions* (au sens statistique) de prix et de salaire. Si changer de fournisseur

[1] Stiglitz [350] p. 524

[2] Stiglitz [350] p. 475. Stiglitz remarque que l'information parfaite est tellement évidente aux yeux de la plupart des modèles traditionnels qu'ils omettent de la mentionner parmi leurs hypothèses.

[3] L'emprunteur connaît mieux sa solvabilité que le prêteur, l'employé connaît mieux ses prestations que son employeur, le vendeur connaît mieux la qualité du produit que l'acheteur, l'assuré connaît mieux son risque que l'assureur, le manager connaît mieux l'effet de ses décisions que les actionnaires…

[4] L'insatisfaction de Stiglitz vis-à-vis du modèle économique dominant dépasse le cadre de l'hypothèse d'information parfaite. Mais sa stratégie de théoricien de l'économie l'a conduit à conserver les autres propriétés du modèle. Il s'explique : « It is not easy to change views of the world and it seemed to me the most effective way of attacking the paradigm was to keep within the standard framework as much as possible » ([350] p.519).

implique des coûts d'information, une firme pourra conserver sa clientèle en demandant un prix supérieur au prix d'équilibre.

- L'hypothèse que les marchés boursiers sont efficients veut que les prix des actions reflètent toute l'information nécessaire. En réalité, le système de prix n'agrège l'information qu'imparfaitement.
- « Under the imperfect information paradigm, markets are almost never Pareto efficient »[1].
- La mobilité de la main d'œuvre souffre de l'asymétrie entre la connaissance que l'employeur a des qualités réelles de son employé et la faible information des autres entreprises le concernant. Il devient alors risqué pour ces dernières de briser le lien entre l'employé et son employeur actuel.
- Sur certains marchés, le prêteur ou l'acheteur doivent opérer un examen critique du risque ou de la qualité du service, ce qui représente une charge considérable. Un tel marché est condamné à la concurrence imparfaite. Généralement, le résultat final verra l'imperfection toucher à la fois la concurrence et l'information.
- Les agents imparfaitement informés tentent de dégager des informations en analysant le comportement des agents mieux informés pour deviner leurs plans et leurs informations. Par exemple, l'insider, vend-il à tel prix ? Les achats, ventes, prêts, emprunts, souscriptions de police d'assurance deviennent ainsi des sources d'information concernant leurs auteurs… qui le savent bien. En conséquence, ceux-ci adaptent leur comportement dans tel ou tel sens en fonction du message qu'ils souhaitent faire passer. Ces altérations feront dérailler les comportements hors des routes de l'optimum. Un exemple est l'entrepreneur qui garde plus de parts de sa société pour communiquer au marché le signal que les affaires sont florissantes, alors que la maximisation de son utilité impliquerait une modification de son portefeuille.
- Les deux *théorèmes du bien-être* (cf. sous-chapitre 4.2.3) ne résistent pas à l'information imparfaite. Celle-ci est toujours une source de *pouvoir de marché* ; le bien-être en souffre nécessairement. L'information imparfaite génère également des transferts de bien-être intempestifs. La mauvaise distribution de l'information fait des gagnants et des perdants. Dans nombre de cas, gagner la course à la vitesse de l'information procure une rente.
- « Under the older, perfect information theory, it made no difference whether firms raised capital by debt or equity, in the absence of tax distortions. This was the central insight of the Modigliani-Miller theorem »[2]. L'observation empirique indique au contraire que la structure financière influence les décisions. La dette est une formule de financement très répandue, qui accroît le risque de faillite. Les ratios de solvabilité et de liquidité du bilan de l'entreprise importent également pour une autre raison : l'attitude des banques à son égard. Le rationnement du crédit résultant du risque d'insolvabilité dément la thèse néoclassique de l'optimisation intertemporelle de l'épargne et de l'investissement, qui remonte à Irving Fisher.

STIGLITZ ET LA RIGIDITE DES PRIX

L'information imparfaite, l'incertitude et l'aversion pour le risque constituent un mélange détonnant ; Stiglitz montre qu'il explique la rigidité des prix et des salaires de mille et une manières.

En 1989 paraît l'article « Toward a Theory of Rigidities » de Greenwald et Stiglitz. Les auteurs partent de la constatation que sur le cycle conjoncturel, la fluctuation des quantités est plus prononcée que celle des prix. Certaines explications invoquent les coûts d'ajustement des prix,

[1] Stiglitz [350] p. 486
[2] Stiglitz [350] p. 508. Le théorème de Modigliani et Miller sera examiné au sous-chapitre 9.6.6.

mais les coûts d'ajustement des quantités leur sont généralement supérieurs. Pourquoi les firmes préfèrent-elles l'ajustement le plus coûteux ? La réponse se trouve dans la volonté de se prémunir contre l'incertitude.

Les auteurs modélisent le cas d'une entreprise dont les actions portent sur un vecteur de variables de décision **x** (dont font partie les différents prix et les différentes quantités) et sont fonction d'un vecteur de variables exogènes **z**. Les grandeurs **x**° et **z**° sont les valeurs d'origine de ces vecteurs. Les coefficients μ et η introduisent un aspect aléatoire à l'effet des vecteurs **x** et **z** sur le profit π. On a :

$$\pi = \pi(x°, z°, \mu(\mathbf{x} - \mathbf{x°}), \eta(\mathbf{z} - \mathbf{z°})) \qquad (7.59)$$

Le vecteur **x** que la firme fort rétive au risque préférera, sera celui qui minimise la variance du profit σ_π avec la condition que $E(\pi) > \pi_0$ (condition minimale de profit espéré). L'entreprise ne laissera donc pas flotter au gré du marché la variable dont l'effet sur le profit est le plus aléatoire. Cette variable s'en trouvera donc plus inerte.

Greenwald et Stiglitz montrent que sur le marché du travail, du fait du lien entre le salaire et la productivité, l'effet d'une variation de l'effectif est plus facile à prévoir que celui d'une variation du salaire. Ils montrent que similairement, sur le marché des produits, la valeur de l'entreprise en fin de période est plus aléatoire en cas de prix s'ajustant au marché qu'en cas de quantité s'ajustant au marché.

Tenant compte de l'influence de l'*aversion pour le risque,* les auteurs peuvent conclure: « Thus, the utility gain from a small (but optimal) price adjustment might not be sufficient to cover the fixed costs of making such a change. At the same time, the utility gain from a larger (but also optimal) quantity adjustment might exceed the perhaps larger fixed cost associated with quantity (employment) adjustments »[1].

[1] Greenwald & Stiglitz [122] p. 369

TITRE VIII- LA MACROECONOMIE DYNAMIQUE

Les modèles étudiés au titre VII sont des modèles statiques en ce sens que l'écoulement du temps n'y est pas une variable essentielle. Les deux problématiques macroéconomiques principales sont la croissance et le cycle conjoncturel, deux phénomènes qui n'ont de signification que par l'écoulement du temps. Le titre VIII est consacré à l'étude de ces phénomènes, toujours dans une optique plus ou moins chronologiques. Les théories expliquant la croissance et le cycle conjoncturel sont très nombreuses. Le présent exposé est évidemment très loin d'être exhaustif.

Le chapitre 8.1, très court, est consacré à la théorie de l'investissement de Ramsey, qui date de 1928 ; sa présence ici se justifie par l'influence qu'elle exercera sur certaines théories de la croissance dans les années soixante. Les chapitre 8.2 et 8.3 sont consacrés respectivement aux théories de la croissance et à celles du cycle conjoncturel.

8.1- L'EPARGNE OPTIMALE SUIVANT RAMSEY

Nous avons évoqué plus haut l'article de Ramsey « A Mathematical Theory of Saving » paru en 1928 ; revenons-y. C'est par sa réponse à la question « how much of its income should a nation save ? » que cet article devint célèbre. Le modèle se place dans une optique dynamique. Ramsey cherche à définir une règle qui, appliquée tout au long d'un intervalle de temps, conduit la société à l'optimum. Il postule implicitement que la société est capable de mettre en œuvre sa règle optimale, mais ne tente pas d'expliquer par quel moyen une économie de marché y parviendrait.

Ramsey adopte nombre d'hypothèses simplificatrices :

- L'économie compte un nombre constant d'agents. Leurs fonctions d'utilité par rapport à la consommation et au travail ne changent pas. « Enjoyments and sacrifices at different times can be calculated independently and added »[1].
- L'innovation technologique est absente. L'amélioration de la production ne peut provenir que de l'accumulation quantitative du capital.
- Abstraction est faite de la distribution du revenu. La satisfaction totale de l'ensemble des membres de la société dépend uniquement du total de la consommation sociale et du total du travail effectué.
- Le modèle considère le travail, le capital et la consommation comme s'il s'agissait de grandeurs homogènes.
- L'Etat établit aujourd'hui un plan d'épargne qui courra sur une très longue période, qu'appliqueront des générations successives Les motifs qui guident l'épargne et la consommation transcendent les générations. Une génération égoïste ne viendra pas consommer ce qui fut accumulé par les prédécesseurs pour ne rien laisser aux successeurs. La génération de départ élabore un plan d'épargne-investissement, suivant la règle que nous devons justement établir ; rien ne viendra troubler ce plan.

On a les variables a(t), c(t) et x(t) qui représentent respectivement la quantité de travail, la quantité de capital et la consommation au temps *t*. La fonction de production est donc *f(a,c)*. Le taux d'intérêt égale la productivité marginale du capital.

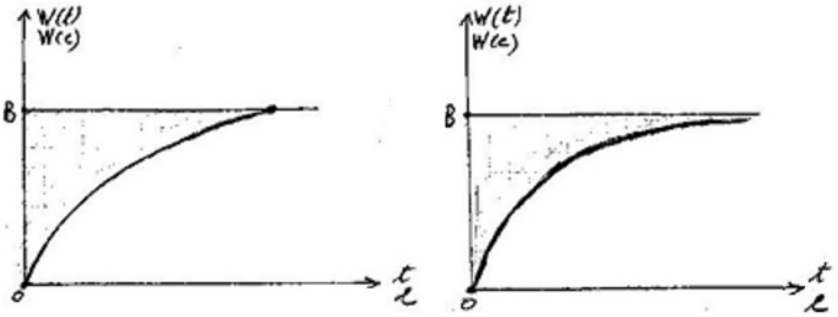
Au temps *t*, l'utilité TOTALE de la consommation sociale se note U(x) et la désutilité TOTALE du travail V(a). L'utilité MARGINALE de la consommation sociale vaut $u(x) = dU(x)/dx$ et la désutilité MARGINALE du travail vaut $v(a) = dV(a)/da$. Le bien-être total au temps *t* vaut $W = U(x) - V(a)$. Ramsey veut déterminer le taux d'épargne qui maximise la valeur cumulée de *W* sur tout l'intervalle de temps considéré.

W est une fonction de *c* ; en fait, la production ne dépend que de cette variable, *a* étant donné. *W* est donc également une fonction de *t*, puisque *c* est fonction de *t*. Le capital s'accroît à chaque période du montant non consommé. La relation entre *W* et *c* est au centre du modèle. Comme le montre le graphique 8.1, Ramsey estime que l'accumulation de *c* n'augmente pas indéfiniment *W*. Deux raisons expliquent conjointement l'existence d'un maximum que *W* ne transgressera pas :

- l'utilité marginale décroissante de la consommation u(x) : la production crée de moins en moins de satisfaction.
- la productivité marginale décroissante du capital $\delta f/dc$: l'accumulation du capital augmente la production de moins en moins.

[1] Le modèle ne comporte que des données agrégées, y compris lorsqu'il s'agit d'utilité ou de désutilité.

Graphique 8.1 : l'épargne optimale selon Ramsey



Le maximum de bien-être, Ramsey le nomme « *Bliss* » (béatitude). Notons-le par la lettre *B*. Deux cas sont possibles : soit une valeur finie de *c* permet d'atteindre W = B ; soit, *W* tendra asymptotiquement vers *B*. Dans le modèle, cette distinction est peu significative. Le but est évidemment d'atteindre *B*, mais épargner un maximum pour l'atteindre le plus rapidement possible n'est pas nécessairement optimal, car l'épargne élevée réduit l'utilité dont la société profite au début de la période. Augmenter l'épargne *dc/dt* au temps *t* produit un effet contradictoire sur la courbe *W* : à la hausse du fait de la production accrue ; à la baisse du fait de la consommation moindre.

La béatitude atteinte, le taux d'épargne optimum deviendra nul car $(\delta f/\delta c).u(x)$ ne peut plus être positif. On ne peut pas dépasser l'optimum. Le modèle est dynamique en ce sens qu'il se préoccupe du chemin menant à l'optimum, mais l'optimum lui-même est conçu statiquement.

Notre problème consiste à atteindre *B* (ou l'atteindre quasiment) en un temps fini, tout en minimisant la surface hachurée sur le graphique[1], qui est finie, même dans le cas asymptotique. Les équations d'équilibre sont :

$$dc/dt + x = f(a,c) \qquad (8.1)$$
$$v(a) = (\delta f/\delta a).u(x) \qquad (8.2)$$
$$du(x(t))/dt = -(\delta f/\delta c).u(x(t)) \qquad (8.3)$$

L'équation (8.1) énonce simplement que la production-revenu égale la consommation plus l'investissement. L'équation (8.2) demande l'égalisation de la désutilité marginale du travail et de l'utilité supplémentaire générée par la dernière unité de travail, une condition d'optimisation classique. L'équation (8.3) implique que la consommation *x* augmente tant que la productivité marginale du capital et l'utilité marginale de la consommation sont positives, c'-à-d jusqu'à ce que la béatitude soit atteinte.

Quelques manipulations algébriques amènent à l'équation suivante ; au temps *t*, on a:

$$(dc/dt).u(x) = \{B - (U(x) - V(a))\} \qquad (8.4)$$

Nous avons donc notre règle : « rate of saving multiplied by marginal utility of consumption should always equal bliss minus actual rate of utility enjoyed »[2]. Ramsey insiste sur le fait étonnant que le taux d'intérêt ne joue aucun rôle dans la formule.

---

[1] Cette surface mesure la perte de satisfaction cumulée due à la non réalisation de la béatitude. Elle vaut : $\int_0^\infty (B-U(x)+V(a))\,dt$

[2] Ramsey [293] p. 547

On peut constater que la règle de Ramsey met toutes les générations concernées sur le même pied, puisqu'aucun escompte ne vient pondérer les satisfactions plus reculées dans le futur. Il se refuse à considérer la satisfaction de la génération présente comme prioritaire et écrit à ce propos : « it is assumed that we do not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible and arises merely from the weakness of the imagination »[1]. Le choix du non-escompte donnera lieu à un débat entre économistes dans les décennies suivantes.

Dans ce débat, la dernière contribution importante -décisive, pourrait-on dire- est celle de **Tjalling Koopmans** qui démontra qu'au-delà du débat éthique, le non-escompte des satisfactions futures est mathématiquement intenable. Il construit son raisonnement en trois phases :

1- Axiomatiser ce que serait l'attitude rationnelle d'un Etat appelé à prendre la décision relative au plan d'épargne.
2- Montrer qu'un fonction d'utilité[2] peut être déduite de ces axiomes, permettant de classer les plans d'épargne possibles selon ses préférences.
3- Montrer que cette fonction d'utilité escompte inévitablement les satisfactions futures, ce que Koopmans qualifie d'*impatience*.

L'intervalle de temps que ce modèle prend en considération est infini, ce qui en constitue une caractéristique essentielle. Le but est de définir la fonction $U = U[u(x_1),u(x_2),u(x_3)\ldots]$ où x_1, $x_2\ldots$ sont les consommation des différentes périodes, où u(.) est l'utilité immédiate de ces consommations et où *U* est la fonction d'utilité intertemporelle qui nous intéresse.

Les principaux postulats de Koopmans sont :

- La fonction d'utilité doit avoir la propriété de *continuité*
- La fonction d'utilité est sensible à des avantages ou sacrifices immédiats et elle ne peut privilégier la concentration de la satisfaction en une période reculée.
- L'indépendance entre les périodes. Si l'Etat préfère la consommation x_1 suivie d'une séquence infinie X à la consommation x_1' suivie de cette même séquence X, il préférera nécessairement aussi x_1 suivi d'une séquence infinie X' à x_1' suivi de X'. Koopmans reconnaît que « one cannot claim a high degree of realism for such a postulate because there is no clear reason why complementarity of goods could not extend over more than one time period »[3].
- La *stationnarité* : on ne modifie pas la préférence entre deux plans d'épargne en les décalant tous deux d'une période vers le futur pour intercaler un même x_1 en première période.

Koopmans, après une démonstration mathématique complexe, arrive à la conclusion que moyennant quelques conditions supplémentaires[4], le principe d'impatience est bien sous-jacent dans la fonction *U*. A mon avis, la grande faiblesse de la démonstration est qu'elle repose sur un intervalle de temps infini.

---

[1] Ramsey [293] p. 543

[2] La démarche de Koopmans est à ce titre similaire à celle de von Neumann et Morgenstern vis-à-vis de l'*utilité attendue*. Comme il s'agit d'une fonction d'utilité ORDINALE, elle n'est unique qu'à une transformation linéaire près.

[3] Koopmans [1960] p. 292

[4] Ces conditions mathématiques semblent peu significatives du point de vue économique.

L'intérêt de cette double analyse de Ramsey et de Koopmans est très théorique. Pratiquement, il est impossible pour un planificateur de baser ses décisions sur l'utilité des consommateurs, qui n'est pas mesurable, contrairement à la consommation proprement dite.

L'hypothèse de l'invariance de l'utilité ressentie sur de très longues périodes est irréaliste. On a critiqué à bon droit les comparaisons interpersonnelles d'utilité, mais le cas est plus grave lorsqu'il s'agit de générations qui connaissent des degrés d'opulence différents et qui ont donc une autre conception de ce qu'est « beaucoup » ou « peu ».

Il est peu probable qu'une économie déjà industrialisée au départ trouve un avantage à accumuler du capital pendant une aussi longue période (plusieurs générations selon Ramsey), en l'absence de progrès technique. Il ne faudra pas longtemps pour que le rapport capital sur travail connaisse la saturation. Dans la réalité économique, l'accumulation sert principalement à incorporer les innovations technologiques.

Même abstraction faite de l'infinité du temps envisagé, sa très longue durée rend l'exercice peu significatif. Le monde évolue vite, dans ses dimensions politique, internationale, culturelle, technologique, démographique, écologique… Et ces évolutions sont imprévisibles pour une large part. Vouloir déterminer le taux d'épargne optimal au-delà d'un terme de dix ans relève de la gageure. Ramsey et Koopmans ont négligé le facteur INCERTITUDE. La prise en compte de celui-ci impose une limite temporelle étroite à ce type d'analyse[1].

L'un des arguments de Koopmans en faveur de la période infinie est qu'au terme d'une période finie, il existe un stock de capital qui servira aux générations ultérieures et qu'il est difficile d'intégrer objectivement ce capital dans la fonction d'utilité de la consommation. Cette difficulté est bien réelle mais n'est-elle pas un beau problème pour un économiste ? Koopmans semble plus intéressé par les problèmes mathématiques.

[1] Peut-être cette incertitude est-elle la vraie justification de l'escompte du futur. Plus les événements sont lointains, plus ils sont incertains et plus leur prise en compte risque de dérégler les choix.

8.2- LA CROISSANCE ET LE PROGRES TECHNIQUE

8.2.1. Instabilité de la croissance dans le modèle Harrod-Domar

HARROD : PREMIERE THEORIE DE LA CROISSANCE

La plupart des théories macroéconomiques antérieures avaient un caractère statique. Tout intéressantes qu'elles fussent, leurs limites plaident pour le développement d'un nouveau modèle de théorie macroéconomique, de type dynamique. Tel est le chemin que l'article « An Essay in Dynamic Theory » (1939) prétend inaugurer et qui sera poursuivi dans l'ouvrage « Towards a Dynamic Economics » (1948).

Harrod commence par présenter une équation qu'il qualifie de truisme :

$$G.C = s \qquad (8.5)$$

G est le taux de croissance annuel valant ΔY/Y où *Y* est le revenu national. *C* est le coefficient marginal de capital valant ΔK/ΔY où *K* est le capital total (fixe plus circulant : les stocks en font partie). La variable *s* représente le taux d'épargne valant S/Y où S est l'épargne totale. En remplaçant les variables de 8.15B par leur ratio puis en simplifiant, on obtient en effet l'égalité évidente S = ΔK.

Il présente ensuite l'éqauation-sœur, qu'il qualifie d'équation fondamentale :

$$G_W = \frac{s}{C_r} \qquad (8.6)$$

Les variables sont les mêmes à quelque nuance près. *s* n'a pas changé. *C* est remplacé par C_r. Harrod la définit ainsi : "the value of the capital goods required for the production of a unit of increment of output". Alors que *C* était une simple constatation ex-post, C_r est une donnée technologique. *C* et C_r sont nets de la dépéciation du capital. Quant à G_W, c'est le taux d'équilibre (« warranted rate »), défini ainsi : " …that rate of growth, which if it occurs will leave all parties satisfied that they have produced neither more nor less that the right amount. Or to state the matter otherwise, it will put them into a frame of mind which will cause them to give such orders as will maintain the same rate of growth"[1]. Les taux G et G_W sont des moyennes sociales entre des taux qui variant d'une entreprise à l'autre.

Comme *G* résulte de décisions prises indépendamment par de très nombreux individus, il n'égalera G_W que par un heureux hasard. On voit immédiatement que si $C > C_r$, $G < G_W$ et vice-versa.

Les équations 8.5 et 8.6 s'appliquent à des périodes de plusieurs années consécutives. Bien-sûr, C et C_r et donc G et G_W sont variables d'une année à l'autre[2], mais comme l'objectif de Harrod est de déterminer un chemin de croissancer équilibré durable (à taux constant), c'est la moyenne de ces taux sur la période qui l'intéresse. Malgré son manque de réalisme, l'hypothèse de C_r constant s'impose donc. Comme le remarque Harrod, la constance du *coefficient de*

[1] Harrod [132] p. 16. La notation des variables de 1948 diffère légèrement de celle de 1939. Dans la présente section, j'adopte celle de 1948, y compris dans les citations de 1939.

[2] Par exemple, au sortir d'une dépression, lorsque beaucoup d'équipement sont inutilisés, Cr pourra être très bas.

capital implique celle du taux d'intérêt (dont les variations pourraient inciter à modifier l'intensité capitalistique) ainsi que la *neutralité du progrès technique*[1].

Le point essentiel selon Harrod est que contrairement à l'équilibre statique, l'équilibre dynamique est instable. « Suppose an excessive output, so that G exceeds G_w. The consequence will be that C, the actual increase of capital goods per unit increment of output, falls below C_r, that which is desired. There will be, in fact, an undue depletion of stocks or shortage of equipment, and the system will be stimulated to further expansion. G, instead of returning to G_w, will move farther from it in an upward direction"[2]. La conséquence est symétrique lorsque G tombe en dessous de G_w, par exemple lorsque ce dernier est augmenté sous l'effet d'une hausse de la propension à épargner.

La surproduction est fondamentalement paradoxale. Lorsque les producteurs ne parviennent pas à écouler toute leur production ou à utiliser intégralement leur équipement, on dit qu'il y a surproduction. Mais c'est précisément lorsque G est inférieur à G_W que cette situation se produit. « … a condition of general over-production is the consequence of producers in sum producing too little »[3]. La surproduction va de pair avec une croissance faible. Un paradoxe symétrique existe dans le cas inverse, lorsque les producteurs se sentent à l'étroit dans leur capacité de production.

Harrod ne prétend pas que les écarts entre G et G_W suffisent à l'élaboration d'une théorie du cycle conjoncturel, mais il considère les deux aspects comme liés : « I am afraid that a proper understanding of the relation between the requirements of a steady advance and what the market can provide is very much mixed up with the trade cycle problem »[4].

La démonstration de l'instabilité ci-avant considérait que C et s sont indépendants de G. Harrod admet leur dépendance par rapport au niveau du revenu mais pas de son taux de variation. Supposons un déséquilibre où $G > G_w$; un accroissement de s ou une réduction de C_r pourraient ramener G_w au niveau de G ; un tel scénario, est-il plausible ? La réduction de C_r paraît hautement improbable. Concernant s, Harrod démontre qu'il devrait s'accroître dans une proportion invraisemblable pour qu'il y ait effectivement compensation.

Les variables C et Cr mettent en rapport ΔK et ΔY. Elles sont donc en prise directe avec le *principe d'accélération*, un concept important de la théorie du cycle (cf. infra), selon lequel les variations de Y donnent lieu à des variations amplifiées de K. Harrod craint que l'équation (8.6) exagère l'influence de ce principe. Une partie des investissements échappent en effet à ce déterminisme. Ces investissements plus autonomes comptent notamment ceux qui sont motivés par des perspectives de longue période, par la réduction du coût de la production existante ou par l'adaptation de celle-ci aux goûts fluctuants des consommateurs. Parmi ces investissements, Harrod distingue :

- ceux qui dépendent du niveau du revenu (selon une fraction k) et non de son taux de croissance.
- Ceux, notés K, qui sont indépendants à la fois du niveau et de l'accroissement du revenu

Ces investissements n'élèvent pas le taux de croissance nécessaire à l'équilibre. Harrod les élimine de l'équation (8.6), tant au dénominateur qu'au numérateur. Dans l'équation (8.7), C_r

[1] Selon Harrod, le progrès technique neutre est celui qui ne modifie pas le coefficient de capital. Nous reviendrons sur cette notion au sous-chapitre 8.2.3.
[2] Harrod [132] p. 22
[3] Harrod [132] p. 24
[4] Harrod [133] p. 77

n'inclut plus les éléments des capitaux *k.y* et *K* (où *y* est le revenu de la période). Plus *k.y* et *K* sont importants, plus bas sera le taux G_w.

$$G_W = \frac{s - k - K/y}{C_r} \qquad (8.7)$$

Nous avions déjà deux taux de croissance, le taux REEL et le taux D'EQUILIBRE. Harrod en introduit un troisième : le *taux de croissance naturel* (G_n). Il s'agit du taux maximum rendu possible par la croissance démographique, l'accumulation du capital, les inventions techniques… Le taux naturel est totalement indépendant du taux d'équilibre. Contrairement à celui-ci, il implique le plein emploi des ressources. A moyen et long termes, il est une limite qui est infranchissable par le taux réel *G*, qui peut toutefois l'excéder pendant la phase ascendante du cycle.

On s'attendrait de prime abord à ce que l'économie bénéficie d'une situation où G_W est supérieur à G_n ; elle serait ainsi stimulée, tirée vers le haut. Mais paradoxalement, l'effet serait néfaste. L'économie ne peut avancer plus vite que G_n. Si au moment de la reprise, on a $G > G_n$, l'accession consécutive au plein emploi ramènera G à G_n. Si à ce moment, G_W est supérieur à G_n, il sera forcément supérieur à *G*, ce qui- comme nous le savons maintenant- est source de dépression. Dans l'article de 1939, Harrod donne une explication légèrement différente qui met l'accent sur la valeur de G_W qui assure le plein emploi, notée G_p. Si G_p est au-dessus du taux naturel G_n, il y a une tendance chronique à la dépression, qui abaisserait G_w en-dessous de son niveau de plein emploi. « But this reduction of the warranted rate is only achieved by having chronic unemployment »[1].

Selon Harrod, l'instabilité de l'équilibre renforce la nécessité d'une politique conjoncturelle visant à combattre les oscillations. Outre cette politique anticyclique, une politique à long terme peut se justifier visant à empêcher G_p d'outrepasser G_n. « The ideal policy would be to manipulate the proper warranted rate (NDLR: G_p) so that it would be equal to the natural rate »[2]. Faisant partie de *K*, les travaux publics auraient cet effet.

DOMAR : CROISSANCE ET PLEIN EMPLOI

L'économiste américain Evsey Domar publie en 1946 et 1947 deux articles fort semblables à celui de Harrod, dont il n'avait pas connaissance à l'époque. Keynésien comme Harrod, Domar envisage la question du plein emploi sur la longue période. L'hypothèse de Keynes que l'emploi ne dépend que du revenu n'est correcte qu'à court terme. A plus long terme, l'emploi dépend du ratio du revenu sur la *capacité productive*. Ce concept de capacité productive est essentiel ; il inclut nombre de facteurs comme la technologie, les ressources naturelles, les équipements productifs, la main d'œuvre.

L'investissement en capital a un effet dual : d'une part, il génère du revenu, d'autre part, il accroît la capacité productive. « The emphasis on this dual character of the investment process is the essence of this paper's approach to the problem of employment. If investment increases productive capacity and also creates income, what should be the magnitude of investment, or at what rate should it grow, in order to make the increase in income equal to that of productive capacity? »[3].

[1] Harrod [132] p. 30.
[2] Harrod [132] p. 32.
[3] Domar [78] p. 39

Déterminons l'équation de cet équilibre. Les variables sont l'investissement *I*, le revenu *Y*, la propension marginale à épargner *s* et la productivité sociale potentielle de l'investissement α. S'il faut un supplément de $3 de capital pour obtenir un accroissement du produit de $1, $\alpha = 0{,}33$. La partie de ce capital qui se rapporte à des transferts de ressources ou du capital neuf entraînant la mise au rebut d'anciens équipements est déduite. Cette grandeur est nette de dépréciation.

Il faut égaliser le taux de croissance du revenu et celui de la capacité productive.

- L'équation de l'accroissement du revenu est $\Delta Y = \Delta I.1/s$. 1/s n'est autre que le multiplicateur. Domar remarque que l'investissement ne suffit pas pour accroître le revenu ; il faut un investissement qui, lui-même, s'accroît.
- L'équation de l'accroissement de la capacité productive est $I.\alpha$.

La solution pour le taux d'accroissement commun est :

$$\frac{\Delta I}{I} = \frac{\Delta Y}{Y} = s.\alpha \qquad (8.8)$$

On reconnaît la formule de Harrod, puisque α est en quelque sorte l'inverse de *C*. Pour maintenir le plein emploi sur la longue période, l'investissement et le revenu doivent croître de concert à un taux égalant le produit de la propension marginale à épargner avec la productivité sociale potentielle. Ce taux d'accroissement est évidemment cumulatif comme des intérêts composés, revenant à un accroissement exponentiel du revenu. Selon Domar, son modèle n'implique pas que les grandeurs *s* et α soient considérées comme constantes, mais il cite une étude empirique indiquant une stabilité à long terme de α.

Un taux de croissance réel de l'investissement et du revenu inférieur à α.s entraînerait une sous-utilisation du capital et donc du chômage de la main d'œuvre. Et l'existence du capital inactif décourage la propension à investir. Si la propension marginale à épargner est très élevée, le maintien du plein emploi s'avérera très compliqué, sauf en cas de dynamisme en matières de technologie, de ressources naturelles et de démographie.

L'effet dual de l'investissement est à la base de ce dilemme: « if sufficient investment is not forthcoming today, unemployment will be here today. But if enough is invested today, still more will be needed tomorrow »[1].

8.2.2. La réponse néoclassique : stabilité

SOLOW : ACCUMULATION ET SUBSTITUTION ENTRE LES FACTEURS

D'inspiration purement keynésienne, le modèle d'Harrod-Domar ne répond pas aux canons de l'économie néoclassique. La réponse néoclassique vient en 1956, sous la plume de Solow dans son article « A Contribution to the Theory of Economic Growth ». Cet article se positionne clairement en opposition au modèle d'Harrod-Domar, dont il résume la conclusion ainsi : « even for the long run the economic system is at best balanced on a *knife edge* of equilibrium growth ». La critique fondamentale est celle-ci : « But this fundamental opposition of warranted and natural rates turns out in the end to flow from the crucial assumption that production takes place under conditions of fixed proportions. There is no possibility of substituting labor for capital in production. If this assumption is abandoned, the knife edge notion of unstable balance seems to go with it »[2].

---

[1] Domar [78] p. 49

[2] Solow [338] p. 65. Aujourd'hui encore, le terme *knife edge* est couramment utilisé par les économistes pour caractériser les modèles de type Harrod-Domar.

La substitution entre facteurs est au centre du paradigme néoclassique. L'outil par excellence pour en rendre compte est la fonction de production, plus précisément ici la fonction de production agrégée (cf. chapitre 4.4), plus précisément encore, la *fonction de production intensive*, évoquée au chapitre 5.3 (cf. figure 5.9), dont le présent article constitue en fait le baptême[1]. Solow présente donc ici un modèle de croissance économique, qu'il prétend avoir les mêmes hypothèses que celui d'Harrod, à la seule différence que les facteurs sont substituables.

Il y a deux facteurs de production : le travail (quantité *L*) et le capital (quantité *K*) ; la terre est absente. Les rendements d'échelle sont constants[2]. La production *Y* est nette de dépréciation. Elle consiste en un produit homogène qui sert de substance au capital aussi bien qu'à la consommation. Une hypothèse importante est le maintien permanent du plein emploi. La fonction de production est évolutive avec le temps *t*[3]. La population active croît à un taux constant *n* (composé) à partir d'un effectif de L_0. Le capital s'accroît du fait de l'application d'un taux d'épargne *s* au revenu Y(t). Les équations de départ sont donc :

$$Y = F(K,L) \qquad (8.9)$$

$$L(t) = L_0.e^{nt} \qquad (8.10)$$

$$\frac{dK}{dt} = s.Y \quad ou \quad \frac{dK}{dt} = s.F(K, L_0.e^{nt}) \qquad (8.11)$$

L'équation (8.11) indique le taux auquel le capital doit croître pour assurer le plein emploi. La fonction de production intensive a comme abscisse la variable k = K/L. La variable *k* est donc représentative de la proportion entre les facteurs de production. En différenciant cette équation par rapport au temps, en remplaçant *K* et *L* par leurs valeurs des équations (8.10) et (8.11) et en arrangeant, Solow aboutit à ce qu'il appelle l'équation fondamentale :

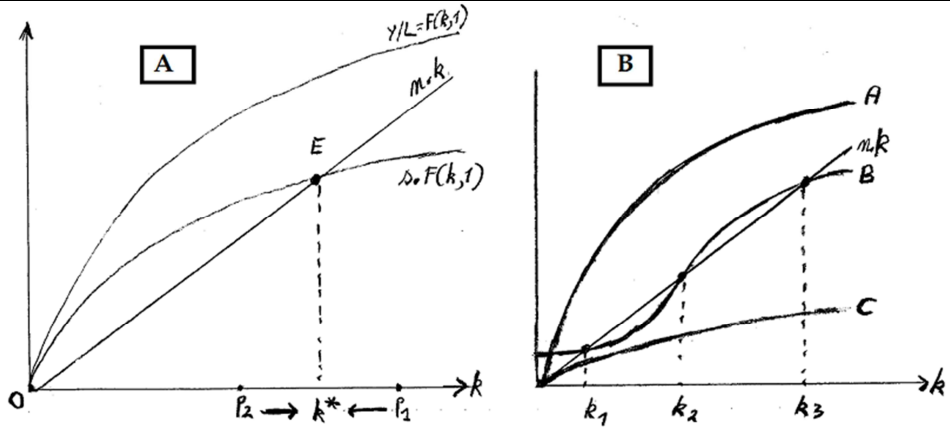
$$dk/dt = s.F(k,1) - nk \qquad (8.12)$$

L'objectif est de déterminer la valeur d'équilibre de *k* (notée k*). L'équilibre advient lorsque dk/dt vaut zéro (position de repos). Il faut donc qu'il y ait égalité entre s.F(k,1) représentant le taux d'accumulation du capital et *n.k* représentant le taux d'accroissement de la main d'œuvre. Sur la figure 8.2, cela correspond à l'intersection entre ces deux courbes. On est dans le cadre d'une croissance semi-stationnaire[4] : *Y*, *K* et *L* augmentent au même rythme, en l'occurrence au taux *n*.

L'équation différentielle (8.12) révèle que l'équilibre *E* est stable. Si, k > k*, n.k > s.F(k,1) et dk/dt devient négatif, ce qui signifie que *k* va diminuer jusqu'à rejoindre k* et inversement si k < k*.

[1] Le spectre de cette fonction hantait déjà l'article de J. Robinson qui avait lancé la controverse du capital.
[2] Dans ses illustrations algébriques, Solow prend souvent l'exemple d'une fonction de Cobb-Douglas.
[3] Solow travaille en temps continu, d'où la présence des exponentielles dans certaines équations.
[4] Cf. Cassel (sous-chapitre 6.1.1).

Graphique 8.2 : la croissance équilibrée selon Solow



La courbe Y/L ne joue aucun rôle, si ce n'est que la courbe s.F(k,1) en est déduite. Solow se permet même de l'omettre dans ses graphiques. Je l'inclus pour faciliter la compréhension.

Le graphique 8.2 est atemporel. La croissance n'y apparaît pas explicitement. Il faut imaginer qu'à chaque période *t*, il y a un graphique identique, mais avec une différence sous-jacente : *L* étant plus élevé, *Y* et *K* le sont aussi ; cela n'apparaît pas sur le graphique qui n'affiche que des variables dont *L* est le dénominateur.

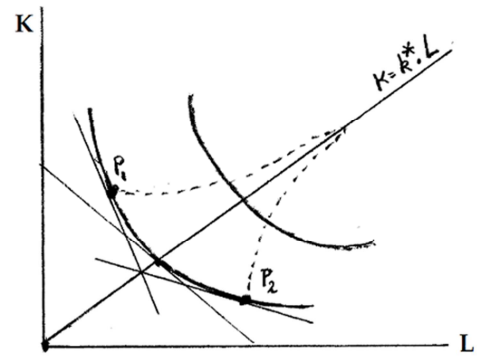
Supposons que l'économie, en position d'équilibre *E*, soit sujette à une hausse du taux d'épargne. La courbe s.F(k,1) s'élève et l'intersection avec *nk* se déplace vers la droite. Le nouvel équilibre aura une intensité capitalistique k* supérieure, mais le taux de croissance vaudra toujours *n*. Ce n'est qu'au cours de la transition d'un équilibre vers l'autre que le taux de croissance aura été supérieur.

Sur le graphique *A*, la courbe *s.F(k,1)* avait la forme idéale pour la démonstration. Le graphique *B* montre 3 autres formes possibles de la courbe *s.F(k,1)*, notées *A*, *B* et *C*. La courbe *B* (rendements d'échelle variables) est coupée trois fois par la droite *n.k*. Les équilibres k_1 et k_3 sont stables, mais k_2 est instable. Pour une intensité capitalistique inférieure à k_2, *k* tendra vers k_1 ; si *k* est supérieur à k_2, il tendra vers k_3. Les courbes *A*[1] et *C* ne sont pas coupées par *n.k* ; il n'y a donc pas de croissance équilibrée. L'économie *A* est tellement productive que la croissance démographique ne suit pas les besoins ; l'économie *C* est si peu productive que le plein emploi ne peut éviter la baisse du revenu par tête. Mais Solow conclut que lorsque les hypothèses néoclassiques habituelles sont satisfaites, aucun hiatus n'est possible entre le taux de croissance d'équilibre et le taux naturel.

L'hypothèse du plein emploi implique que les rémunérations des facteurs s'ajustent au niveau qui équilibre les marchés. La figure 8.3 montre cet ajustement sur un champ d'isoquants. La droite de pente k* est le lieu des combinaisons entre les facteurs, qui satisfont à la proportion d'équilibre. Elle implique un rapport de prix déterminé w/q (pente du segment d'isocoût) où *w* est le salaire réel et *q* le loyer fictif d'une unité de capital (qui correspond implicitement à un taux d'intérêt). En P_1, trop de capital est utilisé ; le salaire réel doit diminuer pour inciter les entreprises à substituer du travail au capital, ce qui fait baisser la pente des isocoûts. Le Point P_2 est dans la situation inverse. Les lignes pointillées indiquent le chemin parcouru par l'économie pour rejoindre la croissance équilibrée.

[1] La figure 8.2-B donne l'impression que les fonctions *A* et *nk* se croiseront à la droite du graphique. Cela tient à son caractère approximatif.

Graphique 8.3 : croissance équilibrée : rôle de la rémunération des facteurs



Solow admet avoir évacués les éléments keynésiens de son modèle. Il reproche à Harrod d'avoir mélangé la macroéconomie keynésienne avec l'analyse dynamique de la croissance. Selon lui, ces deux objets doivent être étudiés séparément. « Underemployment and excess capacity or their opposites can still be attributed to any of the old causes of deficient or excess aggregate demand but less readily to any deviation from a narrow 'balance' »[1].

Solow reconnaît ainsi implicitement que la différence essentielle entre son modèle et celui de Harrod ne se trouve pas dans les coefficients fixes ou variables comme il l'affirmait au début de l'article mais dans son hypothèse du plein emploi, contradictoire avec le taux de croissance capricieux de Harrod. Comme le note Hoover, commentateur des deux modèles, Solow se focalise sur l'adéquation entre le taux de croissance réel et le taux naturel ; cet aspect est secondaire chez Harrod en comparaison avec la dualité entre le taux réel et le taux d'équilibre.

LE RESIDU DE SOLOW

Est-ce là vraiment une théorie de la croissance ? La production se calque sur l'accroissement de la population ; le revenu par tête demeure constant, la société ne s'enrichit donc pas. Jusqu'ici, il ne fut pas question du progrès technique, dont chacun pressent qu'il est un facteur décisif de la croissance. Dans la section consacrée aux « extensions » du modèle, Solow lui consacre enfin une page sur les trente que compte l'article. Chez Harrod, le *taux de croissance naturel* inclut le progrès technique.

Solow se rattrapera par la suite en consacrant plusieurs publications au progrès technique. Voyons l'article « Technical Change and the Aggregate Production Function » (1957). Il définit ainsi l'objet de son article : « The new wrinkle I want to describe is an elementary way of segregating variations in output per head due to technical change from those due to changes in the availability of capital per head"[2].

Solow adapte la fonction de production, qu'il suppose homogène de degré un, en y ajoutant un facteur représentant le progrès technique, le facteur A(t). Il montre que si le progrès technique est *neutre*, c'est-à-dire s'il ne modifie pas le taux marginal de substitution entre les facteurs de production, la fonction de production prend la forme :

$$Q = A(t).f(K,L) \qquad (8.13)$$

Solow fait intervenir deux nouvelles variables : $w_K = \frac{\partial Q}{\partial K}.\frac{K}{Q}$ $et\ w_L = \frac{\partial Q}{\partial L}.\frac{L}{Q}$.

[1] Solow [338] p. 91.
[2] Solow [339] p. 312

Ce sont respectivement les parts relatives du capital et du travail dans le partage du revenu national. En différentiant (8.13) par rapport au temps et en divisant par *Q*, on obtient :

$$\frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + w_K\frac{\dot{K}}{K} + w_L\frac{\dot{L}}{L} \qquad (8.14)$$

Solow revient ensuite à la fonction intensive où les grandeurs sont divisées par *L*. Il est alors utile de noter Q/L par *q* et K/L par *k*. Comme $\dot{q}/q = \dot{Q}/Q - \dot{L}/L$, on a :

$$\frac{\dot{q}}{q} = \frac{\dot{A}}{A} + w_K\frac{\dot{k}}{k} \qquad (8.15)$$

Telle est l'équation fondamentale que Solow va soumettre à une épreuve statistique, visant à déterminer la valeur de *A*, c'est-à-dire l'importance du progrès technique dans la croissance. Les statistiques (pour la période 1909-1949 aux Etats-Unis) lui renseignent l'évolution de la production, la part des revenus du capital dans le revenu national et l'évolution de la valeur du capital. Aucune statistique n'existe évidemment pour *A*, qui est un concept abstrait, mais il peut être déterminé par l'équation (8.15) dont il est le seul terme inconnu, d'où l'appellation le *résidu de Solow*.

Voici la conclusion de ces calculs : « It is possible to argue that about one eighth of the total increase is traceable to increased capital per man hour, and the remaining seven eighth to technical change »[1]. Le progrès technique compte sensiblement plus que l'accumulation du capital dans le progrès économique.

SWAN : CROISSANCE DE LA POPULATION, DU CAPITAL ET DU REVENU

La même année mais malheureusement pour lui dix mois plus tard, l'économiste australien Trevor Swan publie sa théorie de la croissance conçue indépendamment de celle de Solow mais néanmoins fort semblable.

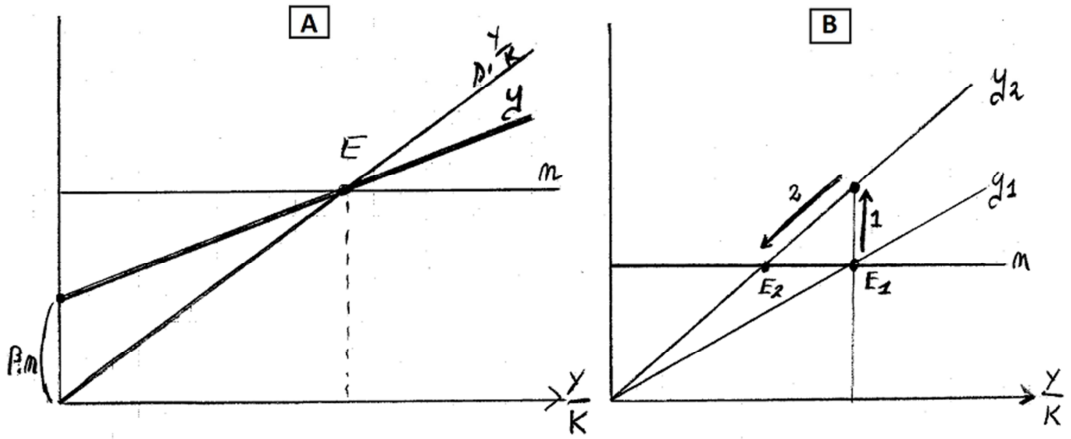
Swan base son modèle sur la fonction de Cobb-Douglas $K^{\alpha}.L^{\beta}$ avec $\alpha+\beta=1$. Les hypothèses sont grosso modo les mêmes que chez Solow. Sa théorie est parfaitement résumée par la figure 8.4. On y voit en abscisse, non le rapport capital/travail comme chez Solow mais le rapport production/capital. En ordonnée, les trois taux de croissance : le taux d'accroissement démographie *n*, exogène, le taux de croissance du capital qui vaut évidemment *s.Y/K* et le taux de croissance de la production, le plus important, noté *y*. la formule de ce dernier est :

$$y = \alpha.s.Y/K + \beta.n \qquad (8.16)$$

Le premier terme du membre de droite est la contribution du capital à la croissance et le deuxième la contribution du travail. L'équilibre a lieu lorsque les trois taux de croissance s'égalent. Lorsque *s.Y/K* vaut *n*, l'équation (8.16) vaut également *n*. Cet équilibre est stable. Swan en conclut : « The given rate of growth of labour thus determines the equilibrium rate of the whole economy, while the saving ratio determines the output-capital ratio at which equilibrium will occur »[2].

[1] Solow [339] p. 316.
[2] Swan [352] p. 336

Graphique 8.4: la croissance équilibrée selon Swan



*La figure **B** indique l'effet d'un accroissement de s. Dans un premier temps, le taux de croissance augmente pour rejoindre la nouvelle courbe y. Puis, le long de cette courbe, il diminue pour revenir à n. Le nouveau taux Y/K est plus bas.*

Comme chez Solow, c'est une croissance semi-stationnaire ; mais Swan introduit le progrès technique, *neutre*, à la fin de l'article. Un taux de progrès technique, noté *m* est introduit dans l'équation (8.16) qui devient $y = \alpha.s.Y/K + \beta.n + m$. Swan montre que dans ce cas :

$$y = n + m/(1-\alpha) \qquad (8.17)$$

Le progrès technique a donc sur le taux de croissance un impact supérieur à *m*. La cause en est que la contribution du capital à la croissance augmente, car le taux de croissance de celui-ci se libère de la contrainte de l'égalisation avec *n*. Graphiquement, la courbe *y* monte en altitude d'une valeur *m*, le taux de croissance d'équilibre correspond à l'intersection de cette courbe avec *s.Y/K* qui trouve place au-dessus de la droite horizontale *n*.

Swan semble désolé de constater que tant en croissance cassellienne qu'avec le progrès technique, la propension *s*, absente des formules (8.16-8.17), n'a pas d'effet définitif sur le taux de croissance. Un accroissement de *s* voit son effet positif contrebalancé par la baisse du ratio Y/K, comme le montre la figure 8.4-B. L'impuissance du taux d'épargne se retrouvait également chez Solow. Swan se console en suggérant que le taux de progrès technique est peut-être dépendant de l'accumulation du capital. L'accumulation du capital peut également stimuler la croissance, pense-t-il, lorsqu'il existe une main d'œuvre inemployée que le nouveau capital met au travail. Là, Swan néglige le fait que le chômage de la main d'œuvre va généralement de pair avec le sous-emploi du capital.

8.2.3. Progrès technique et fonction de production

INFLUENCE DU PROGRES TECHNIQUE SUR LA REPARTITION DU REVENU

Hicks est l'un des premiers économistes à s'être intéressé aux effets sociaux du progrès technique, précisément dans le chapitre de la « Theory of Wages » où il introduisait l'*élasticité de substitution*.

En situation concurrentielle, une invention ne peut être profitablement adoptée que si en fin de compte elle accroît le dividende national. En effet, elle ne sera rentable pour l'entrepreneur que si elle lui permet d'obtenir le même produit avec moins de ressources. Des ressources sont donc libérées pour qu'une entreprise, la même ou une autre, mette sur le marché un supplément de produit.

Mais il est peu probable que l'invention augmente le produit marginal de tous les facteurs dans la même proportion. "'labour-saving' inventions increase the marginal product of capital more than they increase the marginal product of labour; 'capital-saving' inventions increase the marginal product of labour more than that of capital; 'neutral' inventions increase both in the same proportion"[1]. Rappelons-nous qu'en situation concurrentielle, la rémunération se règle suivant sa productivité marginale.

Une invention économisant le travail ne DOIT donc pas diminuer le produit marginal du travail et par là la part ABSOLUE de ce facteur dans le revenu national, bien que cette éventualité soit possible. Par contre, une invention économisant le travail ou économisant le capital réduit toujours la part RELATIVE du facteur économisé.

Hicks est convaincu que les inventions économisant le travail prédominent dans la réalité économique. L'explication est la suivante. L'épargne cumulée des générations accroît l'offre de capital plus que n'augmente celle du travail. Le travail tend donc à devenir le facteur cher ; l'incitation à l'économiser est donc plus forte[2].

Cette dernière remarque amène à distinguer les *inventions induites* et les *inventions autonomes*. Les premières résultent de la volonté de réduire le coût du facteur dont le prix relatif s'est accru ; compte tenu de la remarque ci-avant, elles sont plus souvent labour-saving. Les secondes étant indépendantes du prix des facteurs, la balance entre les deux biais doit plus ou moins s'équilibrer. Donc au total, l'ensemble des inventions penchera du côté de celles qui économisent le travail. Certaines inventions induites par le renchérissement du travail, plus efficaces, auraient été rentables même au rapport de prix antérieur ; ce sont elles que les salariés doivent le plus craindre.

Hicks remarque encore : « since an enlarged absolute return is more likely to stimulate an increase in the supply of capital than an increase in the supply of labour, autonomous inventions may have the secondary effect in encouraging the accumulation of capital »[3]. Le travail en bénéficierait alors indirectement.

En définitive, Hicks est confiant quant à la préservation de la part absolue du travail dans le revenu national. Par contre, « it is difficult to feel the same degree of optimism in the matter of relative shares »[4].

Harrod n'est pas satisfait de la définition hicksienne de la *neutralité du progrès technique*. Il lui reproche de la rendre dépendante de facteurs extrinsèques à l'invention elle-même. Les productivités marginales des facteurs dépendent effectivement de l'élasticité de substitution dans d'autres secteurs de production ainsi que de l'élasticité de l'offre des facteurs.

Harrod avance sa propre définition : « I define a neutral advance as one which, at a constant rate of interest, does not disturb the value of the capital coefficient ; it does not alter the length

---

[1] Hicks [151] p. 122. Les économistes qualifient le progrès technique de *neutre-Hicks* si le rapport des productivités marginales n'a pas changé ALORS QUE L'INTENSITE DES FACTEURS, QUELLE QU'ELLE SOIT, EST RESTEE CONSTANTE.

[2] En l'absence de progrès technique, l'accroissement de l'offre relative d'un facteur sur une longue période finit par abaisser l'élasticité de substitution, ce qui est défavorable à ce facteur. Des inventions qui économisent l'autre facteur tendent à à compenser cette évolution en augmentant l'élasticité de substitution.

[3] Hicks [151] p. 128

[4] Hicks [151] p. 130

of the production process »[1]. Ce qui importe selon Harrod, ce n'est pas l'invention isolée, mais le flux d'inventions au niveau agrégé. Il est possible que le flux agrégé satisfait le critère même si ce n'est la cas d'aucune invention particulière.

Si le taux d'intérêt est constant, un flux d'inventions *neutres-Harrod* n'altère pas la répartition du revenu national entre les facteurs. Il assure également l'égalité entre le taux de croissance de *Y* et celui de *K*. Harrod écrit : "As I have chosen to approach the dynamic problem by asking what rate of increase of capital would be consistent with certain rates of increase of other parts of the system, it has seemed simplest to define a neutral stream of inventions as one which shall require a rate of increase of capital equal to the rate of increase of income engendered by it"[2]. Pour que l'économie connaisse une croissance semi-stationnaire, le progrès technique doit être *neutre-Harrod*.

Dans l'optique de Harrod, une invention nécessitant que le capital croisse à un taux supérieur à celui du revenu est dite *labour-saving*.

Supposons une économie dont la population et la main d'œuvre augmentent à un taux annuel régulier de *n* et dont le taux d'intérêt est constant. Le progrès technique est absent. Quel est le taux d'épargne annuel $s = \Delta K/Y$ requis ? Il est égal au coefficient de capital K/Y multiplié par *n*. Par exemple si K/Y = 4 et *n* = 0,01, un taux d'épargne de 4% sera requis. Inversons maintenant les paramètres : la population est constante mais il y a un progrès technologique neutre régulier au taux annuel λ ; le taux d'épargne requis se calcule de façon similaire ; $s' = \lambda.K/Y$. Et si l'économie connaît simultanément une croissance démographiques et un progrès technique neutre, tous deux réguliers, le taux d'épargne requis vaudra $s+s'+ss'$, soit $s+s'$ puisque ss' est négligeable.

On peut classer les progrès techniques selon la façon dont ils influencent la fonction de production. L'expression la plus générale d'une fonction de production intégrant le progrès technique est :

$$Q = F(K,L,t) \qquad (8.18)$$

Comme Solow, désignons le niveau de la technique au temps *t* par la variable A(t). Deux cas particuliers sont :

- Le *progrès augmentant le travail* :

$$Q = F(K,A(t)L) \qquad (8.19)$$

- Le *progrès augmentant le capital* :

$$Q = F(A(t)K,L) \qquad (8.20)$$

Une *fonction augmentant le travail* est comparable à une fonction qui serait sans progrès technique et où la main d'œuvre augmenterait au taux A(t). On utilise les termes *unités de travail efficace* pour le produit A(t).L.

Si A(t) connaît une croissance à taux constant λ, il augmente exponentiellement et la fonction (8.19) devient $F(K,e^{\lambda t}L)$ en temps continu.

Les fonctions dont le progrès technique est *neutre-Hicks* sont du type de l'équation (8.13). Le progrès technique *neutre-Harrod* est nécessairement du *type augmentant le travail*. Les types

[1] Harrod [133] p. 23. Le *coefficient de capital* est le rapport K/Y.
[2] Harrod [133] p. 26

de fonction (8.13), (8.19), (8.20) ne sont pas mutuellement exclusifs. Bien qu'il s'agit de cas exceptionnels, des fonctions peuvent satisfaire même les trois conditions simultanément. Pour que le progrès technique soit neutre à la fois selon Hicks et selon Harrod, l'élasticité de substitution entre *K* et *L* doit valoir l'unité. C'est le cas de la fonction Cobb-Douglas.

Un progrès technique *neutre-Hicks* d'ampleur λ a pour effet de déplacer la *frontière des prix des facteurs* (graphique 5.11) vers le nord-est de telle façon que chaque point est remplacé par un point pour lequel tant *w* (l'ordonnée) que *r* (l'abscisse) sont multipliés exactement par 1+λ.

PROGRES TECHNIQUE INCORPORE

Nous avions quitté Solow au sous-chapitre 8.2.2 alors qu'il constatait la part assez faible de l'accumulation du capital dans l'explication de la croissance. Cette constatation lui inspire réticence. Il revient à la charge dans plusieurs écrits entre 1959 et 1963 avec une nouvelle idée : si le progrès technique s'incorpore dans le capital, l'accumulation de celui-ci regagne en importance. Voici comment il introduit le concept de *progrès technique incorporé* : "Shifts in technology can be divided into two kinds, which may be described as disembodied and disembodied. Disembodied improvements in technique are purely organizational; they permit more output to be produced from unchanged inputs, with no investment required. Embodied improvements in technique permit an increased output from an 'ultimately' smaller input, but require the construction of new capital goods before the knowledge can be made effective"[1].

L'article « Investment and Technical Progress » (1959) réexpose son précédent modèle, qui a entretemps inspiré d'autres économistes. En la supposant du type *Cobb-Douglas* à rendements d'échelle constants, avec un progrès technique neutre-Hicks à taux constant λ (donc exponentiel), la fonction de production s'écrit :

$$Q(t) = Be^{\lambda t}L(t)^{\alpha}K(t)^{1-\alpha} \qquad (8.21)$$

Appelons-la la fonction *T* (comme « traditionnelle »). Le progrès technique y est « disembodied ». *Q* augmente au taux λ même si *K* et *L* sont constants, ce qui surprend. Il n'y a aucune interrelation entre λ et le taux d'investissement. Solow regrette : "The striking assumption is that old and new capital equipment participate equally in technical change. This conflicts with the casual observation that many if not most innovations need to be embodied in new kinds of durable equipment before they can be made effective"[2].

Solow va lui opposer une fonction que j'appellerai *I* (comme « incorporé »). Voyons comment il la construit. L'idée de base est la distinction entre les équipements des différentes dates de construction que Solow appelle « vintages ». Après *t*, une deuxième variable temporelle, notée *v*, est introduite pour désigner ces dates de construction-installation. Ainsi, I(v) est l'investissement au temps *v* (y compris le remplacement) et $K_v(t)$ représente l'ensemble des machines produites en *v* qui sont opérationnelles en *t*. Leur taux de dépréciation est δ, ce qui correspond à une durée de vie de 1/δ. Chaque nouvelle génération d'équipement incorpore le progrès technique de son millésime qui la rend (1+λ) fois plus productive que la précédente. A chaque période, les salariés sont redistribués des équipements déclassés vers ceux de la nouvelle génération. Solow travaille en temps continu : les périodes sont donc infiniment petites.

[1] Solow [341] p. 216. Solow ne se risque pas à diagnostiquer lequel des deux types de progrès technique est plus courant dans l'économie réelle.
[2] Solow [340] p. 91

Il y a une fonction de production Cobb-Douglas $Q_v(t)$ par génération d'équipement et la production totale est l'addition (par intégration) de ces fonctions partielles, ce qui donne :

$$Q(t) = B.e^{-\delta(1-\alpha)}.L(t)^{\alpha}.J(t)^{1-\alpha} \qquad (8.22\text{-A})$$

$$o\grave{u}\, J(t) = \int_{-\infty}^{t} e^{\left(\delta+{}^{\lambda}/_{1-\alpha}\right).v}.I(v)dv \qquad (8.22\text{-B})$$

J(t) est une espèce d'équivalent-capital. Solow compare les résultats de la fonction *I* avec ceux de la fonction *T*, qu'il retouche de façon à remplacer *K* par une expression contenant les variables I(v) et δ. A l'aide de données empiriques, Solow procède à une estimation, très approximative- il l'avoue- de λ : la fonction *I* donne une valeur plus élevée que la fonction *T*. Ce n'est pas étonnant, puisque le progrès technique n'y agit que par le vecteur plus étroit des équipements récents.

Le but de Solow est de comparer l'importance du taux d'accumulation dans les fonctions *T* et *I*. A l'aide d'exemples chiffrés basés sur des valeurs plausibles de principaux paramètres, il envisage deux hypothèses :

- I(v) est constant puis à partir du temps (peut-être mal nommé) t_0, il est doublé.
- I(v) est constant, puis il augmente annuellement de 5% à partir de t_0.

Après dix ans à ce régime, de combien Q(10) est-il supérieur à Q(0) ? De combien aurait été l'accroissement de *Q* sans ce supplément d'accumulation ? Clairement, le supplément d'accumulation a plus d'effet avec la fonction *I* qu'avec la fonction *T*. Solow peut conclure que la fonction *I* « makes the extra capital formation and technical progress share almost equally in the increase in output »[1].

Dans un article plus empirique paru en 1962, Solow se pose la question : « How much fixed investment is necessary to support alternative rates of growth of potential output in the United States in the near future ? »[2] Le terme « potentiel » est important ; il exclut l'inactivité des facteurs de production ; il faut néanmoins tenir compte de cette inactivité car les statistiques ne donnent que la production réelle. Solow reconstruit une fonction de type *I* mais légèrement modifiée. La production potentielle P(t) est ici donnée par l'équation :

$$P(t) = F\big(L(t), J(t)\big) \qquad o\grave{u}\, J(t) = \sum_{v=-\infty}^{t} (1+\lambda)^{v}.B(t-v).I(v) \qquad (8.23)$$

B(t-v) exprime la proportion des équipements construits en *v* qui sont opérationnels en *t*. L'*équivalent-capital* J(t) est l'addition du nombre de machines opérationnelles de chaque vintage, pondéré par le facteur d'accroissement de la productivité. On reconnaît ici une fonction *capital-augmenting*, car λ est entièrement intégré dans J(t).

Solow remarque une contradiction : si les facteurs *K* et *L* sont substituables selon les critères habituels, il est impossible qu'une partie du facteur *K* soit inactive tant que sa productivité marginale en valeur reste positive, car une fois acquis, il ne coûte plus rien et son usage ne peut donc qu'être avantageux. La solution proposée par Solow consiste à concevoir différemment la substitution entre *K* et *L* : l'installation de nouveaux équipements est l'occasion pour la firme de revoir son intensité capitalistique, mais pendant la durée de vie du matériel, la substitution s'avérera impossible. Solow ne recourt toutefois pas à ce type de substitution dans l'article de 1962, car il est trop complexe pour une application empirique.

[1] Solow [340] p. 99
[2] Solow [342] p. 76

Solow n'est pas seul sur la piste de ce capital de moins en moins *malléable*. Hahn et Matthews écrivent à ce sujet : « It may be however that while an entrepreneur is able to choose between techniques with varying capital-labour ratios at the time when he is installing a new machine he loses his freedom of action once the machine has been bought: each machine is designed to be worked by a crew of men and the crew cannot thereafter be changed. In the terminology of Johansen (1959), there is then 'ex ante substitutability' between labour and capital, but no 'ex post substitutability'; or in that of Phelps (1963), capital is 'putty-clay' (putty ex ante, clay ex post), instead of being pure putty »[1]. Dans ces modèles, le cas échéant, le matériel est déclassé à la suite de son obsolescence et non plus du fait de son usure physique. Les équipements les plus anciens cessent d'être rentables car leur utilisation requiert trop de personnel au vu du salaire qui augmente avec le revenu par tête. Contrairement au progrès technique incorporé, la qualité *putty-clay* d'une technologie n'implique pas la croissance économique ; elle est compatble avec des modèles de production stationnaires.

8.2.4. La croissance vue par les Cambridgiens

Les Cambridgiens s'expriment sur le sujet dès 1956 avec le livre « Accumulation of Capital » de Joan Robinson et l'article « A Model of Economic Growth » de Kaldor paru l'année suivante. En 1962, Robinson reprécise sa pensée dans un nouveau livre « Essays in the Theory of Economic Growth », que nous examinons ici.

Ce livre aborde non seulement la croissance économique mais également la méthodologie d'analyse de cette question complexe ; il se montre critique envers les tentatives antérieures. Robinson range les modèles de croissance en deux catégories :

- Les modèles FERMES. Il s'agit principalement des modèles néoclassiques. L'équilibre se découvre par la résolution d'un système d'équations simultanées. "The equations may determine a path through time- say a continuous accumulation of capital, or a particular pattern of fluctuations. But the time through which such a model moves is, so to speak, logical time, not historical time"[2].

- Les modèles OUVERTS. Ils fonctionnent en temps HISTORIQUE. Le présent se situe entre un passé irrévocable et un futur inconnu, alors que les équations des modèles fermés déterminaient le passé autant que le présent et le futur. La causalité joue un rôle important. L'équilibre résulte des actions des agents et non d'un système d'équations. Robinson s'insurge également contre les modèles qui greffent des relations causales de type keynésien sur des équilibres à la walrassienne qui ne connaissent que la simultanéité.

Les modèles ouverts n'offrent pas la facilité de fonctionner avec un nombre restreint de variables bien définies, homogènes. Cette remarque de Robinson caractérise bien son propre ouvrage. A chaque chapitre, le lecteur est confronté à une multitude de scénarios, chacun valable dans une situation bien précise. Si le réalisme y gagne, la lecture peut s'avérer parfois un peu désespérante.

CROISSANCE ET ACCUMULATION

Robinson classe son modèle parmi les modèles KEYNESIENS, dont elle distingue plusieurs types selon la motivation qu'ils attribuent à l'accumulation du capital ; le modèle de Harrod est ainsi rangé parmi les « capital stock-adjustment mechanism » ; le sien intègre la catégorie « animal spirits » : « In our model, the inducement to invest is conceived in terms of a desired rate of

[1] Hahn & Mattews [126] p. 838
[2] Robinson [304] p. 24

growth rather than a desired stock of capital »[1]. Mais elle y intègre des éléments microéconomiques néo-ricardiens. Le système des prix relatifs tout entier, en ce compris le *salaire réel*, dépend de seulement deux facteurs : les conditions techniques et le taux de profit. La demande exercerait également une influence s'il y avait des facteurs rares (la terre) ou si les rendements d'échelle n'étaient pas constants, deux phénomènes exclus par les hypothèses simplificatrices. Le *salaire nominal* détermine le niveau absolu des prix. La question se pose alors de ce qui détermine le taux de profit[2]. Robinson considère qu'aucune des théories existantes n'est convaincante et parle d'un « huge blank in traditional economic teaching »[3]. Elle avance sa propre explication : c'est le taux d'accumulation qui détermine le taux de profit.

$$I/K => S/K => (s_w.w.L + s_c.r.K)\ /K => r \qquad (8.24)$$

Le ratio I/K n'est autre que le taux d'accumulation ; l'identité keynésienne entre l'épargne (S) et l'investissement (I) sert de courroie de transmission[4]. Compte tenu des propensions marginales à épargner des salariés (s_w) et des capitalistes ($s_c > s_w$), le taux de profit $r = P/K$ est la variable d'ajustement par l'entremise du système de prix qui détermine leurs revenus respectifs $w.L$ et $r.K$. « The main weight of the equalisation of saving to investment (at normal prices) falls upon the distribution of income between classes. The level of normal prices has to be such as the rate of profit is such that the distribution of income is such that the ratio of saving to the stock of capital is equal to the rate of accumulation »[5].

Une nouvelle question vient donc : qu'est-ce qui détermine le taux d'accumulation ? Ici également, Robinson constate un blanc dans la science économique. Aucune des variables souvent citées, qu'elle passe en revue, ne la convainc. Elle en conclut que le moteur se trouve dans les caractéristiques sociologiques de la société considérée, plus ou moins propices au dynamisme. Robinson reconnaît toutefois une relation réciproque entre le taux de profit et le taux d'accumulation. Le profit dépend de l'accumulation puisqu'il est la somme de l'investissement et de la consommation des rentiers[6]. Parallèlement, il sert d'incitant à l'accumulation.

La figure 8.5-A décrit cette relation réciproque. La courbe *A* montre le taux de profit réellement généré par les différents taux d'accumulation. La courbe *I* indique le taux d'accumulation qu'induiraient les différents taux de profit. On peut diviser ce graphique en quatre parties :

- Zone I : taux d'accumulation excessif à corriger par des plans d'investissements plus modérés
- Zone II : taux d'accumulation trop bas à corriger par des plans d'investissements plus ambitieux
- Zone III : économie courant à la ruine, incapable de soutenir un taux d'accumulation déjà extrêmement bas.

---

[1] Robinson [304] p. 87

[2] Le *taux de profit* est omniprésent dans l'ouvrage : il s'agit toujours du taux de profit NORMAL, celui qui émerge de la concurrence généralisée et des transferts de capital entre secteurs de production, que d'autres auteurs appellent *taux d'intérêt*.

[3] Robinson [304] p. 11

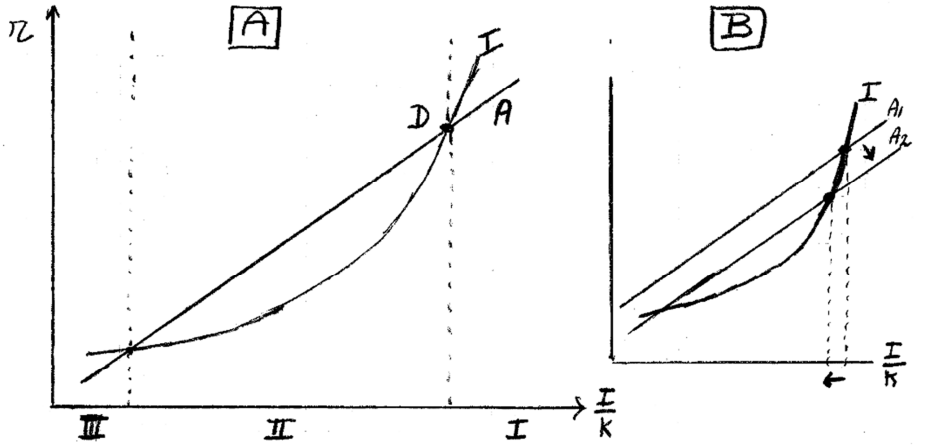
[4] Robinson rappelle qu'une part non négligeable de l'épargne est effectuée au sein même des entreprises (bénéfices non distribués).

[5] Robinson [304] p. 12. Selon Robinson, le modèle de Harrod est sur un *knife edge* pour n'avoir pas endogénéisé le taux de profit.

[6] Par hypothèse, Robinson néglige l'épargne des salariés.

- Le point D « … represents a rate of accumulation which is generating just the expectation of profit that is required to cause it to be maintained »[1]. Robinson l'appelle le *taux d'accumulation désiré*, car au point D, les firmes accumulent effectivement au taux désiré.

Graphique 8.5: le taux d'accumulation désiré



Comme Harrod et Solow, Robinson recherche une croissance équilibrée, stable dans le temps, caractérisée par l'absence de frictions entre les secteurs de l'économie : tous les éléments croissent au même rythme sur le long terme. Deux types de conditions doivent être satisfaites :

1- Une double condition technique : la distinction entre deux secteurs de l'économie est importante : celui qui produit des biens de consommation et celui qui produit des équipements[2]. Ils doivent chacun représenter une part de la production totale en accord avec le taux d'accumulation. La structure d'âge des équipements utilisés doit également être en accord avec le taux d'accumulation pour avoir un renouvellement sans à-coups. Le taux d'accumulation du passé doit donc déjà avoir tendu vers l'équilibre depuis un certain temps.
2- Le *taux de croissance possible* résulte de l'évolution de la population active et du progrès technique. Le taux désiré doit correspondre au taux possible.

Lorsque toutes les conditions d'équilibre sont satisfaites, l'économie est dans un *âge d'or*. « I used the phrase 'a golden age' to describe smooth steady growth with full employment (intending thereby to indicate its mythical nature) »[3]. A côté de l'*âge d'or*, Robinson conçoit et décrit divers chemins de croissance qui en diffèrent par l'une ou l'autre failles, auxquels elle colle des étiquettes imagées : *âge d'or boiteux* (en-dessous du plein emploi), *âge de plomb* (taux de chômage croissant), *âge d'or restreint* (progrès technique insuffisant), *âge de platine galopant* (croissance accélérée par un boom des investissements), *âge d'or bâtard* (taux d'accumulation réfréné pour le salaire réel trop élevé). A cette occasion, une multitude de scénarios sont évoqués qu'il est impossible de résumer ici.

L'impact d'une variation de la propension à épargner est extrêmement complexe. Robinson envisage une multitude de scénarios. Elle se contente d'une analyse de statique comparative entre économies caractérisées par des propensions différentes, pour ne pas la perturber par des effets de choc. Soit une propension à épargner plus élevée, qui est le fait des rentiers puisque l'épargne des salariés est considérée comme négligeable par hypothèse. Plus d'épargne signifie

[1] Robinson [304] p. 48
[2] Pour faciliter l'exposé, je les appellerai comme Marx, secteur I (équipements) et secteur II (consommation).
[3] Robinson [304] p. 52. Ce terme correspond grosso modo au *steady state* déjà rencontré.

dans l'optique keynésienne moins de revenu et donc ici moins de profit. Comme le montre la figure 8.5-B, le déplacement de la courbe de A1 à A2 réduit le taux de profit pour tout taux d'accumulation donné. L'incitation à investir s'en ressent et finalement, on assiste au résultat paradoxal que le taux d'accumulation désiré baisse. Mais dans le cas où le *taux d'accumulation possible* est inférieur au *taux désiré*, la diminution de ce dernier peut ne pas affecter le taux d'accumulation réel.

L'élévation de la propension à épargner peut néanmoins avoir un effet positif. La baisse induite du taux de profit stimule le niveau de mécanisation (rapport K/L) pour un niveau de développement technique donné. Robinson imagine deux économies dans un *âge de platine galopant*, dont la croissance élevée du secteur I permet d'augmenter l'emploi. L'économie qui épargne le plus mettra plus de temps pour atteindre le plein emploi, mais elle y arrivera avec des rapports K/L et Y/L plus élevés. Et cet avantage sera conservé lors du processus de décélération qui les amènera ensuite à l'*âge d'or*.

LE PROGRES TECHNIQUE

Robinson s'intéresse ensuite au progrès technique. Aux hypothèses précédentes sont ajoutées les simplifications suivantes : la population active est constante et homogène ; les biens de consommation constituent un ensemble dont la composition (mais pas la quantité) est invariable; le progrès technique n'affecte que le secteur II, par le fait qu'à chaque période sortent des usines du secteur I des équipements qui produisent plus de biens de consommation par travailleur occupé ; la production de toutes les générations d'équipements requiert la même consommation de travail et de capital ; le salaire nominal est constant, ce qui signifie que la hausses du salaire réel se réalise par la baisse du prix des biens de consommation. A pleine capacité, une usine du secteur II travaille avec un effectif de main d'œuvre donné que le progrès technique n'affecte pas.

Robinson a sa propre définition de la *neutralité* du progrès technique :

- Le progrès est *neutre* si la *valeur monétaire*[1] et la durée de vie de la nouvelle usine restent constantes par rapport à l'ancienne, alors que la production de l'usine augmente. Cela suppose que le rapport K/L n'a pas changé (où *K* est mesuré à son coût réel)[2].

- Concernant le progrès technique *biaisé*, Robinson l'appelle *capital-saving* si, pour une production supérieure, les usines dotées de la nouvelle technologie ont une *valeur monétaire* moindre (pour une durée de vie constante) ; lorsque leur *valeur monétaire* est supérieure, le progrès technique est dit *capital-using*.

En cas de progrès neutre, le rapport entre la production du secteur I et celle du secteur II n'a pas changé si on mesure la production du secteur I en unités de capacité de production du secteur II.

La nouvelle technologie est dite *supérieure* si elle accroît la production A LA FOIS par tête et par unité de *coût réel* de l'usine. Dans le cas contraire, l'amélioration n'est que *partielle*. Une nouvelle technologie résultant d'un progrès *neutre* est toujours *supérieure*. Les innovations provenant d'un progrès *baisé* ne sont pas nécessairement *supérieures*. Un progrès technique qui

[1] Le *coût réel* de l'usine est l'addition de son capital fixe et de son capital circulant valorisés avec le salaire réel pris comme numéraire. La *valeur monétaire* de l'usine est le coût réel multiplié par le salaire nominal. Quand il est question de coût réel ou de valeur monétaire de l'usine, il est toujours sous-entendu « au taux de profit en vigueur ».

[2] Commentant son critère, Robinson écrit : « Thus Hicks's criterion of neutrality is satisfied. His criterion, therefore, comes to the same thing as ours » ([304] p.113).

économise du capital peut valoir la peine même s'il réduit la production par tête, pour autant que la baisse de la production par tête soit moindre que celle du coût de l'usine. Un progrès technique qui use du capital se justifie même si la hausse la production par tête est moindre que celle du coût de l'usine, pour autant que la baisse du coût salarial par unité produite compense la hausse du coût du capital par unité produite.

En quoi consiste un *âge d'or* dans ces conditions ? Le taux de progrès technique doit être stable : de période en période, la production de biens de consommation par tête augmente à un taux identique. Le progrès technique est neutre. Le rapport capital sur travail reste invariable. Les parts respectives du capital et du travail dans le revenu national demeurent constantes : le taux de profit est stable mais il s'applique sur un capital croissant ; celui-ci augmente dans la même proportion que la production de biens de consommation ; le salaire réel croît à ce même taux. L'économie demeure sereinement à l'équilibre. Robinson rappelle que le concept d'*âge d'or* n'a pas pour vocation de représenter une réalité, mais de mettre en lumière les forces qui gouvernent l'économie en croissance.

Le progrès technique biaisé fait varier le *degré de mécanisation*, qu'il faut comprendre comme un synonyme du rapport K/L. Ce processus ne doit pas être confondu avec le choix des techniques variant par leur degré de mécanisation auquel les producteurs sont confrontés à chaque stade de l'évolution et qui a occupé les économistes lors de la controverse du capital. Dans ce choix, seul le rapport entre le salaire et le taux de profit permet de départager les techniques.

L'effet du *progrès technique biaisé* sur l'emploi et la prospérité est complexe à analyser. Supposons un progrès technique *capital-using*. La société n'évitera l'émergence d'un chômage technologique que si les producteurs de biens de consommation accroissent les sommes qu'ils destinent à l'investissement, car à budget constant, ils renouvelleraient les usines de façon incomplète. Pour maintenir l'emploi global, il faut un transfert de main d'œuvre du secteur II vers le secteur I. Si l'économie était en *âge d'or* avant l'innovation, on peut supposer qu'elle rejoindra un nouvel âge d'or caractérisé par un rapport K/L plus élevé, mais une période de transition s'intercalera entre les deux âges d'or, pendant laquelle la nouvelle technologie fera l'objet d'une adoption progressive par les entreprises.

L'effet du *progrès technique biaisé* sur la répartition du revenu entre les classes dépend de l'évolution du taux de profit. Soit une innovation *capital-using* : à taux de profit constant, la part des salariés dans le revenu baisse puisque la masse de capital croît plus que le revenu. Comme le profit vaut la consommation des rentiers plus l'investissement, son évolution dépend d'éventuels changements qui affectent ces grandeurs, tels une modification de la politique de distribution des dividendes ou de la propension marginale à consommer des rentiers.

PASINETTI COMPLETE LE MODELE

La même année 1962, l'article de Pasinetti « Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth » défend le même principe de façon plus succincte et plus algébrique. Compte tenu des propensions à épargner des salariés (s_w) et des capitalistes (s_c), il y a toujours une répartition de revenu entre ces classes qui assure l'égalité entre l'épargne *S* et l'investissement *I*, où *I* est plus spécifiquement le niveau d'investissement qui assure le maintien du plein emploi à long terme compte tenu du progrès technique et de la croissance démographique. Cette répartition est représentée par le rapport P/Y (où *P* est le profit et *Y* le revenu total) et il lui est associé un taux de profit P/K (où *K* est la valeur du capital) déterminé. Pasinetti calcule :

$$\frac{P}{Y} = \frac{1}{s_c - s_w}\ \frac{I}{Y} - \frac{s_w}{s_c - s_w}, \qquad \text{(8.25-A)}$$

$$\frac{P}{K} = \frac{1}{s_c - s_w}\ \frac{I}{K} - \frac{s_w}{s_c - s_w}\ \frac{Y}{K}, \qquad \text{(8.25-B)}$$

Pour éviter que la part soit des salaires ou des profits, soit nulle ou négative, Pasinetti impose en outre aux propensions à épargner des salariés et des capitalistes les restrictions suivantes : $s_w < I/K$ et $s_c > I/K$.

Pasinetti estime devoir reformuler cette théorie pour en corriger une faille. Si on prend en considération l'épargne des salariés, que Kaldor et Robinson avaient tendance à négliger, il faut tenir compte de ce qu'une partie des profits leur est attribuée : $P = P_c + P_w$. Les équations 8.25-A et 8.25-B restent valables si dans le membre de gauche, on remplace *P* par P_c. L'équation 8.25-A indique alors la distribution du revenu entre les classes. Par contre, pour le taux de profit, il importe de connaîtrez P/K plutôt que P_c/K.

Si l'on suppose qu'en équilibre dynamique de longue période, le taux rémunérant l'épargne des salariés correspond au taux de profit général, les équations (8.25-A) et (8.25-B) deviennent beaucoup plus simples :

$$P/Y = 1/s_c \,.\, I/Y \qquad (8.26\text{-A})$$
$$P/K = 1/s_c \,.\, I/K \qquad (8.26\text{-B})$$

Ces formules révèlent une dissymétrie entre les capitalistes et les salariés : la propension à épargner des premiers est une variable déterminante pour la répartition entre le salaire et le profit ainsi que pour le taux d'intérêt, contrairement à celle des salariés[1]. Cela est dû à ce qu'en équilibre dynamique de longue période, le rapport entre *P* et *S* est le même pour les deux classes, ce que démontre Pasinetti. Et cette relation est déterminée uniquement par le comportement d'épargne des capitalistes, car toute leur épargne provient des profits. Pour les travailleurs qui épargnent sur leur salaire, toute valeur de s_w est compatible avec le rapport entre *P* et *S*, car ils peuvent épargner sur leur salaire et consommer sur leurs profits.

Pasinetti remarque que les auteurs post-keynésiens, pour conclure à l'importance déterminante de l'épargne des capitalistes, se sont toujours sentis obligés de formuler l'hypothèse drastique que l'épargne des salariés est négligeable. « The novelty of the present analysis has been to show that the relation is valid independently of any of those assumptions. It is valid whatever the saving behaviour of the workers may be »[2].

Pour analyser la stabilité de l'équilibre résultant de son analyse, Pasinetti résout l'équation différentielle :

$$\frac{d}{dt}\left(\frac{P}{Y}\right) = f\left(\frac{I}{Y} - \frac{S}{Y}\right), \qquad f(0) = 0,\ f' > 0 \qquad \text{(8.27)}$$

En définitive, la seule condition de stabilité se révèle être l'évidente inégalité :

$$s_c > 0 \qquad (8.28)$$

---

[1] En fait, s_w (où l'indice *w* indique le bénéficiaire du revenu et non le type de revenu) influence la distribution du revenu entre les deux classes, mais pas entre les deux catégories de revenu.

[2] Pasinetti [274] p. 275

Dans une économie socialiste, l'Etat remplace les capitalistes dans les équations ci-dessus. Comme il ne consomme pas, on a $s_c = 1$. Les équations (8.26-A) et (8.26-B) sont donc remplacées par $P/Y = I/Y$ et $P/K = I/K$. En équilibre dynamique de longue période, l'investissement égale nécessairement le profit et le salaire égale nécessairement la consommation. L'Etat ne doit donc pas se préoccuper de la propension à épargner des citoyens, si ce n'est pour s'assurer que globalement la société n'épargne pas au-delà de ce que requiert le taux de croissance naturel, ce qui générerait du sous-emploi.

KALDOR ET LA FONCTION DE PROGRES TECHNIQUE

Dans l'article « Capital Accumulation and Economic Growth » (1961), Kaldor pense pouvoir dépeindre la croissance à l'aide de quelques *faits stylisés*, c'est-à-dire des tendances lourdes, exprimées sans une attention portée vers les détails. Les faits stylisés qu'il propose sont les suivants :

1. Croissance continue de la production globale et de la productivité du travail à un rythme assez stable ; la croissance de la productivité ne fait montre d'aucun signe d'essoufflement.
2. Hausse continue du capital par travailleur, quelle que soit le moyen de le mesurer.
3. Taux de profit constant sur la longue période, du moins dans les économies développées
4. *Ratio capital sur production* stable à long terme, sans trend à long terme. Ceci implique la presque-identité des taux de croissance du capital et de la production.
5. Forte corrélation entre la part des profits dans le revenu et la part de l'investissement dans la production. La part des salaires dans le revenu est stable, ce qui implique qu'ils augmentent au même taux que la productivité.
6. Variabilité du taux de croissance de la productivité du travail d'un pays à l'autre, mais les proportionnalités relevées ci-dessus transcendent ces différences.

Kaldor écrit : « None of these 'facts' can be plausibly 'explained' by the theoretical construction of neoclassical theory »[1]. Vu la théorie de la productivité marginale, on s'attendrait à ce que l'accumulation du capital s'accompagne d'une baisse tendancielle du taux de profit (explicitement prévue par Wicksell), d'un accroissement du *rapport capital sur produit* et à un affaiblissement progressif de la croissance de la productivité du travail. Il faut reconnaître qu'avant l'article de Solow de 1957 qui lui était consacré, le progrès technique était le parent pauvre de l'économie néoclassique. Depuis lors, les économistes ont tenté de combler cette lacune. Il est vrai qu'il aura fallu longtemps avant que la majorité des économistes néoclassiques renoncent à l'idée que l'accumulation du capital est le principal moteur de la croissance.

Kaldor lui reproche à Solow de ne pas vraiment rompre avec le passé. La théorie de la productivité marginale est juste amendée pour pouvoir assimiler le progrès technique. La fonction de production reste l'élément central ; simplement, l'état changeant de la technologie lui fait subir des sauts vers le nord-est du graphique. « As depicted in figure (8.6-A), at any point of time, *t*, there is assumed to be a unique relationship between capital and output, which conforms to the general hypothesis of diminishing productivity, but this relationship is constantly shifting with the passage of time ».[2]

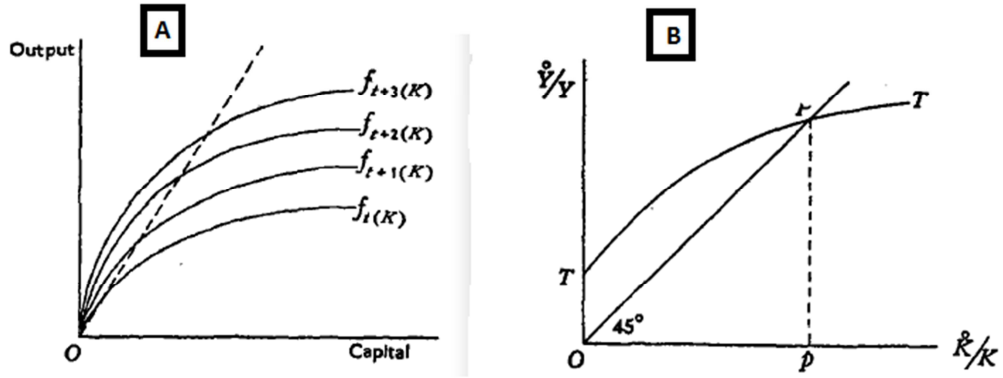
L'hypothèse de la *neutralité du progrès technique* signifie que les tangentes aux fonctions de production successives ont la même pente, à l'intersection de tout rayon tiré à partir de

[1] Kaldor [174] p. 179
[2] Kaldor [174] p. 204

l'origine. C'est cette hypothèse qui permet à la théorie néoclassique de rencontrer certains faits stylisés. Kaldor estime improbable que le progrès technique obéisse à une règle aussi rigide. Il adresse en outre quelques autres critiques :

1- Cette théorie suppose que l'économie peut à tout moment adapter sans délai n'importe quelle quantité de capital à n'importe quelle quantité de travail alors que ce n'est possible qu'en modifiant la nature qualitative des équipements. En l'absence inévitable de cette flexibilité infinie, l'économie évolue non le long de la courbe mais à l'intérieur et la pente de la courbe ne peut donc pas refléter la rémunération d'un facteur.
2- Le progrès technique n'est pas quantifiable. Un état donné de la connaissance technologique ne peut être défini qu'implicitement. La volonté d'isoler le déplacement de la courbe du mouvement le long de cette courbe repose sur un raisonnement qui présuppose sur ce qu'il veut prouver. L'hypothèse d'une fonction Cobb-Douglas rend inévitable que le résidu soit attribué au déplacement de la courbe.
3- Implicitement, la connaissance technique est traitée comme un facteur de production. Mais son caractère d'élément rare non appropriable et non quantifiable ne permet pas qu'elle soit rétribuée à son propre produit marginal. La bonne méthode consiste à se contenter des « vrais » facteurs dans la fonction de production, le capital et le travail, et à reconnaître à cette fonction la propriété de rendements d'échelle croissants[1]. Mais il est alors illégitime de supposer que les facteurs sont rémunérés suivant leur productivité marginale, vu la théorie de l'épuisement du produit.
4- Isoler les déplacements de la courbe a d'autant moins de sens que l'innovation technologique n'est pas une simple fonction du seul écoulement du temps. Elle-même dépend de l'accumulation du capital (mouvement le long de la courbe) pour se concrétiser.

Graphique 8.6 : Kaldor et le progrès technique



Ce dernier point est essentiel. « Each society has only a limited capacity to absorb technical change in a given period »[2]. L'accroissement de la production généré par l'accroissement du capital sera plus ou moins élevé, pas tellement selon le progrès de la connaissance mais selon le rapport entre le taux d'accumulation et la capacité de mettre en œuvre les innovations. Kaldor bâtit la *fonction de progrès technique* pour exprimer cette relation intégrant l'accumulation du capital et le progrès technique sans tenter de les séparer. Comme le montre la figure 8.6-B, tant que le taux d'accumulation est inférieur à *p*, le taux de croissance de la production est supérieur au taux d'accumulation, puis il devient inférieur. L'altitude de la courbe indique le dynamisme de la société, son esprit d'innovation. Même pour une accumulation nulle, la production peut

[1] L'assimilation du progrès technique aux rendements d'échelle croissants me paraît peu convaincante, car il n'est pas lié à la taille du processus de production.
[2] Kaldor [174] p. 207.

croître d'une part grâce aux investissements de remplacement, d'autre part parce que certains progrès ne nécessitent pas d'investissement.

8.2.5. La règle d'or et le retour de Ramsey

PHELPS ET LA REGLE D'OR

Au début des années 1960, quelques économistes, Allais, Robinson, Phelps et d'autres, découvrent et énoncent, indépendamment l'un de l'autre, un théorème que Phelps baptise *règle d'or de l'accumulation*, dans son article « The Golden Rule of Accumulation ; A Fable for Growthmen » (1961), que nous allons résumer à présent. L'article est rédigé comme un conte.

Dans un pays imaginaire nommé SOLOVIA, les autorités se préoccupent de la croissance de leur économie. Le progrès technologique et l'expansion démographique étant perçus comme des données exogènes, la variable sur laquelle la politique peut influer se révèle être le *rapport capital sur travail*. Le gouvernement estima qu'il convenait de fixer un taux d'investissement constant d'année en année et lança un concours parmi les citoyens pour récompenser celui qui trouverait le taux d'épargne optimal. Ce fut un certain OIKO NOMOS qui remporta le prix.

Oiko formule les hypothèses suivantes :
- Solovia a toujours été, est et restera toujours dans un *âge d'or*.
- Le taux de croissance d'âge d'or g étant naturel, donc indépendant du taux d'accumulation, il faut une substituabilité parfaite entre les facteurs pour que le ratio capital sur produit (K/Y) s'adapte pour égaliser dK(t)/dt avec g, quelle que soit la valeur du taux d'épargne s.

Ecoutons les explications d'Oiko à la foule rassemblée : « Each generation in a boundless golden age of natural growth will prefer the same investment ratio, which is to say the same natural growth path. In deciding which growth path is best from its standpoint, a generation will look only at the amount of consumption which each path offers it"[1]. Après la démonstration de ce lemme, Oiko en arrive au théorème fondamental : « Along the optimal golden age path, under conditions of natural growth, the rate of investment is equal to the competitive rate of profits »[2]. C'est cette égalité qu'Oiko appelle la *règle d'or de l'accumulation*.

Dans un âge d'or gouverné par la règle d'or, chaque génération investit pour le compte des suivantes la proportion du revenu identique à celle que les générations précédentes ont investie pour son compte.

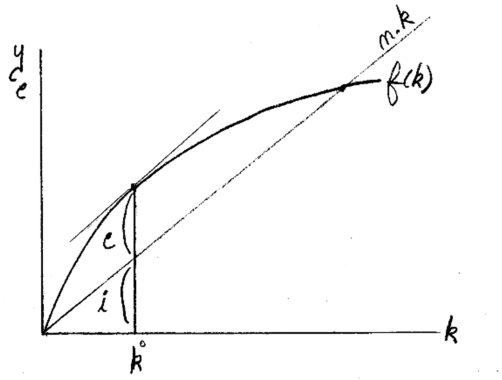
Que faire pour rejoindre le sentier de croissance correspondant à la règle d'or et comment s'assurer quelle est notre position vis-à-vis de ce sentier demanda la foule à Oiko. La solution est simple : au sentier de la règle d'or est associé un rapport K/Y unique. Si notre ratio est plus élevé que celui de la règle d'or, nous devons consommer plus ; dans le cas contraire, nous devons freiner notre consommation.

Attention au fait que l'objectif de la *règle d'or* consiste à maximiser non la production, mais la consommation. Accroître l'épargne peut stimuler la production, ce qui favorise la consommation, mais comme l'épargne est par définition de la non-consommation, il y a un juste équilibre à trouver. C'est ce qu'indique la figure 8.7. Y apparaissent la fonction de production intensive, qui indique la production par tête, ainsi que la droite d'investissement au

[1] Phelps [279] p. 640
[2] Phelps [279] p. 641

taux naturel (n). L'écart entre les deux correspond évidemment à la consommation et il est maximum lorsque la tangente à f(k) est parallèle à la droite d'investissement, soit pour un rapport K/L valant k°.

Graphique 8.7 : la règle d'or de l'accumulation



Dans un article ultérieur (1965), Phelps corrigera deux erreurs du présent article. D'une part, il est possible de concevoir un équilibre de règle d'or dans un environnement à coefficients de production fixes, comme celui du modèle Harrod-Domar (isoquants à angle droit : pas de substitution entre les facteurs). D'autre part, l'éventail des fonctions de production compatibles avec la règle d'or doit être restreint : « a positive-investment GR path can exist only if technical progress can be described as purely 'labor augmenting' »[1].

Phelps ajoute : « But no proof of the 'optimality' of the GR path was given nor was any suggestion of its optimality seriously intended. Society need not confine itself to golden age paths (should they exist) nor aim to achieve golden age growth asymptotically". Mais il démontre quand-même que les sentiers de croissance qui accumulent durablement le capital au-delà de l'intensité capitalistique exigée par la règle d'or sont nécessairement dynamiquement inefficients. Il démontre encore qu'un sentier de croissance de règle d'or existe toujours, indépendamment de la substitution entre les facteurs, si les conditions suivantes sont réunies :

- Un input de travail est nécessaire à la production ;
- La main d'œuvre augmente à un taux constant ;
- La dépréciation du capital est à taux constant ;
- Le progrès technique est purement labor-augmenting (à un taux constant).

CASS, KOOPMANS ET RAMSEY

En 1965, **Cass** et **Koopmans** (séparément) combinent les analyses de Solow et de Phelps avec l'analyse plus ancienne de Ramsey. Comme chez Ramsey, l'optimum requiert la maximisation non de la production ou de la consommation, mais de l'utilité que cette dernière procure. Koopmans reprend la problématique de Ramsey telle qu'il l'avait adaptée, avec un escompte des satisfactions futures sur un intervalle de temps infini (cf. chapitre 8.1). Pour bien comprendre la problématique, il faut se mettre dans la peau d'un planificateur social qui a la responsabilité de déterminer le taux d'accumulation (et donc de croissance) pour l'éternité ; la question est : comment maximiser l'utilité intertemporelle générale en ne sacrifiant pas certaines générations au profit des autres. Les éléments du modèle sont :

[1] Phelps [280] p. 794

- Un bien unique (servant à la production et la consommation) est produit à partir des facteurs *K* et *L* dans le cadre d'une fonction de production *F* homogène de degré un. La technologie est constante.
- La main d'œuvre croit à un taux *n*. En temps continu, on a : $L_t = L_0.e^{nt}$. Contrairement à Cass qui complique son modèle avec la distinction entre les ménages qui consomment et les individus qui travaillent, pour Koopmans, un travailleur égale un consommateur. L(t) indique donc le nombre d'individus dont l'utilité au temps *t* doit être prise en compte.
- $u(c_t)$ est l'utilité de la consommation immédiate par tête ; on a : $c_t = C_t/L_t$ (répartition égalitaire au sein de chaque génération). La fonction u(.) est invariable dans le temps. Sa concavité stricte « implies that we attribute greater weight to the marginal unit of per capita consumption of a poor generation as compared with a rich one »[1]. Pour éviter qu'une génération tombe sous le seuil de subsistance, Koopmans impose que $u(c_t)$ tend vers $-\infty$ lorsque c_t tend vers zéro.
- Les satisfactions futures sont escomptées au taux ρ, pour la raison indiquée au chapitre 8.1. L'utilité immédiate est affectée d'un coefficient $e^{-\rho t}$.

Les chemins de croissance *atteignables* sont ceux pour lesquels se vérifie la condition :

$$\dot{k_t} + c_t = f(k_t) - n.k_t \qquad (8.29)$$

$c_t > 0, \quad k_t > 0, \quad k_0$ donné

L'équation (8.29) montre qu'une fraction de la production sert à offrir aux nouveaux travailleurs (de plus en plus nombreux) le même capital par tête que celui à la disposition des travailleurs en place.

Le critère d'optimalité est :

$$\int_0^\infty e^{-\rho t}.u(c_t)\,dt \qquad (8.30)$$

Sont *éligibles*, les sentiers de croissance pour lesquels l'expression (8.30) a une valeur finie. Le sentier optimal maximise cette expression.

L'analyse de l'optimalité aboutit à la conclusion suivante :

1- Si $\rho > 0$: le sentier optimal tendra vers l'accumulation *k'* combinée avec la consommation *c'* : si sur la figure 8.7, on remplace la droite *nk* par $(n+\rho)k$, sa pente va augmenter et la tangence à f(k) lui sera parallèle pour une intensité capitalistique *k'* plus à gauche que k°. En conséquence, l'intensité capitalistique, la production et la consommation optimales seront inférieures à leur niveau de la règle d'or.
2- Si $\rho = 0$, la règle d'or tendra à être le sentier optimal et elle le sera précisément si la valeur initiale de *k* et k°.
3- Si $\rho < 0$, l'expression (8.30) n'a pas de maximum fini ; il n'y a pas de sentier de croissance optimal.

La prise en compte de l'accroissement de la population "immediately raises a new ethical question : this is whether one should maximize (…) an integral over discounted per capita utility $u(c_t)$ where $c_t = C_t/L_t$ is consumption per head or an integral over a discounted sum $L_t u(c_t)$ of individual utilities"[2]. Autrement dit, faut-il égaliser le poids des différentes générations ou le pondérer en proportion du nombre d'individus ? Dans ce dernier cas, qui a (modérément) la préférence de Koopmans, on pourrait considérer que $\rho = \rho' - n$, où ρ' est le

[1] Koopmans [195] p. 241.
[2] Koopmans [196] p. 6

taux d'impatience pur et *n* le taux d'accroissement de la population. Mais il faut alors imposer la condition peu intuitive que ρ' >n, faute de quoi, l'optimum n'existerait pas.

8.2.6. La croissance endogène

Le modèle de Solow reste la référence des économistes néoclassiques en matière de croissance. Toutefois, ce modèle, pris tel quel souffre de lacunes de deux types :

1- Ses enseignements ne concordent pas avec l'observation empirique. Rien dans le modèle ne permet d'expliquer les énormes différences entre les taux de croissance des différents pays, ni pourquoi tant de pays pauvres peinent à sortir de leur sous-développement industriel. D'après le modèle, le commerce international et les mouvements internationaux de capitaux devraient mener à une certaine *convergence* des économies du globe : la même technologie étant accessible à tous, les économies devraient toutes tendre vers une même intensité *k* avec la seule différence que les pays initialement arriérés y accéderaient via une transition qui les tirerait vers le haut.

2- Le modèle élève le progrès technique (la variable *A*) au titre de facteur principal de la croissance ; mais simultanément il se contente de le prendre comme exogène ; autrement dit, il ne l'explique pas.

Les économistes néoclassiques ont cherché des solutions à cette double difficulté. Une première piste explorée consiste à considérer que le capital se compose de deux parties : le CAPITAL PHYSIQUE et le CAPITAL HUMAIN. Les dépenses qui augmentent la compétence et la capacité de travail individuelles des travailleurs, comme les dépenses de santé, les dépenses d'éducation, sont des investissements dans un *capital humain* qui s'accumule. Augmenté du capital humain, le capital voit son importance prévaloir dans l'explication de la croissance, au détriment du progrès de la connaissance technique.

La théorie de la *croissance endogène*, apparue dans les années 1980, explore l'autre piste : l'endogénisation du progrès technique. Rivale de la précédente, elle semble plus novatrice en s'écartant davantage du modèle de Solow, mais elle reste néanmoins bien ancrée dans l'orthodoxie néoclassique.

ARROW ET L'APPRENTISSAGE PAR LA PRATIQUE

Avec son article « The Economic Implications of Learning by Doing » (1962), Arrow fait figure de précurseur de cette théorie. De multiples facteurs expliquent l'accroissement du savoir technique et les modèles ne peuvent les inclure tous simultanément. Dans cet article, Arrow s'intéresse à l'un d'entre eux : l'*apprentissage par la pratique*. Non pas au niveau de la compétence individuelle qui serait un capital humain, mais du savoir accessible à tous les candidats producteurs.

Le début de l'article présente la motivation de la théorie de la croissance endogène : « Nevertheless a view of economic growth that depends so heavily on an exogenous variable, let alone one so difficult to measure as the quantity of knowledge, is hardly intellectually satisfactory »[1].

Pour construire un modèle de l'apprentissage par la pratique, encore faut-il disposer d'un indicateur de l'expérience totale acquise par une société. Arrow sélectionne à cet effet l'investissement brut cumulé. Chaque installation de machine est une expérience qui augmente

[1] Arrow [14] p. 155.

le savoir commun. On en arrive naturellement à un modèle à progrès technique incorporé avec vintage. Les équipements sont identifiés par un numéro de série séquentiel *G*. L'expérience au temps *G* est entièrement incorporée dans les équipements *G* ; l'expérience des périodes suivantes n'affectera plus les équipements déjà installés. Le modèle ne prend en compte que l'apprentissage qui bénéficie à la conception des outils de production, pas à celui qui en améliore l'utilisation. Arrow reconnaît qu'il s'agit d'un manque, qui mériterait d'être comblé par des études complémentaires.

Arrow construit une espèce de fonction de production dont les deux inputs sont le travail *L* et l'investissement cumulé *G*. Cette fonction de production virtuelle révèle des *rendements d'échelle croissants*. En effet, la présence de *G* introduit une externalité positive : la machine *G* incorpore gratuitement un supplément de productivité par rapport à G-1, G-2… Comme l'avantage bénéficie à tous les producteurs sans qu'ils doivent l'acquérir sur le marché, le travail et le capital peuvent être rémunérés à leur productivité marginale et cette non-convexité ne trouble pas la concurrence.

Nous ne détaillerons pas le modèle, qui est assez complexe parce qu'il embrasse pas mal de dimensions du système économique ; notamment, les investissements sont décidés sur base de l'anticipation des revenus nets futurs actualisés et Arrow reprend de Muth l'hypothèse des *anticipations rationnelles*. Rapportons simplement cette conclusion : « It should be made clear that all that has been demonstrated is the existence of a solution in which all variables have constant rates of growth, correctly anticipated. The stability of the solution requires further study »[1]. Comme dans le modèle de Solow, le taux de croissance se révèle indépendant du taux d'épargne.

Le modèle montre que la croissance du salaire est plus élevée si la population croît plus vite. Cette constatation paradoxale s'explique par le fait qu'en cas de plein emploi, une main d'œuvre plus abondante permet une introduction plus rapide des équipements les plus récents.

Arrow compare la solution émanant des forces du marché avec l'optimum théorique qui adviendrait si un planificateur déterminait *G* de façon à maximiser une fonction d'utilité sociale basée sur la consommation intertemporelle avec escompte. Les deux solutions arrivent au même taux de croissance, mais le taux d'investissement de la solution du marché est moindre.

PAUL ROMER ET LA RECHERCHE ET DEVELOPPEMENT

L'économiste qui a le mieux associé son nom à la *croissance endogène* est Paul Romer[2]. Parmi les nombreuses publications qu'il lui a consacrées, « Endogenous Technical Change » (1990) s'intéresse à l'autre source d'accroissement du savoir technique que celle étudiée par Arrow : la *recherche et développement* (R&D). Il intègre le *capital humain* dans son modèle mais dans une optique différente de celle qui fut mentionnée ci-avant.

La thèse défendue s'appuie sur trois prémisses :
1- Le changement technologique joue un rôle central dans la croissance
2- L'innovation technologique répond à des stimuli du marché.
3- Une technologie est un bien qui a ceci de particulier que sa production occasionne un coût initial et son exploitation par la suite plus aucun coût.

Pour bien comprendre les spécificités de la technologie, il importe de pouvoir classer les biens :

[1] Arrow [14] p. 166
[2] A ne pas confondre avec le *nouveau keynésien* David Romer.

- D'une part, les caractéristiques techniques des biens les rendent *rivaux* si l'utilisation par un agent diminue la possibilité de les utiliser par les autres agents ; sinon ils sont *non-rivaux*.
- D'autre part, les institutions sociales et légales les rendent *exclusifs* si un monopole d'utilisation est accordé à un agent (brevet) ; sinon ils sont *non-exclusifs*.

Les *biens économiques* sont rivaux et exclusifs contrairement aux *biens publics* qui sont non-rivaux et non-exclusifs. La recherche fondamentale est un bien public ; le capital humain est un bien économique, car l'apprentissage d'un deuxième travailleur coûtera environ autant que celui du premier. La technologie est un bien non-rival partiellement exclusif. La productivité d'un facteur de production de ce type est potentiellement illimitée et seule une part de cette productivité peut être rétribuée sur le marché. La fonction de production qui intègre un input non-rival révèle forcément des rendements d'échelle croissants, elle ne peut être concave. S'il paye ses inputs rivaux à leur productivité marginale, le producteur n'a plus de ressources à affecter à l'input non rival. Un équilibre général concurrentiel est donc impossible si l'input non-rival est ne fût-ce que partiellement exclusif. Romer résout la difficulté en faisant une entorse à la concurrence parfaite qui régente la plupart des modèles : il inclut une espèce de concurrence monopolistique. Solow n'eut pas à faire face à ce problème parce qu'il considérait la technologie comme un bien public, donc entièrement disponible et gratuit. Des trois prémisses de Romer, Solow n'avait pas fait sienne la deuxième.

Exposons brièvement le modèle. Quatre inputs productifs interviennent :
- Le capital physique K
- Le travail L
- Le capital humain H, de productivité γ.
- La technologie A

Il y a trois secteurs économiques que je symboliserai par la même lettre que leur production :
- Le secteur de la recherche et développement : A
- Le secteur des biens intermédiaires : X
- Le secteur des biens finaux : Y : une partie de ceux-ci est consommée et l'autre réinvestie.

Le modèle recrée une mécanique économique globale simplifiée comme suit : le secteur A produit des designs d'équipements, identifiés par l'indice i. Le progrès technologique se matérialise par le fait que de nouveaux designs sont produits en permanence. A(t) est donc le nombre de ces designs existant au temps t. Ces designs sont achetés sous forme de licences (à terme infini) au prix P_A par le secteur X. Chaque producteur de X produit en monopole un bien intermédiaire correspondant à un design i en quantité x_i. Les équipements i sont offert en location aux firmes Y au prix p_i. La valeur d'une unité de i correspond à la somme des revenus futurs actualisés.

Diverses simplifications sont introduites :
- L'offre globale de L et celle de H sont fixes dans le temps
- La dépréciation et l'obsolescence des équipements sont ignorées
- Le produit fini est pris comme numéraire. $P_Y = 1$.
- La consommation des inputs par les secteurs est simplifiée :
    - A consomme A et H_A.
    - X utilise A et Y
    - Y utilise L, H_Y et X

- On a : $dA/dt = \gamma H_A A$. La production de connaissance est linéaire par rapport au stock accumulé et par rapport au capital humain. Cette linéarité, peu habituelle, traduit une productivité marginale non décroissante de ces inputs. Romer le justifie ainsi : « The

specification here, in which unbounded growth at a constant rate is feasible, was chosen because there is no evidence from recent history to support the belief that opportunities for research are diminishing » (84).[1]

- La variable *i* est traitée comme continue plutôt que discrète, ce qui d'un point de vue économique est un non-sens, mais a l'avantage de permettre le traitement mathématique avec le calcul intégral.

La connaissance *A* a donc un double impact sur la production :

- l'accroissement de la productivité des recherches ultérieures
- l'élargissement des équipements de production.

La fonction de production *Y* est de type Cobb-Douglas arrangé :

$$Y(H_Y,L,X) = H_Y^{\alpha}.L^{\beta}.\int_0^{\infty} x_i^{1-\alpha-\beta}\,di \qquad (8.31)$$

L'intégrale additionne le x_i. Cette additivité exclut les effets de complémentarité et de substitution entre les équipements de production, ce qui n'est évidemment pas réaliste.

Romer définit l'équilibre comme un sentier d'évolution des prix et quantités respectant ces conditions :

1- Les ménages déterminent leur épargne en considérant le taux d'intérêt *r* comme donné.
2- Les détenteurs du capital humain le répartissent entre la recherche (H_A) et la production (H_Y) en considérant donnés w_A (la rétribution du capital humain), P_A et *A*.
3- Les producteurs de *Y* déterminent leur input de *L* et H_Y en considérant les prix donnés.
4- A l'opposé des précédents, les producteurs de *X* déterminent leur prix, de façon à maximiser leur profit, en considérant *r* et leur demande comme donnés. La décision de démarrer la production x_i se prend en considérant P_A comme donné.
5- L'offre et la demande s'équilibrent sur tous les marchés.

Romer démontre mathématiquement que cet équilibre est compatible avec une croissance en *steady state* où *A*, *K* et *Y* croissent exponentiellement à un taux constant. Plus précisément, ces variables ainsi que la consommation *C* croissent toutes à un taux *g* où :

$$g = \gamma.H_A \qquad (8.32)$$

La démonstration est complexe car elle fait intervenir de nombreux paramètres : la demande de x_i par les firmes *Y*, la maximisation du profit par les firmes *X*, l'égalisation de P_A avec l'actualisation des revenus futurs $x_i.p_i$, l'optimisation intertemporelle de l'utilité des ménages tenant compte du taux d'intérêt, de l'escompte des satisfactions futures et du taux marginal de substitution intertemporel… Les simplifications du modèle ne permettent pas de différencier l'offre des producteurs de *X*. Ils offrent donc tous une quantité identique x°, qui ne varie pas avec *t*, car la croissance de *K* s'exprime uniquement par le nombre de x_i.

Comme Romer le constate lui-même, le modèle se comporte comme un modèle néoclassique traditionnel avec un progrès technique « labor and human capital augmenting ».

Un taux d'intérêt bas favorise la croissance. Son élévation réduit la valeur des designs d'équipements, du fait de l'escompte ; la motivation d'investir dans l'activité R&D se réduit et donc également le taux de croissance *g*.

[1] Romer [308] p. 84.

Dans le présent modèle, la capital humain *H* est exogène. Romer reconnaît qu'il serait souhaitable d'endogénéiser cette variable. Le modèle met en lumière que lorsque *H* augmente, la part de H_A par rapport à H_Y augmente par la même occasion, ce qui est favorable à la croissance comme l'indique l'équation (8.32). Sous un certain seuil de *H*, la croissance disparaît même, car la trop faible augmentation de *A* génère des revenus futurs escomptés insuffisants pour compenser le sacrifice que représente l'investissement.

Romer pense que d'une façon générale, les sociétés allouent trop peu de capital humain au secteur de la recherche. L'externalité positive n'étant pas (totalement) payée, la motivation d'investir dans ce domaine est émoussée. Une politique de subsides encourageant la recherche et l'accumulation de capital humain devrait rapprocher l'économie de l'optimum social.

Romer considère que l'intégration économique mondiale permet le partage du capital humain. Elle est donc favorable à la croissance de tous les pays, y compris ceux qui, selon d'autres modèles, pourraient s'en dispenser parce qu'ils disposent d'un large marché intérieur.

8.3- LE CYCLE CONJONCTUREL

L'étude des crises et de leur retour cyclique date du XIXe siècle avec le Français Clément Juglar et l'Ukrainien Mikhaïl Tugan-Baranovsky qui firent œuvre de pionniers.

Les théories avancées depuis plus d'un siècle pour expliquer le cycle économique sont nombreuses et variées. Nous nous contenterons d'aborder trois écoles :
- Le surinvestissement et tout particulièrement le principe d'accélération
- L'explication monétaire de l'école autrichienne
- Le cycle économique réel

8.3.1. Le cycle causé par le surinvestissement

LE PRINCIPE D'ACCELERATION : AFTALION ET CLARK

En 1913, l'économiste français Albert **Aftalion** consacre au cycle conjoncturel un ouvrage de 750 pages intitulé « Les crises périodiques de surproduction »[1]. L'ouvrage, qui se veut complet recherche les causes du cycle et en fait la description sous tous ses aspects avec force statistiques.

Aftalion rappelle ce qui à notre époque apparaît comme une évidence, peut-être plus qu'à la sienne, à savoir que le cycle forme un tout, que les crises ne constituent pas des événements accidentels. L'essentiel, c'est l'alternance des phases. Au vu de cette vérité, le titre de son ouvrage ne paraît d'ailleurs pas des plus judicieux.

Aftalion fait reposer sa théorie sur deux piliers, chacun lié à une théorie économique existante :
- Les fluctuations cycliques de la valeur d'usage. Les variations de la production entraînent dans leur sillage celles de la valeur d'usage, du fait de l'utilité marginale décroissante.
- Le délai entre le lancement d'un investissement et l'offre aux consommateurs de la production résultante. Une production très *détournée*, au sens de Böhm Bawerk, comme celle des économies modernes implique une attente fort longue avant la satisfaction les consommateurs.

La combinaison de ces deux facteurs s'avère explosive. S'il y a surproduction et baisse des prix, cela signifie qu'on a trop produit par rapport au besoin : la valeur d'usage diminue. L'utilité sociale des biens produits ne justifie plus l'effort réalisé en vue de cette production. On s'attendrait plutôt à voir baisser les prix au moment où la production s'élève, c'est-à-dire pendant l'expansion ; or, à ce moment, ils augmentent. L'accroissement productif caractérisant l'expansion n'est pas incompatible avec une pénurie relative de biens de consommation. Au début de l'expansion, de nombreux investissements sont mis en chantier, mais leurs « fruits » ne seront mûrs que plus tard. On produit intensément, mais ce sont principalement les biens capitaux qui profitent de cet essor. La pénurie de biens de consommation est encore aggravée par la baisse du chômage qui accroît le pouvoir d'achat de la population.

Il y a donc pénurie relative dans la première phase de l'expansion. Puis nombre de nouvelles unités de production deviennent opérationnelles ; la production des biens de consommation finit par devenir excessive. Maintenant, face à la surcapacité, les investissements chutent ; malgré la baisse de la production totale, les biens de consommation sont en surproduction, d'autant plus que la baisse des investissements met des salariés au chômage. Ce sous-investissement

[1] L'auteur avait déjà exposé les idées principales du livre dans un article paru en 1909.

engendrera une sous-capacité de production. La pénurie menace. Le temps est venu pour la reprise. Un nouveau cycle peut commencer.

Aftalion défend sa thèse de la baisse générale des valeurs d'usage en période de dépression. N'est-elle pas contradictoire avec le postulat largement admis de l'insatiabilité des besoins ? La satiété partielle et temporaire est possible. Comme on tend à satisfaire les besoins dans l'ordre décroissant d'intensité, les derniers besoins non satisfaits peuvent être peu intenses. Sans baisse du coût de production, la surproduction menace. L'insatiabilité se fonde d'ailleurs plus sur une diversification progressive des besoins que sur une boulimie primaire ; elle agit donc à long terme et n'empêche pas la satiété temporaire.

Contrairement à la valeur d'échange, la valeur d'usage n'est pas une donnée relative. Parler d'une hausse ou d'une baisse générales des valeurs d'usage n'est donc pas incongru.

Aftalion réfute l'explication alternative selon laquelle les variations cycliques des prix auraient une cause une cause monétaire. Les statistiques montrent que les années où la circulation monétaire hausse alors que les prix baissent ne sont pas rares, pas plus que les années où cette circulation baisse alors que les prix haussent.

La théorie quantitative annonce une baisse des prix et non une hausse, lorsque la circulation des marchandises s'accroît plus que celle de la monnaie. La circulation accrue des marchandises augmente le besoin de monnaie. D'un point de vue strictement monétaire, la valeur de la monnaie devrait être à son maximum au moment où, dans la réalité, les prix haussent. Aftalion ne conteste pas la validité de la théorie quantitative. Simplement, elle trouve sur sa trajectoire un facteur plus fort qu'elle : les variations de la valeur d'usage des marchandises.

Les fluctuations des prix sont un facteur d'instabilité. Les investissements sont décidés à une période où les prix montent. Mais leurs produits seront tôt ou tard vendus à des prix bradés qui portent atteinte à la solvabilité des entreprises.

On ne peut comprendre le cycle économique si l'on n'observe que la production totale. Il faut dissocier la production de biens capitaux de celle des biens de consommation. Le délai de réalisation des investissements produit une pénurie de biens de consommation. Leur valeur augmente donc. Cette valorisation rejaillit sur les biens capitaux qui interviennent dans leur production[1]. La valeur des biens capitaux est donc à son maximum lors de l'expansion. Les variations de la valeur d'usage trouvent leur cause ultime du côté des biens de consommation. Le besoin de biens capitaux n'est qu'un « besoin dérivé ». Mais les variations de prix induites par les biens de consommation sont accentuées lorsqu'elles affectent les biens de production.

Deux concepts essentiels chez Aftalion sont la *sous-capitalisation*, typique de la dépression et la *surcapitalisation* caractéristique de l'expansion. Selon la période, on investit non seulement peu ou beaucoup, mais trop peu ou trop. L'appareil productif sur-réagit à la baisse ou à la hausse des valeurs d'usage et des prix. Le problème vient de ce que les prix et les profits actuels se déterminent dans l'ignorance du capital en cours de constitution pendant la période d'expansion. Ils sont donc de mauvais indicateurs, abusant les entrepreneurs qui continuent à investir lorsque le capital social a atteint un niveau suffisant. En fin de dépression, la même cause agit symétriquement. Il manque d'un régulateur qui avertirait les entrepreneurs que la pénurie de capital menace et qu'il est temps de reprendre les investissements, malgré le bas prix des produits qui sévit encore. Aftalion compare le cycle économique avec le chauffage d'une pièce à l'aide d'un poêle. On continue à alimenter le poêle trop longtemps, parce que la

[1] Rappelons-nous Menger : les biens d'ordre supérieur tirent leur valeur des biens de consommation.

température souhaitée n'advient que lentement. Ensuite, la chaleur excessive dissuade de l'alimenter ; on ne s'y résoudra que trop tard, lorsque la température de la pièce sera devenue trop basse.

De ces excès dans la réaction, Aftalion conclut que « la cause directe de la surproduction et de la surcapitalisation, ce sont sans doute les erreurs des hommes (…) Le profit n'est pas seul à constituer en ces circonstances un symptôme trompeur. C'est l'état des besoins lui-même qui égare et aveugle les producteurs ». Il n'est toutefois pas sûr qu'une meilleure information résoudrait le problème : « Il a déjà été montré, à propos des prix, comment la spéculation, au lieu de prévoir la baisse prochaine que pourra déterminer l'augmentation vraisemblable des approvisionnements, croit au contraire à la persistance de la hausse des cours et contribue jusque dans les derniers instants de la prospérité à en exagérer l'intensité »[1].

Comment évolue la répartition du revenu pendant le cycle ? Pendant la phase de prospérité, la demande augmente plus que la production. Le prix s'accroît par rapport au coût de production. Toutes les formes de revenu augmentent : salaire, intérêt (taux procyclique), rente, profit. Mais le profit augmente plus que les autres ; sa part dans le revenu national augmente. Les revenus fixes peuvent être malmenés si leur hausse ne compense pas celle des prix. Le salaire horaire[2] ne devrait pas souffrir, car les prix des biens de consommation ouvrière sont parmi les plus stables. En outre, les salariés travaillent plus d'heures et les chômeurs retrouvent un salaire : la situation globale de la classe salariée s'améliore certainement. En période de dépression, la situation est exactement inverse.

En conclusion, supprimer le cycle conjoncturel semble difficilement réalisable. Mais il est souhaitable que l'Etat échelonne ses commandes de façon à le modérer. Il est aujourd'hui difficile de savoir si les crises seraient moins graves dans un régime socialiste mais il est peu probable qu'elles disparaitraient, car l'allongement du procès de production caractérise toutes les économies développées, pas seulement le capitalisme.

En 1917, John Maurice **Clark** (le fils de John Bates Clark) publie l'article « Business Acceleration and the Law of Demand: a Technical Factor in Economic Cycles ». Le terme « accélération » apparaît enfin, bien que l'idée était déjà présente chez Aftalion.

Les faits à expliquer sont :
- Les fluctuations des prix et quantités sont plus amples sur les marchés de biens de production que sur ceux des biens de consommation.
- Les fluctuations du prix des matières premières précèdent celles des produits finis.

Le passage de la demande de biens de consommation à la demande de biens capitaux est un processus économique qui intensifie ce qu'il transmet et qui est capable de transformer une simple baisse du taux de croissance de la première en une baisse véritable de la seconde. La même accélération joue également à la hausse.

La demande des biens capitaux comporte une double composante :
- Le remplacement des équipements usagés de façon à maintenir le parc existant

[1] Aftalion [1] t. II pp 359, 361, 363

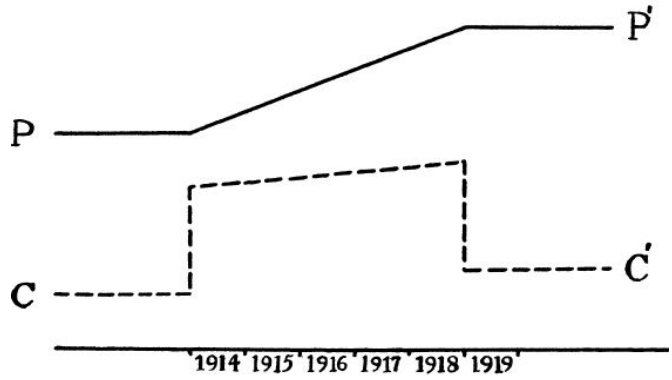
[2] Le salaire se calque sur la productivité marginale en valeur. Concernant la productivité physique, Aftalion la perçoit contracyclique, de par la loi des rendements décroissants. En outre, en prospérité, les firmes remettent parfois des équipements obsolètes en activité pour répondre au pic de la demande, alors qu'en dépression, la compression des coûts est le mot d'ordre.

- L'extension ou l'amélioration de la capacité de production[1].

La première composante dépend du niveau de la demande des produits de consommation et la seconde des variations de cette même demande. Il en découle des soubresauts.

Clark donne l'exemple numérique suivant : soit une industrie produisant le bien *P* à l'aide de l'équipement *C*. La demande de *P* est stable, mais en 1914 et pendant les quatre années qui suivent, elle augmente annuellement de 10% du montant de 1913, puis redevient stable. Supposons que l'équipement *C* s'amortit en 20 ans. Le graphique 8.8 montre l'évolution comparée des demandes de *P* et *C*. En 1914, la demande de *C* triple parce que celle de *P* augmente de 10% ; elle se maintient à ce niveau et augmente même légèrement jusqu'en 1918 inclus. En 1919, elle revient à 3/2 de son niveau de 1913, ce qui représente une chute de 4/7 par rapport à 1918[2]. Le bond et la chute de la courbe *C* en 1914 et 1919 résultent de variations mineures de la courbe *P* ces mêmes années.

Graphique 8.8 : La demande dérivée : influence des deux composantes[3]



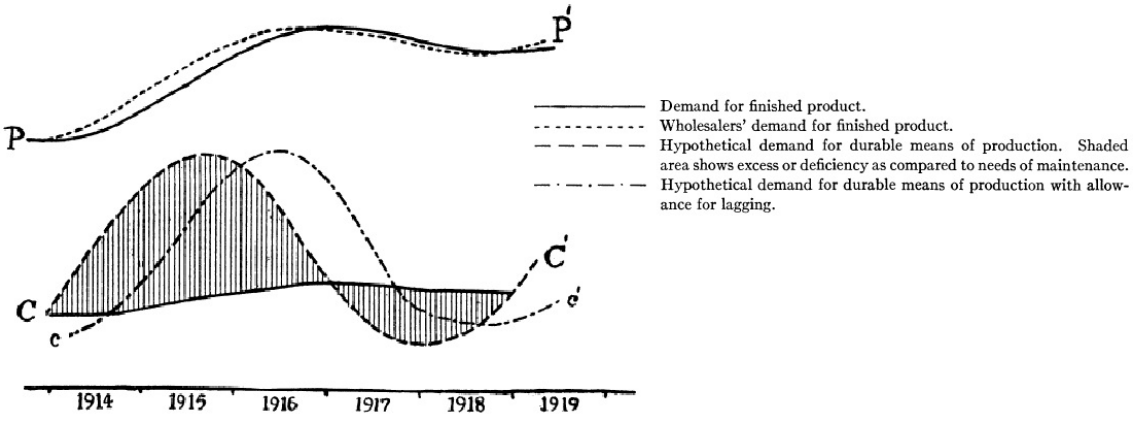
Comme le montre la figure 8.9, les points d'inflexion de la courbe *P* correspondent à des retournements de la courbe *C*. Comme conséquence, les retournements de *C* précèdent ceux de *P* et la courbe *C* donne l'illusion d'être en avance sur la courbe *P* et d'être la meneuse. La courbe *c* est l'équivalent de la courbe *C* lorsqu'un délai de réalisation de l'investissement est pris en compte ; il est en effet peu réaliste de présumer comme dans l'exemple ci-dessus que l'adaptation de l'appareil productif à la demande des consommateurs n'exige ni délai de réaction ni délai de réalisation.

[1] Dans les faits, vu le progrès technique, beaucoup d'investissement chevauchent cette distinction, qui n'en perd pas pour autant sa raison d'être.

[2] Clark suppose que la capacité installée augmente dans exactement la même proportion que le produit. La composante remplacement vaut toujours 1/20 de la valeur du capital installé.

[3] Clark [55] p. 221

Graphique 8.9 : la demande dérivée : aspect temporel[1]



Plus élevé sera le taux de dépréciation, moins l'accélérateur sera déstabilisant, car l'investissement d'expansion représentera une part plus faible de l'investissement total.

Aftalion imputait l'accélération de la demande dérivée à la sur-réaction des décideurs d'investissements ; Clark met en cause les comportements divergents de l'investissement de remplacement et de l'investissement d'extension. Aftalion avait déjà évoqué la seconde cause comme facteur complémentaire appuyant la première.

LES CRITIQUES CONTRE L'ACCELERATEUR

En 1931, Ragnar **Frisch** publie un article critique contre la théorie de Clark, plus précisément contre l'exposé qu'il en fait dans un livre publié en 1923. Dans cet ouvrage, Clark semble laisser entendre un peu imprudemment que si le taux de croissance de la consommation baisse, il s'ensuivra nécessairement une baisse absolue de la demande dérivée. Or selon Frisch, il s'agit d'une possibilité, non d'une nécessité. La question est d'importance. Si Clark avait raison, il disposerait d'une théorie qui explique les retournements conjoncturels. En effet, la limitation des facteurs de production disponibles finirait toujours par faire baisser le taux de croissance de la consommation à un moment ou à un autre. S'ensuivrait une chute des investissements engendrant la crise.

L'article de Frisch est très mathématique. Il démontre que si la consommation s'accroît à un rythme qui diminue progressivement, la demande de capitaux peut évoluer de multiples façons et notamment emprunter chacune de ces trois voies : (1) rester constante, (2) croître en tendant asymptotiquement vers une valeur, (3) tendre vers une constante au travers d'oscillations amorties ; l'itinéraire dépend des divers paramètres : le taux de dépréciation, le rapport entre le capital et la production de biens de consommation, le taux de croissance de la consommation. Frisch indique le rapport mathématique qui doit prévaloir entre ces paramètres pour que la production de biens capitaux baisse effectivement.

Dans sa réponse, Clark reconnaît que la baisse de la production de biens capitaux n'est pas une règle générale (l'article de 1917 ne le prétendait d'ailleurs pas). Il estime toutefois que le cas présenté par Frisch pour contrer sa théorie implique un déclin si faible du taux de croissance de la consommation qu'il ne peut représenter qu'une restriction à la règle tout-à-fait bénigne.

Deux ans plus tard, Frisch publie son article « Propagation and Impulse Problems in Dynamic Economics ». De nouveau, il élabore un modèle « clarkien » qui n'aboutit pas à un résultat

[1] Clark [55] p. 225

cyclique. Il envisage plusieurs facteurs qu'on pourrait ajouter dans le modèle pour générer les oscillations souhaitées. Parmi eux, les délais d'Aftalion : « … activity at a given moment does not depend on the decisions taken at that moment, but on decisions taken at earlier moments. By this, we introduce a new element of discrepancy in economic life that may provoke cyclical oscillations ».[1] Il intègre cette dimension dans un nouveau modèle qui aboutit effectivement à des oscillations. Le problème est que ces oscillations s'amortissent très rapidement. Elles ne pourraient correspondre au schéma réel du cycle conjoncturel que si de nouvelles impulsions revenaient constamment relancer la propagation. Le problème de la PROPAGATION est résolu ; concernant l'IMPULSION, Frisch reprend l'idée de Slutsky que des chocs erratiques en série sont capables de générer des oscillations ; les innovations techniques pourraient très bien en être la source.

L'autre censeur de l'accélérateur est l'économiste néerlandais **Tinbergen**, qui publie en 1938 l'article « Statistical Evidence on the Acceleration Principle ». Il distingue deux versions de ce principe :

- Sous sa FORME RIGIDE, le principe présume que le stock de capital fluctue en exacte proportion de la production de biens de consommation
- Sous sa FORME MODEREE, il se contente d'établir entre ces deux variables une relation, par exemple de type linéaire.

Les implications de la forme rigide sont très fortes, tellement qu'il est peu probable qu'elles puissent être observées dans le mode réel. La forme rigide est donc peu réaliste. Tinbergen se montre très sceptique à son égard, même sur le plan théorique, pour les raisons suivantes :

- Elle revient à affirmer que « a constant part of productive capacity is idle and that entrepreneurs never increase production of consumer goods before having increased correspondingly their capacity »[2].
- Il y a une limite naturelle au désinvestissement. L'investissement net[3] ne peut devenir négatif au-delà de la dépréciation. Il est donc impossible qu'une chute extrême de la consommation soit répercutée intégralement sur l'investissement.
- La relation présumée stable entre le capital et la production ne résisterait pas à des progrès techniques imposant un changement soudain de ce ratio.

Tinbergen soumet les deux versions à l'épreuve statistique. Pour quelques industries, il effectue une régression entre les fluctuations de l'équipement et celles de leur production. Il obtient un coefficient de régression d'environ 0,5, ce qui exclut la forme rigide. Mais la corrélation n'est bonne que dans un nombre limité de cas[4]. Sa conclusion : le principe d'accélération n'apporte qu'une aide limitée pour comprendre les fluctuations de l'investissement. Il explique certes l'amplification des fluctuations de la demande dérivée, mais cette amplitude n'est en réalité que la moitié de ce qui pourrait être attendu.

Comme Tinbergen le reconnaît lui-même, ce test statistique n'est valide que pour autant que la PRODUCTION de biens de consommation est un indice correct de la DEMANDE de ces biens. Car le principe d'accélération, c'est bien sur la DEMANDE de biens de consommation qu'il est censé s'exercer.

---

[1] Frisch [113] p. 181.
[2] Tinbergen [359] p. 166
[3] L'investissement net (positif ou négatif) égale l'investissement brut moins l'amortissement (dépréciation).
[4] Comme elle l'avait été pour les chemins de fer aux Etats-Unis que Clark avait pris comme illustration statistique.

MODELES COMBINANT LE MULTIPLICATEUR ET L'ACCELERATEUR

Trois grands auteurs keynésiens, déjà bien connus du lecteur, ont proposé des modèles produisant des oscillations cycliques, à partir de la combinaison entre le multiplicateur keynésien et l'accélérateur. Harrod a ouvert la voie (1936)[1] ; Samuelson (1939) et Hicks (1950) ont suivi. Comme concept, l'accélérateur se situe clairement dans le camp keynésien.

Le principal mérite de **Harrod** est l'originalité. L'analyse contenue dans son ouvrage « The Trade Cycle » est moins convaincante que celles de Samuelson et Hicks. Un autre mérite que saluera Hicks est sa volonté de compléter la macroéconomie traditionnelle, de caractère statique, par une macroéconomie dynamique qui se centrerait sur l'évolution des variables dans le temps et qui embrasserait la croissance économique et les fluctuations conjoncturelles. Son article de 1939 étudié au sous-chapitre 8.2.1 s'inscrivait dans cette ligne.

Dans deux articles publiés en 1939, **Samuelson** est le premier à élaborer un modèle mathématique combinant le multiplicateur et l'accélérateur. Il imagine une économie simplifiée dont les paramètres et les caractéristiques sont ceux-ci :

- *c* est la propension marginale à consommer. Y_t est le revenu national au temps *t*.
- C_t est la consommation au temps *t*. On a : $C_t = c.Y_{t-1}$; il y a donc un délai d'un an.
- I_t est l'investissement[2] au temps *t* et *v* est le coefficient d'accélération. On a : $I_t = v.(C_t - C_{t-1})$.
- Le revenu national Y_t est l'addition de $I_t + C_t + G_t$ où G_t représente les dépenses gouvernementales. On a donc :

$$Y_t = G_t + c(1+v)Y_{t-1} - c.v.Y_{t-2} \qquad (8.33)$$

Samuelson pose d'abord que c = 0,5 et v = 1 et qu'il n'y avait pas de dépense gouvernementale dans le passé. A partir de *t* = 1, le gouvernement dépense chaque année G_t = \$1, par *deficit spending*, ce qui assure que C_t et I_t ne souffrent de cette dépense. En *t* = 1, il n'y a ni investissement ni consommation : Y_1 vaut donc \$1. Grâce à la formule (8.33), on peut calculer Y_t jusqu'à l'infini. L'évolution des variables Y_t, I_t, C_t, G_t apparaît dans le tableau ci-dessous repris de Samuelson.

[1] Il peut sembler étonnant que Harrod soit déjà en mesure de publier un ouvrage en phase avec la théorie keynésienne l'année même de la parution de la *Théorie Générale*, alors que la rédaction a dû l'occuper les années précédentes. Cela s'explique par la proximité entre Keynes et Harrod, dès avant 1936.

[2] Il s'agit de l'investissement NET. Le remplacement ne joue pas un rôle direct dans ce modèle.

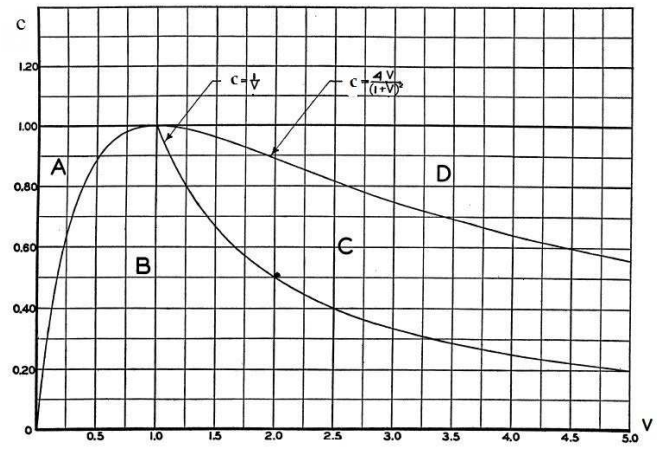
(Unit: one dollar)

Period	Current governmental expenditure	Current consumption induced by previous expenditure	Current private investment proportional to time increase in consumption	Total national income
1	1.00	0.00	0.00	1.00
2	1.00	0.50	0.50	2.00
3	1.00	1.00	0.50	2.50
4	1.00	1.25	0.25	2.50
5	1.00	1.25	0.00	2.25
6	1.00	1.125	−0.125 *	2.00
7	1.00	1.00	−0.125	1.875
8	1.00	0.9375	−0.0625	1.875
9	1.00	0.9375	0.00	1.9375
10	1.00	0.96875	0.03125	2.00
11	1.00	1.00	0.03125	2.03125
12	1.00	1.015625	0.015625	2.03125
13	1.00	1.015625	0.00	2.015625
14	1.00	1.0078125	−0.0078125	2.00
.. ..	...	.....	.....	.....

On constate que le revenu national connaît bien des fluctuations cycliques[1]. En réalité, l'évolution de Y_t dépend des valeurs prises par *c* et *v*. Par exemple, quand v = 0, c'est-à-dire sans le phénomène d'accélération, il n'y a pas d'oscillation. Samuelson dessine la figure 8.10, qui répartit les combinaisons de *c* et *v* quatre zones, selon le comportement de *Y*.

- Zone A : Y_t tend asymptotiquement vers $G_t/(1-c)$.
- Zone B : oscillation amortie vers cette même valeur $G_t/(1-c)$.
- Zone C : oscillation explosive (d'amplitude croissante)
- Zone D : croissance explosive (éventuellement exponentielle)[2].

Graphique 8.10; évolution de revenu national en fonction de s et v



Samuelson insiste sur l'importance de l'hypothèse relative à la fonction de consommation $C_t = c.Y_{t-1}$. Sans le délai d'une période (donc si C_t valait $c.Y_t$), il n'y aurait pas d'oscillations possibles[3].

[1] Les valeurs négatives que prend I_t à certaines périodes sont dues aux hypothèses peu réalistes : pas d'investissement ni de consommation avant $t = 1$.

[2] Décroissance explosive si l'impulsion initiale est négative.

[3] Hicks confirmera cette condition des oscillations, mais en l'argumentant : le délai comprend deux phases : (1) gagner le revenu et puis seulement l'encaisser ; (2) l'encaisser et puis seulement le dépenser. Il est fort probable que le délai soit plus long pour les revenus non salariaux que pour ceux des salariés.

Le modèle est évidemment très simplificateur. Par exemple, il ne tient pas compte d'une influence que le revenu pourrait avoir sur la propension marginale à consommer et sur le coefficient d'accélération.

Dans son ouvrage « A Contribution to the Theory of the Trade Cycle » (1950), **Hicks** développe une théorie cycle conjoncturel, complète et aboutie.

Le point de départ est le même que celui de Samuelson, avec quelques nuances. Le revenu comporte trois éléments : la consommation, l'*investissement induit* et *l'investissement autonome*. L'investissement induit est celui qui résulte du phénomène d'accélération. Chez Samuelson, tout l'investissement était induit. Hicks lui adjoint un investissement autonome, c'est-à-dire exogène. L'investissement autonome joue plus ou moins le même rôle que la dépense publique chez Samuelson. La principale innovation hicksienne est celle-ci : l'accélération dérive l'investissement, non plus de la variation de la consommation mais de la variation du revenu total.

Les équations deviennent donc :

$$C_t = (1\text{-}s)Y_{t-1} \; ; \; I_t = v.(Y_{t-1} - Y_{t-2}) \; ; \; Y_t = A_t + (1\text{-}s+v)Y_{t-1} - v.Y_{t-2} \qquad (8.34)$$

Où I_t est l'investissement induit, A_t l'investissement autonome et *s* la propension marginale à épargner ($s = 1 - c$).

Hicks effectue le même exercice que Samuelson, si ce n'est que le déclencheur est ici un investissement autonome additionnel et qu'il est ponctuel[1] (en première période). Comment le revenu (déterminé selon l'équation 8.34) se comporte dans le temps, compte tenu des valeurs de *s* et *v* ? Le rôle de *v* s'avère déterminant, au travers de trois valeurs pivot : le pivot bas (qui est fonction de *s*), le pivot moyen (toujours égal à un) et le pivot haut (également fonction de *s*). Avec *v* sous le pivot bas, le revenu converge ; entre le bas et le moyen, il est sujet à des oscillations amorties ; avec $v = 1$ exactement, il subit des oscillations constantes ; entre le moyen et le haut, les oscillations sont explosives ; au-delà de pivot haut, le revenu décolle explosivement.

Hicks admet que ces calculs ne sont valides que moyennant certaines hypothèses simplificatrices peu réalistes :

- Symétrie entre l'accélération suivant une hausse du revenu et celle qui suit une baisse. Comme l'avait remarqué Tinbergen, il est matériellement impossible qu'une baisse de l'investissement net aille au-delà de la dépréciation : le désinvestissement se fera par non-remplacement du matériel usagé ; il s'étalera sur une plus longue durée mais sera moins profond[2].
- L'investissement induit se déroule entièrement dans la période qui suit immédiatement celle où la hausse causale de la production a eu lieu. Il n'est donc ni étalé ni postposé.

Pour comprendre comment le système génère les fluctuations conjoncturelles, il faut d'abord prendre connaissance d'une autre dimension de la théorie hicksienne : la croissance économique qui est un trend sur lequel vient se greffer la conjoncture. Hicks met en regard la croissance perturbée par la conjoncture avec une croissance équilibrée, qui sert de référence ; il s'agit d'une croissance semi-stationnaire ; toutes les variables croissent au même taux constant.

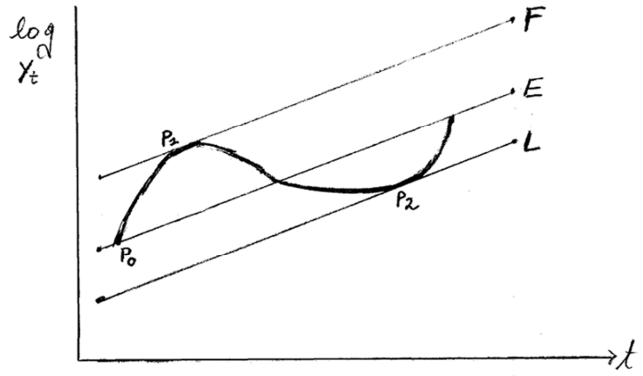
[1] L'exemple chiffré de Samuelson portait sur une dépense additionnelle récurrente, mais il avait également analysé le cas d'une dépense ponctuelle pour conclure que les deux variantes avaient le même effet.

[2] Hicks précise toutefois que la limite de Tinbergen vaut pour le capital fixe, pas pour le capital circulant.

Non pas qu'Hicks suppose qu'une telle croissance soit possible dans la réalité économique, mais en théorie, elle est cohérente. Hicks démontre que dans ce contexte, il existe un taux de croissance *g* qui maintient l'équilibre entre l'épargne et l'investissement pour autant que I_t, S_t et A_t croissent tous trois à ce même taux constant. Un taux *g* différent existe pour chaque niveau de A_t.

Une analyse graphique est la plus parlante. Commençons par dessiner sur le graphique 8.11 la croissance équilibrée au taux *g*. Elle devrait avoir la forme courbe de la fonction exponentielle $(1+g)^t$. Hicks suppose que la croissance perturbée fluctuera autour d'elle, et ce d'une façon qu'il nous reste à sélectionner parmi les différentes évolutions possibles mises en lumière ci-avant. Logiquement, les écarts absolus entre la courbe d'équilibre et la courbe perturbée devraient également croître au taux *g* ; des écarts relatifs identiques donneront des écarts absolus différents. Pour éviter ces complications, Hicks dessine, non le revenu mais son logarithme, ce qui a pour effet de remplacer les courbes par des droites et de les rendre parallèles. A noter que dans cette perspective, la valeur de *v* correspondant au pivot moyen n'est plus 1 mais $(1+g)^2$.

Graphique 8.11 : le cycle conjoncturel selon Hicks



Dans la réalité économique, la conjoncture est un cycle qui se perpétue sans qu'il y ait à long terme une tendance à l'amortissement ou à l'explosivité. L'explication qui vient en premier est que *v* égale le pivot moyen $(1+g)^2$. Hicks trouve toutefois hautement improbable qu'à travers les vicissitudes du temps, cette variable reste imperturbablement fixée à une même valeur. Une autre explication possible est celle de Frisch : *v* est compris entre 0 et $(1+g)^2$. Les oscillations s'amortissent vite, mais de nouvelles impulsions relancent toujours la mécanique conjoncturelle. Hicks reproche à cette conception d'être peu explicative.

Après ces prolégomènes, nous pouvons dévoiler l'explication hicksienne. L'accélérateur *v* est supérieur à $(1+g)^2$. Le cycle devrait donc être explosif. Mais il se heurte à un plafond et à un plancher qui le maintiennent dans une amplitude moyenne. Est-il réaliste de penser que $v > (1+g)^2$? Hicks montre que cette condition revient au même que celle-ci : que $i > a+g$, où $i = (I_t+A_t)/Y_t$ et $a = A_t/Y_t$. Les statistiques ne permettent pas de distinguer l'investissement autonome de l'investissement induit, mais des statistiques américaines auxquelles Hicks se réfère indiquent que *i* est sensiblement supérieur à *g*, de quoi laisser une marge appréciable pour *a*.

Les trois idées forces de l'explication hicksienne sont :

1- Les valeurs de *s* et *v* sont telles qu'elles rendent les oscillations explosives.
2- La disposition des ressources constitue un plafond à Y_t. L'économie ne peut aller au-delà du plein-emploi.
3- La limitation de la chute des investissements induits (cf. Tinbergen) empêche la descente de Y_t en-deçà d'un certain plancher

Sur le graphique, la ligne E est celle de la croissance équilibrée ; la ligne F représente le plafond correspondant au plein emploi des ressources. La ligne L correspond au plancher.

Supposons que l'économie était en croissance équilibrée (point P0 sur le graphique) et qu'au temps t0, l'investissement autonome s'accroisse de façon ponctuelle. Vu l'explosivité de l'accélérateur, le revenu s'écarte de la ligne *E* par le haut. Arrivée au plein emploi[1], l'économie peut rester un temps sur la ligne *F* mais l'accélérateur transforme l'arrêt de la croissance au-delà de *g* en une baisse de l'investissement induit ; le revenu hausse moins vite que *g* et il pourrait même connaître une baisse absolue si *g* n'est pas très élevé. Malheureusement, la baisse ne s'arrête pas à la ligne *E*, car l'investissement induit positif ayant quasiment disparu, l'investissement total est maintenant rabaissé à l'investissement autonome moins la dépréciation. Cet investissement réduit détermine un nouveau revenu d'équilibre par la formule du multiplicateur traditionnel[2] ; il est sur la ligne *L* que l'économie rejoint finalement en P2. Par hypothèse, A_t croît également au taux *g* ; la production et le revenu augmentent. L'accélérateur se réveille, d'où le revenu décolle de la ligne *L*. « We are now almost back where we started. When the new upswing reaches the upper equilibrium E, it cannot possibly stop there because output is necessarily expanding more rapidly than it would do along E. The boom will therefore continue, apparently until it reaches the ceiling ».[3]

Le mécanisme décrit ci-dessus génère un cycle d'une régularité parfaite. Le cycle de la réalité économique est irrégulier parce que les conditions de cette régularité ne se rencontrent pas dans le monde concret : taux de croissance constant du revenu, de l'investissement autonome et du plafond de plein emploi, multiplicateur constant, accélérateur constant, régularité du cycle de l'investissement induit.

Hicks insiste sur le caractère non-monétaire de sa théorie : les facteurs réels suffisent à générer le cycle. Mais, reconnaît-il, cela n'empêche pas les facteurs monétaires de l'aggraver, le cas échéant. Le début de la crise voit croître la préférence pour la liquidité. La déflation fait tomber l'économie plus bas que la ligne *L* sur le graphique.

La politique économique que Hicks préconise consiste à stabiliser l'investissement autonome. Maximiser celui-ci en permanence pour que la ligne *E* se confonde avec la ligne *F* ne rencontre pas son assentiment, car cette politique risquerait d'anémier l'investissement induit, ce qui handicaperait l'adaptation de l'offre à la demande.

D'AUTRES THEORIES DU SURINVESTISSEMENT

Il n'est évidemment pas possible de passer en revue les diverses théories du cycle conjoncturel basées sur l'accélération. Voyons encore deux théories du surinvestissement qui ne recourent pas à l'accélération.

Nous avons déjà rencontré l'économiste marxiste polonais Michal **Kalecki** aux sous-chapitres 5.1.5 et 5.5.7. En 1933, il publie « Essai d'une théorie du mouvement cyclique des affaires ». Cet ouvrage a un double mérite. D'une part, il expose une explication complète du cycle

[1] Il y a un plafond pour l'industrie des équipements et un autre pour celle qui fabrique les biens de consommation. Si le rapport de prix se modifie en faveur de la première, le transfert de ressources en provenance de la seconde prolongera le boom mais pas indéfiniment.
[2] Le multiplicateur n'est plus ici dopé par l'investissement induit, un gonflement que Hicks appelait *supermultiplicateur*.
[3] Hicks [153] p. 105

conjoncturel. D'autre part, il révolutionne la macroéconomie en énonçant maints principes qui figureront dans la *Théorie Générale* de Keynes…trois ans plus tard. Malheureusement pour lui, l'ouvrage passa inaperçu, car il est écrit en polonais. Par la suite, Kalecki s'installa en Angleterre et s'intégra dans l'école cambridgienne. Sa théorie du cycle est encore exposée dans divers articles postérieurs en langue anglaise, dont « A Theory of the Business Cycle » (1937), que nous examinons à présent.

Commençons avec l'équilibre macroéconomique à court terme, c'est-à-dire à l'intérieur d'une période ; nous passerons ensuite à l'enchaînement des équilibres de période en période. Par hypothèse, dans le modèle kaleckien, l'épargne est entièrement le fait des capitalistes. L'aphorisme : "the capitalists earn what they spend, while workers spend what they earn", qu'on prête à Kalecki caractérise bien sa conception. Le revenu des capitalistes inclut leur consommation *C* et l'investissement *I*. Leur décision d'augmenter l'un ou l'autre accroît leur revenu, alors que le revenu des salariés dépend des décisions prises en dehors d'eux. « … the spending of capitalists 'forces' a capitalists' income which is equal to that spending »[1]. De la même façon, l'investissement "force" l'épargne, idée centrale du multiplicateur keynésien.

Donc si la répartition du revenu des capitalistes entre consommation et épargne est réglée par un paramètre du type propension marginale à consommer, en déterminant *I*, les capitalistes déterminent l'ensemble des paramètres de l'équilibre à court terme : leur revenu, leur consommation, leur épargne, l'emploi, le revenu des salariés et la consommation de ceux-ci, et donc le revenu national. A noter que chez Kalecki, comme chez Keynes, les variables sont exprimées en *unités de salaire* et non en monnaie. En résumé, l'équilibre macroéconomique à court terme dépend du niveau d'investissement :

$$Y = F(I), \text{ avec } F' > 0 \qquad (8.35)$$

Si *I* détermine l'équilibre à court terme, qu'est-ce qui détermine *I* ? Ici, nous entrons dans l'analyse dynamique, car l'investissement de la période présente résulte de décisions prises lors des périodes antérieures. Selon Kalecki, la valeur présente de *I* est purement un héritage du passé ; elle ne dépend d'aucune autre variable présente. Distinguons donc les variables D_t (investissements décidés pendant la période *t*) et I_t (investissements réalisés pendant la période *t*).

L'axe temporel est divisé en périodes. On sait que I_{t+1} est déterminé par D_t. Mais de quoi dépend D_t ? Comme chez Keynes, c'est le surplus de l'efficacité marginale du capital *r* sur le taux d'intérêt *i* qui incite à investir. Selon Kalecki, le volume des investissements s'établira au niveau qui égalise la différence *r* - *i* avec la prime compensant le risque[2]. Nous avons donc un premier résultat provisoire :

$$D_t = f(r-i), \qquad f' > 0. \qquad (8.36)$$

Nous pouvons encore préciser cette relation en expliquant de quels facteurs dépendent *r* et *i*.

L'efficacité marginale du capital augmente avec les perspectives de profit et diminue avec le prix du capital. Selon Kalecki, les anticipations s'inspirent principalement de la situation économique présente ; vu l'équation (8.35), le profit anticipé croît donc avec *I*.
Comme *I* influence également le prix du capital, mais dans un sens qui diminue *r*, on en conclut que l'efficacité marginale du capital est une fonction de *I* dont le sens est ambivalent a priori.

Après *r*, Kalecki montre que *i* est aussi fonction de *I*, et ce par deux canaux :

[1] Kalecki [177] p. 79
[2] Selon Kalecki, le niveau de risque augmente avec le montant de l'investissement.

- D'abord un canal direct : en augmentant la confiance des investisseurs, *I* pousse l'intérêt à la baisse.
- Ensuite par un chemin plus tortueux : l'investissement fait augmenter le revenu national nominal (non plus en unités de salaire) ; il pousse donc à la hausse la demande de monnaie de transaction, ce qui, comme chez Keynes, fait hausser le taux d'intérêt, car la hausse de la masse monétaire n'est généralement pas suffisante

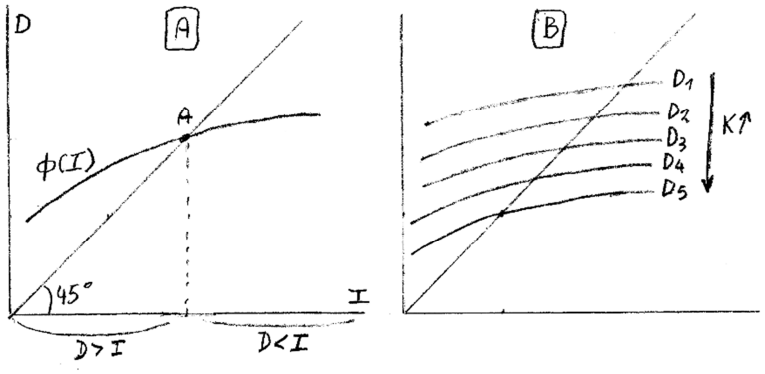
D'après Kalecki, l'augmentation de *I* fait d'abord baisser *i* jusqu'à un certain niveau après lequel elle le fait remonter.

Par l'équation (8.36), *D* est fonction de deux variables qui dépendent toutes deux de *I*. Nous arrivons donc à notre équation fondamentale :

$$D = \Phi(I) \qquad (8.37)$$

Selon Kalecki, cette relation doit avoir la forme apparaissant sur la figure 8.12-A. Pour de faibles valeurs de *I*, la croissance de *D* est favorisée par le moindre coût marginal de la production des biens d'investissement et le fait que *i* est encore dans sa phase décroissante ou peu croissante. Le niveau d'investissement *A* qui assure la stabilité n'est normalement pas supérieur à celui du plein emploi.

Graphique 8.12 : la décision d'investir selon Kalecki



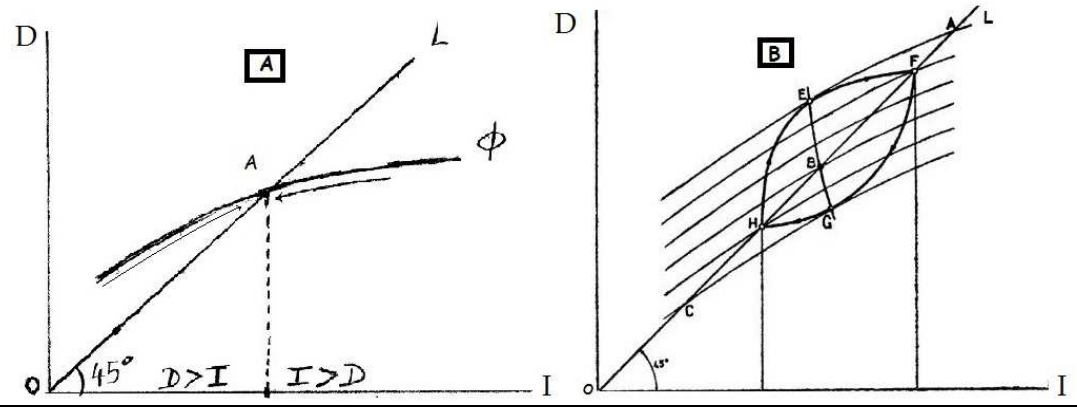
Nous avons présumé implicitement que le volume du capital était donné. En réalité, ce volume influence la relation entre *D* et *I* parce qu'il affecte tant *r* que *i*. Comparons deux économies identiques, notamment par le niveau de *I* mais disposant de capitaux (K) inégaux :

- Une même valeur de *I* signifie un même montant global de profit, mais quand celui-ci est réparti sur un plus grand nombre d'unités de capital, *r* diminue.
- L'augmentation de *K* érode la rentabilité de l'entreprise et réduit donc la part relative des capitalistes dans le revenu national. Comme par hypothèse, le profit est donné avec *I*, le revenu national doit avoir augmenté ; comme ci-dessus, cela fait hausser le taux d'intérêt.

Pour ces deux raisons, l'économie avec plus de capital, ceteris paribus, décidera moins d'investissements nouveaux. C'est ce qu'indique la figure 8.12-B, qui trace la courbe Φ pour cinq valeurs différentes du stock de capital.

Venons-en maintenant à la dynamique sur plusieurs périodes. A gauche de *A*, tant que D > I, *I* croîtra le long de la droite *L* et ce jusqu'à ce que le point *A* soit atteint. A droite de *A*, tant que I > D, *I* décroîtra le long de la droite *L*. Après ajustement sur quelques périodes, l'économie rejoindra immanquablement le point *A* d'égalité entre *D* et *I*, qui est un équilibre stable pour *K* donné (figure 8.13-A).

Graphique 8.13 : le cycle selon Kalecki



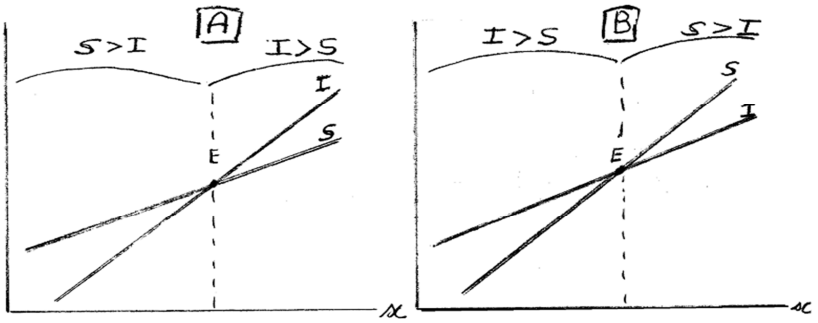
Tenons compte maintenant de ce que l'investissement de chaque période accroît *K*, en même temps que la dépréciation le réduit. Chaque période, un certain volume d'investissements *W* est nécessaire pour maintenir le capital, du fait de la dépréciation. Sur la figure 8.13-B, la courbe EG relie ces niveaux pour les différentes valeurs de *K* ; *W* est évidemment fonction croissante de *K*. Si $I_t > W_t$, K_t augmente et inversement. Pour notre étude du cycle, prenons comme point de départ le point *E* de la figure. En l'absence de variation de *K*, l'économie se dirigera vers le point *A* d'égalisation entre *D* et *I*. Mais à droite de *E*, I > W, et donc *K* augmente, ce qui abaisse la courbe Φ. L'économie aboutira ainsi au point *F* où *I* n'a plus de raison de croître. L'investissement positif continue à faire baisser cette courbe. L'économie se retrouve alors en situation ou I > D, ce qui la pousse vers la gauche où ils pourront s'égaliser. Les mouvements simultanés de la courbe et le long de la courbe amènent l'économie au point *G*, où l'égalité entre *I* et *W* arrête la descente. Mais l'économie continue sa course vers la gauche, espérant trouver le point d'équilibre temporaire *C*. Dans cette zone, l'investissement est insuffisant pour assurer le maintien du capital. La courbe Φ remonte donc. De retour à *H*, la boucle est bouclée. L'économie oscille autour de son point d'équilibre sans pouvoir l'atteindre s'il n'est pas son point de départ.

Kalecki conclut ainsi : « We see that the question, 'What causes the periodical crisis ?' could be answered shortly: the fact that the investment is not only produced but also producing. Investment considered as capitalists' spending is the source of prosperity, and every increase of it improves business and stimulates a further rise of spending for investment. But at the same time investment is an addition to the capital equipment and right from birth it competes with the older generation of this equipment. The tragedy of investment is that it calls forth the crisis because it is useful. I do not wonder that many people consider this theory paradoxical. But it is not the theory which is paradoxical but its subject-the capitalist economy »[1].

En 1940, Nicholas **Kaldor** publie son article "A Model of the Trade Cycle". Le graphique 8.14 présente l'épargne et l'investissement en tant que fonctions du niveau d'activité *x* (exprimé en emplois). Les courbes *S* et *I* donnent respectivement l'épargne et l'investissement EX-ANTE, c'est-à-dire projetés aux différents niveaux d'activité ; toutes deux sont croissantes. Sur la figure A, la pente de *I* est supérieure, sur la figure B, la pente de *S* est supérieure.

[1] Kalecki [177] pp. 95-96

Graphique 8.14 : fonctions d'épargne et d'investissement linéaires



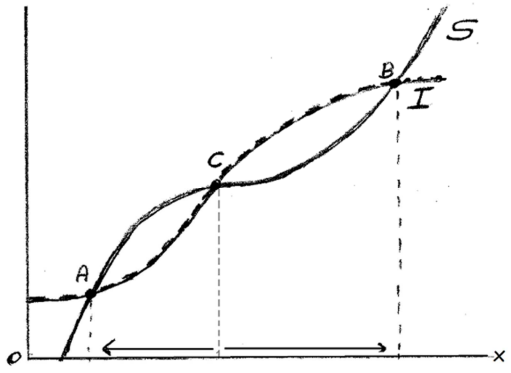
L'équilibre correspond évidemment à la conciliation des projets d'investissement et d'épargne, autrement dit à l'intersection de *I* et *S*. La supériorité de *I* sur *S* stimule l'activité, la supériorité de *S* sur *I* la freine. Il en découle que l'équilibre de 8.24-A est instable car l'activité est stimulée quand elle est au-delà de son niveau d'équilibre et freinée quand elle est en-deçà. Au contraire, l'équilibre de 8.24-B est stable. La figure 8.14-A décrit une situation catastrophique : l'économie est toujours poussée, soit vers l'hyperinflation, soit vers la stagnation. La figure 8.14-B décrit au contraire une situation idéale.

Kaldor considère que le graphique 8.14 est irréaliste. La situation 8.24-A est trop pessimiste pour être vraie, la situation 8.24-B trop parfaite. Pour sortir de cette contradiction, Kaldor préconisée de remplacer les fonctions de la figure 8.14 par des fonctions non-linéaires. Les formes qu'il suggère apparaissent sur le graphique 8.15 ; il les justifie ainsi :

- Courbe *I* : dI/dx est faible aux extrémités de la courbe : pour une activité faible, à cause des capacités en excès ; pour une activité élevée, vu les coûts croissants et la difficulté d'emprunter.
- Courbe *S* : dS/dx est élevé aux extrémités. Lorsque le revenu est très bas, l'épargne est sacrifiée à la consommation et peut virer à la désépargne. Lorsque le revenu est très élevé, la part du revenu épargnée augmente considérablement.

Il y a non plus un mais trois points d'intersection entre *I* et *S*, soit un équilibre multiple. L'équilibre *C* est instable, contrairement aux équilibres *A* et *B*. Donc l'économie tendra tôt ou tard vers l'un de ces deux équilibres, dont les niveaux d'activité diffèrent sensiblement.

Graphique 8.15 : fonctions d'épargne et d'investissement non linéaires



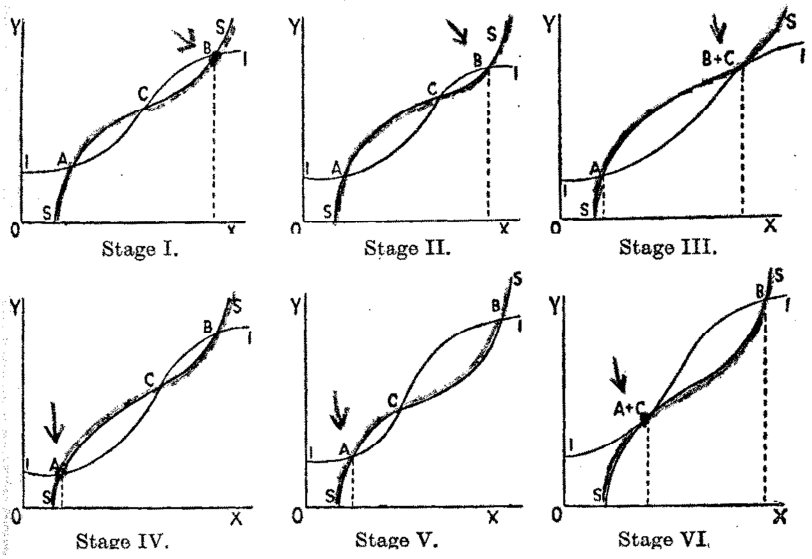
En fait, ce graphique décrit un équilibre de courte période. Le niveau d'activité déterminé par l'équilibre au temps *t* impacte la situation qui prévaudra en t+1, en modifiant la localisation des courbes *I* et *S*.

Un niveau d'activité plus élevé :

- Elèvera la courbe *S*, car plus de revenu donnera plus de consommation et plus d'épargne.
- Abaissera la courbe *I* en réduisant l'éventail des opportunités d'investissement rentables.

On l'aura deviné, les courbes *I* et *S* voyageront en fonction de l'équilibre atteint et ces déplacements feront bouger l'équilibre, ce qui se traduira sur l'abscisse par des fluctuations du niveau d'activité. C'est ainsi que le cycle des affaires se trouve expliqué assez simplement par la forme non linéaire des courbes *I* et *S*. Le graphique 8.16 en expose le déroulement.

Graphique 8.16 : le cycle conjoncturel selon Kaldor



Les flèches indiquent la localisation de l'économie aux différents stades

On part du point d'équilibre B. Le niveau d'activité élevé fait monter la courbe *S* et abaisse la courbe *I* jusqu'à ce que les deux courbes soient tangentes en un point B+C. Il s'agit d'un équilibre instable où S > I à gauche et à droite, ce qui ne peut que pousser à la réduction du niveau d'activité. L'économie arrive à l'équilibre A ; l'activité déprimé fait monter la courbe *I* et abaisse la courbe *S* jusqu'à ce qu'elles deviennent tangentes au point A+C. Cet équilibre est instable avec I > S à gauche et à droite, ce qui pousse à l'accroissement du taux d'activité, jusqu'à l'équilibre *B*. L'économie est de retour à la case de départ.

Ce scenario repose sur trois hypothèses :
- Au niveau d'activité moyen, dI/dx > dS/dx
- Aux niveaux d'activité extrêmes, dS/dx > dI/dx
- Aux points A et B, le volume de l'investissement doit être suffisamment bas ou suffisamment élevé pour que s'ensuive le déplacement des courbes *I* et *S*.

Les deux premières conditions semblent généralement satisfaites. Un doute peut subsister, selon Kaldor, quant au fait que l'investissement net devienne négatif à l'équilibre A. « Hence the forces making for expansion when we start from a state of depression are not so certain in their operation as the forces making for a down-turn when we start from prosperity… »[1].

La PERIODE du cycle dépend de la vitesse du coulissement entre les équilibres A et C ainsi que de la vitesse de déplacements des courbes *I* et *S*, influencée par la durée de production des biens capitaux et par leur durabilité. L'AMPLITUDE du cycle dépend de la distance entre les équilibres

[1] Kaldor [173] p. 186

A et C. Des facteurs exogènes comme le progrès technique, les habitudes d'épargne et de consommation expliquent les variations d'amplitude d'un cycle à l'autre.

Comme Kaldor le constate lui-même, son modèle n'est pas sans ressemblance avec celui de Kalecki.

8.3.2. Le cycle causé par les phénomènes monétaires

Tout comme, d'ailleurs, les théories du surinvestissement, les théories monétaires du cycle ont une histoire ancienne. Mais elles connaissent leur heure de gloire entre les deux guerres, avec Hawtrey en Angleterre et avec l'école autrichienne représentée par Hayek[1] et Mises. Le présent sous-chapitre examinera d'abord la théorie monétaire du cycle qu'Hayek exposa dans deux essais, « Monetary Theory and the Trade Cycle » (1929) et « Prices and Production » (1931). Nous élargirons l'analyse à la conception de la monnaie chez Hayek. Le sous-chapitre abordera ensuite la critique de la théorie hayekienne par Kaldor et la controverse entre les deux économistes qui s'ensuivit (1939-1942).

HAYEK : CREDIT ET DETOURS DE LA PRODUCTION

La théorie hayekienne du cycle repose sur deux piliers :
- L'opposition wicksellienne entre le taux d'intérêt naturel et le taux monétaire.
- La tradition autrichienne des détours de la production.

Si l'octroi de crédits et la création monétaire associée sont excessifs, le taux monétaire tombe en deçà du taux naturel. Le taux d'intérêt trop bas et le crédit facile poussent à l'accroissement de l'intensité capitaliste de la production, autrement dit à une production plus détournée. Les aléas monétaires augmentent et réduisent tour à tour les détours de la production ; ce processus entraîne toute l'économie dans son sillage. Voilà le cycle hayekien brièvement résumé.

Après une dépression, lorsque l'économie reprend, les banques se laissent aller à l'excès de crédit. von
Mises en impute la responsabilité au pouvoir politique qui, par démagogie, préfère l'inflation à la dépression. Plus pondéré que son collègue, Hayek conçoit que le phénomène peut également être endogène. Pressées par la concurrence entre elles, les banques commerciales veulent profiter de l'accroissement de la demande de crédit des entreprises et n'hésitent pas à réduire leur coefficient de réserve monétaire pour ne pas rater le train. Ce faisant, elles peuvent accroître l'encours de crédit, en diminuant leur taux d'intérêt si nécessaire.

Hayek introduit dans « Prices and Production » ses fameux *triangles* qui lui permettent de visualiser les détours de la production et donc de comparer leur évolution selon que les capitaux proviennent de l'épargne ou du crédit monétaire (figure 8.17). Il s'agit d'un empilement vertical de rectangles de largeur croissante, formant grosso modo un triangle rectangle. Le rectangle du bas représente la production de biens de consommation. Plus on monte, plus les biens sont d'un ordre supérieur. La hauteur du rectangle correspond à la période de production absolue, exprimée comme une variable discrète dont l'unité est le stade (pour simplifier). Le supplément de largeur d'un rectangle par rapport à celui juste au-dessus correspond aux facteurs primaires ajoutés à ce stade de la production. Les nombres inscrits dans les rectangles indiquent la valeur (et non la quantité physique) que le produit atteint au stade considéré ; ils sont repris de

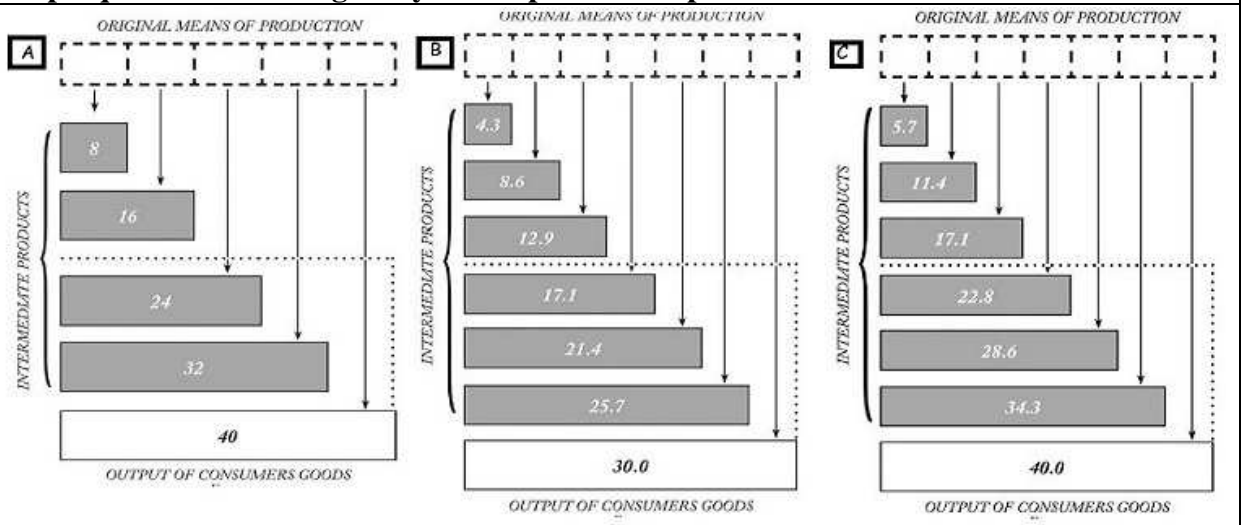
[1] Hayek quitta l'Autriche pour professer à la *London School of Economics* à partir de 1930. La LSE comptait sur lui comme fer de lance dans la rivalité qui l'opposait à Cambridge.

l'exemple chiffré donné par Hayek. Le capital social vaut le total des rectangles autre que celui des biens de consommation (base du triangle). Hayek introduit les simplifications suivantes :

- égalité de l'accroissement à chaque stade (sinon, ce n'est plus un triangle)
- l'intérêt est ignoré. La valeur d'un rectangle par rapport au précédent devrait s'accroître non seulement de l'adjonction de facteurs primaires supplémentaires, mais aussi de l'intérêt sur les stades antérieurs (cf. Wicksell).
- le passage d'un stade à l'autre implique toujours un échange monétaire entre deux entreprises distinctes.

Dans ces conditions, la période de production moyenne vaut la moitié de la hauteur du triangle

Graphique 8.17 : le triangle hayekien : production plus détournée



Au départ, l'appareil productif de l'économie se présente tel que décrit par par la figure 8.17-A. Supposons que la demande de biens capitaux par les entreprises augmente, de telle façon que le rapport capital/ produit passe de 2 à 3. Le financement de ces investissements peut emprunter deux voies :

1- <u>l'épargne volontaire</u> (figure 8.17-B) : la société remplace une partie de la production des biens de consommation par la fabrication de biens capitaux. Ce déplacement implique nécessairement l'augmentation de la hauteur du triangle. Par hypothèse, la masse monétaire et sa vitesse de circulation sont restées constantes, tout comme l'offre de facteurs primaires. Le revenu NOMINAL global restant constant, les valeurs des différents stades diminuent parce qu'elles sont exprimées en monnaie dont la valeur unitaire s'accroît ; le prix des biens de consommation bénéficie de la productivité accrue des détours nouveaux et baisse plus encore que la valeur du triangle inférieur ; le revenu REEL hausse donc, même si cela ne transparaît pas dans les valeurs nominales.

2- <u>le crédit monétaire</u> (figure 8.17-C) : le revenu nominal augmente des 40 correspondant à l'accroissement du capital total ; le produit physique est le même qu'à la figure 8.17-B, mais sa valeur est plus élevée, car le prix des biens de consommation a moins diminué que dans le cas précédent.

Entre ces deux voies, il existe une différence fondamentale. L'évolution résultant de l'épargne est irréversible, car elle résulte d'un choix volontaire des investisseurs. Au contraire, dans le cas du crédit, le sacrifice de l'épargne est consenti involontairement par les consommateurs

subissant une hausse des prix[1]. Hayek met en avant le concept d'*épargne forcée*. "(This sacrifice) comes about not because (consumers) want to consume less, but because they get fewer goods for their money income. There can be no doubt that, if their money receipts should rise again, they would immediately attempt to expand consumption to the usual proportion"[2]. Les détours de la production feraient alors marche arrière. Le changement de structure de la production n'est donc pas irréversible.

Le crédit, l'épargne et les variations du taux d'intérêt associées influencent la structure de la production en modifiant le rapport entre la base et la hauteur du triangle hayekien. Ainsi, une baisse du taux d'intérêt fera passer le centre de gravité plus en hauteur. Pratiquement, l'ajustement de l'appareil productif dépend de la direction prise par l'investissement, plus en aval ou plus en amont ; mais il passe aussi par le transfert d'équipements *non spécifiques* d'un stade à l'autre. Chaque équipement a une productivité marginale potentielle dans les stades où il peut intervenir et prendra la direction qui conduira à l'égalisation de ces productivités marginales. Si l'économie connaît une poussée de l'intensité capitalistique, les entreprises des stades antérieurs verront leur prix de vente monter, ce qui accroîtra la productivité marginale de leurs biens capitaux, provoquant un afflux de nouveaux biens.

Après avoir rassemblé les pièces du puzzle, nous pouvons maintenant reconstituer le cycle conjoncturel. Tout commence avec des moyens supplémentaires que le crédit met à la disposition des entreprises, moyennant une baisse du taux d'intérêt. Les prix des biens d'ordre supérieur s'élèvent, ainsi que leur part dans la production qui devient donc de plus en plus détournée. Le cycle est dans sa phase d'expansion. Dans un premier temps, l'essor des productions d'ordre supérieur se nourrit du crédit, et n'obère pas la production des industries en aval. Toutefois, après un certain temps, il détournera des ressources de la production de biens de consommation qui s'en trouvera réduite. Et cela au moment où le revenu disponible des facteurs production, dopé par la croissance du revenu nominal fait hausser la demande de ces biens de consommation. Par le jeu de l'offre et de la demande, leur prix s'élève par rapport à celui des biens de production.

La production de biens de consommation, devenue plus profitable, attire les ressources à elle : la production devient moins détournée ! C'est le retournement conjoncturel. L'emploi chute dans les processus longs mais l'embauche ne commence qu'à pas timides dans les processus courts. Le chômage s'accroît. Hayek constate une asymétrie : la transition vers une production plus détournée ne cause pas de problèmes ; la transition vers une production moins détournée débouche sur la crise.

Paradoxalement, l'octroi de crédit à l'investissement aboutit en définitive à la même situation que si ce crédit avait été octroyé directement à la consommation. Cette évolution ne pourrait être évitée qu'en continuant à augmenter la masse monétaire par toujours plus de crédits aux entreprises ; mais il est évident qu'un tel processus ne peut durer indéfiniment.

Pour assurer un progrès économique équilibré, il faut éviter que la consommation augmente sans accompagnement de l'épargne. La production doit augmenter proportionnellement aux différents stades. Le fait qu'en dépression, une partie du capital reste oisif ne signifie pas qu'il y ait pléthore de capital et excès d'épargne. Que du contraire ! Il faudrait épargner plus (et donc consommer moins) afin d'investir dans des productions qui rendraient rentable l'utilisation de ce capital apparemment excédentaire.

---

[1] Dans l'exemple ci-dessus, le prix a haussé par rapport à son niveau d'équilibre qui était lui-même en baisse.

[2] Hayek [144] p. 243

Le crédit à la consommation est à proscrire absolument. Le crédit à l'investissement est nuisible lorsqu'il fait paraître rentables des processus plus détournés que ce qui résulte de l'état naturel. D'une façon générale, Hayek est réticent aux stimulants artificiels.

LA THEORIE MONETAIRE DE HAYEK

Hayek critique les théories monétaires qui prévalaient à son époque, principalement celle de Fisher. Il est un des premiers économistes néoclassiques à se départir de la vision dichotomique. Il affirme qu'une variation du stock monétaire affecte les prix relatifs et le revenu réel, et ce de façon directe, c'est-à-dire autrement que par le biais de l'effet sur le niveau général des prix. Contrairement à Wicksell qui voulait favoriser la stabilité des prix par l'égalité entre le taux d'intérêt naturel et le taux courant, Hayek estime que ces deux objectifs sont inconciliables. La stabilité des prix n'est compatible avec des variations de la production que si le stock monétaire suit ces variations, ce qui bouscule l'équilibre entre l'offre et la demande de capitaux et donc le taux d'intérêt.

Il y a toutefois un théoricien monétaire qui recueille l'admiration de Hayek : Cantillon, pour sa théorie selon laquelle l'effet d'une variation du stock monétaire dépend de son itinéraire : par quelles poches est-elle entrée. Ces variations, presque toujours, modifient les rapports de prix entre les différents stades de la production. Il en résultera une déformation de la structure de la production et une modification de son volume.

Le dernier chapitre de « Prices and Production » s'intitule « Pour ou contre une monnaie élastique ». Selon Hayek, l'*élasticité du système monétaire* est la facilité avec laquelle la masse monétaire répond à certains stimuli. Une élasticité élevée a, certes, l'avantage de la souplesse. Mais comme l'a montré la section précédente, l'instabilité économique peut lui être imputée. Plutôt que la *monnaie élastique*, Hayek défend ce qu'il appelle la *monnaie neutre*, c'est-à-dire celle qui n'induit pas d'écart entre le taux d'intérêt naturel et le taux courant. Comme on l'a dit, la monnaie neutre ne mène pas nécessairement à la stabilité des prix, objectif que Hayek ne préconise pas. Selon lui, la meilleure manière d'assurer la neutralité monétaire est l'invariabilité du stock monétaire. Il écrit : "It would appear rather that the fall of prices proportionate to the increase in productivity, which necessarily follows when, the amount of money remaining the same, production increases, is not only entirely harmless but is in fact the only means of avoiding misdirections of production"[1].

Hayek reconnaît néanmoins un motif qui justifie valablement l'ajustements du stock monétaire. C'est l'adaptation à ce qu'il appelle le *coefficient de transaction monétaire*. Celui-ci mesure la part relative des flux productifs visualisés dans les triangles, qui se réalise sous la forme d'un échange monétaire. Ce ratio dépend de l'organisation générale des affaires, notamment du degré d'intégration verticale des entreprises[2].

Concluons avec ces citations :
- "bankers need not be afraid to harm production by overcaution".
- "It is probably an illusion to suppose that we shall ever be able entirely to eliminate industrial fluctuations by means of monetary policy"[3].

[1] Hayek [144] p. 283
[2] Dans un essai ultérieur (1939), il acceptera également que la masse monétaire s'adapte à l'offre des facteurs primaires.
[3] Hayek [144] p. 298

LA CONTROVERSE ENTRE HAYEK ET KALDOR

Cette controverse se déroule en trois phases : en 1939, **Kaldor** publie « Capital Intensity and the Trade Cycle », un article critiquant la conception du cycle exposée dans « Prices and Production ». Peu de temps plus tard, Hayek publie son essai « Profits, Interest and Investment ». Dans quelle mesure est-il et se veut-il une réponse à Kaldor ? L'explication du cycle qui y apparaît s'écarte apparemment de celle de « Prices and Production », mais Hayek prétend que les deux essais concordent. En 1942, Kaldor revient à a charge, pointant la discordance entre les deux essais de Hayek et critiquant le second.

Dans « Prices and Production », l'intensité capitalistique de la production varie procycliquement. Selon Kaldor, c'est l'inverse qui est vrai.

- Concernant le CAPITAL EXISTANT, en période de boom, les recrutements font monter le nombre d'emplois sur le capital existant ; en dépression, des équipements restent sous-utilisés, faute de bras.
- Concernant le NOUVEAU CAPITAL, le taux d'intérêt élevé et le salaire réel modéré, caractéristiques du boom, favorisent les technologies intenses en travail. En dépression, c'est l'inverse.

Certes, la confiance qui inonde l'économie pendant l'expansion exerce un effet semblable à une baisse du taux d'intérêt, puisqu'elle abaisse la prime de risque, mais la tendance à la baisse de l'intensité capitalistique l'emporte[1]. Quoi qu'il en soit, au vu de la marge assez étroite de variation du taux d'intérêt, Kaldor déduit que les fluctuations cycliques de l'intensité capitalistique sont mineures.

Par unité d'accroissement de la capacité productive, il faut plus de capital investi lorsque l'intensité capitalistique est élevée que lorsqu'elle est basse. D'une certaine façon, les investissements à haute intensité capitalistique freinent l'accroissement de la capacité productive, alors que les investissements à basse intensité capitalistique l'élèvent. Compte tenu de l'évolution contracyclique de l'intensité capitalistique, Kaldor conclut que la capacité productive augmente plus que l'investissement durant le boom et moins que lui durant la dépression.

C'est cet accroissement débridé de la capacité productive pendant le boom qui causera le recul de l'investissement et le retournement conjoncturel, pour deux raisons :

- La croissance du revenu entraîne celle de la propension marginale à épargner, ce qui empêche la demande des biens de consommation de suivre l'offre ;
- Les investissements du boom aboutissent à une pénurie de main d'œuvre.

En définitive, le boom ne cesse pas parce que l'intensité capitalistique est devenue excessive, comme l'affirmait Hayek, mais à cause de son insuffisance.

Kaldor critique également le concept d'épargne forcée et rappelle la position keynésienne : toute épargne est volontaire[2].

---

[1] Il n'est peut-être pas inintéressant de connaître l'avis de **Harrod** sur cette question : « What probably happens is that, while the rate at which production as a whole becomes more capitalistic is higher in the boom, just because there is more replacement and new investment going on, the amount by which methods of production become more capitalistic per unit of new investment is less in the boom than in the slump » ([131] p. 94).

[2] Comme nous l'avons vu au sous-chapitre 5.1.1, Keynes ne croyait pas que des investissements pouvaient être financés par le système bancaire sans qu'une épargne correspondante soit suscitée.

Dans l'essai de **Hayek** « Profits, Interest and Investment », l'intensité capitalistique devient contracyclique, comme chez Kaldor, et la monnaie et le crédit sont à peine mentionnés. Cet essai semble moins intéressant que « Prices and Production » et la place nous manque pour l'exposer en détail.

Le concept central est ce que Hayek appelle – un peu improprement selon Kaldor- l'*effet Ricardo* : "Its substance is contained in the familiar Ricardian proposition that a rise in wages will encourage capitalists to substitute machinery for labour and vice versa"[1]. Il s'agit bien entendu du salaire REEL, sujet aux variations du prix des biens de consommation. Précisément, ces prix jouent un rôle important dans le scénario.

Avec l'effet Ricardo, la hausse du prix du bien final décourage les longs détours de la production pour deux raisons :
- La hausse du prix des biens de consommation fait baisser le salaire réel.
- L'augmentation du prix du bien final, lorsque le coût reste constant, accroît la profitabilité des détours courts plus que celle des détours longs[2].

Précisément, dans la deuxième phase du boom, lorsque la consommation s'est déjà bien redressée, la pression de la demande de biens de consommation pousse leur prix à la hausse. Dès lors, les entreprises de ce secteur vont combiner leur équipement avec plus de travail et acheter des machines moins chères qui économisent une quantité de travail moindre. Les entreprises du haut du triangle hayekien commencent à souffrir de ces choix, ce qui entraîne du chômage. Le retournement conjoncturel est amorcé ; il s'accentue lorsqu'en conséquence baissent la demande et le prix des biens de consommation.

La reprise viendra parce que l'investissement ne peut pas tomber plus bas que les nécessités du renouvellement du capital. La reprise et le début de l'expansion sont caractérisés par des investissements importants, tant en largeur qu'en profondeur. La situation reste saine tant que l'offre des biens de consommation suit leur demande. Ce ne sera pourtant pas le cas indéfiniment. Hayek montre que le taux de profit trop bas stimule l'augmentation du revenu et défavorise l'industrie des biens de consommation. S'ensuit une pression de la demande de biens de consommation sur l'offre, une augmentation de leur prix et un accroissement du taux de profit. L'épargne n'arrive plus à suivre l'investissement. La demande de biens de consommation se fait pressante. Nous revoici au point où nous avons pris le cycle en marche.

Je en mentionnerai pas les critiques de Kaldor (1942) à l'encontre de cette théorie, exposée ici trop brièvement ; elles sont plus véhémentes que celles de 1939 à l'encontre de « Prices and Production ». Dans son commentaire sur cette critique, Hayek persiste à prétendre qu'il n'a qu'une seule théorie, qui a juste un peu évolué.

Kaldor reconnaît une continuité sur deux points :
- Ce qu'il appelle l'*effet d'accordéon*. L'intensité capitalistique connaît un cycle de gonflement-dégonflement lié au cycle conjoncturel, synchronisé avec lui dans « Prices and Production » et inversé dans « Profits, Interest and Investment ». Cette appellation *effet*

[1] Hayek [141] p. 8

[2] Pour bien le comprendre, il faut considérer l'application d'un facteur primaire à des instants différents, comme on la retrouvait chez Böhm Bawerk ou chez Wicksell. Imaginons une production impliquant une application de travail valant 100 unités deux ans avant la commercialisation et une application de 100 six mois avant la commercialisation. Si le taux d'intérêt vaut 6%, la première doit rapporter un profit de 12 et la seconde un profit de 3. Si maintenant le prix augmente de 2%, la première rapportera 14, soit un accroissement de 17% et la seconde 5, soit un accroissement de 67%.

d'accordéon est devenue l'étiquette par laquelle est souvent désignée la théorie hayekienne du cycle.
- C'est l'excès de consommation qui retourne la conjoncture vers le bas et non son insuffisance comme le pensent la majorité des économistes.

Kaldor ne croit pas à l'effet d'accordéon. Il assène un argument qui me paraît imparable et évident. Du fait de l'innovation technologique, la mécanisation connaît un trend croissant à long terme. A aucun moment, les firmes n'ont avantage à lui faire violence pour des raisons temporaires. Kaldor montre également, que même dans l'hypothèse improbable où l'effet se manifesterait, il ne pourrait atteindre l'ampleur nécessaire pour conduire au cycle conjoncturel. Selon lui, les variations du taux d'investissement au cours du cycle reflètent des mouvements de sa largeur, non de sa profondeur.

LA DENATIONALISATION DE LA MONNAIE

Comme nous avons esquissé la théorie monétaire de Hayek, une digression s'impose pour évoquer sa proposition de mettre fin au monopole de l'Etat sur l'émission de monnaie fiduciaire. Faisons un bond de quarante ans, jusqu'en 1976 quand paraît son essai « Denationalisation of Money : The Argument Refined ». Pour Hayek, virulent opposant de toujours à l'intervention de l'Etat dans l'économie, cette proposition est une évidence. Il s'étonne qu'elle n'ait jamais été examinée sérieusement auparavant.

Les banques commerciales qui le souhaitent dispenseraient des crédits, géreraient les comptes de leurs clients et émettraient des billets de banque dans une monnaie qui est leur propre création ; il coexisterait donc de nombreuses monnaies indépendantes. Les monnaies publiques existantes disparaîtraient rapidement, incapables de faire jeu égal avec des monnaies mieux gérées parce que privées[1].

Supposons que nous pouvons observer le système en fonctionnement et décrivons-le. La « banque Hayek » émet des « ducats » ; à cette fin, elle dispose de deux canaux : elle achète au public intéressé d'autres monnaies et elle consent des prêts en ducats. A l'origine, elle avait déterminé un panier de produits, le plus probablement des matières premières cotées, qui servira de référence pour sa valeur tout au long de son existence. Le robinet des émissions s'ouvre plus ou moins grand selon l'évolution du cours de la monnaie par rapport à la valeur de ce panier. L'objectif de la banque est de garder ce rapport constant. Le public est abondamment informé des écarts, notamment par la presse. Si la banque Hayek applique une politique d'émission trop laxiste par des taux d'intérêt débiteurs trop bas, le public se méfie de sa monnaie ; par le jeu de l'offre et la demande le cours du ducat baissera par rapport à celui des autres monnaies. Si la banque Hayek a une obligation primordiale, c'est bien d'échanger les ducats qui lui reviennent ; l'incapacité de l'assumer implique la faillite. Une telle perspective suffit à insuffler les bons réflexes aux banquiers. Hayek croit à la discipline imposée par les marchés. En plus du droit d'émettre de la monnaie, il préconise la dérégulation de l'activité bancaire.

Le système monétaire prôné par Hayek fonctionne sans banque centrale. Quant au concept de politique monétaire, il devient sans objet.

---

[1] On connaît la *loi de Gresham* selon laquelle « les mauvaises monnaies chassent les bonnes », entendant par là que les acteurs économiques voudront écouler les mauvaises monnaies dans leurs transactions afin de conserver les bonnes. C'est un peu le scenario inverse que prédit Hayek, mais il rappelle que la loi de Gresham s'applique à des monnaies entre lesquelles existe une parité fixe.

Tous les salariés ne reçoivent pas leur salaire dans la même unité monétaire. Les commerçants sont motivés pour accepter un large éventail de monnaies. L'étiquetage des prix est, certes, un peu compliqué. Le paiement électronique facilite les transactions.

On ne peut exclure qu'une monnaie mal gérée défaille complètement. Les droits et obligations à long terme libellés dans cette monnaie ne s'effaceraient pas pour autant ; les tribunaux y veilleraient. Par contre, les détenteurs de liquidités seraient délestés de leur avoir.

Le but ultime de cette réforme est de prévenir l'inflation, que Hayek considère comme le mal absolu. L'inflation rend approximative la comptabilité des entreprises et donc les actions fondées sur elle. Elle favorise l'accroissement des dépenses publiques au-delà des recettes, puisque l'emprunt sera remboursé au rabais. Or l'histoire monétaire se révèle être la sempiternelle répétition d'un scenario de dépréciation des monnaies. Parlant de l'inévitabilité de l'inflation sous le régime actuel, il écrit : « But I believe that it will lead to the destruction of our civilisation unless we change the political framework »[1]. L'avantage d'opérer toutes ses transactions avec une seule monnaie monopolistique a un coût que le public trouvera excessif lorsqu'il le comprendra. Avec la concurrence entre les monnaies, l'inflation ou la déflation généralisées se révéleront impossibles. Pastichant Adam Smith, Hayek écrit : « Good money can come only from self-interest, not from benevolence »[2].

Comme la politique monétaire, la théorie quantitative est également frappée d'obsolescence. Hayek reste fidèle à sa conception monétaire de « Prices and Production », si ce n'est qu'il ne croit plus qu'une monnaie peut être neutre. Il reproche aux monétaristes d'ignorer l'effet de la monnaie sur les prix relatifs et sur la structure de la production.

Pour contrer les effets indésirables de l'inflation sur la répartition des revenus, Milton Friedman avait préconisé l'indexation des contrats, envisageant non seulement les salaires mais tous les revenus fixes, y compris l'intérêt des fonds d'Etat. Hayek dénonce cette solution. En allégeant certains maux causés par l'inflation, on la rend plus tolérable, ce qui affaiblira la motivation pour l'extirper.

A plus d'une reprise, Friedman exprimera son scepticisme à l'égard de la réforme monétaire proposée par Hayek. Il doute que le public abandonne les monnaies auxquelles il est habitué.

8.3.3. Le « cycle économique réel »

La théorie du « cycle économique réel » (real business cycle : RBC) s'est imposée comme l'une des théories macroéconomiques les plus influentes ; elle date du début des années quatre-vingts. Ses promoteurs appartiennent à la mouvance des *nouveaux classiques*. Avant ça, Lucas, principal porte-parole des nouveaux classiques, avait proposé sa propre conception du cycle économique, qui appartiendrait plutôt à la catégorie des théories monétaires du cycle.

LUCAS ET LE CYCLE DES AFFAIRES

Lucas et le marché du travail

Dans son premier article à succès, « Real wages, Employment and Inflation » (1970), coécrit avec Leonard Rapping, Lucas s'intéresse à l'emploi et au chômage, l'une des variables essentielles du cycle conjoncturel.

[1] Hayek [142] p. 84
[2] Hayek [142] p. 131

Contrairement à beaucoup d'économistes, les auteurs n'adhèrent pas de la prétendue rigidité du salaire nominal. L'hypothèse d'un marché du travail concurrentiel leur paraît la plus proche de la réalité. C'est à un modèle parfaitement walrassien qu'ils nous convient.

Le modèle est basé sur trois équations : la demande de travail, l'offre de travail et le chômage. Ces équations font ensuite l'objet d'un test économétrique. Nous ne les exposerons pas ici, mais nous nous contenterons d'exposer la conception de l'offre de travail et la conception du chômage qui sous-tendent les équations.

Lucas et Rapping l'envisagent l'offre de travail comme un problème d'optimisation intertemporelle, mettant le ménage devant le double choix : loisir ou travail, travail aujourd'hui ou travail dans le futur. Il doit maximiser l'utilité intertemporelle U(C,C*,N,N*), moyennant une contrainte budgétaire. Les variables sont la consommation (C), les heures de travail (N), le salaire nominal (W), le niveau général des prix (P), le montant des actifs détenu (A) et le taux d'intérêt (r). Le modèle est à deux périodes et l'astérisque désigne les variables pour la période future, forcément anticipées. La fonction *U* est censée manifester une forte substituabilité tant intra- que intertemporelle.

L'offre de travail présente *N* varie positivement avec W/P et négativement avec W*/P(1+r), avec P*/P(1+r) et avec A/P. La comparaison entre le salaire présent et le salaire futur anticipé est essentielle. Une hausse du salaire présent perçue comme temporaire incitera à travailler davantage actuellement et moins dans le futur ; une baisse temporaire aura l'effet inverse.

Venons-en à l'explication du chômage. Désignons par w_t le salaire d'équilibre qui égalise l'offre et la demande de travail. Lorsqu'une personne déclare rechercher activement un emploi, cela signifie qu'elle accepte de travailler à ce tarif. Le problème est que ce tarif n'est pas une donnée connue de tous. Supposons une baisse de la demande de travail ; le salaire d'équilibre baisse également, mais les chercheurs d'emploi, insuffisamment renseignés, ignorent le changement et continueront à chercher au tarif précédent. La prise de conscience prend du temps. Plus précisément, le modèle prête aux chercheurs d'emploi un salaire de réserve valant l'anticipation du salaire d'équilibre établie la période précédente, soit w^*_{t-1}. Si $w_t = w^*_{t-1}$, pas de problème. Mais si $w_t < w^*_{t-1}$, une masse de gens chercheront un emploi qu'ils ne peuvent trouver ; ils forment le noyau dur du chômage, noyau auquel il faut ajouter le chômage frictionnel.

Vu le caractère intertemporel des choix posés par les offreurs de travail, leur offre n'est pas homogène de degré zéro par rapport au salaire nominal courant et au niveau des prix, une situation généralement qualifiée d'*illusion monétaire* ; dans ce cas-ci, elle répond pourtant à des motivations rationnelles.

Les auteurs concluent : "Measured unemployment (more exactly, its non-frictional component) is then viewed as consisting of persons who regard the wage rates at which they could be employed as temporarily low and who therefore choose to wait or search for improved conditions rather to invest in moving or occupational change. The view that non-frictional unemployment is, in this sense, 'voluntary' does not of course imply that high measured-unemployment rates are socially costless"[1]

[1] Lucas & Rapping [240] p. 748.

La nature du cycle conjoncturel

Lucas présente sa conception du cycle dans l'article « Understanding Business Cycles » (1977). Le message est que pour comprendre le cycle, il faut passer d'une théorie du déséquilibre à une théorie de l'équilibre[1]. « Keynes wrote as though the 'involuntary' nature of unemployment were verifiable by direct observation, as though one could somehow look at a market and verify directly whether it is in equilibrium or not »[2]. La rigidité du salaire nominal, cause du déséquilibre, a été acceptée comme un postulat, alors que du point de vue de la science économique, il s'agit d'une singularité.

Venons-en à l'explication du cycle. Elle se fonde sur :
- L'équilibre macroéconomique avec information imparfaite, déjà abordé au sous-chapitre 7.4.2.
- Le concept de choc (déjà présent chez Frisch).

Considérons le vendeur d'une marchandise œuvrant sur un marché concurrentiel et n'ayant donc aucune influence sur le prix. Il observe son prix et élabore des prévisions de son évolution future. Cette information, combinée avec le prix de ses inputs guide ses choix quant à la quantité à offrir aujourd'hui et ultérieurement. Supposons que le prix présent augmente. Si notre vendeur perçoit cette hausse comme temporaire, il en profitera pour beaucoup produire à court terme afin de privilégier le loisir par la suite. S'il la perçoit comme permanente, soit il ne changera rien, soit il réduira légèrement sa production présente et future. L'emploi fluctue donc avec les variations de prix temporaires et est insensible aux variations définitives. La cause en es la forte substituabilité intertemporelle entre le travail et le loisir déjà fondamentale dans le modèle Lucas-Rapping.

Face aux variations de prix, les décisions d'investissement suivent un schéma inverse : une augmentation du prix boostera l'investissement si elle est perçue comme permanente et l'affectera très peu si elle est perçue comme temporaire. Comme le cycle connaît une fluctuation synchronisée de l'investissement et de l'emploi, Lucas en déduit que la plupart du temps, les variations du prix sont une combinaison d'éléments permanents et d'éléments temporaires. Il ne suffit donc plus au producteur d'observer les prix ; il faut décoder leur évolution, pour discerner la composante temporaire et la composante permanente qui ne sont pas observables telles quelles. Compte tenu de toutes ces difficultés, Lucas considère que face à une hausse du prix, le comportement rationnel se manifestera par un accroissement du travail et du capital investi.

Les producteurs, salariés et investisseurs sont en outre confrontés à la difficulté de distinguer les variations du niveau général des prix des modifications de prix relatifs.
Examinons maintenant les chocs, c'est-à-dire les modifications de l'environnement qui induisent des modifications des prix, souvent imprévisibles. Il en est de deux sortes : les *chocs réels* et les *chocs monétaires*.

Les chocs réels se composent des progrès technologiques et de l'évolution des goûts des consommateurs. Il y en a en permanence. Le plus souvent, ils n'affectent qu'un marché ou une partie de l'économie. Cette multitude de chocs auront tendance à se compenser, ce qui les élimine des causes potentielles du cycle économique.

[1] C'était déjà le sens de l'article de Lucas-Rapping, concernant le marché du travail.
[2] Lucas [239] p. 12.

L'analyse des chocs monétaires amène Lucas à les considérer comme les fauteurs du trouble économique. S'ils sont mal décodés, ils affecteront l'emploi et l'investissement de la façon expliquée ci-dessus. En outre, ils frapperont l'économie transversalement et le mouvement des prix épousera celui de la production.

La propagation des chocs en un cycle conjoncturel repose sur une condition : "It is essential to this argument that general price movements not be perceived as such they are occurring (…) Of the many sources of risk of importance to him, the business cycle and aggregate behavior generally is, for most agents, of no special importance, and there is no reason for traders to specialize their own information systems for diagnosing general movements correctly".

Supposons un signal monétaire menant à une augmentation générale des prix, qui a été mal décodé. Il s'ensuit des investissements non justifiés. L'accroissement des capacités retarde la hausse générale des prix qui révélerait aux agents leur erreur. La correction prendra du temps et nécessitera probablement un sous-investissement durant une période relativement longue. On pourrait contre-argumenter que les décisions d'investissement se fondent généralement sur des considérations à long terme largement soupesées. Mais Lucas, au contraire, insiste sur la nécessité d'hyperréactivité qu'implique l'environnement concurrentiel : « A quick response to what seems to others a weak 'signal' is often the key of a successful investment. The agent who waits until the situation is clear to everyone is too late… »[1].

Puisque les chocs monétaires causent des fluctuations réelles de l'économie, souvent délétères, la conclusion qui s'impose en matière de politique économique est qu'une politique de stabilisation de la masse monétaire à long terme vaut mieux que des tentatives de la manipuler pour contrer des déséquilibres supposés.

N'est-il pas paradoxal qu'une telle théorie émane de Lucas ? Que l'information soit imparfaite, c'est compréhensible. Mais la rationalité des anticipations motive les agents à s'informer le mieux possible. L'idée que les entrepreneurs ne se préoccupent pas du cycle et sont incapables de situer son évolution est très contestable. Au départ la théorie des nouveaux classiques affirmait la neutralité de la monnaie ; l'imperfection de l'information était juste une petite exception. Ici, l'exception devient tellement importante qu'elle produit le cycle conjoncturel.

LES MODELES RBC

De Lucas, les théoriciens du *cycle économique réel* vont conserver :
- L'environnement d'équilibre général
- Le concept de *choc*
- La forte substituabilité intra- et intertemporelle du travail et du loisir.

Par contre, ils remplaceront les chocs MONETAIRES par des chocs REELS, principalement des chocs technologiques.

Pour les économistes des RBC, la croissance et la conjoncture forment une unité. La conjoncture n'est pas un phénomène indépendant de la croissance qui vient se greffer sur elle, elle est une de ses caractéristiques. Commentant leur modèle, ils le qualifient souvent de « growth model ». Le modèle de croissance de Cass-Koopmans leur sert de socle auquel ils ajoutent une cause de fluctuations.

[1] Lucas [239] p. 23

Kydland et Prescott : le modèle de base

Le premier modèle RBC est présenté dans l'article « Time to Build and Aggregate Fluctuations » de **Kydland** et **Prescott**, publié en 1982. Les termes « time to build » font référence au processus de construction du capital qui s'étale sur plusieurs périodes, comme chez Aftalion.

La technologie s'exprime dans une fonction Cobb-Douglas à rendements d'échelle constants. Comme souvent chez les nouveaux classiques, la dimension économétrique du modèle n'est pas séparée de la dimension théorique, ce qui a pour effet de multiplier variables et coefficients.

Le capital se compose du capital fixe *k,* qui se déprécie à un taux δ*,* et des stocks *y*, dans une proportion σ (0< σ<1) et avec des possibilités de substitutions *v* (0< *v* < ∞). Le progrès technique est exprimé par le facteur λ. La part du travail (en quantité *n*) dans le PNB est θ (0< θ< 1). C'est ainsi qu'on obtient cette fonction de production plutôt complexe (où l'indice *t* est ignoré pour simplifier) :

$$f(\lambda, k, n, y) = \lambda n^{\theta}\left[(1-\sigma)k^{-\nu} + \sigma y^{-\nu}\right]^{-(1-\theta)/\nu} \qquad (8.38)$$

Ce produit se répartit entre la consommation *c* et l'investissement *i*, ce qu'indique l'équation (8.39) :

$$c_t + i_t \leq f(\lambda_t, n_t, k_t, y_t) \qquad (8.39)$$

L'investissement en capital fixe i_t est lui-même la somme de composantes de maturités diverses s_{jt} où *j* (qui varie de 1 à *J*) indique le nombre de périodes avant la mise en opération.

L'utilité (intertemporelle) des ménages, que ceux-ci voudront maximiser, n'est pas plus simple :

$$\sum_{t=0}^{\infty} \beta_t \, u(c_t, \alpha(L) l_t) \qquad (8.40)$$

Un facteur d'escompte β est affecté à l'utilité instantanée ; celle-ci dépend de la consommation c_t et du loisir (non travail) l_t. Mais ce dernier se voit affecté d'un coefficient α(L) qui fait intervenir la substituabilité intertemporelle du loisir. L'idée sous-jacente est que l'utilité du loisir à telle période n'est pas indépendante de la quantité de loisir des autres périodes ; par exemple, un ménage ayant eu beaucoup de loisir la période précédente (pendant laquelle il aura par exemple pu beaucoup bricoler dans sa maison) supportera plus facilement d'avoir peu de loisirs présentement. Le coefficient α(L) est lui-même le résultat d'une équation avec d'autres paramètres, qu'il serait fastidieux d'énoncer ici. Le temps de travail n_t vaut évidemment $1-l_t$.

Le progrès technique comporte une composante non fluctuante λ°, qui nous intéresse peu et une composante fluctuante, elle-même constituée d'une composante permanente λ_{1t} et d'une composante transitoire λ_{2t}. C'est ce qu'indiquent les équations :

$$\lambda_t = \lambda_{1t} + \lambda_{2t} + \lambda^{\circ} \qquad (8.41\text{-A})$$
$$\lambda_{1,t+1} = \rho . \lambda_{1,t} + \varepsilon_{1t} \qquad (8.41\text{-B})$$
$$\lambda_{2,t+1} = \varepsilon_{2t} \qquad (8.41\text{-C})$$
$$\pi_t = \lambda_t + \varepsilon_{3t} \qquad (8.41\text{-D})$$

L'équation (8.41-B) montre comment un choc ε_{1t} maintient ses effets pendant les périodes suivantes, tout s'estompant ; ρ est compris entre 0 et 1, mais selon les auteurs, il est proche de 1, ce qui signifie une persistance élevée. Par contre, les effets du choc ε_2 sont purement ponctuels. Ces deux types de chocs rendent λ aléatoire, mais en outre, leur effet n'est pas observable, du moins pas avec précision : les agents, pour fonder leurs décisions doivent se contenter d'observer, au lieu de λ_t, une variable brouillée π_t. Le troisième choc ε_3 exprime ce brouillage. Les chocs sur lesquels le modèle s'appuie pour produire les fluctuations cycliques

sont donc donnés par les vecteurs $\varepsilon_t = (\varepsilon_{1t},\varepsilon_{2t},\varepsilon_{3t})$, indépendants entre eux et qui suivent une loi normale de moyenne nulle.

A titre de simplification, le modèle ne met pas en scène les firmes qui maximisent leur profit à côté des ménages qui maximisent leur utilité. Toutes les décisions sont prises conjointement ; il faut s'imaginer un planificateur tout-puissant capable de simuler l'équilibre général walrassien. Son but est de maximiser l'espérance mathématique de l'utilité explicitée par la formule (8.40), moyennant les contraintes découlant du contexte exposé ci-dessus (de façon très résumée). Ce faisant, il assure l'équilibre général du système puisqu'en l'absence d'externalités, l'équilibre général et l'optimum de Pareto coïncident. Les individus étant considérés comme identiques (et immortels), l'optimum du ménage représentatif (équation 8.40) correspond à l'optimum social et donc à l'équilibre du système.

Dans sa quête de l'optimum, le planificateur doit prendre à chaque période quatre décisions : n_t, s_{Jt}, y_{t+1} et c_t[1]. Concernant les deux premières, l'information pertinente à sa disposition comprend k_t, s_1 à s_{J-1}, π_t, l'historique de λ avant λ_t ainsi que a_t qui est une sous-variable de $\alpha(L)$. Par contre, c_t et y_t sont décidés un peu plus tard, quand la production de la période *t* est déjà connue et donc λ_t également. Déterminer les quatre règles de décision est un problème de *programmation dynamique*, mais malheureusement, il est impossible de le résoudre analytiquement, car il n'est pas de nature linéaire-quadratique (association d'une fonction d'objectif du second degré avec des contraintes linéaires). Kydland et Prescott vont contourner cette difficulté en construisant une approximation linéaire-quadratique du système au voisinage de l'équilibre stationnaire. Les règles de décision seront déduites des équations de cette économie approximative. Voici comment les auteurs annoncent cette stratégie : « Our approach is to determine the steady state for the model with no shock to technology. Next, quadratic approximations are made in the neighborhood of the steady state. Equilibrium decision rules for the resulting approximate economy are then computed. These rules are linear, so in equilibrium the approximate economy is generated by a system of stochastic difference equations for which covariances are easily determined »[2].

Les auteurs veulent tester si leur modèle est compatible avec les données empiriques de l'économie américaine pendant la période 1950-1979 (118 périodes trimestrielles). La procédure idéale consisterait à voir si à partir de chocs paramétrés de façon réaliste, le modèle est capable de générer un sentier de croissance assez proche de ce que furent les séries chronologiques empiriques de la production, l'emploi etc. Les auteurs se rendent compte que trop d'hypothèses sont posées pour qu'une telle procédure soit fiable La validation empirique du modèle sera donc conduite autrement.

On a affaire à deux environnements : l'environnement modèle et l'environnement empirique. Dans chacun d'eux, on retrouve ces mêmes variables macroéconomiques : la production en termes réels, la consommation, l'investissement, les stocks, la capital installé[3], l'emploi en heures prestées, la productivité et le taux d'intérêt. Chacune de ces variables fait l'objet d'une série chronologique de 118 périodes dans les deux environnements. On reconstitue pour chacune de ces séries sa correspondante sans chocs (« smoothed series »). On peut donc calculer pour chaque période de chaque série l'écart entre la série primaire et la série hors choc.

---

[1] Rappelons que le modèle de croissance néoclassique est monoproduit, en conséquence de quoi *s*, *y* et *c* sont constitués de la même substance.
La substance de s, y et c est la même ; il y a juste une question de répartition.

[2] Kydland & Prescott [205] pp. 1354-1355.

[3] Les biens de consommations durables sont intégrés dans le capital et retirés de la consommation où ils sont remplacés par l'usage fictif de leurs services.

Ces écarts sont la forme sous laquelle le cycle se révèle. Pour chaque variable, on calculera ses caractéristiques statistiques, qui feront l'objet de la comparaison entre les deux environnements :
- la moyenne des écarts (en pourcentage)
- la corrélation de la variable avec la production en termes réels
- autocorrélation (seulement pour la production en termes réels)

Nous avons avancé un peu trop vite dans l'explication. Avant ces traitements statistiques, il générer les séries chronologiques dans l'environnement modèle. Cela se fait par des simulations à partir de combinaisons de paramètres. Le but est de dégager s'il existe une combinaison de paramètres pour laquelle les caractéristiques statistiques des deux environnements sont assez proches.

Un travail essentiel est ce que les auteurs appellent la *calibration du modèle*. Les paramètres sont divisés en deux catégories : les paramètres ESTIMES et les paramètres LIBRES. Les premiers font l'objet d'une estimation sur base de statistiques économiques connues ou d'analyses microéconomiques ; les seconds verront plusieurs valeurs testées dans les simulations. Parmi les paramètres estimés, on trouve notamment $J = 4$, soit un temps moyen d'installation du capital d'un an. La part du travail θ vaut 0,64. D'autres estimations sont la dépréciation du capital δ (10%/an), le taux d'escompte subjectif temporel (β^{-1}-1) de 4%, le temps des ménages partagé entre les activités de marché (n_t) et hors marché (l_t) suivant une clé 1/3+2/3.

Finalement, les paramètres libres sont les variances de $\varepsilon 1$, $\varepsilon 2$ et $\varepsilon 3$ ainsi que la substituabilité entre les stocks et le capital fixe *v*, des sous-paramètres de $\alpha(L)$ que sont α_0 et η et l'aversion du risque γ (qui intervient dans une équation secondaire liée à 8.69). Ils ne sont d'ailleurs pas complètement libres : certaines limites sont fixées ainsi que des relations à respecter, par exemple entre β, γ et le taux d'intérêt. De même, les variances des trois ε doivent assurer que la variance de la production soit égale dans les deux environnements.

Venons-en aux résultats du test. Un ensemble de paramètres a été sélectionné par les auteurs qui estiment que son comportement statistique est certes imparfait mais honorable. Voici en quelques mots le bulletin du modèle :
- l'autocorrélation de la production est correctement simulée
- la variabilité de l'investissement est sensiblement plus forte que celle de la consommation, comme dans la réalité.
- la variabilité des heures prestées est supérieure à celle de la productivité, comme dans la réalité, mais dans une mesure insuffisante.
  Concernant cet aspect du modèle, deux paramètres s'avèrent déterminants pour éviter le résultat absolument erroné où la variation de la production dépendrait plus de celle de la productivité que de celle des heures prestées. D'une part, le délai d'installation du capital doit être suffisant ($J = 1$ n'aurait pas suffi), d'autre part, la substituabilité intertemporelle du loisir (paramètre α) doit être élevée.

Charles **Plosser** est l'un des économistes les plus éminents de l'école RBC. Son article « Understanding Real Business Cycle » (1989) met en exergue quelques caractéristiques de ce type de modèle.

Dans l'otique des RBC, les variables macroéconomiques sont le résultat des actions des agents visant à optimiser leur situation, moyennant contraintes. « As such, the models have an explicit

and firm foundation in microeconomics »[1]. On sait que les microfondations de la macroéconomie sont un thème récurrent chez les *nouveaux classiques*.

Ces modèles sont purement réels. Les variables monétaires n'y jouent aucun rôle.

Tous les biens présents et futurs, y compris le loisir sont considérés comme des *biens normaux* (par opposition aux *biens inférieurs* : cf. supra). En cas d'accroissement TEMPORAIRE de la productivité, un EFFET REVENU poussera à la baisse du travail offert et à l'accroissement de la consommation présente. Toutefois, le travail présent devrait augmenter, car l'effet revenu sera dominé par un EFFET DE SUBSTITUTION poussant l'individu à profiter du moment propice pour travailler plus et investir. Si l'accroissement de la productivité est PERSISTANT, la consommation et le loisir présent l'emporteront sur le travail et l'investissement, car il n'y a plus d'effet de substitution

A certaines périodes, la mauvaise conjoncture se manifeste par une baisse de la production et des heures travaillées. La cause en est que les agents ont ajusté leur comportement à un choc exogène. « It is important to stress that there are no market failures in this economy, so Robinson Crusoe's response to the productivity shifts are optimal and the economy is Pareto-efficient at all points in time »[2].

La croissance et le cycle conjoncturel constituent un seul et même processus ; c'est donc une erreur d'avoir deux théories pour les expliquer. L'évolution de la production par tête suit un chemin aléatoire, « which means that it possesses no affinity for any particular mean. Random walks are also referred to as stochastic trends because while they may exhibit growth, they do not fluctuate about any particular deterministic path. If shocks to productivity are permanent, each one determines a new growth path »[3]. Aucune tendance au retour ne suit nécessairement un choc technologique. Plosser a coécrit en 1982 avec Nelson une étude décortiquant mathématiquement les statistiques macroéconomiques américaines (1860-1970) qui démontre avec une probabilité très élevée, que les principales séries chronologiques ignorent tout trend déterministe.

Pour que les modèles RBC soient cohérents, le progrès technique doit refléter un caractère *neutre-Harrod*.

Dans son article « Theory Ahead of Business Cycle Measurement » (1986), **Prescott** commente également la théorie des RBC.

Comme l'ont montré Cass et Koopmans, l'équilibre concurrentiel dynamique converge vers un sentier de croissance stable lorsque la technologie connaît un accroissement exponentiel (ce qui est le cas si le taux annuel est constant). On a besoin de chocs sur le paramètre technologique pour expliquer le cycle économique.

Les modèles RBC mettent en avant les chocs technologiques. Il est donc souhaitable de disposer d'un instrument de mesure empirique de ces chocs. Prescott estime que le *résidu de Solow* convient très bien à cette fin. Il en examine les caractéristiques statistiques pour l'économie américaine entre 1955 et 1984 (données trimestrielles). Dans cette analyse, il estime

[1] Plosser [287] p. 53
[2] Plosser [287] p. 56. Robinson Crusoé personnifie l'agent représentatif dans cet article.
[3] Plosser [287] p. 59.

trouver la confirmation que les chocs technologiques suivent un chemin aléatoire et font montre d'une persistance élevée.

Prescott calcule un écart-type de 0,763% pour les chocs technologiques ainsi évalués. Selon le modèle Kydland-Prescott, un choc de cette importance entraîne un écart-type de la production de 1,48, ce qui se révèle inférieur à la valeur de 1,76% constatée empiriquement. Une telle variabilité de la production nécessiterait des chocs technologiques sensiblement supérieurs. Cette différence résulte de la sous-estimation de la variabilité des heures prestées dans le modèle, déjà mentionnée.

Prescott conclut son article avec des recommandations de politique économique. « The policy implication of this research is that costly efforts at stabilization are likely to be counterproductive. Economic fluctuations are optimal responses to uncertainty in the rate of technological change ». Certes, cela ne signifie pas que le taux de progrès technologique est nécessairement optimal. Peut-être peut-il même être accru par une politique économique adéquate. Dans cette optique, il est souhaitable que la théorie économique améliore sa compréhension des facteurs déterminant le taux de progrès technique. « If policies adopted to stabilize the economy reduce the average of technical change, then stabilization policy is costly »[1].

Modèles améliorant le réalisme

Nombre de modèles ont été construits pour tenir compte de tel out el aspects de la réalité mal pris en compte par le modèle de base.

La première amélioration concerne le *co-mouvement*, c'est-à-dire le fait que la plupart des variables macroéconomique, en tête desquelles la production, varient transversalement à travers les différents secteurs de l'économie. A l'opposé, les chocs technologiques, la plupart du temps, ne sont pas transversaux, mais propres à une industrie ou un petit nombre d'entre elles. Il faut donc expliquer comment un choc local peut se propager à quasiment toute l'économie. Le modèle de base, monoproduit, ne pouvait expliquer ce phénomène.

Le titre de l'article « Real Business Cycles » de **Long et Plosser** (1983) est la première apparition de cette appellation. Les auteurs y abordent la problématique du co-mouvement dans un modèle multiproduits, que nous nous contenterons de survoler.

La stratégie des auteurs diffère de celle de Kydland et Prescott ; elle consiste à montrer par des simulations que le modèle RBC est capable de générer des séries chronologiques où se manifestent les deux caractéristiques essentielles du cycle conjoncturel que sont le co-mouvement et la persistance. Long et Plosser préfèrent introduire des hypothèses irréalistes dans leur modèle pour rendre le problème des règles de décision linéaire-quadratique ; c'est ainsi qu'est postulée une dépréciation annuelle du capital de 100%. Diverses complications présentes chez Kydland-Prescott, comme le « time to build », les stocks, sont ignorées ici. Par contre, l'injection de chocs technologiques aléatoires et l'utilité intertemporelle à maximiser sous contrainte de la fonction de production sont communs à la plupart des modèles RBC.

Il y a N secteurs de production et donc N produits. La consommation C_t, l'emploi en heures L_t, les chocs technologiques λ_t et la production Y_t sont à présent des vecteurs à N éléments. Chacun des N biens a de nombreux usages. Toute production d'un bien recourt à d'autres biens comme

[1] Prescott [288] p. 21

inputs et tout bien, outre son éventuel usage pour la consommation, intervient dans la production de nombreux autres biens.

Le capital n'intervient pas comme variable agrégée dans la fonction de production, mais y est remplacé par tous les inputs individuels qui le composent. La fonction de production s'écrit :

$$\mathbf{Y}_{t+1} = F(\mathbf{L}_t, \mathbf{X}_t, \boldsymbol{\lambda}_{t+1}) \qquad (8.42)$$

La nouveauté dans cette fonction est la matrice $\mathbf{X}_t$ de dimension N x N, comportant les éléments x_{ijt} qui renseignent la quantité du bien *j* allouée à la production du bien *i* au temps *t*. Il ne faut pas l'interpréter comme une matrice de coefficients de production fixes, mais simplement des coefficients résultant des décisions prises au temps *t.* Les hypothèses du modèle vont justement dans le sens d'une très grande substituabilité entre les inputs. Autrement dit, il y a beaucoup d'éléments non nuls dans la matrice.

Dans le cas d'une fonction Cobb-Douglas à rendements d'échelle constants, l'équation (8.42), pour le produit *i*, pourrait s'écrire :

$$Y_{i,t+1} = \lambda_{i,t+1} L_{it}^{b_i} \prod_{j=1}^{N} x_{ijt}^{a_{ij}}, \qquad i = 1, 2, \ldots, N \qquad (8.43)$$

Où b_i et les a_{ij} sont les exposants correspondant aux θ et 1-θ de Kydland-Prescott ; leur somme doit valoir un ; ils représentent la part du coût du produit *i* que représente le travail *L* ou l'input *j*.

Sur cette base, les auteurs déduisent les règles de décision relatives à la consommation et au loisir et bien-sûr à la répartition de la consommation, du travail et de l'investissement entre les *N* industries. Observant ces règles, ils constatent : « The amounts of a commodity (or time) allocated to each of its productive employments and to positively valued consumption are all increasing functions of the total available amount of the commodity (or time) (…) It implies that if the output of commodity *i* is unexpectedly high at time *t*, then inputs of commodity *i* in all of its employments will also be unexpectedly high at time *t* »[1]. On peut se demander si une telle souplesse n'est pas excessive par rapport à la réalité économique.

La variable **λ** n'a pas les mêmes propriétés que chez Kydland-Prescott ; elle est plus abstraite. Elle vise simplement de soumettre la production à des chocs sans nécessairement simuler le comportement qu'on attribue au progrès technique : le progrès technique de Kydland-Prescott est un type particulier de choc aléatoire, mais pour les besoins de la cause, les auteurs privilégient ici un type encore plus aléatoire. Les valeurs de $\boldsymbol{\lambda}_t$ à travers le temps sont indépendantes entre elles ; leur distribution probabiliste est immuable. Le but est d'assurer que les co-mouvement et la persistance que les auteurs espèrent générer soient imputables au système interne de propagation (dont la matrice $\mathbf{X}_t$ est le cœur) et non à d'éventuelles corrélations dont feraient preuve les chocs.

Pour éclairer le co-mouvement, les auteurs ont divisé l'économie en six secteurs : l'agriculture, les mines, la construction, la manufacture, le secteur du transport et du commerce ainsi que les services. A partir des statistiques de l'économie américaine de l'année 1967, ils ont calculé les a_{ij} de la formule (8.43). Ces a_{ij} sont disposés en une matrice appelée **A**. De la fonction de production et des règles de décision, les auteurs déduisent :

[1] Long & Plosser [233] pp. 48-49.

$$\mathbf{y}_{t+1} = \mathbf{A}.\mathbf{y}_t + \mathbf{k} + \boldsymbol{\eta}_{t+1} \qquad (8.44)$$

où **y** et **η** sont les logarithmes respectivement de la production **Y** et des chocs technologique **λ** et où **k** est une constante liée aux préférences. Toutes ces variables sont des vecteurs sauf **A** qui est une matrice. Cette équation montre que si un produit augmente fortement de $\mathbf{y}_{t-1}$ à $\mathbf{y}_t$, la production des biens pour lesquels il est un input important haussera de $\mathbf{y}_t$ à $\mathbf{y}_{t+1}$..

Long et Plosser soumettent chacun des six secteurs à une impulsion unitaire pour en observer la propagation. Un développement de l'équation (8.44) leur permet de pister l'effet en provenance de chaque secteur vers chacun des autres. Ils observent ainsi que le secteur qui subit l'impulsion connait habituellement une fluctuation plus ample que les autres ; les autres secteurs réagissent avec un certain retard, de façon progressive sur deux ou trois périodes. Les secteurs miniers, agricoles et manufacturier ont connu des fluctuations plus amples que les autres, comme dans l'économie réelle. Les résultats en matière de propagation et de persistance s'avèrent satisfaisants.

Quoi que leur modèle génère une fluctuation cyclique de la variable travail (n_t chez Kydland-Prescott et L_t chez Long et Plosser), le chômage en est le grand absent. La faiblesse de l'emploi qui s'y manifeste en basse conjoncture ne signifie pas qu'une partie des salariés est déclassée mais que tous réduisent le nombre d'heures de travail.

Dans son article « Indivisible Labor and the Business Cycle » (1985), **Gary Hansen** commence par cette constatation: "Equilibrium theories of the business cycle, such as Kydland ant Prescott (1982) and Lucas (1977), have been criticized for failing to account for some important labor market phenomena. These include the existence of unemployed workers, fluctuations in the rate of unemployment, and the observation that fluctuations in hours worked are large relatively to productivity fluctuations"[1]. Il observe également que l'hypothèse d'une substituabilité intertemporelle du loisir élevée est contredite par des études microéconomiques.

Hansen veut montrer qu'il est possible de construire un modèle RBC qui intègre ces aspects de la réalité et qu'un tel modèle se comporte mieux du point de vue statistique. Dans le modèle de Hansen, le travail est *indivisible*, c'est-à-dire que chaque employé doit travailler le nombre d'heures qui est la norme ; la seule alternative est de ne pas travailler. La perspective est binaire : travail à temps plein ou chômage complet.

Hansen justifie son hypothèse de l'indivisibilité. D'une part, statistiquement. Selon certaines études, 55% de la variance de la quantité de travail agrégée serait due à la variation du nombre de travailleurs employés, 20% à la variation des heures de travail[2] par employé et le reste à la covariance.

Du point de vue théorique, deux explications sont possibles :
- Sur le plan de la DEMANDE : la productivité marginale du travail d'un individu serait croissante les premières heures avant de décroître. La fonction de production serait alors partiellement convexe, partiellement concave.
- Sur le plan de l'OFFRE : les coûts fixes liés au travail (longs déplacements par exemple) peuvent inciter l'individu à préférer les situations extrêmes de travail à temps plein ou de

[1] Hansen [130] p. 309
[2] Selon la terminologie en vigueur, il y a *ajustement à la marge intensive* lorsque la quantité de travail s'ajuste par les heures de prestation ; il y a *ajustement à la marge extensive* lorsqu'elle s'ajuste par le nombre de personnes employées.

non-travail à une situation intermédiaire. Il s'agit d'une *non*-convexité dans les préférences. C'est cette situation que postulera le modèle de Hansen.

L'article présente deux modèles RBC en tous points identiques, sauf que dans le premier, le travail est divisible alors qu'il est indivisible dans le second. Le premier est dans la ligne de ce que nous avons vu ci-avant. Nous limiterons ici l'exposé au second dans ce qu'il a de spécifique ; nous verrons ensuite la comparaison de leurs performances.

D'emblée, une difficulté surgit. Il s'agit d'un modèle d'équilibre général. Mais nous savons que l'existence de cet équilibre nécessite la convexité de l'ensemble des consommations possibles (y compris celle de loisir). Or l'indivisibilité du travail contrevient à cette convexité. La difficulté est contournée par un artifice : les transactions portent sur des loteries. Le contrat de travail, c'est une probabilité α de travailler une fraction h_0 ou $1-\alpha$ de ne pas travailler. Le choix du ménage porte sur la valeur de α, mais comme tous les ménages sont supposés identiques, une valeur unique de α émerge du marché. A chaque période, un tirage au sort départage ceux qui travailleront et ceux qui chômeront. Une autre caractéristique étonnante de ce marché est que tous, chômeurs comme travailleurs, percevront le même salaire. Hansen remarque que son système revient au même que la situation pas inhabituelle où les entreprises ne payeraient que les heures réellement prestées et où tous les salariés contracteraient une assurance individuelle complète contre le chômage. Hansen démontre que dans ce cas, contracter cette assurance serait le choix normal de tous les intéressés.

Hansen insiste sur un résultat particulier de son modèle : l'élasticité de substitution intertemporelle du loisir est infinie au niveau agrégé, peu importe la valeur de cette élasticité chez les individus qui composent l'économie. Ceci permet au modèle de renoncer à cette hypothèse contestée que les ménages ont une élasticité de substitution élevée à cet égard.

Hansen opère des simulations avec ses deux modèles pour générer des séries chronologiques ; il compare ensuite les caractéristiques statistiques des séries, d'une part entre les deux modèles et d'autre part avec ces mêmes caractéristiques des données réelles de l'économie américaine. A part la dualité des modèles testés, la démarche est proche de celle de Kydland et Prescott.

Les deux principales constations sont :
- L'économie à travail indivisible connaît des fluctuations de magnitude supérieure à celles de l'économie à travail divisible. Elles sont toutefois encore inférieures à celles de l'économie réelle, mais Hansen pense que ce dernier écart pourrait être dû à des erreurs de mesure.
- Chez Kydland et Prescott, les variations des heures travaillées étaient trop faibles par rapport à celles de la productivité. Cette faiblesse est confirmée dans le présent modèle à travail divisible. Par contre, le rapport de la variation des heures travaillées sur la variation de la productivité est sensiblement supérieur dans l'économie à travail indivisible. En fait ce ratio pour l'économie réelle se situe entre les valeurs dans les deux économies artificielles. Selon l'auteur, ceci est aisément explicable : l'économie réelle combine les ajustements à la marge intensive (heures supplémentaires, temps partiel) et à la marge extensive. Idéalement, le modèle devrait également les associer.

Dans leur article « Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations » (1992), **Christiano et Eichenbaum** s'attaquent à une autre lacune de ces modèles : ils surestiment largement la corrélation positive[1] entre le volume des heures

[1] Les auteurs font référence à Dunlop et Tarshis qui avaient amené à corriger le sens de cette corrélation. Ici, l'ampleur de la corrélation est revue à la baisse.

travaillées et le salaire réel (ainsi que la productivité du travail). Alors que les études empiriques révèlent une corrélation proche de zéro. Le salaire n'apparaît certes pas parmi les variables macroéconomiques testées par Kydland et Prescott, mais leur modèle rend la productivité très procycliques ; or le salaire réel est censé s'aligner sur la productivité marginale.

Trop centrés sur la technologie, les modèles RBC précédents ne faisaient varier l'emploi (en heures ou en hommes) que par le seul canal de la demande de travail, correspondant à la courbe de productivité marginale. Lorsqu'un choc positif élève cette courbe, elle rencontre la courbe d'offre de travail à la fois plus haut (salaire) et plus à droite (emploi). Si les variations de la demande de travail étaient accompagnées de variations de l'offre allant dans le même sens, les fluctuations de l'emploi seraient amplifiées et celles du salaire modérées. Pour introduire ces variations de l'offre, les auteurs ont besoin de l'irruption d'un autre type de chocs qui se combinent avec les chocs technologiques ; ils envisagent plusieurs « candidats » : changements dans la fiscalité, changements démographiques, innovations dans l'offre de monnaie, dépenses gouvernementales erratiques. C'est ce dernier type de choc que Christiano et Eichenbaum retiennent.

La contribution de Christiano et Eichenbaum s'inscrit dans la droite ligne de celles de Hansen, Kydland et Prescott. La particularité essentielle est l'intervention d'une variable g_t représentant la consommation publique (à l'exclusion des investissements publics). Les chocs technologiques demeurent ; il y a donc deux sources de chocs au lieu d'une.

L'effet des dépenses publiques sur la production et l'emploi s'inspire d'une conception alambiquée et peu intuitive exposée par Barro dans un article en 1981. Comme les modèles précédents, celui-ci se place dans la peau d'un planificateur qui maximise une fonction d'utilité moyennant la contrainte de la fonction de production. L'utilité augmente avec les services de la consommation ct et diminue avec la perte de loisir occasionnée par le travail. Les services de la consommation sont explicités par l'équation :

$$c_t = c_{pt} + \alpha g_t \qquad (8.45)$$

La variable c_{pt} représente la consommation privée directe et le paramètre $1 \leq \alpha < 1$ représente l'impact des dépenses publiques g_t sur l'utilité marginale de la consommation privée. Si $\alpha = 1$, consommations privée et publique sont de parfaits substituts ; si $\alpha = 0$, la consommation publique obère intégralement les services de la consommation. Probablement, α est compris entre 0 et 1. Les modèles RBC vus précédemment ne reconnaissaient aucun effet aux dépenses publiques ; implicitement, cela revenait à présumer que $\alpha = 1$. Lorsque $\alpha < 1$, "agents respond to an increase in government consumption as if they had suffered a reduction in their wealth"[1]. Ce qui incitera à travailler plus puisque le loisir est un *bien normal*. Voilà comment les dépenses publiques affectent positivement la production et l'emploi.

Comme le montrent les équations (8.46-A) et (8.46-B), les chocs des dépenses gouvernementales sont aléatoires, mais sont caractérisés par de la persistance et par une relation avec les chocs technologiques qui permet une certaine synchronisation :

$$g_t = g_t^{\circ}.\lambda_t \qquad (8.46\text{-A})$$
$$\ln(g_t^{\circ}) = (1-\rho).\ln(g^{\circ}) + \rho.\ln(g^{\circ}_{t-1}) + \mu_t \qquad (8.46\text{-B}) \text{ où:}$$

- λ_t représente la technologie dans la fonction de production (neutre-Harrod) et inclut les soubresauts habituels malgré une certaine persistance,

[1] Christiano & Eichenbaum [49] p. 435.

- $g_t°$ représente les dépenses gouvernementales pures, débarrassées de cette influence,
- $g°$ est la moyenne des g_t
- μ_t représente l'« innovation » dans les dépenses gouvernementales.

Les auteurs écrivent : "With this specification, the factors giving rise to permanent shifts in government consumption are the same as those that permanently enhance the economy's productive ability"[1].

Christiano et Eichenbaum testent quatre modèles théoriques pour déterminer lequel s'approche le plus de la réalité statistique : travail divisible et α nul, travail divisible et α unitaire, travail indivisible et α nul et travail indivisible avec α unitaire. Il s'avère que le passage de $\alpha = 1$ à $\alpha = 0$ améliore la performance prédictive du modèle. Notamment, la volatilité de l'emploi est gonflée, dans une même proportion que par le passage du travail divisible au travail indivisible. « Nevertheless, even when both effects are operative, the model still underpredicts the volatility of n_t relative to y_t »[2]. La corrélation dérangeante entre la productivité et l'emploi est réduite lorsqu'$\alpha = 0$, alors que l'indivisibilité du travail n'y remédie pas. Mais cette réduction est insuffisante pour éviter la surestimation, qui- selon les auteurs- serait le talon d'Achille des modèles RBC.

La théorie du cycle économique réel s'enrichit de nombreux modèles qui intègrent les facteurs évacués par les simplifications propres au modèle de base. Notre analyse se limite aux extensions principales. En voici une dernière : la monnaie était absente de tous les modèles étudiés. Dans leur article « Money, Credit, and Prices in a Real Business Cycle » (1992), **King et Plosser** notent : "These models are often viewed as incomplete or wrong because they do not generate the widely emphasized, but not easily explained, correlation between the quantity of money and real activity. This paper integrates money and banking into real business cycle theory"[3]. Voyons d'abord comment ils ont complété le modèle standard et ensuite les résultats statistiques.

A côté du secteur industriel, toujours sujet à des chocs technologiques aléatoires, opère un secteur financier doté d'une fonction de production propre, qui a un prix, un coût et des inputs propres. L'output de ce secteur, la gestion des transactions, est un input supplémentaire dans le secteur industriel. Il entre également comme variable dans la fonction d'utilité des ménages. Le temps dont disposent ceux-ci ne se décompose plus seulement en loisir et travail ; il y a aussi le temps consacré à la gestion, que les services offerts par le secteur financier ont précisément pour objet de réduire et qui doit être optimisé. A côté des intermédiaires financiers qui offrent leurs services, le gouvernement émet une monnaie fiduciaire qui leur est un substitut (certes imparfait). Outre fonction de gestion des paiements, le secteur financier prête aux entreprises et se finance par des dépôts ; les ménages répartissent leur épargne entre les dépôts bancaires et l'investissement direct. Les auteurs recourent à nombre de simplifications qu'il serait fastidieux d'énumérer ici ; je mentionne toutefois celle qui fait des dépôts (en valeur réelle) une fonction linéaire du volume des transactions.

Les auteurs n'introduisent la monnaie fiduciaire dans le modèle que dans une deuxième phase, avant laquelle les transactions ne se pratiquent que de compte à compte en valeur réelle. Dès cette phase préalable, les auteurs constatent que les fluctuations de la production entraînent dans

[1] Christiano & Eichenbaum [49] p. 434
[2] Christiano & Eichenbaum [49] p. 442. Les variables n_t et y_t sont respectivement l'emploi (en heures) et la production.
[3] King & Plosser [188] p. 363

leur sillage celles du crédit et des transactions et, partant, celles des dépôts. Cette constatation est importante pour les conclusions qui seront tirées du modèle.

La fonction de demande de monnaie (fiduciaire) est une demande d'encaisse réelle :

$$l_t = l_t(y_t, R_t, \rho_t, w_t) \qquad (8.47)$$

Elle est entièrement déduite du processus d'optimisation par les ménages ; elle augmente avec le revenu réel y_t, générateur de transactions, diminue avec l'intérêt des obligations R_t, augmente avec le coût des opérations bancaires ρ_t ainsi qu'avec le salaire réel w_t, coût d'opportunité du temps consacré aux transactions.

« For clarity, the bulk of the discussion is conducted under the assumption that the treasury-central bank maintains a policy of controlling the issue of nominal currency so that the stock of currency ($C_t = P_t c_t$) is an exogenous random variable »[1]. Dans ce cas,
le niveau général des prix est déterminé par l'égalisation de l'offre de monnaie avec la demande :

$$P_t = C_t / l_t \qquad (8.48)$$

Le modèle attribue à un choc technologique positif imprévu le résultat d'accroître la production et- plus étonnamment- de diminuer le rendement réel des actifs. Quoi qu'il en soit, ces deux facteurs, via les équations 8.47 et 8.48 concourent à accroitre la demande de monnaie et donc à diminuer le niveau des prix. Le tableau est donc celui d'un niveau des prix contracyclique. Les auteurs reconnaissent le caractère contestable de ce résultat et suggèrent même les moyens de résoudre la contradiction :

- Paramétrer les chocs technologiques de telle façon qu'avec une persistance plus élevée, les chocs positifs améliorent le rendement des actifs
- Supposer que les autorités mènent une politique tendant à accroître l'offre de monnaie de base lorsque le revenu augmente.

King et Plosser testent empiriquement leur modèle. A propos de la corrélation entre agrégats réels et agrégats nominaux, leurs conclusions sont :

- Dans l'économie réelle, le revenu réel est plus corrélé avec la monnaie INTERNE qu'avec les autres agrégats monétaires. Ceci cadre parfaitement avec le modèle, compte tenu du rôle important qu'il attribue aux intermédiaires financiers.
- La corrélation entre la monnaie EXTERNE et le revenu réel est certes plus faible mais pas négligeable. Ceci contredit le modèle et met en doute le bien fondé de l'hypothèse que l'émission de monnaie fiduciaire est autonome.

Pour tester l'explication des prix dans leur modèle, les auteurs développent une équation plus complexe à partir de 8.81 et 8.82 sous une forme log-linéaire ; ils en calculent les coefficients par une régression sur les données empiriques. Les coefficients obtenus (et notamment leur signe) ne semblent pas contredire les relations du modèle. L'impact négatif du niveau d'activité sur le niveau des prix se trouve ainsi confirmé indirectement, ce qui paraît assez douteux.

En conclusion, le niveau général des prix est corrélé avec le volume de monnaie EXTERNE alors que le niveau d'activité est corrélé avec le stock de monnaie INTERNE. Mais les sens des deux déterminations sont opposés. La masse monétaire externe DETERMINE le niveau des prix. Le

[1] King & Plosser [188] p. 369. Les auteurs analyseront ensuite une structure financière plus complexe et plus proche de la réalité distinguant les banques des autres intermédiaires financiers et combinant monnaie de base, monnaie interne et monnaie externe ainsi que le contrôle indirect du stock monétaire au moyen de réserves obligatoires.

volume de la monnaie interne EST DETERMINE PAR l'activité. King et Plosser semblent accepter le caractère endogène de la monnaie interne. Comme ils l'annoncent dans l'introduction, leur article aboutit à "a class of models that can account for the correlation between money and business cycles in terms that most economists would label reverse causation"[1].

LES CRITIQUES CONTRE LES MODELES RBC

Critiques générales

Les *nouveaux keynésiens* sont évidemment d'ardents critiques des modèles du *cycle économique réel*. Dans son ouvrage « Macroéconomie approfondie », **David Romer** note les lacunes suivantes :

- Les chocs technologiques sont l'élément essentiel. Mais il est généralement difficile d'isoler dans la réalité les innovations à la base de ces chocs. Romer l'illustre par la grande dépression des années trente, dont on ne voit pas quel choc négatif pourrait en être l'origine.

- Le *résidu de Solow* ne semble pas être un bon indicateur des chocs technologiques, car des études montrent qu'il est corrélé à maints facteurs non liés à la technologie.
  A cette critique, Rebelo, partisan des RBC, répond que certains modèles RBC se sont préoccupés de ce problème : "Introducing mechanisms such as capacity utilization and markup variation in RBC models has two effects. First, these mechanisms make true technology shocks less volatile than TFP. Second, they significantly amplify the effects of technology shocks. This amplification allows models with these mechanisms to generate output volatility similar to the data with much smaller technology shocks"[2].

- Une substitution intertemporelle du travail très élastique est au centre du mécanisme de propagation. Les études empiriques ne confirment pas l'ampleur conférée à cette variable par les modèles RBC.

- Les modèles RBC impliquent la neutralité de la monnaie.

- Le *calibrage* pose question. Les modèles sont trop élastiques pour que leur test empirique soit parfaitement crédible. Il y a trop de liberté dans le choix des paramètres et des formes fonctionnelles. Les équations peuvent s'adapter trop facilement aux données empiriques[3].

L'article « Real Business Cycles : A New Keynesian Perspective » (1989) de **Mankiw** sort l'artillerie lourde. On y retrouve des critiques mentionnées ci-avant, notamment le mystère entourant les chocs technologiques : "Advocates of real business cycle theories have trouble convincing skeptics that the economy is subject to such large and sudden changes in technology. It is a more standard presumption that the accumulation of knowledge and the concurrent increase in the economy's technological opportunities take place gradually over time"[4]. Mankiw reconnaît que le résidu de Solow est à la fois très variable et procyclique. Mais il en impute la cause à des comportements hors fonction de production, comme la *rétention de main d'œuvre*. Or, sans le résidu de Solow, il n'y a pas d'indice de chocs technologiques. C'est

---

[1] King & Plosser [188] p. 363

[2] Rebelo [296] p. 222. « TFP » signifie *total factor productivity*, qui sert fréquemment de synonyme au *résidu de Solow*.

[3] La très paradoxale confirmation empirique par King et Plosser que le niveau de prix serait contracyclique, n'illustrerait-elle pas cette faiblesse ?

[4] Mankiw [251] p. 83

encore plus vrai des chocs négatifs ; l'observateur attentif qu'est Mankiw n'en décèle nulle trace.

Que les ménages choisissent de varier régulièrement leur offre de travail est également douteux. Comment expliquer des fluctuations aussi considérables alors que celles du salaire, facteur de motivation, sont très modérés ? Face à une variation temporaire du salaire, le ménage devrait redistribuer son loisir dans le temps de façon hyper-réactive. C'est à la fois peu plausible et non corroboré par les études sur ce sujet. En outre, la planification optimale d'une action impliquant un revenu doit nécessairement tenir compte du taux d'intérêt. L'introspection suffit à se convaincre que nous établissons notre offre de travail individuelle sans nous en soucier. "While economists can easily convince laymen and students that the quantity of apples demanded depends on the price of apples, it is much harder to convince them that labor supply depends on the real interest rate"[1].

Mankiw s'en prend également au caractère Pareto-optimal du cycle dans toutes ses phases: "Of all the implications of real business cycle theories, the optimality of economic fluctuations is perhaps the most shocking. It seems undeniable that the level of welfare is lower in a recession than in the boom that preceded it"[2]. L'idée que cette régression soit imputable au déclin des capacités technologiques est peu crédible. Puisque les modèles RBC considèrent l'optimalité de l'économie comme permanente, toute tentative de l'Etat de mener une politique stabilisatrice est vouée à l'inefficacité ; l'entrave à la main invisible ferait même empirer la situation. Ce qui fait dire à Mankiw : « Moreover, to the extent that it trivializes the social cost of observed fluctuations, real business cycle theory is potentially dangerous ».

Le modèle intersectoriel de Long et Plosser est également visé. Les chocs technologiques sectoriels ne peuvent survenir dans beaucoup de secteurs simultanément ; sinon, par la loi des grands nombres, il se neutraliseraient mutuellement. Pour contrer cet argument, certains défenseurs des RBC avancent que les déplacements de la main d'œuvre d'un secteur subissant un choc négatif vers d'autres secteurs nécessitent un délai d'ajustement. Le chômage caractéristique de la récession serait alors une vague de chômage frictionnel causé par un afflux de chocs négatifs, que les chocs positifs n'épongeront qu'avec retard. Mankiw relève deux constatations empiriques qui contredisent ce scénario ;

- Que le chômage résulte du souhait de changer d'emploi suppose que son pic coïncide avec celui des offres d'emploi insatisfaites ; or l'inverse est vrai.
- Empiriquement, la mobilité intersectorielle du travail s'avère fortement procyclique.

Controverse relative aux effets des chocs technologiques sur l'emploi

En 1999 paraît un article qui s'est révélé particulièrement influent : « Technology, Employment and Business Cycle : Do Technological Shocks Explain Aggregate Fluctuations ? » L'auteur, Jordi **Gali**, nouveau keynésien, conteste les résultats traditionnels des modèles RBC, tout particulièrement à l'aide de cette idée neuve : loin d'augmenter les heures ouvrées, les chocs technologiques leur imposent une baisse temporaire, suivie d'un retour graduel à la normale. L'article comporte deux parties : un modèle théorique aboutissant à ce résultat et une analyse des données empiriques qui le conforte et contredit donc les conclusions des études RBC.

Voyons d'abord le modèle théorique ; il inclut des éléments absents des modèles RBC mais chers aux nouveaux keynésiens. L'économie est soumise à la fois à des chocs technologiques et des chocs monétaires. On est en concurrence monopolistique : il y a une entreprise par produit.

[1] Mankiw [251] p. 86
[2] Mankiw [251] p. 83

La rigidité nominale des prix est évidemment présente : les firmes fixent leur prix avant la survenance des chocs et ne l'adaptent plus. Le travail fait l'objet d'une double variable : le nombre d'heures et le niveau d'effort, ayant chacune sa rémunération propre. Le rôle de la variable *effort* est de permettre une variation de la productivité en réponse à un choc non technologique. L'effort apparaît comme un facteur dans la fonction de production et comme une source de désutilité dans la fonction de bien-être. Pour le choc technologique et le choc monétaire, on a respectivement les équations :

$$\lambda_t = \lambda_{t-1}.E(\eta_t) \qquad (8.49\text{-A})$$
$$M_t = M_{t-1}.E(\varepsilon_t + \gamma\eta_t) \qquad (8.49\text{-B})$$

La variable λ_t est l'index technologique habituel de la fonction de production et η_t est son taux d'accroissement par rapport à la période précédente, qui suit une *loi normale*. L'évolution de la masse monétaire *M* comporte une composante aléatoire ε mais tente également de satisfaire les besoins supplémentaires générés par η, avec une pondération γ. Je simplifie beaucoup l'exposé et à ce titre, je me permets de ne pas retranscrire la fonction de bien-être ni la fonction de production. En manipulant ces fonctions et d'autres qui les complètent, Gali déduit pour chacune des quatre variables macroéconomiques, *p* (prix), *y* (production), *n* (heures) et *x* (productivité), une équation de réaction (à l'équilibre) où elle est fonction à la fois de ε et η. Les coefficients de ε et η dans ces équations, et tout particulièrement leur signe, dépendent évidemment des paramètres des équations de base : fonction d'utilité, fonction de production et chocs.

Gali combine plusieurs de ces paramètres en une variable φ, plutôt abstraite, qu'il caractérise ainsi : « short run increasing return to labor »[1].

L'interprétation des équations de réaction nous apprend que :
- Un choc MONETAIRE positif non anticipé a un effet définitif : l'accroissement de *p*. Il a également un triple effet transitoire : il pousse *y* et *n* à la hausse. Quant à *x*, il l'augmente si φ >1.
- Un choc TECHNOLOGIQUE positif entraîne une hausse permanente de *y* et de *x*. Le niveau des prix *p* baissera de façon permanente si le facteur d'adaptation de la masse monétaire γ n'est pas trop élevé. Quant à *n*, dans les mêmes conditions, il subira une baisse transitoire.

Ce dernier résultat, le plus novateur, mérite un commentaire éclairant sa rationalité. Si la monnaie est fortement exogène et si le niveau des prix est resté inchangé, les encaisses réelles ne varient pas et la demande non plus ; les entreprises ne modifient donc pas leur production ; mais avec une productivité accrue, moins de main d'œuvre suffit.

En calculant les covariances entre Δy et Δn ainsi qu'entre Δy et Δx, Gali conclut que *n* est procyclique pour autant que le choc monétaire soit suffisamment important par rapport au choc technologique ; *x* est également procyclique pour autant que φ >1. Il s'attaque ensuite à la corrélation entre Δx et Δn, de l'aveux-même de Christiano et Eichenbaum, le talon d'Achille des RBC. Ici, selon Gali, la *covariance inconditionnelle* habituellement utilisée dans les modèles ne permet pas de voir clair ; il recourt à ce qu'il appelle les *covariances conditionnelles* ; il y en a deux : celle qui envisage le choc technologique comme unique impulsion et celle qui envisage le choc monétaire comme unique impulsion. Gali pense qu'une estimation réaliste situe φ entre un et deux et γ entre zéro et un. Dans ces conditions, les covariances conditionnelles du modèle de Gali concluent :
- Choc technologique : corrélation négative entre *x* et *n*

[1] Gali [114] p. 253

- Choc monétaire : corrélation positive entre *x* et *n*.

Ce résultat contredit diamétralement celui des théories du cycle économique réel ; Christiano et Eichenbaum avaient ajouté les chocs non technologiques, précisément pour abaisser la corrélation positive excessive produite par les chocs technologiques.

La deuxième partie de l'article traite les données réelles (Etats-Unis 1948-1994). Afin de déterminer lequel des modèles elles valident, les statistiques doivent être traitées pour démêler les parts respectives des chocs technologiques et des chocs non technologiques dans le co-mouvement des heures et de la productivité à analyser : « My approach involves the use of a structural VAR model, identified by the restriction that only technology shocks may have a permanent effect on the level of productivity »[1]. Gali pense que des chocs technologiques ne peuvent pas être transitoires.

Le traitement statistique est assez complexe et je passe sous silence les hypothèses à caractère technique ; ce sont les résultats qui nous intéressent. Dans le modèle à deux variables (avec les heures non détrendées), la corrélation entre les variations des heures et celles de la productivité fait l'objet des estimations suivantes :

- Corrélation inconditionnelle : -0,26
- Corrélation conditionnelle à un choc technologique : -0,82
- Corrélation conditionnelle à un choc non technologique : 0,26

Ce résultat confirme les prédictions du modèle keynésien de Gali et infirme celles de Christiano et Eichenbaum. Les données relatives aux autres pays du G7, également analysées, mènent à un résultat similaire à l'exception du Japon.

De ces chiffres ressort également que tester un modèle à l'aide de la corrélation inconditionnelle peut induire en erreur. Gali applique ce précepte à la relation entre les variations du PNB et celles de l'emploi en heures. Il calcule les deux corrélations conditionnelles (chocs technologiques et non technologiques). Pour les chocs technologiques, l'indispensable corrélation entre ces deux variables s'avère singulièrement affaiblie ; de quoi douter du rôle des chocs technologiques comme initiateurs du cycle conjoncturel.

L'article « What happens after a technology shock » de Christiano, Eichenbaum et Vigfusson (2003) répond à celui de Gali. Les auteurs répètent leur confiance dans la conclusion des modèles RBC selon laquelle les chocs technologiques ont un effet positif sur l'emploi. Pour le démontrer, ils appliquent aux données réelles un traitement statistique de la même veine que celui de Gali, qui départage les influences respectives des chocs technologiques et non technologiques. L'hypothèse d'identification est la même : seuls les chocs technologiques ont un effet permanent. Il y a toutefois une différence qui s'avère déterminante pour le résultat : Gali corrélait les VARIATIONS de la productivité avec celles des heures travaillées. Ici, c'est le NIVEAU de ces variables qui fait l'objet de l'étude de corrélation. Et effectivement, selon qu'on travaille avec leurs variations ou leurs niveaux, l'effet d'un choc technologique sur les heures travaillées sera inversé. Reste à savoir laquelle des deux démarches est la plus adéquate. Les auteurs argumentent en faveur de la spécification du niveau.

L'article de Gali n'était pas le premier à lancer le débat sur l'effet des chocs technologiques, mais il a servi de catalyseur à son amplification. La toute fin du XXe et le début du XXIe siècles voient l'émergence d'une littérature considérable sur l'importance des chocs technologiques, où voisinent les articles pro- et anti-RBC. L'article "Technology Shocks and

[1] Gali [114] p. 255. Le *vecteur autorégressif* (VAR) est une méthode statistique pour capturer les interdépendances entre séries temporelles.

Aggregate Fluctuations : How Well Does the RBC Model Fit Postwar US Data ?" de Gali et Rabanal (2004) a précisément pour objet de passer cette littérature en revue. Les auteurs remarquent : "It is not easy to summarize in a few words the wealth of existing evidence nor to agree on definite conclusions of a literature that is still very much ongoing. Nevertheless, it is safe to state that the bulk of the evidence reviewed in the present paper provides little support to the initial claim of the RBC literature on the central role of technological change as a source of business cycles"[1].

[1] Gali & Rabanal [115] p. 4

TITRE IX- QUELQUES THEORIES ET QUELQUES DEBATS IMPORTANTS

9.1- L'ECONOMIE DU BIEN-ETRE SELON PIGOU

9.1.1. Introduction : le dividende national

Héritier de Marshall, Pigou étudia d'abord la philosophie morale. Devenu économiste, la préoccupation première ne le quitta pas. C'est donc naturellement qu'il lança son économie du bien-être, fruit du savant dosage d'utilitarisme, de marginalisme et de préoccupation sociale qui transparaît dans « The Economics of Welfare » (1920).

Le *bien-être ECONOMIQUE*, objet de l'ouvrage, n'est évidemment qu'une facette du bien-être général, celle qui peut être exprimée en unités monétaires. Le montant monétaire qui exprime le bien-être économique d'une société, Pigou l'appelle le *dividende national*.

Comment Pigou convertit-il les satisfactions constitutives du bien-être économique en unités monétaires ? Le prix de demande paraît l'instrument idéal, c'est-à-dire que l'étalon des satisfactions est le prix que l'agent est disposé à débourser pour les obtenir.

Relatant la théorie de Pigou, beaucoup d'économistes identifient le *dividende national* avec le *revenu national* et Pigou semble aller dans ce sens. La première caractéristique du dividende national est d'être un flux[1] de produits et services. Se pose alors la question de savoir quels biens appartenant à ce flux général doivent être comptés dans le dividende national. La réponse est « only those, that are actually sold for money ». C'est là la règle qui prévaut dans la comptabilité nationale moderne. A partir de ce principe, Pigou donne une leçon de comptabilité nationale et son ouvrage a sans doute contribué à affermir cette discipline encore balbutiante à l'époque. Pigou remarque que cette règle donne lieu à des paradoxes : ainsi, si un homme se marie avec sa cuisinière, cela contribue à réduire le revenu national, puisqu'il ne paye pas sa femme pour les travaux ménagers qu'elle effectue.

Pour calculer le revenu national, on ne peut évidemment pas additionner les valeurs des biens de production produits dans l'année avec celles de biens de consommation de la même période, puisque les prix de ces derniers incluent le recouvrement des avances relatives aux premiers. Il y aurait double emploi. La solution proposée par Fisher consiste à ne prendre en compte que les biens de consommation, mais à leur valeur totale[2]. A cette méthode s'oppose celle préconisée par Marshall : tous les biens seraient inclus dans le revenu national, mais à leur *valeur ajoutée*, c'est-à-dire qu'est déduite du prix de vente, la somme déboursée auprès des autres entreprises pour acquérir les biens de production[3]. Toutefois, la valeur des équipements durables n'est déduite qu'à concurrence de l'usure mesurée par l'amortissement comptable. Pigou appuie la conception marshallienne, qui s'est d'ailleurs imposée[4].

Pigou insiste sur la distinction entre la CREATION de revenu et le TRANSFERT de revenu. Les transferts de revenu n'influencent pas le revenu national ; il s'agit des revenus obtenus par une

---

[1] Par définition, le FLUX est une quantité par unité de temps. L'habitude en économie est de mesurer des flux annuels. Ainsi procède la comptabilité nationale.

[2] En fait, Fisher va même plus loin : il voudrait ne compter les biens de consommation durables qu'à concurrence du service rendu (par exemple les kilomètres parcourus plutôt que le véhicule).

[3] Il reste alors la rémunération des facteurs travail, capital et terre et le profit de l'entrepreneur, avant taxation.

[4] Elle est plus logique, car une économie B produisant la même quantité de biens de consommation qu'une économie A mais ayant accumulé les moyens d'en produire plus dans un futur proche doit être considérée comme plus riche.

personne ou par l'Etat, sans contrepartie d'un service rendu, par exemple les donations, les impôts, les intérêts de la dette publique…

Malgré cette leçon de comptabilité nationale, le dividende national de Pigou ne correspond pas au revenu national. Il inclut en effet le *surplus du consommateur* (cf. supra : Marshall). Or le revenu national est valorisé au prix du marché. La valorisation du flux au prix de demande ne va pas sans causer de problèmes. Pigou le reconnaît, l'absence de données relatives aux courbes de demande réelles rend irréalisable la mise en pratique de ces principes. A cette difficulté pratique s'ajoute un problème théorique : le *surplus* que le consommateur retire de sa consommation d'un panier de n produits n'est pas égale à la somme des n surplus, car l'utilité de la consommation d'un bien n'est pas indépendante de celle des autres biens.

Un chapitre assez compliqué est consacré aux variations du dividende national. Il n'est pas toujours évident de déterminer le sens de sa variation et encore moins son ampleur. Le critère appliqué par Pigou est celui-ci : le dividende national augmente lorsque les agents donneraient plus pour les biens entrants que pour les biens sortants entre t et $t+1$. Mais cette règle n'est valable que si les goûts restent constants. L'arrivée de nouveaux produits crée également un problème qu'il n'est pas possible de résoudre de façon totalement satisfaisante.

Pigou s'interroge dans quelle mesure le dividende national exprime correctement le bien-être économique. Diverses questions sont posées, notamment :

- Avant une innovation technologique, tel besoin ne pouvait pas être satisfait. Il n'y avait donc pas de demande pour sa satisfaction. L'innovation technologique a créé une demande nouvelle. L'effet de la consommation du bien correspondant sur le bien-être économique, est-il réellement positif ? Pigou estime que oui.
- La consommation est motivée en partie par le désir de différenciation sociale, le fait d'épater l'entourage. La manière dont est distribué le revenu influencera donc la satisfaction procurée par un flux de biens donné.
- La désutilité engendrée par la production des biens (désutilité du travail, abstinence) réduit le bien-être alors qu'elle n'est pas décomptée du dividende national. Pigou juge possible mais peu probable la situation paradoxale où l'accroissement du dividende national impliquerait la diminution du bien-être à cause de la désutilité

Pigou s'intéresse évidemment à la maximisation du dividende national. Cette maximisation implique qu'aucun transfert de ressource productive ne peut l'accroître, autrement dit que la répartition de ces ressources entre les entreprises et les secteurs égalise les valeurs de leurs produits marginaux[1]. En dehors de l'optimum, le transfert de facteurs vers un autre emploi peut donc être productif. Toutefois des obstacles peuvent freiner ces mouvements, principalement l'insuffisance de l'information et l'existence d'un coût de transfert. Si ces coûts dépassent l'avantage du transfert, le maintien d'une situation infra-optimale se révélera la meilleure solution.

9.1.2. La distribution et la redistribution des revenus

Dès le départ, Pigou énonce un principe fort : « The old 'law of diminishing utility' thus leads securely to the proposition: Any cause which increases the absolute share of real income in the hands of the poor, provided that it does not lead to a contraction in the size of the national

[1] Plus précisément des valeurs de *produits marginaux nets sociaux*, concept qui sera précisé au sous-chapitre 9.1.3.

dividend from any point of view, will, in general, increase economic welfare »[1]. Le *motif de différentiation* ne contredit pas ce principe, car il joue essentiellement entre les membres de la classe aisée. Pigou en conclut que le niveau d'inégalité de la société réelle est excessif et que des transferts des riches vers les pauvres élèveraient le bien-être, moyennant la condition indiquée ci-avant.

Majoritairement, les salariés font partie de la classe défavorisée, Pigou analyse les effets d'une augmentation des salaires sur le dividende national. Il commence par définir ce qu'il appelle le *salaire équitable*. Pour qu'il soit équitable, le salaire doit répondre à deux conditions :

1- il doit égaler la valeur du produit marginal. S'il est inférieur, Pigou parle d' « exploitation du travail ». Les écarts salariaux ne peuvent être justifiés que par des différences de productivité marginale.
2- le travail doit être réparti entre les emplois de façon à maximiser le dividende national. Sinon, le produit marginal serait trop arbitraire.

En concurrence parfaite, le salarié rémunéré en dessous de son produit marginal est théoriquement assuré de l'obtenir en cherchant un autre emploi. Néanmoins, « In so far, however, as movements of workpeople are hampered by ignorance and costs, a monopolistic element is introduced into the wage bargain. Consequently, there is created *a range of indeterminateness,* within which the wages actually paid to any workman can be affected by individual "higgling and bargaining." »[2]. Or justement, si le marchandage devient déterminant, les employeurs seront favorisés, car ils ont plus l'habitude de négocier et parce que le désavantage subi en n'engageant pas un travailleur est moindre que celui de n'être pas engagé. L'organisation collective des travailleurs peut compenser ce handicap.

En soi, l'exploitation du travail n'altère pas la répartition des emplois et n'affecte donc pas directement le dividende national. Toutefois, des effets indirects de son abolition pourraient bénéficier au dividende national :

1- Le secteur concerné par l'augmentation de salaire attirera des travailleurs venant d'emplois où le produit marginal est inférieur.
2- Les employeurs seraient incités à chercher le profit par l'innovation plutôt que par le marchandage salarial.
3- Souvent, l'exploitation est une forme de subside aux entrepreneurs les moins compétents.

L'intervention de l'Etat pour augmenter les salaires INEQUITABLES se justifie ; la préférence de Pigou va à des mesures spécifiques pour rencontrer les diverses situations particulières plutôt qu'à la prise de mesures générales. Pigou discute ensuite l'idée assez répandue qu'il est toujours avantageux d'accroître les bas salaires, même lorsqu'ils sont EQUITABLES. L'objectif serait d'assurer des moyens de subsistance décents à tous les travailleurs. L'imposition d'un salaire minimum légal est un exemple de mesure allant dans ce sens. Pigou reconnaît deux avantages possibles : l'incitation des employeurs à améliorer l'organisation ainsi qu'une meilleure motivation des salariés pouvant accroître la productivité marginale du travail. Mais il craint que cette mesure amène au chômage des travailleurs dont la rémunération dépasse la productivité marginale ; par la voie d'une contraction de l'emploi, le dividende national baisserait.

---

[1] Pigou [283] p. 89. Pigou pratique l'addition des utilités à la Bentham. De façon étonnante, le parétien Lerner réexposa le même point de vue dans « The economics of control », mais il avait pris soin de le justifier par la loi des grands nombres : sur un grand nombre de personnes, a priori, la prédisposition caractérielle des riches et des pauvres à profiter d'un accroissement de revenu devrait être plus ou moins égale.

[2] Pigou [283] p. 557.

Se pose également la question si cette hausse forcée de bas salaires au dessus du produit marginal ferait augmenter le revenu global du groupe de travailleurs concernés et celui de la classe salariée dans son ensemble. Le groupe concerné par la hausse verrait son effectif décroître mais le revenu individuel augmenter. Leur revenu global croîtrait si l'élasticité de la demande de ce type de travail est inférieure à l'unité.
Quant au revenu global de la classe ouvrière, il serait soumis à deux effets néfastes :

- le produit serait rendu plus coûteux par la hausse des bas salaires. La classe salariée pourrait être confrontée à un renchérissement de ses moyens de subsistance.
- La rémunération des autres facteurs, principalement le capital, chuterait, ce qui réduirait le volume de leur utilisation et donc, in fine, le produit marginal du travail.

En fait, ces deux désavantages cités par Pigou ne peuvent pas s'additionner, si ce n'est partiellement. Ce qui est payé par les autres facteurs ne doit plus l'être par les consommateurs et inversement.

Après les manipulations du salaire, voyons les effets de l'aide directe aux pauvres : le versement d'allocations, le logement social, la gratuité des soins de santé… Pigou critique deux arguments couramment entendus, qui mettent en doute la capacité effective de ces mesures de redistribuer le revenu en faveur des pauvres :

- *Les dépenses des riches sont nécessaires pour maintenir l'emploi des pauvres*. En fait, pour maintenir l'activité économique, il importe peu que la demande vienne des riches ou des pauvres.
- *S'ils reçoivent des revenus subsidiaires, les employés accepteront de travailler pour des salaires inférieurs*. Les faits contredisent cette affirmation. Plus aisés, on devient plus exigeants.

La perspective des taxes décourage-t-elle l'esprit d'entreprise ou d'épargne des riches et la perspective de l'aide décourage-t-elle l'effort des pauvres ? La réponse de Pigou est mitigée, dépendant des modalités d'imposition et des conditions d'accès à l'aide sociale. Nous n'entrons pas dans le détail de cette analyse. Voyons maintenant les effets directs des transferts de revenu sur le dividende national.

Si en vertu d'un transfert, la consommation des riches baisse de 100 pendant que celle des pauvres augmente d'autant, cela n'affectera pas le dividende national. Par contre, si la contrepartie en tout ou en partie est une réduction de l'épargne des riches, la formation du capital sera atteinte et le dividende national avec elle.

Des dépenses publiques favorables à l'éducation et à la santé des pauvres et surtout à celles de leurs enfants sont un des meilleurs investissements. « In the actual world, there is no easy way in which capitalists can ensure that any considerable part of the return on money invested by them in the capacities of the poor shall accrue to themselves »[1]. Le libre jeu des forces du marché ne conduit donc pas à l'optimum en cette matière. L'intervention de l'Etat s'impose, mais ses modalités doivent lui permettre de contrôler l'utilisation des fonds transférés ; l'allocation d'un pouvoir d'achat incontrôlé aux pauvres pourrait être gaspillée à d'autres fins que l'investissement le plus profitable : l'éducation des enfants.

9.1.3. Les externalités

Le fait d'être le premier à avoir étudié la question des *externalités* a apporté à Pigou une certaine célébrité. Sa façon d'aborder le sujet a toutefois été critiquée. Notamment, l'angle d'attaque est assez étroit. Pigou ne prononce d'ailleurs jamais le terme *externalité* ou un de ses

[1] Pigou [283] p. 746

synonymes. La voie empruntée consiste à décliner le concept clé bien connu de *produit marginal* en deux concepts frères dérivés de lui: le produit marginal net SOCIAL (PmNS) et le produit marginal net PRIVE (PmNP)[1] ; les externalités sont définies implicitement comme l'écart entre eux.

- Le PmNS est le supplément du produit net total, rapporté par un ajout marginal de ressources, « no matter to whom any part of this product may accrue ».
- Le PmNP est le supplément du produit net total, rapporté directement par un ajout marginal de ressources à l'agent qui les a investies.

Dans beaucoup de cas, l'affectation d'un supplément de ressources n'apporte un produit qu'au producteur concerné. Le PmNP équivaut alors au PmNS. Certains investissements permettent à des tiers d'engranger des bénéfices sans obérer le produit obtenu par l'investisseur de la ressource : ce sont les *économies externes*. Le PmNS dépasse alors le PmNP. Inversement, d'autres investissements font supporter à des tiers un dommage sans compensation : ce sont les *coûts externes*. Le PmNP dépasse alors le PmNS. Les dégâts causés par la pollution sont l'exemple le plus courant et le plus couramment cité des coûts externes. Le concept d'*externalité* recouvre les économies externes et les coûts externes. A noter que Pigou n'y inclut que les effets objectifs, non les conséquences psychiques.

Ce sont les PmN SOCIAUX des différentes ressources productives employées dans l'économie qu'il faut égaliser pour maximiser le dividende national. Mais le libre jeu des forces du marché conduit à l'égalisation des PmN PRIVES. Les externalités nuisent donc à la maximisation du dividende national. Par rapport à l'optimum, l'investissement dans les emplois provoquant des coûts externes tendra à être excessif ; dans les emplois générant des économies externes, il tendra à être insuffisant.

Pigou examine différents types d'externalités, selon que le tiers bénéficiaire ou préjudicié soit un agent lié à la production (par exemple le propriétaire foncier d'un fermier qui a amélioré le sol de la terre louée), le consommateur, une entreprise concurrente alimentant le même marché ou un voisin du producteur, confronté à des nuisances comme des fumées.

Outre les externalités, Pigou pointe une deuxième cause de non-maximisation du dividende national : le **monopole**. Comme nous l'avons vu plus haut, le monopoleur a intérêt à restreindre artificiellement la production pour accroître le prix, ce qui lui procurera un profit net. Le produit marginal (tant le PmNP que le PmNS) des facteurs investis dans une industrie monopolistique tend à rester irrémédiablement supérieur à ce qu'il est dans les industries concurrentielles, entravant par là une répartition optimale des facteurs.

Face aux externalités et au monopole, Pigou en appelle à l'intervention de l'Etat et des collectivités locales. Trois modalités d'intervention sont envisageables :

1- L'imposition de **taxes** sur les produits des industries dont le PmNP dépasse le PmNS, pour y décourager la production (Pigou donne l'exemple de la production de boissons alcoolisées) et l'allocation de **subventions** pour encourager les industries dont le PmNS est supérieur au PmNP.
2- La règlementation : Pigou la préconise surtout dans le domaine de l'urbanisme et de la protection de l'environnement. Selon lui, aucune « main invisible » n'est à l'œuvre pour établir une organisation satisfaisante en ces matières.

[1] Indépendamment de cette distinction, Pigou précise sa conception du *produit marginal net*. Non seulement, on a fait varier d'une unité la quantité d'un facteur, mais on a introduit les modifications organisationnelles pour qu'avec une unité en plus ou avec une unité en moins, la production soit optimale dans les deux cas, eu égard aux ressources disponibles.

3- Les entreprises publiques : les pouvoirs publics deviennent eux-mêmes entrepreneur. Ils peuvent ainsi augmenter directement la production dans des secteurs où le PmNP est trop bas par rapport au PmNS. C'est assez courant dans des secteurs comme la distribution d'eau, de gaz, d'électricité.

Les taxes et subsides conviennent pour des industries concurrentielles. Le monopole peut être contré par l'entreprise publique ou la règlementation des prix.

Les interventions publiques sont désirables pour autant que les pouvoirs publics se montrent à la hauteur. Pigou pointe quatre faiblesses potentielles dans leur chef :
1- Les hommes politiques ne sont pas toujours formés et préparés à la gestion industrielle
2- Les élections peuvent amener des changements importants dans la composition des organes politiques, d'où un manque de continuité dans la vision et une perspective parfois arrêtée aux échéances électorales.
3- L'aire géographique sur laquelle œuvre un pouvoir politique n'a pas nécessairement l'extension qui convient pour une gestion optimale. Par exemple, les communes sont souvent trop petites.
4- Les agences étatiques sont soumises aux pressions des firmes privées, sous la forme du lobbying, voire de la corruption.

En outre, concernant les entreprises publiques :
1- le pouvoir politique pourrait être tenté de leur fournir des avantages qui faussent la concurrence avec les entreprises privées et de maintenir en vie des activités qui n'atteignent pas le seuil de rentabilité.
2- Le privé est plus guidé par l'appât du gain, le public par la crainte de la perte ; la gestion publique est donc moins adaptée à l'innovation et à la prise de risque.

Malgré cette énumération d'insuffisances de la gestion publique, la conclusion de Pigou est nuancée. Concernant le contrôle public sur les activités privées, il estime que les quatre difficultés ci-dessus peuvent être résolues par un système institutionnel adéquat. Il est favorable à l'attribution des fonctions de gestion à des comités constitués de professionnels, où les organes purement politiques n'exercent qu'une supervision. Il se refuse également à disqualifier l'entreprise publique où il voit d'un bon œil l'instauration d'un management copié sur les méthodes du privé. L'émulation suscitée par la présence d'entreprises privées peut aussi rendre les entreprises publiques plus performantes.

La critique de Coase

Dans son article « The Problem of Social Cost » publié en 1960, Ronald Coase analyse le problème des externalités à contrepied de Pigou. Les taxes et la régulation pigoviennes ne lui agréent pas.

Coase estime que les externalités sont un problème de réciprocité. La prévention du tort que l'agent A cause à l'agent B, cause elle-même du tort à l'agent A ; il faut donc trouver le moyen d'éviter le tort le plus grave des deux.

Selon Coase, l'analyse de la question doit reposer sur une double distinction :
1- La loi et la jurisprudence, obligent-t-elles, oui ou non, l'agent causant des coûts externes à indemniser les victimes ? C'est ce qu'il appelle le *droit initial*.
2- est-il possible et permis, oui ou non, que les agents concluent des contrats redistribuant les *droits initiaux* moyennant contrepartie financière ?

La seconde distinction est fondamentale. Coase examine d'abord la situation lorsque aucune cause extérieure n'empêche les parties de négocier un compromis. Il donne l'exemple d'un éleveur dont les bêtes causent des dégâts aux cultures du fermier voisin, dégâts qui sont une fonction croissante du nombre de bêtes. Supposons que le droit initial impose d'indemniser le fermier. Si en renonçant à cultiver une parcelle de son champ (voire la totalité), le fermier perd moins de recettes que l'éleveur n'en obtient en maintenant son troupeau, une marge de marchandage se dégage pour un compromis par lequel l'éleveur paye au fermier moins que le dégât qu'il aurait causé et où le fermier obtient une meilleure recette que s'il avait subi un dégât réel, même indemnisé.

Coase montre que si l'obligation d'indemniser n'existe pas, un compromis en sens inverse mais symétrique au précédent sera possible. Au lieu d'avoir l'éleveur qui paierait le fermier pour maintenir son troupeau, ce serait le fermier qui paierait l'éleveur pour que celui-ci réduise son troupeau … pour autant que l'éleveur ne gagne pas plus en maintenant le troupeau que le fermier s'il le réduit. Dans les deux cas, avec ou sans l'obligation d'indemniser, on arrivera à la même taille du troupeau et à la même surface cultivée. Si les parties peuvent négocier un compromis, les *droits initiaux* n'influencent pas l'allocation des ressources, qui sera toujours optimale, en ce sens que la valeur de la production totale sera maximisée. Evidemment la distribution du revenu sera affectée par les droits initiaux.

Dans la réalité, bien souvent, les parties seront empêchées de négocier, pour une de ces deux raisons :

1- les **coûts de transaction** sont trop élevés : recherche des personnes concernées, information, réunions, expertises, vérification de l'application du contrat… Les compromis exposés ci-dessus permettaient d'accroître la valeur de la production ; la négociation ne se justifie que si cet accroissement dépasse le coût de transaction. Lorsque le nombre de parties concernées par un dommage est très élevé, ce coût devient vite prohibitif.
2- le droit ne le permet pas. Ou bien un organe public tranche dans chaque cas si l'activité polluante est autorisée ou non. Ou bien le droit initial devient décisif pour déterminer la rentabilité de l'activité.

Lorsque la négociation ne peut être menée, la maximisation de la valeur de la production ne sera pas nécessairement atteinte. Le résultat dépend de la balance entre :

1- la réduction de la valeur de la production causée par la pollution (les dégâts)
2- la réduction de la valeur de la production polluante confrontée à une interdiction, une régulation ou une obligation d'indemniser (les entraves).

Soit la valeur des dégâts dépasse celle des entraves ; alors l'action des autorités publiques ou le droit initial prescrivant l'indemnisation donneront un meilleur résultat que le droit initial laissant le champ libre à l'activité polluante. Soit la valeur des entraves dépasse celle des dégâts, auquel cas, le droit initial le plus libéral eût été le plus propice à la maximisation de la valeur de la production. Dans ce dernier cas, les voisins immédiats du pollueur eurent été irrémédiablement lésés, mais Coase semble craindre qu'une indemnisation trop accessible ne dissuade pas suffisamment les habitants de s'installer dans les endroits à risque.

A mon avis, une lacune de l'analyse de Coase est sa présupposition que la charge des coûts imposés (indemnités aux victimes des nuisances ou filtres à placer sur les cheminés…) repose nécessairement en dernier ressort sur le producteur. Or moyennant certaines conditions, celui-ci pourra répercuter le coût imposé sur les prix de vente. La probabilité que ces coûts motivent l'abandon de l'activité s'en trouve alors largement réduite. Le principal résultat des coûts imposés sera la dilution de l'inconvénient de la nuisance sur une large population plutôt que sa

concentration sur les voisins immédiats de l'activité polluante, ce qui semble préférable en termes d'équité.

9.2- FAUT-IL CROIRE EN LA THEORIE AUTRICHIENNE ?

La théorie autrichienne du capital fut successivement portée par Jevons, Böhm Bawerk, Wicksell et Hayek. Il nous reste à explorer le quatrième épisode. Celui-ci commence par une controverse. Comme Clark trois décennies plus tôt, Knight monte à l'assaut de ce qu'il qualifie d' « incubus », désignant la "'wage fund theory' of the early classical writers as modified by Jevons, Böhm Bawerk and Wicksell", à propos de laquelle il prétend qu'on ne peut rendre un plus grand service à la science économique que de l'en libérer. Il publie plusieurs articles en 1933, 1934 et 1935 auxquels répondront des articles de Machlup en 1935 et de Hayek en 1936. Kaldor tentera une synthèse du débat en 1937.

9.2.1. Knight contre Hayek à propos du capital

LA PERIODE DE PRODUCTION

Inévitablement lorsqu'il est question de la théorie autrichienne, la *période de production* est une des premières cibles. C'est assez paradoxal car Hayek et Machlup se distancient tous deux de ce concept. Mais tous deux le défendent contre les attaques, car, comme l'écrit Hayek, "Professor Knight, instead of directing his attack against what is undoubtedly wrong and misleading in the traditional statement of this theory (…) seems to me to fall back on the much more serious and dangerous error of his opponents of forty years ago"[1]. Les détours de la production restent à leurs yeux un concept essentiel ; la période de production le sert imparfaitement mais il ne faut pas jeter le bébé avec l'eau du bain.

Comme le montrera Wicksell, la formule de Böhm Bawerk fait l'impasse sur l'intégration des intérêts passés dans le capital. Comme le montrera la controverse des deux Cambridges au début des années soixante, une mesure unidimensionnelle du capital ne peut suffire à établir une relation avec le taux d'intérêt. La forme de la distribution dans le temps importe autant que la durée moyenne. Machlup et Hayek concevaient ces difficultés comme en atteste cet extrait ; partant d'un réarrangement des inputs qui en déplace symétriquement certains vers l'arrière et d'autres vers l'avant sans altérer la moyenne, Machlup écrit : "This shows that not only the average of all time dimensions but also their dispersion may be significant. The 'shape of the investment function' would tell us what the simplest 'average investment period' conceals"[2].

Knight émet à l'encontre de la *période de production* la critique qu'on ne peut en déterminer concomitamment le début et la fin : si on connaît le début, la fin est indéterminée et vice versa. « If we take a certain small increment of consumption, say, drinking a glass of milk, it would never be possible to give any sensible answer to the question when that glass of milk was produced or when its production began. Or conversely, we may take a definite increment of productive activity, such as feeding the cow; then, it would never be possible to say when the result of that increment of activity would be consumed or when the process of consuming them would end »[3].

Hayek considère que l'approche par la ligne de départ et l'approche par la ligne d'arrivée constituent une alternative : l'analyste choisit entre ces deux options. Il est sous-entendu que les deux approches donnent le même résultat. Mais dans l'une comme l'autre options, il faut se préoccuper non du temps effectivement consacré à produire tel ou tel biens dans le passé, mais des « FUTURE TIME INTERVALS between the moments when the factors are or will be invested

[1] Hayek [138] p. 200
[2] Machlup [241] p. 588.
[3] Knight [191] p. 275

and the moment when the product will mature (…) The theory looks forward, not back »[1]. A mon avis, la question centrale n'est pas « forward » ou « backward » mais SIMULATION ou REALITE. Machlup le pressent, car commentant l'option « forward », il écrit : « That it is by its nature hypothetical, representing an average of anticipations which may not be realized is not in contradiction with the fact that it is very real in its effects … »[2].

Tous les professionnels qui pratiquent la comptabilité analytique et estiment des prix de revient savent que leurs calculs sont désincarnés par rapport à la réalité et nécessitent des hypothèses relatives à la durée de vie des machines, à la fraction de leur temps absorbée par tel ou tel produits… Des difficultés surviennent comme le produit A qui sert d'input dans la fabrication du produit B qui est également un input de la production A. La période de production n'y échappe pas. On doit se contenter d'estimations. Dans ces conditions, il serait erroné de penser que la période de production absolue (si elle est déterminable) est nécessairement plus longue que la période moyenne, car celle-ci additionne les durées pondérées de multiples processus pratiqués simultanément dans la réalité.

Je me rallie à la conclusion de Kaldor : « we hope that we have succeeded in demonstrating that the real objections against the Austrian capital theory relate to the MEASUREMENT of the investment period rather than to its RELEVANCE »[3]. Ceci, moyennant les restrictions mentionnées ci-avant.

L'ATTENTE, LE CAPITAL PERMANENT

La théorie autrichienne fait de l'attente l'indice du niveau capitalistique de la production, donc indirectement une mesure du capital lui-même. Knight conteste ce rapprochement : « But additional investment may have the purpose and the result of shortening this waiting period ». L'attente et la valeur du capital évoluent indépendamment l'une de l'autre: « Increased capital investment may either increase or decrease the length of time required for processing of any particular material, from beginning to end or between any two physically definable stages »[4].

Hayek répond en premier lieu par une mise au point. Dans l'histoire économique, l'accroissement du capital s'est généralement accompagné du progrès technique. Au point qu'on aurait tendance à les confondre. Pourtant, le capital peut être accumulé sans amélioration de la connaissance technique, simplement pour généraliser les outils les plus productifs lorsqu'ils ne sont utilisés que par les entreprises de pointe. C'est à cette accumulation du capital pure que la théorie autrichienne rattache l'attente, non au progrès technique. Mais seul le progrès technique est susceptible d'éventuellement réduire la période de production. Machlup, quant à lui, estime que le progrès technique n'influence pas le niveau de détour de la production ; tout dépend de la façon dont l'économie absorbe les facteurs libérés par l'innovation.

Puis, allant au-delà de cette mise au point, Hayek pose la question : si l'attente diminuait, pourquoi augmenterait-on le capital par tête ? On chercherait en vain la réponse chez Knight. Hayek peut donc réitérer la conclusion de Böhm Bawerk : l'accumulation du capital a pour vocation fondamentale d'allonger l'attente. Comme le précise Machlup, c'est le capital par tête qui importe. On retrouve la distinction entre l'élargissement du capital et son

[1] Hayek [138] p. 208.
[2] Machlup [241] p. 597
[3] Kaldor [170] p. 233
[4] Knight [191] pp. 278-279

approfondissement déjà évoquée à plusieurs reprises. L'élargissement est extérieur à la présente discussion.

Un des arguments autrichiens pour justifier l'existence de l'intérêt dans une économie STATIONNAIRE est qu'il est nécessaire pour dissuader les détenteurs du capital de désépargner ; sans l'intérêt, ceux-ci seraient tentés de remplacer leur capital par un revenu consommable. Il n'est donc pas étonnant que Knight et Hayek se soient affrontés sur la *permanence du capital*. Knight soutient que « all capital is inherently perpetual », une thèse déjà présente chez Clark. Le capital perdure et reste égal à lui-même (quantitativement) alors que passent les générations d'équipement qui se succèdent (dans la stationnarité). Sa permanence n'est pas affectée par la décision individuelle d'un capitaliste de consommer son capital. Le capital de l'individu peut diminuer mais pas le capital social : il y aura simplement transfert de propriété de l'équipement concerné. Les individus qui désépargnent ne peuvent pas consommer le capital tel quel, mais seulement la recette de sa revente.

Hayek rétorque que le capital maintenu quantitativement constant est un concept qui n'a de sens que dans l'hypothèse de la stationnarité la plus absolue, en rupture complète avec la vie économique réelle et qu'il est incompatible avec les comportements rationnels dans une économie en changement.

Sans doute une société stationnaire se caractérise-t-elle par un capital perpétuel automatiquement maintenu. Knight semble en outre croire à une dissymétrie entre l'accroissement du capital et sa diminution. Il ne conteste apparemment pas que le capital total diminuera si la société s'abstient de remplacer les biens capitaux usés, mais tout en restant vague sur le sujet, il considère cette éventualité comme propre à des conditions exceptionnelles ; il évoque les crises dont la cause selon lui est totalement indépendante des questions de capital.

Le non-remplacement des équipements ne signifie pas encore que la société a consommé une part de son capital, Le capital comme tel n'est pas consommable. Le non-remplacement signifie alors la régression économique par la décapitalisation. A terme, la consommation devra d'ailleurs diminuer puisque les moyens de l'alimenter régressent. Mais Machlup semble penser que pendant une phase intermédiaire, il serait possible de consommer plus alors que décline le capital.

Ni Knight ni Hayek ne commettent l'erreur de penser que le taux d'intérêt serait nul dans une économie stationnaire, une idée présente chez d'autres économistes comme Schumpeter. Selon Knight, le taux d'intérêt ne peut s'annuler qu'à la condition que tous les biens réels et potentiels dans la production desquels intervient le capital soient des biens libres. Autant dire que l'intérêt nul est impossible. Hayek répond qu'il suffit que l'économie atteigne une position où il devient impossible d'accroître la production en allongeant l'attente[1]. Cette condition est moins exigeante, tout en restant peu probable.

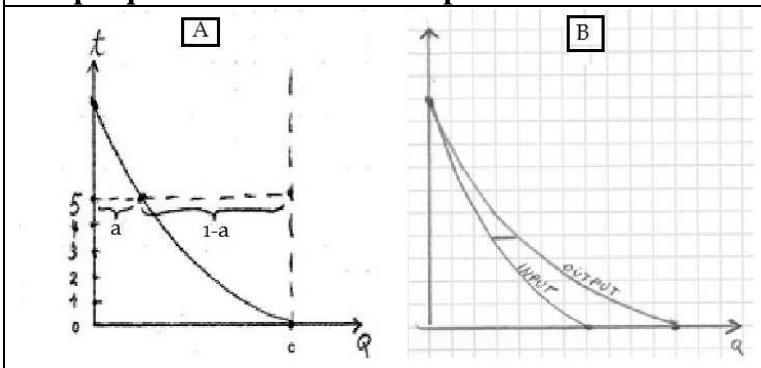
9.2.2. « The Pure Theory of Capital »

Sans doute en partie à cause de la controverse, Hayek ressent le besoin de réexprimer la théorie autrichienne en abandonnant les éléments les plus critiqués. Ce projet aboutit à un ouvrage intitulé « The pure Theory of Capital » publié en 1941. Le but est de sauvegarder l'essentiel. Malheureusement, ce sont les simplifications qui sont sacrifiées ; les idées qui les remplacent sont beaucoup plus complexes. Il en résulte un ouvrage plutôt indigeste.

[1] Cette règle découle logiquement de la formule autrichienne du taux d'intérêt f'(t)/f(t).

La première victime de ce réexamen est la *période de production*, à laquelle Hayek n'était déjà plus attaché dans la controverse. La « Pure Theory » présente ses héritiers : la *fonction d'input* et la *fonction d'output*.

Graphique 9.1: la fonction d'input



Voyons d'abord la *fonction d'input*, sur la figure 9.1-A. Elle est inspirée par le graphique 3.13 de Wicksell. L'ordonnée représente le temps restant avant la disponibilité des biens produits pour la consommation. Le temps zéro correspond donc à la consommation présente. L'abscisse donne la quantité d'input[1] actuellement en maturation dans un produit qui sera rendu consommable au temps *t*. Dans l'exemple présent, on peut donc dire que le produit qui sera commercialisé dans cinq ans est déjà constitué dans une proportion *a* et que dans les cinq ans à venir, il reste à produire une proportion 1 - *a*. La fonction est dessinée convexe par rapport à l'origine. Cela correspond à un flux de services qui s'accroît au fur et à mesure qu'on se rapproche du produit final, l'hypothèse la plus plausible selon Hayek. Si le flux était constant de période en période, la fonction serait rectiligne. Le principe de cette figure est évidemment le même que celui des triangles du graphique 8.17. Mais le point de vue est différent : ici, on observe l'encours de la production à un moment donné ; là, on visualisait l'évolution d'une production.

La *fonction d'output* ressemble très fort à sa sœur. Elle intègre la problématique de l'intérêt composé. L'intérêt du passé fait partie du produit, donc du capital. L'intérêt composé agit un peu comme un agent dilateur de chaque atome d'input et le niveau de grossissement dépend évidemment de la durée de dilatation. C'est ce processus de dilatation qui transforme la fonction d'input de la figure 9.1-B en fonction d'output. Mais Hayek utilise peu cette fonction d'output. Il lui préfère une version en trois dimensions fort complexe que je ne reproduis pas.

Il y a plusieurs types de travail et de matières : comment rendre compte de cette hétérogénéité ? De deux choses, l'une : soit les inputs sont exprimés en valeurs : on peut les additionner, ce qui permet d'illustrer la production par un graphique unique. Soit les inputs sont exprimés en unités physiques, mais alors le processus de production ne peut être illustré que par autant de graphiques qu'il n'y a de facteurs homogènes.

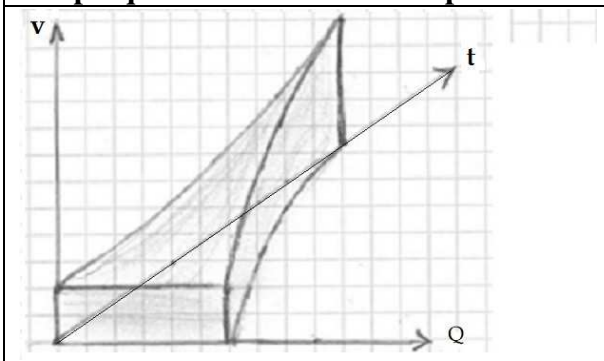
Dans la conception autrichienne primitive, la période de production mesurait la quantité de capital univoquement. Une telle simplicité est maintenant exclue. Voyons comment Hayek met en graphique la valeur du capital intégrant l'intérêt composé. Le plan horizontal représente la fonction d'input de la figure 9.1, mais il le surmonte d'une troisième dimension verticale (*v*).

[1] Le terme *facteurs primaires*, très décrié lors de la controverse, ne se retrouve plus dans la « Pure Theory ». Hayek opère une distinction peu éclairante entre *ressources permanentes* et *ressources non permanentes*. En fait, c'est le terme *input* qui revient le plus fréquemment et je me limiterai à lui dans ce résumé.

L'altitude indique la valeur que l'intérêt composé fait prendre à chaque unité d'input composant la fonction d'input (figure 9.2). Les unités qui sont investies plus longtemps montent plus. L'altitude et l'inclinaison de la surface courbe supérieure dépendent du taux d'intérêt, qui est sous-jacent dans le graphique. La valeur du capital est donnée par le volume du solide.

L'aire de la fonction d'input a plus ou moins la même signification que la traditionnelle *période de production*. Mais deux fonctions d'input de formes différentes et d'aire identique ne serviront pas nécessairement de base à un volume égal. Le rapport du capital avec le temps s'est complexifié.

Graphique 9.2 : la valeur du capital



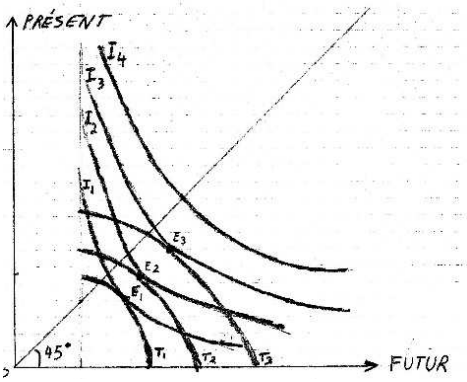
Qu'est-ce qui détermine la forme de la fonction d'input ? Elle dépend de choix technologiques que la firme est appelée à opérer, plus précisément de l'optimisation de ces choix. Hayek explique le principe de cette optimisation de façon très analytique ; j'espère qu'on me pardonnera d'extrêmement simplifier. Dans un premier temps, Hayek met de côté la problématique de la préférence temporelle, de l'épargne et de la valeur du capital disponible. Le choix porte sur des arrangements qui produisent un flux de revenu constant, insensibles aux préférences des agents. Chaque unité de chaque input est investie pour un certaine durée qu'Hayek suppose être variable. En augmentant la durée d'investissement d'un input, on peut augmenter le produit. L'exercice vise à répartir le capital donné de la façon qui maximise le produit, autrement dit à choisir la durée d'investissement idéale de chaque unité d'input. La règle d'optimisation consiste à égaliser les productivités marginales de tous les investissements d'inputs. Et cette productivité marginale commune, plus exactement le taux d'accroissement du produit résultant des variations marginales de l'input, devient le taux d'intérêt.

Cette analyse suppose que l'on puisse modifier la durée d'investissement de tout input en laissant les autres durées constantes, ce qui est assez douteux. Il y a en outre la difficulté que, même si l'hypothèse est vérifiée, "this functional dependence of the size of the product on the investment period of the unit of input concerned is only true for a particular arrangement of all other input and will be different for any other arrangement"[1].

Hayek reconstruit ensuite son modèle en endogénéisant le capital à l'aide de l'épargne[2], dans une analyse d'inspiration fishérienne. Ici de nouvelles simplifications sont posées, notamment qu'un investissement a toujours une période unitaire. Il n'y a donc pas de place ici pour une fonction d'input ou d'output, ce qui déconnecte cet exposé du reste de l'ouvrage.

[1] Hayek [147] p. 152
[2] L'épargne de précaution est exclue. On épargne pour transférer du revenu.

Graphique 9.3 : l'épargne, l'investissement et le taux d'intérêt chez Hayek



Le graphique 9.3 comporte deux familles de courbes :

- les *courbes de transformation* qui sont concaves et indiquent le taux auquel du revenu peut être transféré du présent vers le futur. C'est grosso modo la courbe 3.21 de Fisher. Sa pente reflète la productivité de l'investissement[1][2].
- les courbes d'indifférence qui sont convexes et indiquent les préférences de l'agent entre la consommation future et la consommation présente. Le taux d'impatience s'inscrit dans la pente de la tangente.

La droite à 45° est le lieu des distributions égales du revenu entre le présent et le futur. A sa droite, on épargne, à sa gauche, on désépargne.

En fait, l'équilibre advient à chaque période et le déroulement diachronique des équilibres successifs a été concaténé sur une figure unique[3]. En période 1, la capacité productive est exprimée par la courbe T1. L'équilibre résulte évidemment de la tangence entre T1et une courbe d'indifférence, soit le point E1. Le taux d'intérêt est donné par la pente (identique) des deux courbes en E1[4]. En période 2, grâce à l'investissement précédent, la capacité productive a augmenté, ce qui se marque par le déplacement de la courbe de transformation vers le nord-est[5]. Un nouvel équilibre s'établira de façon identique en E2. Et ainsi de suite…

Quelles sont les évolutions probables sur ce sentier d'équilibre ? Avec le revenu croissant, le taux d'intérêt et le taux d'impatience diminuent, ce qui réduit l'incitation à épargner. L'économie s'oriente vers une situation stationnaire, sans épargne, donc située sur la droite à 45°. La vitesse d'accession à cet équilibre final dépend de la forme des courbes d'indifférences ; dans un cas extrême, on peut ne jamais y arriver lorsqu'il y a une préférence récurrente pour le futur.

Hayek pèse les rôles respectifs de la préférence temporelle et des détours de la production :

- en chemin vers l'équilibre, il y aura toujours un taux d'intérêt positif, quel que soit le taux d'impatience
- le taux d'épargne dépend uniquement des opportunités d'investir et non de l'inclination à épargner

---

[1] Selon Hayek, la concavité de ces courbes remet dans le jeu la variabilité du temps d'investissement, qui serait l'explication essentielle de la productivité marginale décroissante.

[2] Plus exactement, la pente égale l'inverse du facteur d'accroissement.

[3] La constance des goûts est posée, ce qui permet la conservation d'un même champ d'indifférence.

[4] Plus exactement, la pente égale 1/(1+i).

[5] Il s'agit d'investissements NETS, s'ajoutant à la reproduction du capital investi précédemment, qui est entièrement assurée.

- la longueur du chemin vers l'équilibre découle de la forme des courbes d'indifférence. Au stade final, le taux d'intérêt sera nul ou positif se calquant sur le taux d'impatience.

9.3- LE DEBAT SUR LE CALCUL SOCIALISTE

9.3.1. Prélude

Dans le chapitre du Manuel (1906) consacré à l'équilibre, **Pareto** insère une section intitulée « L'équilibre dans la société collectiviste ». L'Etat collectiviste doit résoudre deux problèmes : d'une part, comment distribuer entre les citoyens les biens produits par la société et d'autre part, comment produire ces biens pour que soient maximisées les ophélimités individuelles. Les règles d'optimisation à appliquer par l'Etat collectiviste s'avèrent être celles qui fonctionnent en concurrence parfaite. Pareto pense même que l'Etat collectiviste aura plus de facilité à atteindre l'équilibre que le système concurrentiel. Par contre, ses fonctionnaires pourraient se révéler moins efficaces que les entrepreneurs. D'où une conclusion mi-figue mi-raisin : « …l'économie pure ne nous donne pas de criterium vraiment décisif pour choisir entre une organisation de la société basée sur la propriété privée et une organisation socialiste »[1].

Tant d'indulgence envers le socialisme peut étonner de la part d'un économiste réputé antisocialiste. En fait, Pareto veut montrer que sa théorie de l'équilibre général transcende les systèmes politico-économiques. La théorie pure se situe à un niveau supérieur à celui des actions des hommes.

Deux ans plus tard, Enrico **Barone**, disciple de Walras et Pareto, embraye sur la même idée. Cette fois-ci, c'est un article de quarante-cinq pages qui est entièrement dédié à cette question, comme l'indique son titre : « Le ministère de la production dans l'Etat collectiviste ». D'emblée, l'auteur, pourtant pas plus enclin à la sympathie envers le socialisme que Pareto, précise « … I do not write for or against collectivism » et il tient parole. L'article se compose de deux parties assez symétriques consacrées, la première au régime individualiste concurrentiel[2], la deuxième au régime collectiviste. Dans les deux cas, il démontre l'existence d'un équilibre ainsi que la coïncidence entre cet équilibre et le maximum de bien-être. L'équilibre est démontré par l'égalité entre le nombre d'équations et le nombre d'inconnues. Les équations sont « à la walrassienne », mais assaisonnés à la cassellienne. Comme Cassel (1918), Barone refuse toute notion d'utilité ou de préférence dans ses équations ; la dépendance de l'offre et de la demande par rapport à l'ensemble des prix lui suffit. Il est aussi l'un des premiers à introduire des coefficients techniques variables dans l'équilibre général : ceux-ci sont déterminés de façon à minimiser le coût. Développant son système d'équations, Barone insiste particulièrement sur celles qui minimisent le coût et celles qui égalisent prix et coût.

Sans fonction d'utilité, Barone assimile le bien-être à la somme des produits des quantités vendues (R_i) des *n* biens par leur prix (p_i). Pour les valeurs d'équilibre, cette fonction est baptisée ϕ. On a :

$$\phi = \Sigma_{i1}^{n} R_i.p_i, \qquad i = (1...n) \qquad (9.1)$$

Barone démontre que :

$$\partial\phi/\partial R_i = 0, \forall i \qquad (9.2)$$

A partir de l'équilibre, tout changement, qui perturbe l'égalité du prix avec le coût de production ou écarte le coût de son minimum, entraîne un $\Delta\phi$ négatif.

---

[1] Pareto [273] p. 364

[2] Barone analyse également le cas des monopoles privés, mais ce cas n'est pas au cœur de la problématique.

Au ϕ collectif correspondent des θ pour les individus. On a :

$$\phi = \Sigma_{j1}^{m} \theta_j \qquad j = (1...m) \qquad (9.3)$$ [1]

Les θ sont additifs. Le $\Delta\phi$ négatif dû à un écart par rapport à l'optimum n'implique pas nécessairement une baisse de tous les θ ; certains peuvent augmenter pendant que d'autres baissent, mais le poids des baisses doit l'emporter. L'optimum collectif n'implique donc pas celui de tous les agents.

Finalement, Barone arrive à la conclusion que "…the system of equations of the collectivist equilibrium is no other than that of the free competition. Which only means that with equal resources, the economic quantities of the collectivist equilibrium will be the same as those in the individualist equilibrium"[2].

Comment l'Etat collectiviste atteint-il cet optimum ? Les équations de l'équilibre seraient solubles a priori s'il n'y avait pas le problème extrêmement complexe de la détermination des coefficients de production. Tenant compte de cette difficulté, le ministère de la production n'a d'autres moyens de trouver l'équilibre que le tâtonnement.

Concernant la structure de l'économie socialiste, Barone inaugure le modèle qui servira de référence au cours de la controverse dont il sera question dans la suite. L'économie comporte un véritable marché pour les facteurs de production primaires ainsi que pour les biens de consommation. Entre les deux, la production des biens d'équipements est gérée par le ministère de la production. Les ménages choisissent donc librement leur occupation professionnelle et leur consommation.

L'Etat socialiste doit également apporter une réponse à certaines questions, à propos desquelles Barone émet son avis :

1- Certains facteurs primaires, du type terre, sont détenus par la collectivité. Le revenu des ces facteurs doit être partagé entre les citoyens : deux méthodes sont envisageables :
   - La distribution INDIRECTE : le citoyen bénéficie d'une baisse du prix des biens qu'il consomme du fait que l'usage de ces facteurs n'est pas compté dans le calcul du prix.
   - La distribution DIRECTE : le prix des produits comporte ces coûts, mais les citoyens se voient gratifiés d'une allocation s'ajoutant à leur salaire.

   Barone défend la distribution directe car l'autre méthode suscite un gaspillage desdites ressources.
2- Exactement la même question se pose pour le profit[3] des entreprises. La distribution indirecte aurait l'avantage de faire hausser la production, du moins dans les industries qui ne sont pas à coût croissants. Finalement, Barone prône la répartition directe, car elle augmente la liberté de choix du citoyen en tant que consommateur.
3- L'épargne des citoyens est nécessaire pour financer les investissements. L'Etat collecte seul l'épargne. Un certain niveau d'épargne est suscité par le seul motif de précaution. Si cette épargne spontanée ne suffit pas pour financer tous les investissements prévus, l'Etat peut la compléter de deux façons : soit effectuer un prélèvement sur les revenus collectifs (par exemple, ceux de la remarque 1), soit offrir un intérêt qui a pour but d'élever l'épargne au niveau souhaité, solution qui a la préférence de Barone. Barone pense que le taux d'intérêt

[1] Barone montre qu'en outre, en régime individualiste, $\partial\theta_i/\partial r_{ij} = 0$, $\forall i$.
[2] Barone [24] p. 274
[3] Comme le prix doit égaler le coût de production, le profit est celui gagné par les firmes qui ont une efficacité supérieure à celle du reste de leur industrie.

pourrait être plus élevé en régime socialiste qu'en régime capitaliste, car il n'y aura plus l'épargne quasi-automatique des possesseurs de grosses fortunes.

Barone s'est intéressé au sort de la rente foncière et du profit net, mais il n'envisage pas de façon directe le cas du coût normal du capital, pourtant de masse bien supérieure. Ce coût est indirectement concerné par la troisième remarque, mais doit-on considérer que l'épargne privée est la seule source de capital ? Ce serait étonnant d'un système qui veut se démarquer du capitalisme. Si, comme le commande la rationalité, le prix des produits couvre le coût du capital, n'y a-t-il pas là constitution d'un réservoir de financement public pour les investissements ?

L'une des critiques adressées au capitalisme est le gaspillage de ressources que représente la destruction continuelle de capital par la concurrence qui mise sur l'innovation technologique. Barone estime que le socialisme n'évitera pas ces destructions s'il veut réaliser toutes les innovations que la connaissance technique rend possible.

La conclusion de Barone est celle-ci : "If the Ministry of Production proposes to obtain the collective maximum- which it obviously must, whatever law of distribution may be adopted- all the economic categories of the old regime must reappear, though maybe with other names…"[1].

Bien que d'une rédaction un peu chaotique[2], l'article de Barone reste encore aujourd'hui incontournable pour qui veut s'intéresser à cette problématique.

Une conception similaire se retrouve chez **Cassel**, pourtant pas plus socialiste que les deux Italiens. Il affirme explicitement la transcendance de la théorie de la formation des prix sur les systèmes sociaux. Il définit la société socialiste comme une économie d'échange et monétaire, pratiquant le salariat. « Il n'y a pas de raison pour que la société socialiste donne un champ plus large au principe de gratuité, car ce dernier appartient au communisme et non au socialisme »[3].

Cassel plaide pour le maintien de l'intérêt dans la société socialiste. En régime capitaliste, l'intérêt est intimement lié à la propriété bourgeoise et à l'activité de cette classe qui n'existera plus dans le socialisme. On serait donc tenté de croire que l'intérêt disparaîtra avec elle. Mais derrière ces mécanismes bourgeois, « il y a les caractères essentiels de la vie humaine qui existent, bien que sous des formes variables, dans toute économie d'échange »[4]. Pour induire une répartition rationnelle du capital entre les branches de production, il est indispensable que le prix des produits de consommation inclue « un supplément proportionnel à la disponibilité du capital en sus du coût du travail et des frais divers ». L'intérêt est nécessaire pour limiter la demande de capital qui obère la consommation. Dans la société capitaliste, le volume de l'épargne et donc le sacrifice imposé aux consommateurs est réglé par les marchés. Par contre, dans le socialisme, le taux d'intérêt est une variable sous le contrôle du gouvernement. Pas plus que Barone, Cassel ne s'attend à un abaissement du taux d'intérêt par rapport à son niveau capitaliste.

[1] Barone [24] p. 289.
[2] Cf. Bradley [39] p. 9.
[3] Cassel [46] pp. 178-179.
[4] Cassel [46] p. 367.

9.3.2. La controverse entre les deux Guerres

VON MISES ET LE LANCEMENT DE LA CONTROVERSE

En 1920, l'économiste autrichien, Ludwig von Mises, connu pour ses opinions ultralibérales, publie l'article « Economic Calculation in the Socialist Commonwealth ». L'article constitue une attaque en règle dénonçant l'absence de rationalité économique dans le socialisme.

Avant d'attaquer la rationalité du socialisme en tant que mode de production, c'est une certaine attitude de beaucoup de penseurs socialistes qu'il dénonce. C'est leur incapacité à développer une analyse économique de la société socialiste qui fait problème. Ils ne proposent aucune théorie économique du socialisme. La société socialiste est envisagée de façon très vague et uniquement sous l'angle social, culturel ou politique, jamais économique.

L'article paraît moins de trois ans après la Révolution d'Octobre, mais il ne semble pas y avoir de lien direct. Sont plutôt visés les sociaux-démocrates allemands et autrichiens, notamment Karl Kautsky, Otto Bauer, Rudolf Hilferding, à l'époque où ces partis se revendiquaient encore du marxisme. La polémique entre la loi de la valeur-travail chère aux marxistes et la loi de la valeur subjective était encore vivace. Mises tire à boulets rouges contre la loi de la valeur-travail. Pour Marx, elle était un outil servant à démontrer l'exploitation de l'homme par l'homme, mais certains marxistes semblent vouloir lui faire jouer le rôle de déterminant du prix en régime collectiviste[1]. D'après Mises, une telle prétention est absurde ; il y a cela deux raisons :

- le travail n'est pas le seul facteur de production primaire ;
- le travail n'est pas homogène

Si tant de socialistes tiennent à cette loi, c'est parce qu'elle offre une bouée de sauvetage quand la disparition du marché et de la propriété privée a fait disparaître tous les moyens de calcul. Mises reconnaît toutefois qu'on peut être socialiste sans croire en cette loi.

Mais l'objet principal de l'article est le manque de rationalité économique du système socialiste. Diverses critiques sont émises à l'égard de ce système, mais la plus fondamentale est l'impossibilité du calcul économique. A ce moment, Mises ignore l'existence de l'article de Barone qui traitait précisément de ce sujet et dont l'opinion nuancée ne concorde pas avec la sienne. L'article de Barone ne sera exhumé qu'au milieu des années trente et il servira alors d'argument dans le débat. A noter que le système institutionnel socialiste qu'envisage Mises est le même que celui de Barone : marchés pour les biens de consommation et le travail, économie dirigée pour le reste.

Beaucoup de socialistes s'imaginent que le bureau du plan est à même de diriger l'économie en gérant uniquement des flux en nature. C'est une illusion. Seul le calcul en valeur est compatible avec la rationalité et il implique deux préalables :

- les biens de production participent à la sphère de l'échange ;
- il existe un medium universel dans l'échange.

Ces conditions ne sont pas remplies en régime socialiste. Mises en conclut : "Without economic calculation, there can be no economy. Hence, in a socialist state, wherein the pursuit of economic calculation is impossible, there can be –in our sense of the term, no economy whatsoever"[2]. Sans la réalisation des deux conditions susmentionnées, comment le bureau du plan peut-il déterminer quelle méthode de production est la plus économique, comment peut-il

[1] Mises cite un passage de l'Anti-Dühring de Engels qui va dans ce sens.
[2] Mises [375] p. 105

détecter les processus de production excessivement longs (donc trop gourmands en capital), les gaspillages de matière et de travail ? Sans prix pour les biens de production, il n'est pas possible de calculer le coût des biens de consommation, alors qu'ils doivent être vendus sur un marché. Pour agréger des inputs hétérogènes, la seule solution est l'addition de prix monétaires. Un marché des biens de production suppose la propriété privée des moyens de production de ces industries : comment l'industrie du charbon et celle de l'acier pourraient-elles s'entendre sur un prix sans la propriété privée des aciéries et des mines ? Mises reste un peu vague sur ce point, mais il veut sans doute signifier que sans maximisation du profit, il n'y a pas de base pour fonder un système de prix et que cette maximisation est impensable si elle n'est pas motivée par l'accroissement de la propriété personnelle.

Ceci nous amène au deuxième faisceau de critiques adressées par Mises. Rien ne poussera les responsables à agir pour maximiser le bien-être économique. "It is now universally agreed that the exclusion of free initiative and individual responsibility, on which the successes of private enterprise depend, constitute the most serious menace to socialist economic organization"[1].

Certains auteurs socialistes établissent un parallèle entre les dirigeants des unités de production socialistes et ceux des entreprises privées modernes issus de la révolution managériale qui ont supplantés les patrons capitalistes de la première génération. La comparaison ne tient pas. Toutes les sociétés privées sont gérées par des personnes intéressées au profit. L'efficacité de ces managers n'est donc pas transposable. Les dirigeants des sociétés privées ne sont ni plus intelligents ni plus volontaires que ceux des entreprises publiques. C'est la fonction qui fait l'homme et non l'inverse. L'intéressement au résultat de l'entreprise est indispensable.

Sans cet intéressement, il manque un incitant au dévouement à son entreprise. Les socialistes fondent leur espoir sur des motifs éthiques. C'est contraire à la nature humaine. Un simple aménagement du système de rémunération ne suffit pas pour égaler l'incitation capitaliste très puissante que procure le fait d'être propriétaire ou d'espérer le devenir ou la perspective de laisser l'entreprise en héritage.

Mises pointe deux faiblesses opposées qui guettent la direction des entreprises socialistes :

- le manque d'initiative du fait de la dilution du pouvoir
- la prise de risques excessive parce que les dirigeants ne supportent pas personnellement les pertes

Ce tableau bien sombre présenté avec une pointe d'agressivité ironique est le point de départ d'une controverse qui durera un quart de siècle. Dans un premier temps, le débat se tint surtout entre économistes de langue allemande et peu d'idées neuves furent émises. Certains économistes socialistes comme Otto Neurath argumentèrent en faveur d'une comptabilité de flux en nature.

TAYLOR ET DICKINSON : LA POSSIBILITE D'UN CALCUL SOCIALISTE

Les historiens de la controverse considèrent qu'elle prit un nouveau départ avec les interventions de l'économiste américain Fred Taylor puis de l'économiste anglais H.D. Dickinson, tous deux défendant le socialisme. Ceux-ci s'attachent à montrer comment à partir des deux marchés aux extrémités (facteurs primaires et biens de consommation), un système de prix peut être établi, qui équilibre l'économie dans son ensemble. En fait, c'est ce

[1] Mises [375] p. 116

que fit Barone, mais son article n'est pas encore sorti de l'anonymat à ce moment. L'analyse des deux auteurs anglo-saxons ne dépasse d'ailleurs pas en profondeur celle de Barone.

L'article « The Guidance of Production in a Socialist State » de **Taylor**, qui paraît en 1929, ébauche assez sommairement un modèle socialiste. L'Etat y distribue un revenu monétaire aux membres de la collectivité, suivant une règle que Taylor ne présume pas. Il est calculé de telle façon que la dépense totale de ce revenu par l'ensemble de la population correspond à la somme des prix de tous biens de consommation mis sur le marché. Ainsi, en sélectionnant librement les produits qu'il consomme, le citoyen impose à l'industrie de privilégier les productions qui emportent sa préférence.

La condition d'équilibre sur laquelle Taylor insiste le plus est l'égalité entre le prix des produits et leur coût de production. Il assimile ce coût à la valeur totale des ressources productives que draine leur production. Par « ressources productives », Taylor semble entendre les facteurs primaires. Le calcul du coût implique que les facteurs primaires soient valorisés rationnellement. Taylor s'en tient à un terme vague : suivant leur « importance effective ». S'agit-il de leur productivité marginale ?

Taylor ne pondère pas le coût des facteurs primaires par la durée de leur présence dans le produit en devenir suivant l'idée de Böhm Bawerk. Sans doute sous-estime-t-il aussi la difficulté de déterminer la quantité de facteurs primaires cristallisée dans un produit quand de nombreuses entreprises interviennent dans sa fabrication, qui parfois se rendent des services mutuellement. S'en tenant à un haut niveau de généralité, Taylor ignore ce genre de problèmes.

Le modèle de Taylor lui semble équilibré : du fait de l'égalité entre prix et coût, le consommateur qui paye un bien compense exactement la société pour les ressources qu'il lui a fait consommer.

Taylor revient ensuite sur la valorisation des facteurs primaires, qu'il appelle le « problème de l'imputation ». Déterminer la valeur de chaque emploi de facteurs primaires dans l'économie semble une tâche irréalisable du fait de son ampleur. Mais Taylor propose sa solution : "…the so-called method of trial-and-error, that is the method which consists in trying out a series of hypothetical solutions till one is found which proves a success"[1]. Le bureau du plan affecte aux différents facteurs primaires des valeurs provisoires basées sur une analyse minutieuse ainsi que sur l'expérience ; l'économie fonctionne à l'aide de ces valeurs. Il faut ensuite détecter les déséquilibres et les corriger. La pénurie d'un facteur signifie qu'il a été sous-évalué ; le sous-emploi qu'il a été surévalué.

L'article de Taylor est beaucoup trop court pour embrasser un sujet aussi large. Assez sommaire, sa réputation lui vient de sa proposition du *trial-and-error* qui s'imposera comme une des idées maîtresses de la controverse. L'article de **Dickinson**, paru en 1933, est beaucoup plus fouillé. Son but explicite est de réfuter la thèse de Mises sur l'impossibilité du calcul socialiste et il n'hésite pas à concéder au marché les fonctions qu'il juge nécessaires pour convaincre les tenants de celui-ci.

Le cadre socialiste est celui qu'on retrouvait chez Barone et Mises. Dickinson ajoute néanmoins cette précision : les entreprises sont autonomes, mais intégrées dans une structure de trusts arborescente : toutes coopèrent en étant pleinement informées les unes à propos des autres. Contrairement à leurs consœurs capitalistes, les entreprises sont des « glass houses ». Opportunément, Dickinson attire l'attention sur une distinction importante : l'économie étant

[1] Taylor [356] p. 6

nationalisée, l'Etat assure seul la totalité des activités sociales. Toutefois, l'Etat social et politique et le secteur économique nationalisé sont bien deux domaines distincts fonctionnant selon leurs propres lois[1].

Les prix des biens de consommation peuvent trouver leur équilibre grâce à une tactique consistant à les faire fluctuer de façon à maintenir les stocks à un niveau stable.

L'équilibre est un peu plus compliqué pour les *biens d'ordre supérieur*[2]. Compte tenu des prix des biens de consommation, les entreprises sont en mesure d'établir des courbes de demande pour ces biens, et ce en remontant jusqu'aux facteurs primaires. La quantité disponible de ces facteurs étant connue, leur prix de marché est ainsi fixé au niveau qui en assure le plein emploi. Les prix des facteurs primaires déterminés, on peut calculer le coût de production des différents biens. Sur cette base, l'offre de ces biens peut être définie de la façon suivante : on cesse de produire ceux dont le prix ne couvre pas le coût et on augmente la production de ceux dont le prix est supérieur au coût[3], ce qui entraînera, le cas échéant, un réajustement du prix des facteurs.

Concernant les salaires, la société collectiviste fait face à un choix. Soit payer chaque travailleur à la valeur de marché de son travail, soit considérer celle-ci comme un artifice comptable couvrant des paiements effectués sur base d'un autre critère. Dickinson déconseille cette deuxième option, car la liberté de choix de l'emploi pourrait alors causer des déséquilibres sur le marché du travail.

Après avoir détaillé ces procédures devant conduire l'économie à l'équilibre, Dickinson estime qu'elles pourraient être remplacées par la résolution d'un système d'équations simultanées[4]. Contrairement à Barone, il n'énumère ni ne compte les équations et les inconnues de ce système. Barone ne considérait d'ailleurs pas les équations comme un outil pour trouver l'équilibre, mais comme un argument pour en démontrer l'existence et même la coïncidence avec l'équilibre de concurrence parfaite.

Voyons comment Dickinson envisage la planification des investissements : la première phase est la détermination d'une courbe de demande de capital. Le bureau du plan l'obtiendra en agrégeant les courbes de demandes individuelles des entreprises. A cette fin, il leur sera demandé de fournir un plan d'investissements à option : investissements prévus pour diverses valeurs du taux d'intérêt. Armée de la demande de capital, l'autorité planificatrice est face à une alternative : "two procedures can now be followed, according to whether interest-rate or quantity of accumulation is to be taken as the independent variable"[5]:

- La demande de capital combinée avec le taux d'intérêt permet de déterminer le besoin d'accumulation.
- La demande de capital et l'accumulation disponible déterminent ensemble le taux d'intérêt.

Si le bureau du plan choisit le taux d'intérêt comme variable indépendante, il tiendra compte du rythme du progrès technologique et des autres changements économiques (goûts des consommateurs). S'il préfère agir sur le volume de l'accumulation, Dickinson préconise l'application de la formule de Ramsey.

---

[1] Il en est ainsi dans le capitalisme où les acteurs des deux secteurs sont distincts. Le fait que l'Etat supervise les deux secteurs dans le socialisme n'implique nullement leur fusion.

[2] Dickinson recourt à la terminologie de Menger. Pour séduire les tenants de l'école autrichienne ?

[3] Dickinson estime que le coût externe (externalités) devrait être ajouté au coût de production.

[4] Dans un écrit ultérieur, il opérera une courbe rentrante à propos du système d'équations.

[5] Dickinson [76] p. 244

Ce taux d'intérêt sera surchargé d'une prime de risque qui sera variable d'une entreprise à l'autre et que l'on pourra calculer sur base de distributions de fréquences extraites des statistiques économiques. En régime socialiste, le risque sera amoindri, car les glass houses réserveront peu de surprises, mais il subsistera vu que les acteurs garderont leur liberté sur les marchés des biens de consommation et ceux des facteurs de production.

L'intérêt, la rétribution du facteur terre, les profits dus aux variations imprévues de la demande (déduction faite des pertes) alimenteront ce que Dickinson appelle le « Social Fund ». Celui-ci servira à financer les investissements ainsi que la production de biens sociaux produits par l'Etat politique et social.

La conclusion de Dickinson est presque lyrique : « The beautiful systems of economic equilibrium described by Böhm Bawerk, Wieser, Marshall and Cassel are not descriptions of society as it is but prophetic visions of a socialist economy of the future »[1].

HAYEK FAIT LE POINT

Hayek, prend évidemment le parti de son ami et collègue Mises. En 1935, il publie un ouvrage intitulé « Collectivist Economic Planning » qui fait le point sur le débat. Y sont compilés :

- l'article de Mises (1920)
- deux articles allant dans le même sens écrits par des auteurs moins connus
- une traduction de l'article de Barone (1908) entretemps redécouvert grâce à Schumpeter.
- une liste bibliographique des contributions au débat en provenance des deux camps.
- une introduction (« Nature and History of the Problem ») et une conclusion (« The Present State of the Debate »), de sa propre main.

Dans « Nature and History of the Problem », Hayek remarque que beaucoup, surtout parmi les socialistes, pensent que les théories économiques s'appliquent seulement au capitalisme et qu'on peut se contenter d'aborder le socialisme du point de vue social. Il dénonce cette conception. Elle semble prendre sa source chez Marx. Celui-ci pratiquait l'« historicisme » voulant que chaque période de l'histoire rencontre des problèmes qui lui sont propres. Parlant de Marx, Hayek écrit : "One may search his writings in vain for any definitive statement of the general principles on which the economic activity in the socialist community would be directed"[2].

La *planification* peut être considérée comme un *moyen* pour réaliser cette *fin* qu'est le *socialisme*. La science ne peut juger une fin, en l'occurrence ici une éthique de l'égalitarisme, qui est affaire d'adhésion personnelle. Par contre, l'analyse critique des moyens entre dans son champ.

Dans quelle mesure le socialisme peut s'allier à l'économie de marché, Hayek le considère comme une question à débattre, bien que cette éventualité semble lui inspirer une certaine méfiance. A contrario, un capitalisme planifié lui paraît « unthinkable ».

Concernant l'article de Mises, Hayek n'est pas avare de louanges : "The distinction of having first formulated the central problem of socialist economics in such a form as to make it

[1] Dickinson [76] p. 247

[2] Hayek [145] p. 13. Hayek a assurément raison. Marx fut particulièrement discret sur le socialisme en comparaison avec ses thèmes de prédilection : l'explication philosophique du corps social et de l'histoire, le fonctionnement du capitalisme et la lutte contre celui-ci.

impossible that it should ever again disappear from the discussion belongs to the Austrian economist Professor Ludwig von Mises"[1].

"The Present State of the Debate" analyse comment la position des penseurs socialistes a évolué à la suite de l'attaque de Mises: "But the great majority of the more recent schemes try to get around the difficulties by the construction of alternative socialist systems which differ more or less fundamentally from the traditional types against which the criticism was directed in the first instance and which are supposed to be immune against the objections to which the latter are subject"[2].

Hayek n'est absolument pas convaincu par la solution mathématique proposée par Taylor, Dickinson et d'autres. Il reconnaît que leur argument amène à corriger la critique telle qu'elle fut énoncée par Mises ; mais ses fondements demeurent valables. Parlant du calcul socialiste: "Now it must be admitted that it is not an impossibility in the sense that it is logically contradictory. But to argue that a determination of prices by such a procedure being logically conceivable in any way invalidates the contention that it is not a possible solution, only proves that the real nature of the problem has not been perceived"[3].

La mise en application de ces systèmes serait d'une complexité monstrueuse. Si le planificateur doit réellement reprendre la responsabilité des managers des entreprises privées, il devra s'occuper de mille et un détails. Les exemples donnés par Hayek ne sont pas tous probants, car il fait fi de la collaboration que le planificateur peut obtenir des dirigeants des usines.

Mais un autre problème est encore plus important : la technologie n'est pas simplement une donnée ; elle se construit progressivement. Lorsque plusieurs technologies différentes permettent de fabriquer un bien, la concurrence fait triompher la plus efficace. Comment le planificateur pourra-t-il opérer cette sélection ? En outre, sur le terrain, les ingénieurs trouvent régulièrement des petites améliorations. Selon Hayek, le planificateur devrait également les découvrir.

Le rassemblement des données et la résolution des équations constituent une tâche surhumaine. « The number of these unknowns will be equal to the number of commodities which are to be produced »[4]; or un même produit a deux endroits différents ou a des moments différents ou dans des conditionnements différents, ce sont chaque fois des biens différents. Hayek est ici en dessous de la vérité : dans le système de Barone, par exemple, il y a beaucoup plus d'inconnues que de biens, car les facteurs primaires et les coefficients de production, de même que l'épargne sont également des inconnues.

Certes, Taylor et Cie ne comptaient pas sur la résolution de toutes ces équations mais sur un processus de « trial and error ». Hayek n'y croit pas plus. Le déséquilibre sur un marché nécessitera de modifier des centaines de prix vu l'interdépendance générale. Et il faudra chaque fois tenir compte de l'élasticité de la demande, des liens de substitution et de complémentarité entre produits…

Hayek conclut donc : « No one would want to exclude every possibility that a solution may yet be found. But in our present state of knowledge, serious doubt must remain whether such a

[1] Hayek [145] p. 32
[2] Hayek [146] p. 202
[3] Hayek [146] pp. 207-208
[4] Hayek [146] p. 212

solution can be found"[1]. Si jamais la solution devait être trouvée, les socialistes devraient une fière chandelle à leurs adversaires.

LANGE : L'EQUILIBRE GENERAL

La réponse socialiste vient de l'économiste polonais Oskar Lange, parétien, qui publie « On the Economic Theory of Socialism », un article en deux parties parues respectivement en 1936 et 1937. Le début de cet article fait écho à la conclusion de Hayek, dont il reconnaît le bien fondé au moins partiellement : « Both as an expression of recognition of the great service rendered by him and as a memento of the prime importance of sound economic accounting, a statue of Professor Mises ought to occupy an honourable place in the great hall of the ministry of Socialisation or of the Central Planning Board of the socialist state »[2].

Mises déplorait que l'absence de prix en régime socialiste ne permettait pas de rationaliser le choix entre des alternatives. Lange fait appel à Wicksteed qui distinguait deux notions de *prix* :

1. sens étroit : somme d'argent à offrir pour acquérir une chose vendue
2. sens large : « terms on which alternatives are offered »[3]

Pour résoudre le problème de l'allocation des ressources, un système de prix au sens large suffit. Le bon fonctionnement de celui-ci nécessite :

1. une échelle de préférences guidant le choix
2. connaissance des termes de l'alternative
3. connaissance du stock de ressources disponibles

Il n'y a pas de raison de penser que les facteurs un et trois sont moins présents dans le régime socialiste. La présence du facteur deux est contestée, mais la variable permettant d'exprimer l'alternative est le taux marginal de transformation dépendant des fonctions de productions. « The administrators of a socialist economy will have exactly the same knowledge of production functions, or lack of knowledge, as capitalist entrepreneurs have »[4]. La sentence de Mises vient de sa confusion entre les deux sens du mot *prix*.

Pour contextualiser le sujet, Lange revient sur l'équilibre en concurrence parfaite. La réalisation de celui-ci implique la satisfaction de deux conditions :

- une condition SUBJECTIVE : chaque participant au marché optimise sa situation (qui son utilité, qui son profit). L'égalisation de l'utilité marginale par unité monétaire des produits consommés et celle de la productivité marginale par unité monétaire des facteurs utilisés en sont l'expression.
- Une condition OBJECTIVE : l'offre et la demande s'équilibrent sur tous les marchés des biens et facteurs.

Comment se réalise la condition subjective en régime socialiste ?

1- Les consommateurs tentent toujours de maximiser l'utilité pour un budget donné.

2- Les managers des entreprises, par contre, ne poursuivent pas la maximisation du profit. En lieu et place, ils sont tenus d'appliquer cette double règle :

---

[1] Hayek [146] p. 242.

[2] Lange [206] p. 53. Le ton se veut humoristique, mais la reconnaissance est sincère.

[3] On retrouvait déjà cette distinction dans l'article de Barone, écrit deux ans avant le « Common Sense » de Wicksteed. Pour traiter des prix dans le système socialiste, il employait le terme « équivalent ».

[4] Lange [206] p. 55

- Les entreprises choisissent la combinaison des facteurs et l'échelle de la production qui minimisent le coût moyen[1].
- Chaque branche produit exactement ce qui peut être écoulé à un prix égalant le coût de production

Ensemble, ces deux règles déterminent implicitement le nombre de firmes dans chaque branche. Puisque les prix des facteurs primaires résultent de l'offre et la demande et vu que ceux des produits sont fixés par le ministère, on est en présence de *prix paramétriques* ; l'offre de produits et la demande de facteurs sont déterminées.

3- Les travailleurs choisissent leur emploi de façon à optimiser le rapport entre le salaire et la désutilité.

4- Les ressources productives appartenant à l'Etat, qui en fixe le prix, sont dirigées vers les entreprises capables de les assumer budgétairement.

Les prix paramétriques étant fixés par le ministère, n'ont-ils pas un caractère arbitraire qui prive l'économie d'un optimum ? Rappelant la théorie walrassienne des prix, Lange insiste sur la quasi-unicité du système de prix qui assure l'équilibre simultané sur tous les marchés. Lorsque l'économie aura trouvé son système de prix d'équilibre, il sera aussi objectif qu'en régime capitaliste. Des pénuries et surproductions partielles résulteront inévitablement des erreurs de calcul, mais elles attireront l'attention du planificateur sur les correctifs à apporter pour garder l'équilibre. La réalisation de la condition objective repose sur la fonction paramétrique des prix ; le ministère de la production collectiviste doit donc imposer aux cadres des entreprises et des organes gérant les ressources naturelles l'usage exclusif des prix qu'il prescrit.

En définitive, « the central planning board performs the functions of the market »[2]. Aucune fonctionnalité du capitalisme n'est donc perdue.

Puisque Hayek admet la possibilité théorique d'un calcul socialiste mais qu'il doute de sa faisabilité, le processus d'itération correctrice avancé par Barone, Taylor et Dickinson devient le nœud de la problématique. Lange considère que l'exposé qu'en fait Taylor est parfaitement édifiant. Lange insiste surtout sur la similitude avec le capitalisme. Dans l'optique de l'école de Lausanne, le capitalisme a aussi son trial and error : le tâtonnement walrassien : ce qui marche d'un côté marchera également de l'autre côté.

Le bureau du plan a donc apporté une solution satisfaisante au problème économique sans devoir connaître l'ensemble des courbes de demande de tous les produits ni de résoudre des centaines de milliers d'équations, comme Hayek l'avait suggéré. Les seules équations à résoudre sont celles que les ménages et les unités de production résolvent inconsciemment dans leur quotidien.

Lange évoque également la question de l'accumulation du capital (il utilise bien ce terme). Il y a lieu de distinguer le court et le long termes. A court terme, le volume du capital disponible est fixe. Le taux d'intérêt se placera au niveau qui assure l'équilibre entre l'offre de capitaux par les banques d'Etat et la demande émanant des entreprises. A long terme, le volume du capital peut croître par l'accumulation. Lange semble croire que le volume optimum de capital implique l'annulation de sa productivité marginale nette, un objectif jamais atteint car tous les éléments qui rendent l'économie non stationnaire tendent à repousser ce volume toujours plus

[1] Beaucoup d'études empiriques mettent en doute la prédominance des fameuses courbes en U. Si la courbe est horizontale, le volume de la production ne peut pas être optimisé.
[2] Lange [206] p. 64

loin. La solution à établir porte donc sur le rythme de l'accumulation et non sur le volume du capital à atteindre.

Dans un système où c'est l'épargne des citoyens qui forme l'offre de capital, l'équilibre du marché reflète les préférences intertemporelles. Ce résultat est optimal du point de vue de l'économiste, mais Lange juge son mode de réalisation incompatible avec le socialisme. Il préconise donc l'option alternative : l'accumulation est réalisée au sein du monde des entreprises avant la distribution du dividende social aux citoyens. Comme conséquence, le taux d'accumulation sera arbitraire puisqu'il dépendra d'une décision du ministère. Pour Lange, il s'agit là d'un inconvénient du socialisme qu'il dit d'abord assumer comme tel. Toutefois, dans la deuxième partie de l'article (1937), il avance un contre-argument montrant qu'à cet élément non rationnel du socialisme correspond un élément aussi irrationnel au débit du capitalisme : « Besides, saving is also in the present economic order determined only partly by purely utility considerations, and the rate of savings is affected much more by the distribution of incomes, which is irrational from the economist's point of view »[1]. Lange pointe également à charge de l'épargne capitaliste le mal diagnostiqué par Keynes que la volonté d'épargner peut se muer en baisse du revenu national si la demande globale n'est pas suffisamment soutenue.

Lange, dans sa comparaison des mérites respectifs des deux systèmes épingle deux faiblesses potentielles du socialisme :
- l'efficacité des responsables de l'économie, sera-t-elle à la hauteur ?
- le risque de bureaucratisation de la vie économique.

Les avantages du socialisme qu'il mentionne sont :
- maximisation du bien-être plus poussée lorsque les écarts de revenus sont faibles
- externalités plus faciles à prendre en compte
- réduction des imperfections de la concurrence et du pouvoir des monopoles
- « Whatever the theoretical explanation of the business cycle, that cumulative shrinkage of demand and output caused by a cumulative reduction of purchasing power could be stopped in a socialist economy »[2]. Lange adopte un point de vue original et astucieux au sujet des crises économiques. Elles ont souvent comme point de départ un réel déséquilibre qu'il convient de corriger, mais la plongée de secteurs entiers de l'économie dans le marasme représente une externalité de cette correction qu'il y a lieu d'éliminer.

LERNER : L'EGALISATION DU PRIX AVEC LE COUT MARGINAL

Lerner poursuit le même but que Dickinson et Lange, mais estime qu'eux et pas mal d'autres auteurs prennent le problème par le mauvais bout. Son article « Statics and dynamics in socialist economies » (1937) commence ainsi: « This article is in the main a protest against the developing tradition, in approaching the problems of socialist economics, of starting from the consideration of general equilibrium, instead of going direct to the more fundamental principle of marginal opportunity cost »[3].

Le calcul socialiste pourra être considéré comme performant si le critère suivant est satisfait : « If we so order the economic activity that no commodity is produced unless its importance is greater than that of the alternative that is sacrificed ». Il faut que si le consommateur modifie sa consommation d'une façon qui laisse son utilité inchangée, le sacrifice de ressources par la

[1] Lange [207] p. 127
[2] Lange [207] p. 126
[3] Lerner [221] p. 253

société reste constant. Le sacrifice de l'individu et celui de la société ne peuvent pas diverger, sinon l'individu est incité à prendre des décisions non optimales. Mais comment obtenir ce résultat ? Lerner a trouvé dans la littérature sur ce sujet de multiples formules qui ressemblent à la règle correcte, faisant intervenir des variables comme le coût moyen, le profit normal etc., des formules qui sont elles-mêmes correctes dans certaines situations (concurrence parfaite par exemple), mais qui deviennent erronées dès que les circonstances s'en écartent. La règle correcte est très simple et peut être énoncée en deux versions :

1. Engage tel et tel facteurs de production, tant que son prix est inférieur à sa productivité marginale physique multipliée par le prix du produit.
2. Etends la production tant que le prix du produit est supérieur à son coût marginal[1]

Telle est donc la règle à imposer aux dirigeants des entreprises socialistes. Elle est valable, même en situation de *prix non paramétriques*. Dans ce cas, elle impose aux managers un comportement qui n'est pas celui qu'ils adopteraient spontanément, par lequel ils s'arrêteraient d'engager dès que la productivité marginale multipliée par le prix ne serait plus supérieure au *coût marginal du facteur*. Ce dernier est supérieur à son prix lorsque l'offre n'est pas horizontale. Un tel comportement mènerait l'économie à une position infra-optimale

Cette règle n'exclut pas que l'entreprise puisse gagner du profit au-delà du profit normal. L'absence de profit net est une caractéristique de l'équilibre à long terme, situation qui n'est normalement jamais atteinte ; même si elle était accessible, le profit existerait à court terme en l'absence de prévision parfaite.

Est-ce qu'une usine doit être remplacée dès que sa technologie est dépassée ? Selon Lerner, la réponse est simple : oui, si la valeur actuelle des revenus espérés est supérieure à l'investissement initial. Peu importe que l'investissement précédent soit amorti ou non. La règle ci-dessus suffit en tout état de cause[2]. Le coût d'une usine déjà construite cesse d'être un paramètre pertinent.

De façon un peu confuse, Lerner amène dans ce débat la question de la *tarification au coût marginal*, un thème parétien qu'on connaît surtout par le traitement qu'en fit Hotelling (1938)[3]. Le prix égal au coût marginal ne couvre pas les frais fixes. Les entreprises capitalistes ne peuvent pas se le permettre. D'après Lerner, les entreprises socialistes le peuvent[4] et c'est là un avantage du socialisme sur le capitalisme, car cela permet d'approcher l'optimum de plus près. Lerner ne précise pas l'organisation financière à mettre en place pour soutenir la tarification au coût marginal.

HAYEK : AU-DELA DU CALCUL

En 1940, Hayek publie « The Competitive Solution » qui est une réponse directe à Lange et Dickinson.

---

[1] On comparera cette instruction avec la minimisation du coût moyen imposée par Lange. En fait le numéro de la Review of Economic Studies d'octobre 1936 fait suivre l'article de Lange d'une critique par Lerner et de la réponse de Lange. Lerner y affirme que l'égalisation du prix avec le coût moyen n'est qu'une conséquence de l'application de la véritable règle, l'égalisation du prix avec le coût marginal, lorsque les conditions sont celles de la concurrence parfaite. Dans sa réponse, Lange admet la validité de cette opinion.

[2] Lerner se concentre sur l'aspect RENTABILITÉ mais ignore le volet SOLVABILITÉ. L'entreprise capitaliste qui agirait ainsi pourrait avoir du mal à digérer la perte de capital. On peut concevoir des modes de fonctionnement socialistes qui contournent ce problème.

[3] Cf. sous-chapitre 4.2.3.

[4] Il écrit : « In the socialist State it is much simpler, because there is no need for any particular firm or industry to cover its costs » ([221] p. 269).

Voyons d'abord une série de critiques qui ont trait à la capacité du ministère d'adapter le système de prix aux changements économiques avec toute la souplesse voulue :

- sur le principe, comment le ministère pourrait-il réagir aussi rapidement et adéquatement que les personnes directement intéressées présentes sur le terrain des opérations ?
- Lange et Dickinson basent leur argumentation sur la théorie pure de l'équilibre stationnaire ; le système proposé s'en ressent : il pourrait fonctionner dans une économie stationnaire mais sa vitesse de réaction est insuffisante pour suivre le cours changeant de l'économie. Changer un prix requerra une procédure pour s'assurer que la décision soit pertinente et prise par qui de droit.
- La fixation préalable des prix par le ministère est envisageable pour les produits standardisés, mais beaucoup de produits ne le sont pas. Nombre de contrats définissent un produit spécifique qui ne sera produit qu'une seule fois (construction, appels d'offres…)
- Des conditions exceptionnelles en un lieu donné ou en un temps donné sont susceptibles de justifier un prix particulier, s'écartant du prix normal.
- Les entrepreneurs ne s'intéressent pas qu'aux prix présents ; ils forment des anticipations. Comment intégrer celles-ci dans les prix paramétriques imposés ?

Hayek affiche également son scepticisme vis-à-vis de l'instruction donnée aux cadres des entreprises de produire au moindre coût. D'abord, l'incitant principal à la minimisation du coût, la concurrence, sera absent. Ensuite, on est ici face à une des lacunes de la science économique : elle considère les courbes de coûts comme des « given facts ». Or il faut DECOUVRIR la méthode la moins chère. Et ce sera peut-être un nouvel entrant qui la découvrira ; cet outsider aura alors la charge de convaincre le ministère que le prix doit être abaissé, car sa méthode est moins coûteuse.

Hayek s'interroge également sur les modalités de répartition du capital accumulé entre les entreprises socialistes. Comment sélectionner les investissements et comment tenir compte du risque émanant des projets conçus par ces managers qui manipulent des fonds dont ils ne sont pas propriétaires. Il n'y a d'ailleurs pas que la répartition du nouveau capital ; le capital existant doit pouvoir être redistribué le cas échéant.

L'appréciation du succès des managers se révélera aussi très complexes. Ceux-ci devront convaincre le ministère de la validité de leurs choix technologiques et commerciaux. On s'oriente vers une bureaucratisation.

Des écrits de Lange et Dickinson ressort l'impression que le ministère devra refaire tous les calculs de chefs d'entreprise pour les contrôler. Ces deux auteurs n'ont donc pas totalement abandonné la croyance en un ministère omniscient. Dans l'économie de marché, c'est la concurrence qui agrège la connaissance éclatée entre de nombreux individus. Sans la concurrence, l'agrégation de cette connaissance est impossible.

Dans sa conclusion, Hayek reconnaît quelques mérites à Lange et Dickinson : "As courageous attempts to face some of the real difficulties and completely to remold socialist doctrines in order to meet them they deserve our gratitude and respect"[1].

L'agrégation de la connaissance par le marché, évoquée ci-dessus, est un des thèmes favoris de Hayek et il lui a consacré à peu près à la même époque deux articles qui éclairent son point de

[1] Hayek [139] p. 208. Il faut mentionner la correction de tous les participants à cette controverse qui s'est tenue alors que l'oppression stalinienne battait son plein en Union Soviétique. Hors quelques allusions anecdotiques, le socialisme dont il fut débattu était le socialisme théorique ; les protagonistes considéraient tous que le débat portait sur le potentiel d'un mode de production restant entièrement à construire.

vue : « Economics and Knowledge » (1937) et « The Use of Knowledge » (1945). Le socialisme n'y est pas mentionné, mais sa présence en filigrane est perceptible. En voici les idées forces.

L'optimisation parétienne n'est pas le véritable problème auquel est confrontée la société. Les données permettant de résoudre le calcul économique ne sont pas concentrées dans un esprit mais fractionnées en une multitude de savoirs incomplets voire contradictoires. Le vrai problème est de tirer le meilleur parti de cette connaissance fragmentée.

Il n'y a pas d'activité économique sans planification ; cette planification est dite DECENTRALISEE lorsque chaque agent établit son propre plan et l'adapte aux circonstances extérieures ; elle est dite CENTRALISEE si tous les individus appliquent un plan global édicté par une autorité. Lequel des deux modes de planification est le plus efficace dépend du type de connaissance requis. Par exemple, on pourrait concevoir que pour la recherche scientifique, la gestion centralisée convienne mieux.

A côté de la connaissance scientifique, il existe une connaissance non organisée, plus pratique que théorique, qui peut d'ailleurs se révéler plus profitable à son possesseur que la connaissance théorique, « a knowledge of particular circumstances of time and place »[1], qu'on acquiert sur le terrain. Cette connaissance n'est pas agrégeable en statistiques, sinon au prix d'une perte de contenu importante.

L'homme de terrain dispose de l'information de terrain, mais celle-ci ne lui suffit pas ; elle doit être complétée par une information plus générale très synthétique, dépouillée de la multitude de détails qui caractérisent la connaissance économique. Le vecteur de cette connaissance existe : c'est le système des prix. Hayek donne cet exemple : l'étain est rendu plus cher, soit parce qu'un nouvel usage industriel en augmente la demande, soit parce qu'une source d'approvisionnement s'est tarie. Les producteurs d'étain doivent connaître la cause exacte de ce phénomène et trouveront certainement l'information. Pour les autres, il suffit de savoir qu'à l'avenir, il faudra utiliser l'étain avec plus de parcimonie et cette information leur est communiquée par la hausse de son prix.

Le système de prix doit donc être compris comme un vecteur de transmission de l'information. Ce système assure une économie d'information. Chaque acteur peut se permettre une connaissance partielle. La présence d'un acteur omniscient serait même inutile.

Chaque intervenant sur le marché participe inconsciemment à un système global qui le dépasse. Sans que ce soit délibéré, il y contribue. La volonté consciente n'aurait pu établir une telle « merveille », selon les termes de Hayek.

PREMIERES CONCLUSIONS

Après cette intervention de Hayek, la controverse a cessé en tant que telle. Les intervenants publieront bien des livres ou des articles sur le même thème au cours des décennies qui suivent, mais ces ouvrages cesseront de se répondre et répéteront ce qui a été dit précédemment.

Il faut toutefois mentionner l'article de Lange « The Computer and the Market » datant de 1967. Entretemps, les ordinateurs avaient été inventés et commençaient à se répandre. Lange pense pouvoir déduire de cet avènement une correction de ses interventions antérieures : "Were I to rewrite my essay today my task would be much simpler. My answer to Hayek and Robbins

[1] Hayek [140] p. 80

would be: so what's the trouble? Let us put the simultaneous equations on an electronic computer and we shall obtain the solution in less than a second. The market process with its cumbersome tatonnement appears old-fashioned"[1]. C'était une idée assez répandue dans les années soixante que l'essor des ordinateurs invalide l'objection selon laquelle le trop grand nombre d'équations empêcherait le planificateur de déterminer l'équilibre. Par exemple, Ernest Mandel commenta : « Aujourd'hui, à l'époque des machines à calculer électroniques qui effectuent des milliers d'opérations par minute, cette objection fait sourire »[2].

En réalité, le gros problème n'est pas de résoudre le système d'équations mais de le constituer. Prenons le système de Cassel. Il y a quatre groupes d'équations, dont celui que nous avions numéroté (6.2) au sous-chapitre 6.1.1 et qui indique que la quantité demandée du bien *i* est fonction de l'ensemble des prix :

$$D_i = F_i\,(p_1, p_2, \ldots, p_n) \quad \text{pour } i = (1 \ldots n) \qquad (9.4)$$

En quoi consiste F_i ? Equation linéaire ou curviligne ? Quels sont les coefficients ? En outre, la demande varie en fonction des circonstances. Il faudrait faire intervenir les états contingents (cf. sous-chapitre 6.3.2). Certes, l'informatique peut aider au traitement des données statistiques à partir desquelles les coefficients seront déterminés, mais la difficulté demeure immense. Le nombre de couples prix-quantité constatés empiriquement pour un produit dans un état donné ne sera généralement pas suffisant pour qu'une régression suffise à définir l'équation de la demande avec une certitude suffisante ; une intervention humaine s'avérera donc nécessaire. Tout ce travail doit être multiplié par le nombre d'équations.

Les intervenants socialistes s'estiment satisfaits ; ils ont prouvé que le calcul économique était possible dans le cadre du socialisme, ce qu'avait nié Mises. Il faut toutefois reconnaître qu'ils sont loin d'avoir répondu à toutes les critiques. Notamment, ils ont négligé l'objection relative à l'efficacité des cadres de l'économie socialiste.

En tant que disciples de Walras et Pareto, Lange et Lerner tendent à surestimer la place de l'équilibre statique dans la compréhension de l'économie. Leur équilibre général est sous-tendu par une conception purement passive de la concurrence, qui se résume à l'absence de pouvoir de manipulation des prix, à mille lieux de la concurrence qui anime l'économie réelle. Voici ce qu'en dit Schumpeter :

> "Neither Marshall and Wicksell nor the classics saw that perfect competition is the exception and that even if it were the rule there would be much less reason for congratulation than one might think"[3]
>
> "But in capitalist reality as distinguished from its textbook picture, it is not that kind of competition that counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (…) This kind of competition is as much more effective than the other as a bombardment is in comparison with forcing a door and so much more important that it becomes a matter of indifference whether competition in the ordinary sense functions more or less promptly"[4].

---

[1] Lange [209] p. 158
[2] Mandel [249] t.4 p. 126
[3] Schumpeter [329] p. 78. Outre Marshall et Wicksell, Schumpeter aurait pu mentionner la plupart des grands économistes.
[4] Schumpeter [329] pp. 84-85

Les écrits de Lange, Lerner, Taylor et Dickinson ne laissent aucune place à l'innovation.

Comme leurs contradicteurs socialistes, les libéraux se félicitent également de la tournure du débat. Hayek écrit:

> "It seems then that, on this point, the criticisms of the earlier socialist schemes have been so successful that the defenders, with few exceptions, have felt compelled to appropriate the arguments of their critics and have been forced to construct entirely new schemes of which nobody thought before"[1].

Plus loin, Hayek se demande dans quelle mesure, les modèles de Lange et consort sont encore socialistes. Ces interrogations sont symptomatiques du contexte intellectuel qui a baigné tout le vingtième siècle : l'hégémonie du marxisme dans le camp socialiste[2]. On en arrive à qualifier de « socialistes » des concepts qui sont simplement marxistes. La condamnation du marché anarchique et la préconisation de la planification comme solution sont caractéristiques du marxisme. A priori, ces éléments ne sont pas consubstantiels au socialisme.

9.3.3. Une vision moderne du socialisme de marché

SOCIALISME DE MARCHE ? MAIS ENCORE

On a beaucoup utilisé les termes « socialisme de marché » pour désigner le socialisme de Dickinson, Lange et Lerner. A bon escient ? Des marchés n'y existent que pour les biens de consommation et le travail. Les autres marchés existent virtuellement dans le cerveau du planificateur. Hayek écrit à ce propos :

> "But they both refuse to let prices determined directly in the market and propose instead a system of price fixing by a central authority, where the state of the market of a particular commodity, i.e., the relation of demand to supply, merely serves as an indication to the authority whether the prescribed prices ought to be raised or lowered"[3].

Hayek se demande pourquoi Lange et Dickinson ne laissent pas libre cours au jeu du marché, mais émet immédiatement un doute quant a la faisabilité d'un tel système.

Dans la littérature économique plus récente, le socialisme de marché renaît, mais cette fois avec de marchés réels à la place des marchés virtuels. Les arguments des opposants au système langien semblent avoir convaincu les tenants modernes du socialisme de marché.

Dans la suite de l'article, l'appellation « socialisme de marché » sera réservée à un système social caractérisé d'une part par la prépondérance de la propriété collective des moyens de production et d'autre part, par la décentralisation de toutes les décisions de consommation, de production et d'investissements au niveau d'agents mus par l'incitation au profit.

Dans son article "Post-Lange Market Socialism : an Evaluation of Profit-Oriented Proposals", Yunker recense quelques modèles de socialisme de marché sans viser l'exhaustivité. Il apparaît qu'il serait plus adéquat de parler DES socialismes que DU socialisme. Les constructions sont diverses et mêmes variables à l'infini. Ci-après, je présenterai les propositions de Roemer et Bardhan et j'évoquerai le débat sur la "soft budget control" qu'elles ont fait rejaillir.

---

[1] Hayek [139] p. 183.
[2] Malgré cette hégémonie, n'est-il pas révélateur que le socialisme a été défendu dans la controverse par des non-marxistes.
[3] Hayek [139] p. 185

BARDHAN ET ROEMER : COMMENT REMPLACER LE MARCHE FINANCIER

Ce ne sont pas moins de deux modèles que ces auteurs proposent dans leur article “Market Socialism : A Case for Rejuvenation” (1992). Les deux modèles ont en commun de placer un étage intermédiaire entre l’Etat et les entreprises en déléguant son autorité au système bancaire.

Le premier s’inspire du capitalisme japonais et met en scène des groupes d’entreprises liées qui s’articulent autour d’une banque (“main bank”) qui est le centre nerveux du groupe. Les entreprises sont des sociétés par actions : celles-ci sont détenues par parties par leurs travailleurs, par les entreprises liées et les travailleurs de celles-ci et bien sûr par la banque du groupe ; éventuellement aussi par des entreprises extérieures, des fonds de pension, des autorités locales. L’Etat détient la majorité des actions dans les banques[1].

L’une des préoccupations majeures des auteurs est de soumettre les dirigeants d’entreprise à une évaluation permanente. La banque opère un monitoring constant et tout particulièrement lorsque des actions d’une entreprise subissent une dévalorisation parce que les autres détenteurs tentent de les lui céder. L’alarme est ainsi donnée. La banque prend le contrôle de l’entreprise en difficulté ; les gestionnaires sont démis si leur responsabilité est en cause. La banque recapitalise l’entreprise si les perspectives de survie le justifient ; sinon, elle est mise en liquidation et ses actifs sont répartis entre les autres entreprises du groupe.

Les groupes ont une certaine homogénéité technologique, horizontale ou verticale, ce qui permettrait la mise en commun de ressources, des économies d’échelle, une meilleure spécialisation. Le groupe doit néanmoins être suffisamment large et diversifié pour limiter le risque de la banque. Dans le même ordre d’idées, la banque octroie une partie de ses crédits en dehors du groupe.

Le deuxième modèle réinstitue un marché financier, mais émoussé. Les entreprises émettent des titres à l’intention du public. Tout adulte se voit allouer un portefeuille standard. Les titres sont négociables en bourse contre d’autres titres, mais pas contre de l’argent. Les titres ne représentent pas la propriété des entreprises mais un droit dans le partage de leur profit. Ce procédé évite les gains et les pertes en capital ; les inégalités sur le marché des capitaux se limiteront aux écarts de rentabilité. Il sera impossible, aussi bien d’accumuler des fortunes mobilières que d’être écarté du marché financier. Les citoyens pourraient toutefois mettre leurs titres en nantissement pour obtenir un prêt bancaire.

Pour des raisons pratiques, les auteurs imaginent en fait un système à deux étages : les titres des entreprises sont détenus par des fonds mutuels et ce sont les titres émis par ceux-ci qui sont alloués au public.

Le marché émet donc des signaux relatifs à la santé des entreprises, mais le risque couru par les participants serait limité.

Les titres ne représentant pas la propriété des entreprises, le financement de celles-ci serait assuré par les banques. Les conseils d’administration des entreprises seraient composés de représentants des fonds mutuels et des banques d’investissement.

---

[1] Les auteurs reconnaissent que cette dépendance directe représente un risque d’interférence, mais ils comptent sur divers garde-fous mentionnés plus loin pour contrer les influences néfastes. De plus, l’Etat est actionnaire majoritaire mais pas unique.

Selon les auteurs, le premier modèle convient mieux pour des pays qui ne disposaient pas d'un marché financier très développé dans la phase antérieure.

Assez logiquement, l'indépendance des gestionnaires des banques et des entreprises par rapport à l'interférence du politique est un souci majeur des auteurs tout au long de l'article. Ils prennent soin d'établir des garde-fous :

- le fait que les firmes appartiennent à un groupe. La complicité entre le gouvernement et une entreprise du groupe s'avère plus difficile.
- Les gestionnaires des banques ont une réputation à défendre, qui devrait réduire leur inclination à céder aux pressions politiques pour consentir de mauvais prêts.
- La liaison de la rémunération des gestionnaires des banques à leur performance.
- L'indépendance des gestionnaires des banques et des entreprises peut faire l'objet d'une protection dans la législation voire dans la constitution.

Vu le risque de collusion entre les entreprises pour manipuler les prix à leur avantage, les auteurs préconisent le maintien d'une législation anti-trust.

Les auteurs exposent également leur opinion à propos d'une catégorie de modèles alternatifs : "worker-owned or labor-managed firms", des modèles assez répandus dans la littérature. Deux critiques leur sont généralement adressées :

1. Les travailleurs prenant part au vote des décisions impliquant l'avenir de la firme pourraient négliger les effets au-delà d'un horizon correspondant à leur présence dans l'entreprise.
2. Les projets maximisant le profit par travailleur pourraient être privilégiés au détriment de ceux qui maximisent le profit de l'entreprise, ce qui conduirait à un niveau d'emploi sub-optimal.

LA « CONTRAINTE BUDGETAIRE DOUCE »

Comme on l'a vu, Roemer et Bardhan ont mis un point d'honneur à immuniser la gestion de l'économie de l'influence du politique. Et pourtant, ils se feront encore reprocher leur faiblesse sur ce point appelé à devenir un sujet de débat majeur. L'un des thèmes principaux de ce débat est constitué par la "Soft Budget Constraint" (SBC), un concept introduit par l'économiste hongrois János Kornai en 1980.

Celui-ci définit la SBC ainsi :

> "The "softening" of the budget constraint appears when the strict relationship between expenditure and earnings has been relaxed because excess expenditure over earnings will be paid by some other institution, typically by the State. A further condition of "softening" is that the decision-maker expects such external financial assistance with high probability and this probability is firmly built into his behavior."[1].

Les moyens sont divers : subsides, taxation négociée, crédit à des conditions préférentiels, fixation de prix administrés avantageux. La situation inverse est la "Hard budget Constraint" (HBC): "The budget constraint is hard, if persistent loss is a matter of life or death"[2].

C'est l'expérience passée qui fonde l'espoir de la firme d'être assistée en cas de besoin et cet espoir affecte son comportement de diverses façons :

[1] Kornai [198] p. 4
[2] Kornai [198] p. 8

- sa courbe de demande des facteurs de production perd en élasticité ; elle est moins réactive aux variations de prix, ce qui génère une certaine rigidité de ceux-ci. D'où l'apparition d'une demande excédentaire de certains inputs
- la maximisation de l'effort n'est pas stimulée ; tout particulièrement, sont émoussés les sursauts d'adaptation ou d'innovation dans un contexte difficile
- développement de projets à la rentabilité douteuse.

Les motivations de l'Etat peuvent porter sur la protection de l'emploi, le soutien des entreprises nationales contre la concurrence étrangère, la justice sociale et la solidarité, la recherche de la stabilité et de la sécurité. La mentalité moderne tend à conférer à l'Etat un rôle providentiel ce qui lui transfère une part de la fonction allocative auparavant reconnue au marché.

Kornai présente le résultat de ses recherches empiriques sur l'importance de la SBC dans les économies de la Hongrie, de la Yougoslavie et de la Chine[1], ainsi que dans les économies occidentales : "Socialist economies exhibit a rather extreme degree of budget constraint softness. To a lesser degree and in more restricted segments of the system similar phenomena can be observed in mixed economies as well"[2].

En 1994 paraît l'article "The Politics of Market Socialism" par Schleifer et Vishny. Les auteurs prennent le contrepied de Roemer et Bardhan, affirmant l'impossibilité d'immuniser l'économie de pressions étatiques qui nuisent à l'efficacité. Leur argument pose d'abord que quel que soit le système économique en vigueur, capitalisme ou socialisme, il est impossible qu'un gouvernement démocratique favorise le progrès économique. Rapportant cette prémisse au contrôle accru des firmes par l'Etat dans le système socialiste, ils en concluent que le problème y est plus sensible qu'en régime capitaliste.

> So let us consider the case of a democratically elected socialist government in control of a nation's firms. Most market socialists presume that such a government will strive for efficient resource allocation. How likely is this objective to occur? (…) The futility of trying to insulate public firms from political pressures is best illustrated by the experience of public enterprises in western Europe. The experience with public enterprises suggests grave skepticism about the possibility of insulating public firms from the objectives of the government[3].
>
> The majority voting model predicts that the majority will redistribute resources from the minority to itself even at the cost of reduced efficiency (…) In light of the multiple examples of the tyranny of the majority, the claim that a majority will elect a government committed to economic efficiency is simply false[4].

On est ici à la frontière entre l'économie politique et la sociologie politique. Schleifer et Vishny se réfèrent à la "Public Choice Literature" qui ressortit à la seconde. Il ressort de ces références une conception misanthrope à la fois des électeurs et des hommes politiques. Dans leur réponse Roemer et Bardhan estiment que cette conception politique "displays a rather simple-minded (though popular in mainstream economics) theory of the State"[5].

---

[1] Nous sommes en 1986, avant la transition de ces pays vers le capitalisme.
[2] Kornai [198] p. 21
[3] Shleifer & Vishny [333] p. 171
[4] Shleifer & Vishny [333] p. 177
[5] Bardhan & Romer [23] p. 177.

A mon avis, cette sociologie politique néglige deux aspects de la psychologie des électeurs :

- Les électeurs ont pour la plupart une multitude d'affiliations qui se chevauchent partiellement. Interviennent des critères religieux, locaux, politiques, idéologiques, économiques, ethniques, de sexe, de santé, d'âge, de loisir… De ce fait, il n'y a pas de relation mécanique entre l'appartenance à une classe sociale et la préférence électorale. Elles ne coïncident pas nécessairement : certains riches votent à gauche et, plus nombreux, des membres de catégories modeste votent à droite.
- En outre, les électeurs tiennent à leur niveau de vie et souhaitent en général une situation politique stable et consensuelle. Il y a une contradiction à croire que les électeurs vont plébisciter des partis qui maximisent leur part du gâteau mais en même temps sont insensibles à la taille totale de celui-ci.

La conséquence de cette complexité est le rôle du compromis dans la démocratie moderne, compromis qui suscite une réelle adhésion sociale même s'il ne satisfait parfaitement personne[1].

Schleifer et Vishny en appellent à l'expérience de l'Europe occidentale où des électeurs ont épisodiquement sacrifié le progrès économique à des intérêts particuliers. Pointer leur irrationalité serait une mauvaise interprétation. Les intérêts contradictoires au sein de la société sont tels qu'il peut être rationnel pour une catégorie sociale de refuser un progrès si elle craint de ne pas en profiter suffisamment[2]. On pourrait même en déduire que dans le socialisme, si les fruits de la croissance sont mieux partagés, les attentes de la population d'une politique pratiquant la SBC seront réduites.

En conclusion, l'électorat ne rejettera pas un progrès économique si le ruissellement de ses avantages est attendu. Mais le public n'est pas idiot ; il demande à voir les sillons du ruissellement avant de se prononcer.

[1] La démonstration d'Arrow que la fonction de bien-être social est impossible pèche en ceci qu'elle néglige totalement cet aspect de la question.

[2] Le compromis social se construit par l'opposition, mais il ne serait pas correct de considérer celle-ci comme une guerre sociale.

9.4- LE RISQUE ET L'INCERTITUDE

Dans un premier temps, les économistes ont analysé les comportements des agents dans un environnement empreint de certitude. Il était présumé implicitement que chacun connaissait les contextes futurs et pouvait choisir ses actions en pleine connaissance de cause. L'incertitude n'avait droit qu'à quelques allusions. Inévitablement, l'analyse des comportements en situation incertaine était appelée à devenir un objet d'investigation.

Comme nous le verrons, la théorie de l'incertitude économique est traversée par une ligne de démarcation entre deux courants opposés : selon les mots d'Arrow, « those which use exclusively the language of probability distributions and those which call for some other principle, either to replace or to supplement »[1].

9.4.1. La dispute philosophique

KEYNES : LE DEGRE DE CROYANCE NON MESURABLE

Le « Treatise on Probabilities » (1921) est un traité qui aborde cette matière, non sous l'angle mathématique, mais psycho-logico-philosophique. Keynes souhaite conserver l'appellation « probabilité » pour tous les événements probables sans restrictions. Parmi ces probabilités, certaines sont susceptibles de supporter un traitement mathématique, d'autres non.

Soit une proposition dont nous ne savons pas avec certitude si elle est (ou se révélera) vraie ou fausse. L'individu qui souhaite se faire une opinion à ce sujet peut lui accorder un crédit plus ou moins élevé. C'est là que Keynes fonde sa théorie des probabilités. « I have spoken of probabilities as being concerned with degrees of rational belief »[2]. Keynes insiste sur le caractère rationnel de la démarche. L'individu met en relation la proposition examinée et une connaissance préalable constituée de faits rendus plus ou moins certains par le fait qu'ils ont été intégrés par l'expérience. Cette optique exclut explicitement les croyances irrationnelles. Des individus distincts formeront éventuellement des probabilités différentes à propos d'une même proposition, parce que leur connaissance de base ne concorde généralement pas.

Puisque les probabilités sont un degré de croyance, la question se pose de savoir si et quand elles sont mesurables sur une échelle numérique. Keynes regrette que le calcul des probabilités ait reçu beaucoup plus d'attention que leur logique. A côté des probabilités qui peuvent être exprimées numériquement, il y a les autres ; parmi ces dernières, il n'est même pas toujours possible de les ranger par ordre de probabilité. C'est la nature de certaines propositions qui les rend inaptes à l'expression numérique, non pas le fait que l'esprit manque de l'intelligence nécessaire pour découvrir l'expression numérique. « It has always been agreed that a numerical measure can actually be obtained in those cases only in which a reduction to a set of exclusive and exhaustive equiprobable alternatives is practicable »[3]. Mais précisément, Keynes montre que la division en alternatives équiprobables, ce qu'il appelle le *principe d'indifférence*, est loin d'être toujours valide.

Keynes insiste également sur le fait que des propositions alternatives peuvent différer, non seulement par la balance entre les arguments favorables et défavorables qui affecte différemment leurs probabilités, mais également par la masse d'informations à laquelle le

[1] Arrow [11] p. 410
[2] Keynes [187] p. 21
[3] Keynes [187] p. 70

jugement fait appel ; de ce fait, différentes probabilités peuvent mériter des pondérations inégales.

Keynes critique la principale théorie rivale, la *théorie des probabilités par les fréquences*, largement répandue à son époque. Il cite l'un des tenants de cette théorie: « If the probability of a given event be correctly determined, he says, 'the event will on a long run of trials tend to recur with frequency proportional to their probability' »[1]. S'il est vrai que cette théorie trouve certaines applications valides, néanmoins un nombre important de jugements échappe aux conditions de son applicabilité. « Part of the plausibility of (this) theory is derived, I think, from a failure to recognise the narrow limits of its applicability »[2].

RAMSEY: LE DEGRE DE CROYANCE MESURABLE

Dans son essai « Truth and Probability » (1926), Ramsey s'oppose frontalement à la conception keynésienne. Il s'attache à élaborer ce que Keynes considérait comme impossible: « we must therefore try to develop a purely psychological method of measuring belief. It is not enough to measure probability; in order to apportion correctly our belief to the probability we must also be able to measure our belief »[3]. La solution qu'il propose considère le lien causal de nos croyances vers nos actes: « But when we seek to know what is the difference between believing more firmly and believing less firmly, we can no longer regard it as consisting in having more or less of certain observable feelings; at least, I cannot recognise any such feelings. The difference seems to me to lie in how far we should act on these beliefs »[4].

Ramsey base son raisonnement sur un principe qu'il sait être rejeté par la psychologie moderne mais qui peut avoir un fondement dans le cas de comportements conscients du type de ceux qu'on retrouve dans la vie professionnelle : nous agissons de façon à réaliser nos désirs. En conséquence, nos comportements sont déterminés par ceux-ci et par nos opinions de probabilité : « if *p* is a proposition about which he is doubtful, any goods or bads for whose realisation *p* is in his view a necessary and sufficient condition enter into his calculations multiplied by the same fraction, which is called the 'degree of belief in *p*'. We thus define degree of belief in a way which presupposes the use of mathematical expectation »[5]. Ramsey va même jusqu'à établir une ébauche d'axiomatisation de ce qu'il considère comme un comportement rationnel. Il appelle ces règles « the logic of consistency in partial belief » ; elles sont une extension de la logique formelle. Les lois des probabilités sont directement liées à cette logique de cohérence. L'ensemble des croyances d'un individu forme un ensemble structuré : de telle croyance, on peut déduire telle autre mais pas telle troisième. L'esprit qui ne respecterait pas ces règles s'écarterait des chemins de la rationalité.

La logique de cohérence ne suffit pas pour faire fonctionner l'esprit humain. L'observation et la mémoire sont également génératrices de croyances. « We want our beliefs to be consistent not merely with one another but also with the facts »[6]. L'observation de nouveaux faits fait réviser nos probabilités. Ici également, certains modes de pensée sont plus rigoureux que d'autres. Quelques bonnes pratiques peuvent être édictées, que Ramsey appelle « the logic of truth ». Ramsey recommande la pratique de l'*inférence inductive*, ce qui remet à l'avant-plan la *théorie*

[1] Keynes [187] p. 101
[2] Keynes [187] p. 104
[3] Ramsey [294] p. 166
[4] Ramsey [294] p. 170
[5] Ramsey [294] p. 175
[6] Ramsey [294] p. 191

des probabilités par la fréquence. Notre esprit affecte des probabilités à des croyances en fonction de la fréquence observée[1]. L'inférence inductive n'est pas un outil parfait : il est impossible de prouver que la pratiquer aboutit à un résultat juste, mais s'en priver serait un handicap.

9.4.2. Knight : distinction entre le risque et l'incertitude

L'ouvrage « Risk, Uncertainty and Profit » de Knight (1921) est le premier ouvrage majeur d'économie politique consacré aux effets de l'incertitude sur la structure et le fonctionnement de l'économie. L'auteur regrette le peu d'intérêt que le sujet avait suscité auparavant.

Knight écrit: « The universal form of conscious behavior is thus action designed to change a future situation inferred from a present one »[2]. Comme Keynes et Ramsey, Knight philosophe à propos des modalités de cette inférence. L'esprit réduit la part de hasard inexplicable en groupant les situations possibles en cas homogènes auxquels une probabilité peut être appliquée (classification). Suivant le type de contingence rencontrée, Knight distingue trois modes de formation des probabilités :

- Le raisonnement a priori de type logique et mathématique
- L'inférence statistique présumant l'analogie entre le passé et le futur
- L'estimation, mode de pensée intuitif ; c'est une manière de contourner l'absence de base valide pour classer les occurrences.

Les décisions dans la sphère économique concernent généralement des situations très spécifiques ; les agents doivent alors recourir à l'estimation. Dans les affaires, l'intuition l'emporte sur le raisonnement. Certes, même dans le processus d'estimation, le businessman forme un jugement probabiliste et même un double jugement probabiliste, puisque sur base de son expérience, il forme également une opinion sur sa propre capacité de former des jugements. Toutefois, l'*estimation* n'a d'une probabilité que la FORME ; sa NATURE est différente : « it is a ratio, expressed by a proper fraction. But in fact, it appears to be meaningless and fatally misleading to speak of the probability, in an objective sense, that a judgement is correct »[3].

A ce stade, Knight établit une distinction qui est resté gravée dans l'histoire de la science économique, même s'il serait erroné de prétendre qu'elle fait l'objet d'un consensus : « To preserve the distinction which has been drawn in the last chapter between the measurable uncertainty and an unmeasurable one we may use the term '*risk'* to designate the former and the term '*uncertainty'* for the latter »[4]. C'est bien l'INCERTITUDe qui concentre l'intérêt de Knight. Selon lui, toutes les incertitudes économiques se résument au choix entre une rémunération inférieure plus sûre et une rémunération supérieure plus aléatoire.

Le comportement rationnel visera à réduire l'incertitude. Il y a deux moyens principaux :

- La **consolidation**, qui consiste à regrouper les occurrences possibles en classes. L'exemple le plus évident est celui des assurances.
- La **spécialisation**, par laquelle le pouvoir de décision est attribué en principal aux agents les plus aptes (intellectuellement et psychologiquement) à faire face à l'incertitude. Les différences individuelles en cette matière sont considérables.

---

[1] Ramsey reconnaît qu'on ne peut pas passer directement des fréquences observées à la probabilisation d'une croyance. Le lien n'est qu'indirect et pas toujours précis.
[2] Knight [189] § 7.8
[3] Knight [189] § 7.46
[4] Knight [189] § 8.1

- Knight y ajoute encore un moyen complémentaire : la **diffusion**. Réparti sur un grand nombre de têtes, le risque est plus supportable.

La *libre entreprise* est un système mis en place par la société pour réduire l'incertitude par la spécialisation ; il consiste à faire supporter l'incertitude par une classe particulière : les entrepreneurs. Son évolution historique réduit toujours plus l'incertitude par l'augmentation de la taille des entreprise (consolidation), le passage de l'entreprise individuelle à société d'associés puis la société d'actionnaires (consolidation dans le chef de l'investisseur qui diversifie son portefeuille et diffusion du risque entre les actionnaires).

Knight pense que la fonction d'entrepreneur doit son existence à la présence de l'incertitude. Sans celle-ci, il y aurait simplement des managers, dont l'activité, certes hautement qualifiée, relèverait de la routine. L'entrepreneur paie une rémunération contractuelle aux facteurs travail, capital et terre et s'octroie le résidu. Evidemment, il n'assume l'incertitude de son revenu que moyennant l'obtention du pouvoir de décision qu'il monopolise, à tout le moins en dernier ressort. Les facteurs de production lui délèguent le pouvoir, car ils sont délestés de l'incertitude[1]. Knight parle d'une « céphalisation » de l'entreprise. Toutes les parties y trouvent un avantage. « It is unquestionable that the entrepreneurs' activities effect an enormous saving to society, vastly increasing the efficiency of economic production »[2]. On ne pourrait surestimer l'importance de l'incertitude : Knight fait remonter à elle l'entreprise[3] elle-même et le système salarial.

Dans la réalité économique complexe, peu de revenus sont purement résiduels ou purement contractuels ; la plupart associent les deux composantes, mais dans des proportions diverses. L'entrepreneur reçoit une rémunération contractuelle (salaire, rente ou intérêt), implicite ou officielle, pour ses tâches de routine. Parallèlement, les facteurs de production voient généralement leur revenu varier directement ou indirectement avec les affaires de l'entreprise. Généralement, il n'est pas possible de déterminer les montants respectifs avec précision.

Le niveau du profit[4] qui échoit à l'entrepreneur dépend de sa capacité de négociation-marchandage lorsqu'il acquiert les facteurs de production sur leurs marchés respectifs. Plus nombreux sont les entrepreneurs, moins ils sont en position de force sur ces marchés ; leur rémunération s'en ressentira. Knight pense que la complexité de la gestion augmente sensiblement avec la taille de l'entreprise ; les individus capables de gérer les grandes entreprises, peu nombreux, perçoivent donc une haute rétribution pour leurs services.

Le mot « entrepreneur » revient régulièrement dans cet exposé ; mais qui est au juste l'entrepreneur dans l'économie contemporaine ? Après avoir laissé planer une certaine ambiguïté, Knight développe enfin le sujet. Une fois n'est pas coutume, j'exposerai d'abord ma propre conception et je montrerai ensuite qu'elle est étayée par quelques citations de Knight.

---

[1] Knight insiste sur la nécessité pour l'entrepreneur d'inspirer la confiance à ses partenaires pour que ceux-ci ressentent effectivement une réduction de l'incertitude. Cela se produit essentiellement par la contagion de la confiance que l'entrepreneur a en ses propres capacités.

[2] Knight [189] § 9.24

[3] Cette explication de l'entreprise sera contestée par les auteurs qui produiront des explications alternatives (cf. chapitre 9.8).

[4] Knight préconise l'usage du terme « profit pur » pour le profit résiduel, ce qui permet d'éviter la confusion avec la rémunération complète de l'entrepreneur que la littérature qualifie souvent de « profit ». Le *profit pur* se calcule par la déduction d'un intérêt, d'une rente et d'un salaire compétitifs fictifs, du profit.

L'entrepreneur est une fonction tellement particulière qu'on ne peut la considérer comme un facteur de production en soi. Ne peut occuper cette fonction que l'acteur qui œuvre déjà comme facteur de production, en tant que travailleur, capitaliste ou propriétaire terrien[1]. L'entrepreneur perçoit le profit résiduel, mais on ne peut pas considérer celui-ci comme la rémunération de la fonction entrepreneuriale. C'est heureux pour lui, car ce résidu peut être négatif aussi bien que positif. La science économique le fait d'ailleurs tendre vers zéro en concurrence parfaite[2]. C'est un des facteurs de production qui assume la fonction d'entrepreneur, le plus souvent le capital par actions. Ce qu'on appelle *profit résiduel* n'est alors que la modulation de l'intérêt perçu par ce capital par les résultats de l'entreprise. Cet intérêt inclut évidemment une prime de risque ; c'est ce dernier élément qui rémunère l'incertitude et non le profit résiduel lui-même qui n'est pas la rémunération mais motive la prime de risque. Voyons ce qu'en dit Knight :

> « The apparent separation between control and risk taken turns out, as predicted, to be illusory. The paradox of the hired manager, which has caused endless confusion in the analysis of profit, arises from the failure to recognize the fundamental fact that in organized activity the crucial decision is the selection of men to make decisions, that any other sort of decision-making or exercise of judgment is automatically reduced to a routine function »[3].
>
> « To whom is the responsibility ultimately transferred when the entire conduct and policy of a business are in the hands of a hired manager? The answer is obvious: to the owners of the productive services used in the business »[4].
>
> « It is impossible for entrepreneurship to be completely specialized or exist in a pure form (…) The natural result is a complicated division or diffusion of entrepreneurship, distributed in the typical modern business organization by a hierarchy of security issues carrying every conceivable gradation and combination of rights to control and to freedom from uncertainty… »[5].
>
> « In most cases, it would not be fruitful to attempt an accurate separation of profit from interest (…) The entrepreneur must almost of necessity own some property and the owner of property used in a business can hardly be freed from all risk and responsibility »[6].

Knight lie ensuite cette analyse de l'entrepreneur avec ses considérations précédentes : c'est l'incertitude et non le risque qui produit le profit. « Profit arises out of the inherent, absolute unpredictability of things, out of the sheer brute fact that the results of human activity cannot be anticipated and then only in so far as even a probability calculation in regard to them is impossible and meaningless »[7].

Knight se montre sceptique à l'encontre de la thèse courante de l'*aversion pour le risque*. Le capitalisme jouit de l'avantage majeur que sa classe dominante est joueuse. Bien qu'ayant moins à perdre, les dirigeants d'une économie socialiste se montreraient plus conservateurs.

---

[1] Il peut éventuellement cumuler plusieurs de ces rôles.

[2] Certes, la théorie prévoit un profit résiduel positif en concurrence imparfaite, mais même dans cette situation, la prévention de l'entrée de nouveaux concurrents peut l'annuler.

[3] Knight [189] § 10.11

[4] Knight [189] § 10.14

[5] Knight [189] § 10.16

[6] Knight [189] § 10.22

[7] Knight [189] § 10.33

Selon Knight, les entrepreneurs n'ont pas besoin du profit comme incitant à prendre des risques. Il suggère même qu'à l'échelle de l'économie globale, le profit pur pourrait se révéler négatif. Cet avis n'est étayé par aucun argument ni aucune donnée. Comme il le reconnaît lui-même, il s'agit simplement d'une impression générale.

9.4.3. Théorie de l'utilité espérée

Ramsey, dans l'essai analysé ci-dessus, a montré la voie à la conception de l'incertitude qui dominera l'économie néoclassique. Cette conception s'appuie sur deux idées :

- **Les probabilités subjectives** : dans la majorité des situations où une décision doit être prise, les conséquences n'apparaissent pas à l'individu assorties de probabilité objectives comme ce peut être le cas lors d'une loterie. En l'absence de probabilités objectives, l'agent construit intuitivement des probabilités sur lesquelles il basera ses choix.
- L'idée que les agents agissent de manière à maximiser l'espérance mathématique de leur utilité. C'est la **théorie de l'utilité espérée**. Ses racines remontent au XVIIIe siècle.

LE PARADOXE DE SAINT-PETERSBOURG

En 1713, le mathématicien suisse Daniel **Bernoulli** expose le problème paradoxal suivant : proposons un jeu de pile ou face non truqué. On procède au nombre de jets nécessaires jusqu'à ce que *face* apparaisse une première fois. Le gain du joueur vaut 2^n ducats où n est le nombre de jets nécessaires. Combien serions-nous prêts à payer pour participer à ce jeu ? Le paradoxe vient de ce que l'espérance mathématique de gain vaut l'infini, alors que personne ne paierait cher pour y participer. C'est son frère Nicolas qui avança l'explication en 1738. En fait, l'espérance mathématique du gain n'intervient pas dans le choix ; c'est l'espérance mathématique de l'utilité du gain. Or celle-ci augmente moins que le gain. Ce disant, Nicolas Bernoulli est le premier précurseur de la théorie de l'*utilité marginale décroissante*.

VON NEUMANN ET MORGENSTERN

Dans l'ouvrage « Theory of Games and Economic Behavior » déjà mentionné au sous-chapitre 6.1.2, la *théorie de l'utilité espérée* ne fait l'objet que d'une quinzaine de pages qui sont une simple introduction aux développements qui suivent. Pourtant, ces pages eurent rapidement un retentissement considérable.

Les auteurs s'intéressent au système de préférence d'un individu, comme l'ont fait leurs prédécesseurs parétiens lorsqu'ils ont conçu les courbes d'indifférence. Simplement, ils élargissent les préférences à des situations comportant des alternatives stochastiques. Par exemple, le choix entre une possibilité A certaine et un événement qui sera soit B à une probabilité α, soit C à une probabilité $(1-\alpha)$. Les évènements B et C doivent être mutuellement exclusifs. Est sous-jacente à ce système de préférence, une notion que les auteurs appellent « utilité » avec toutes les précautions d'usage. Dans l'exemple ci-dessus, si l'agent préfère la perspective A (d'utilité a) à la perspective [B ou C] (d'utilité b), cette préférence peut être notée $a > b$[1]. Ceci ne signifie pas encore qu'il y a moyen de quantifier a et b, mais c'est justement la possibilité de cette quantification que les auteurs veulent démontrer.

[1] Attention : a et b ne sont pas des gains mais des perspectives offertes à l'individu, appréciées sous l'angle de l'avantage qu'elles procurent. Les auteurs ne font jamais intervenir directement les gains dans leur modèle : la comparaison porte toujours sur les perspectives elles-mêmes.

Le point de départ du raisonnement est ce double élément :

$a > b$ (9.5-A)

$\alpha.a + (1-\alpha).b$ (9.5-B)

Ces deux expressions signifient respectivement la possibilité de hiérarchiser des utilités et la possibilité de les combiner avec des probabilités. Ce sont les seules véritables hypothèses du modèle.

Le but est de prouver qu'à chaque utilité a, b…, on peut associer un nombre réel ρ qui est la valeur d'une fonction v(a), v(b)… qui respecte la double condition (9.6) :

si $a > b$, alors $v(a) > v(b)$ (9.6-A)[1]

$v(\alpha.a + (1-\alpha).b) = \alpha.v(a) + (1-\alpha).v(b)$ (9.6-B)

Préalablement, les auteurs démontrent que si, pour un système de préférences, plusieurs fonctions respectent cette double condition, elles sont nécessairement liées entre elles par le fait qu'elles sont mutuellement des transformations linéaires. Autrement dit : pour ρ, ρ' respectant (9.6), on a $\rho' = k.\rho + l$. Les auteurs considèrent qu'il s'agit là d'une relation suffisamment restrictive pour que les fonctions v(.) restent significatives[2] ; on en trouve régulièrement de semblables en physique, par exemple pour les températures.

Pour assurer les relations (9.6), il faut postuler certaines propriétés du système de préférences. Autrement dit, il faut établir une base axiomatique. « A choice of axioms is not a purely objective task (…) The axioms should not be too numerous, the system is to be as simple and transparent as possible, and each axiom should have an immediate intuitive meaning by which its appropriateness may be judged directly"[3].

Les axiomes choisis constituent évidemment le cœur de la démonstration.

1- Le système de préférences est complètement ordonné. Pour tout couple d'utilités *a* et *b*, on aura toujours entre elles $a > b$, $a = b$ ou $a < b$.
2- Les préférences sont transitives : si $a > b$ et $b > c$, on aura nécessairement $a > c$.
3- $a > b$ implique $a > \alpha.a+(1-\alpha).b$, ce qui est justifié par le caractère mutuellement exclusif des événements
4- $a > b > c$ implique l'existence d'un α tel que $\alpha.a + (1-\alpha)c > b$. « This is a plausible continuity assumption »[4].
5- $\alpha.a + (1-\alpha).b = (1-\alpha).b + \alpha.a$
6- $\alpha.[\beta.a + (1 - \beta).b] + (1 - \alpha).b = \gamma.a + (1 - \gamma)b$, où $\gamma = \alpha.\beta$. Une application d'alternatives stochastiques successives (ici 2 étapes : α et β) peut toujours correspondre à une application unique (ici : $\gamma = \alpha.\beta$)

L'équation (9.6-B) signifie que la fonction d'utilité d'une perspective stochastique prend comme valeur l'espérance mathématique des utilités des événements possibles pondérées par leur probabilité. Une caractéristique de cette fonction est donc que la valeur qu'elle prend n'est pas influencée positivement ou négativement par l'attitude de l'individu face au risque. Les auteurs s'interrogent si on pourrait les accuser de construire un système visant à occulter l'attitude face au risque mais ils ne trouvent d'indice d'une telle culpabilité dans aucun de leurs axiomes.

---

[1] Le signe « > » a une signification différente dans a >b (simple préférence) et dans v(a) >v(b) (supériorité numérique).

[2] Sans l'intervention des probabilités α et 1-α, toute transformation monotone eût été acceptable.

[3] Von Neumann & Morgenstern [377] p. 25

[4] Von Neumann & Morgenstern [377] p. 27

Terminons avec quelques mises au point. D'abord, les auteurs estiment que leur modèle se marie mal avec les probabilités SUBJECTIVES ; ils le trouvent mieux assorti aux probabilités formées à partir des fréquences constatées sur le long terme. Cette affirmation est toutefois peu étayée. Ils rappellent également cette évidence que leur fonction d'utilité ne peut servir à des comparaisons ou des additions interpersonnelles. Enfin, ils reconnaissent que « we cannot avoid the assumption that all subjects of the economy under consideration are completely informed about the physical characteristics of the situation in which they operate and are able to perform all statistical, mathematical etc., operations which this knowledge makes possible »[1]. Mais pensent-ils, leur théorie des jeux basée sur cette hypothèse explique valablement certains comportements qu'on attribuerait prima facie à l'information imparfaite ; cette dernière interprétation ne serait donc pas nécessairement la plus adéquate.

L'AXIOME D'INDEPENDANCE

A la fin des années quarante et au début des années cinquante, beaucoup d'auteurs s'emparèrent de la théorie de l'utilité espérée pour la critiquer, pour la défendre, pour la réinterpréter ou l'exposer avec d'autres formules mathématiques. Un colloque fut même tenu à Paris.

Parmi les auteurs qui réexposèrent cette théorie, **Samuelson** est l'un des partisans de ce qu'on a appelé l'*axiome d'indépendance*.

Dans son article « Probability, Utility and the Independence Axiom », Samuelson commence par définir :

- Une ***situation de probabilité de revenu simple*** : c'est une loterie qui rapporte des lots X_1, $X_2 \ldots X_n$, respectivement avec des probabilités mutuellement exclusives x_1, $x_2 \ldots x_n$ (dont la somme vaut évidemment 1).
- Une ***situation de probabilité de revenu composée*** : c'est un ticket de loterie dont les lots sont eux-mêmes la participation à d'autres loteries. Par exemple :[(X1, X2 ; ½, ½),(X1, X3 ; ¼ ; ¾) ;⅔,⅓)] offre à deux chances sur trois le gain X1 ou le gain X2 à une chance sur deux ou à une chance sur trois le gain X1 à une chance sur quatre ou le gain X3 à trois chances sur quatre.
- Une ***situation de probabilité de revenu associée*** à une situation composée est la situation simple qu'on obtient en prenant tous les lots de toutes les loteries et en calculant leur probabilité en multipliant la probabilité de la loterie par celle du lot dans la loterie et en additionnant les probabilités ainsi obtenues dans toutes les loteries ou le lot est présent. Dans l'exemple ci-dessus, la probabilité de X1 est (2/3 x 1/2) + (1/3 x 1/4), celle de X2 est (2/3 x 1/2) et celle de X3 est (1/3 x 3/4).

Pour spécifier un comportement rationnel face au risque, Samuelson estime pouvoir se satisfaire de ces deux axiomes :

1- Axiom I of Complete Ordering: All situations can be completely ordered and in terms of their associated prizes alone. (This ordering may be further assumed to be continuous in the probabilities x1, x2…).

2- Axiom II (Strong Independence): If lottery ticket $(A)_1$ is (as good or) better than $(B)_1$, and lottery ticket $(A)_2$ is (as good or) better than $(B)_2$, then an even chance of

[1] Von Neumann & Morgenstern [377] p. 30

> getting $(A)_1$ or $(A)_2$ is (as good or) better than an even chance of getting $(B)_1$ or $(B)_2$[1].

Comme nous l'avons vu aux titres III et IV, les relations de complémentarité et de substituabilité entre les biens affectent l'utilité du groupe de biens. Selon Samuelson, le principe d'indépendance, tout-à-fait illégitime en situation de certitude, devient raisonnable dans le cas stochastique et même indispensable à la construction de ce qu'il appelle « the Bernoulli-Savage theory of maximisation of expected utility ». Vu que, d'une part $(A)_1$ et $(B)_1$ sont mutuellement exclusifs et que, d'autre part $(A)_2$ et $(B)_2$ le sont aussi, on ne voit pas pourquoi le choix entre $(A)_1$ et $(B)_1$ pourrait être contaminé par celui entre $(A)_2$ et $(B)_2$.

Samuelson avait déjà démontré lors du colloque de Paris que ces deux axiomes sont nécessaire pour définir le comportement qui maximise la « Bernoulli expected cardinal-utility magnitude » : $x_1.u(X_1) + \ldots + x_n.u(X_n)$.

Dans leur article « The Expected-Utility Hypothesis and the Measurability of Utility » (1952), **Friedman et Savage** défendent ladite hypothèse voulant que dans des "choices involving risk (…) individual choose in such circumstances as if they were seeking to maximize the expected value of some quantity". Ils proposent de fonder ce principe sur un trio d'axiomes :

1- Le premier axiome correspond aux axiomes 1 et 2 de von Neumann et Morgenstern.
2- Si $\alpha.a + (1 - \alpha)c \leq b$, $\forall\ 0 \leq \alpha < 1$, alors $a \leq b$.
3- $\forall\ 0 < \alpha < 1$, $\alpha.a + (1 - \alpha)c \leq \alpha.b + (1 - \alpha)c$ n'est vrai que si $a \leq b$
   Les auteurs expliquent le sens de cet axiome avec l'exemple suivant. Si un médecin hésite entre plusieurs maladies pour établir son diagnostic, mais que le repos est le meilleur remède pour chacun d'eux, il peut prescrire ce remède, même sans diagnostic exact. En fait, il s'agit d'une forme d'axiome d'indépendance, puisqu'en combinant respectivement une perspective *a* et une perspective *b* avec une perspective *c*, on ne modifie pas la préférence entre elles.

LA COURBE D'UTILITE D'UN REVENU INCERTAIN

L'article « The Utility Analysis of Choices Involving Risk » (1948) de **Friedman et Savage** développe la théorie de l'utilité espérée dans une nouvelle direction : le lien entre le revenu et le comportement face au risque.

Pour effectuer leur analyse de la maximisation de l'utilité espérée, les auteurs construisent un outil graphique. L'objet est une fonction d'utilité : le revenu est placé en abscisse et l'utilité en ordonnée. Il est assez facile d'y placer un revenu fixe. Mais comment placer une perspective aléatoire comportant plusieurs événements possibles ? L'astuce des auteurs consiste à apporter les simplifications suivantes :

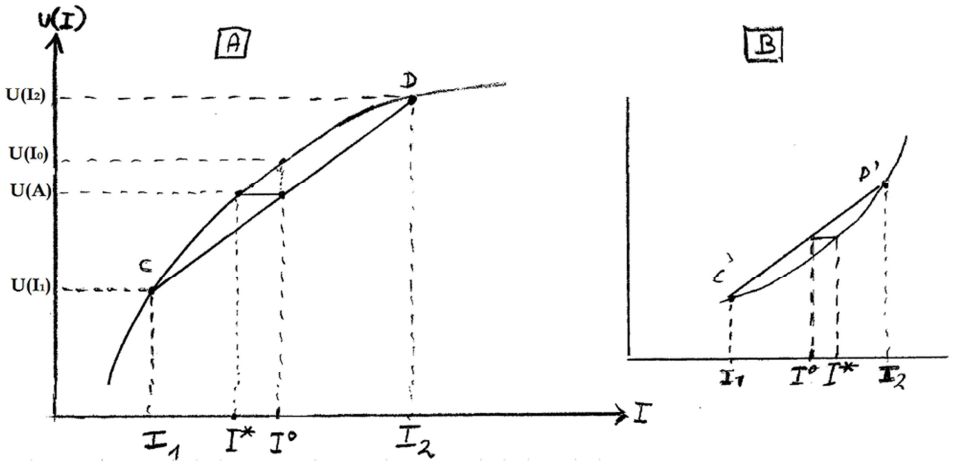
- Les alternatives offertes peuvent être toutes et entièrement converties en un revenu monétaire. Bien-sûr, dans la réalité, la manière dont un revenu s'acquiert peut influencer son attrait autant que son montant, mais la simplification fait avancer considérablement le modèle
- Ne sont considérées que des perspectives dont le choix comporte seulement deux options, dont les attraits ne sont pas égaux. Ici aussi, les auteurs sacrifient délibérément le réalisme à l'efficacité du modèle.

[1] Samuelson [314] p. 672. D'autres auteurs comme Marschak ou Savage ont également donné leur définition de l'axiome d'indépendance. Il y a des nuances entre ces définitions, mais elles vont évidemment dans le même sens.

Voici les éléments et les variables du modèle :

- Le choix se pose entre une perspective incertaine *A* procurant un revenu I_1 à une probabilité α ou un revenu I_2 à une probabilité (1-α) et une perspective *B* d'un revenu certain I_0.
- Compte tenu de la théorie de l'utilité espérée, on a $U(A) = \alpha.U(I_1) + (1-\alpha).U(I_2)$
- L'espérance mathématique du revenu $\alpha.I_1 + (1-\alpha).I_2$ est désignée par $I°(A)$
- L'égalité $I° = I_0$ signifie que le jeu est équilibré. Si le jeu est équilibré, la préférence de l'Agent pour *A* signifierait qu'il est attiré par le risque et la préférence pour *B* que le risque le rebute.
- I* est le revenu certain d'un attrait égal à celui de la perspective *A* ; on a : $U(I^*) = U(A)$.

Graphique 9.4 : l'utilité d'une perspective aléatoire



Sur le graphique, α est supposé égal à 0,5 : l'espérance mathématique est donc située à mi-chemin.

Sur le graphique 9.4, les points *C* et *D* correspondent aux niveaux d'utilité des revenus I_1 et I_2. Pour déterminer l'espérance mathématique de U(A), on tire le segment CD correspondant à la corde de l'arc entre C et D. L'espérance mathématique se trouve sur cette corde, à un niveau dépendant de α. Sur la figure A, on constate que $U(I°) > U(I^*)$. L'utilité de l'espérance de revenu est donc supérieure à l'espérance de l'utilité du revenu. Cette différence provient du fait que le risque lié à la perspective *A* réduit son utilité. Dans le cas où l'agent ressent de l'aversion pour le risque, la courbe est donc convexe. Le revenu I* est inférieur au revenu I° qui correspond à un revenu certain égal à l'espérance mathématique du gain de la perspective *A*. L'écart entre les deux correspond à la prime que l'agent est prêt à payer pour être débarrassé du risque.

Si le revenu I_0 est supérieur à I*, l'agent préférera la perspective *B* ; si $I^* > I_0$, il préférera la perspective *A*.

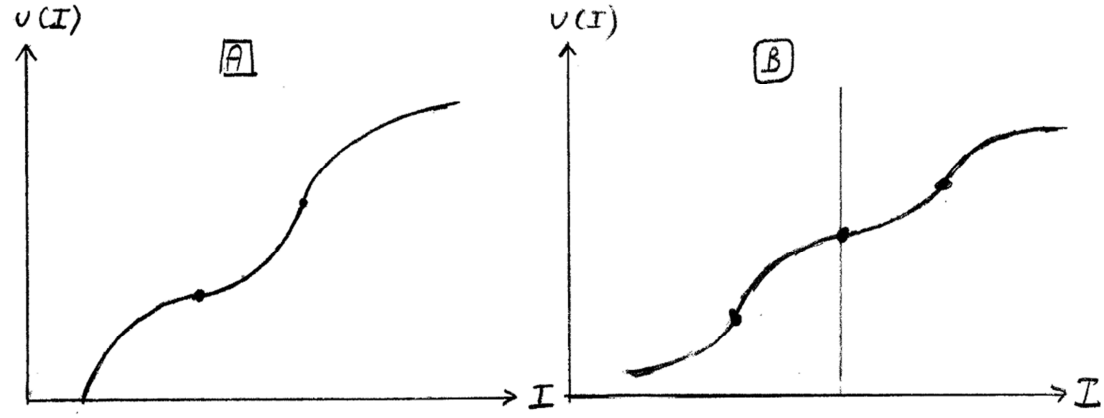
Si l'agent est attiré par le risque, la courbe d'utilité sera concave, comme sur la figure B. Pour un agent indifférent au risque, la fonction sera linéaire.

Les auteurs s'interrogent quelle est la forme de la fonction pour l'homme réel. Ils extraient d'une observation empirique très superficielle (sans sondages, sans statistiques) quelques caractéristiques sur la base desquelles des raisonnements parfois alambiqués mènent à une courbe divisée en 3 secteurs. Elle est d'abord convexe puis concave puis à nouveau convexe, comme sur la figure 9.5-A. Dans un premier temps, la courbe est censée se rapporter à un individu opérant des choix (loterie, assurance) à différents niveaux de revenu. Le principe de la courbe est ensuite étendu à une interprétation sociologique. Les trois secteurs correspondent aux classes sociales dans l'ordre croissant de revenu : la courbe indique alors l'attitude face au

risque, caractéristique de chaque classe. « On this interpretation, increases in income that raise the relative position of the consumer unit in its own class yield diminishing marginal utility, while increases that shift it into a new class, that give it a new social and economic status, yield increasing marginal utility »[1].

Les auteurs reconnaissent que leur courbe est discutable. Un approfondissement futur des connaissances empiriques permettra peut-être de l'affiner. Par exemple, n'y aurait-il pas plusieurs secteurs concaves ? Friedman et Savage avaient également dégagé de leur observation empirique qu'une prime plus élevée est nécessaire pour induire les agents à prendre un risque MOYEN plutôt qu'un PETIT risque ou un GROS risque. Ils reconnaissent que leur courbe ne reflète cette attitude que pour les agents de condition modeste.

Graphique 9.5 : la fonction d'utilité stochastique chez Friedman-Savage et Markowitz



La figure A indique la fonction de Friedman-Savage avec deux points d'inflexion. La figure B montre la fonction de Markowitz avec trois points d'inflexion

Dans son article "The Utility of Wealth" (1952), **Markowitz** "argue(s) that the Friedman-Savage (F-S) hypothesis contradicts common observation in important respects"[2]. Il pointe nombre de situations où des membres des trois classes sociales parieraient différemment de ce qu'affirme cette hypothèse. Notamment, l'idée que les gens très riches ou très pauvres ne parieraient jamais est contraire à l'observation.

Pour échafauder sa propre fonction d'utilité, Markowitz a interviewé son entourage appartenant à la classe moyenne, posant des questions du type « préférez-vous $1 avec certitude ou une chance sur dix de recevoir $10 ? ». Puis la même question avec $10 et $100, $100 et $1000… Et ensuite les mêmes questions avec des pertes. Les conclusions sont les suivantes. Concernant les gains en dessous d'un certain seuil, les gens prennent le risque, mais au-delà s'en tiennent à la perspective certaine. Concernant les pertes, en-dessous d'un seuil, on préfère les petites pertes sûres, mais au-delà, on tente la chance d'échapper à la perte malgré le risque d'en encourir une plus importante.

Markowitz en conclut que le courbe d'utilité doit avoir une forme semblable à celle de la figure 9.5-B : concave, puis convexe puis concave puis convexe. Il y a donc trois points d'inflexion et non deux. Le point d'inflexion central correspond au revenu actuel ou plus précisément ce que Markowitz appelle le *revenu habituel* afin que soit tenu compte d'éventuelles variations récentes et importantes de ce revenu. Markowitz pense que s'il devait interroger des riches, la courbe aurait la même forme mais les deux autres points d'inflexion seraient plus éloignés du

[1] Friedman & Savage [110] pp. 298-299
[2] Markowitz [252] p. 152

revenu habituel ; pour des personnes de condition modeste, ils seraient plus proches. La courbe de Markowitz a encore deux autres caractéristiques :

- « Generally, people avoid symmetric bets »[1]. La pente de la courbe est donc plus forte à gauche du *revenu habituel* qu'à droite.
- Pour éviter le paradoxe de Saint-Pétersbourg, la courbe est bornée par au-dessus et par en-dessous[2].

9.4.4. La critique d'Allais

Maurice Allais est en désaccord avec la théorie de Bernoulli elle-même et avec ses différentes déclinaisons que nous venons d'examiner. Il l'a combattue dans plusieurs articles, dont « Le Comportement de l'Homme Rationnel devant le Risque : Critique des Postulats et Axiomes de l'Ecole Américaine » (1953).

Il commence par exposer les quatre éléments psychologiques essentiels pour analyser les comportements face au risque :

- Elément I : *la déformation psychologique des valeurs monétaires et la courbure de la satisfaction absolue*. Cet élément concorde avec le principe bernoullien mettant en avant la valeur psychologique des gains qui augmente moins que ceux-ci.
- Elément II : *la déformation psychologique des probabilités objectives*. Certains croient en leur bonne étoile alors que d'autres s'estiment poursuivis par la malchance. Seules importent les probabilités subjectives.
- Elément III : *la pondération suivant les probabilités des valeurs psychologiques et la considération des ESPERANCES MATHEMATIQUES de la distribution des probabilités des valeurs monétaires*.
- Elément IV : *la prise en considération de la forme des distributions de probabilités des valeurs psychologiques et en particulier de leur DISPERSION*. Ce facteur a trait au plaisir ou au déplaisir procuré par le jeu en tant que tel. L'agent réticent au risque préférera les densités de probabilité à faible variance. Allais insiste tout particulièrement sur ce facteur qui a tendance à être négligé.

Une des difficultés est que les éléments I et IV sont conceptuellement distincts, mais pratiquement indissociables. C'est à la possibilité de grandes pertes et de grands gains que beaucoup d'agents sont particulièrement sensibles. Pour celui qui désire à tout prix une grosse somme, le jeu est parfois le seul moyen rationnel de se la procurer. Les éléments I à III ne rendent pas compte de cette réalité. « On ne saurait ainsi négliger la dispersion des valeurs psychologiques même dans une première approximation. C'est même à notre avis l'élément spécifique de la psychologie du risque »[3]. Allais insiste également sur le fait que l'existence d'émotions positives ou négatives face au risque n'est en rien un signe d'irrationalité.

Allais formalise sa théorie par l'équation (9.7) :

---

[1] Markowitz [252] p. 154

[2] Cette remarque est importante. Il ne suffit pas de donner le premier rôle aux utilités plutôt qu'aux gains réels pour empêcher que l'espérance mathématique puisse tendre vers l'infini. On peut concevoir des jeux dont l'utilité espérée serait infinie pour un agent dont la fonction d'utilité serait non bornée (pour autant que cela soit possible).

[3] Allais [7] p. 512. On ne peut s'empêcher de dresser un parallèle avec la controverse du capital (chap. 5.3). Dans les deux cas, la théorie néoclassique tire des conclusions de la seule espérance mathématique d'une variable (la durée du processus de production ou l'utilité du gain). Dans les deux cas, elle est contestée par l'affirmation que la forme de la distribution (de l'application des facteurs primaires dans le temps ou la densité de probabilité des gains) importe également.

$$s(V) = f(g_1,...,g_n,p_1,...,p_n) \qquad (9.7)$$

La fonction s(.) indique l'utilité d'une perspective ou d'un gain et *V* est une perspective aléatoire comportant des gains g_i selon des probabilités p_i. Allais propose également une deuxième équation dont la signification est identique mais qui convient mieux pour des variables continues. Considérons la variable γ qui correspond à la satisfaction des gains possibles. On a $\gamma = s(g)$. Sa fonction de densité de probabilité est $\Phi(\gamma)$. On a alors :

$$s(V) = h(\Phi(\gamma)) \qquad (9.8)$$

Par contre, l'équation qui exprime la théorie de l'école américaine est :

$$B(V) = \sum_i p_i . B(g_i) \qquad (9.9)$$

B(.) est l'indicateur d'utilité de von Neumann et Morgenstern, correspondant à la fonction v(.) de l'équation (9.6). D'après Allais, pour ces auteurs, B(.) et s(.) coïncident, abstraction faite des transformations linéaires de v(.). Par contre, il estime ambigüe la position de Friedman, Savage et Samuelson. Leur version de la fonction B(.) semble être un amalgame des éléments I et IV. D'après Allais, sous la plume de von Neumann et Morgenstern, la formule (9.9) décrit le comportement de l'homme moyen, alors pour les autres, elle est le modèle du comportement rationnel.

REFUTATION DE LA FORMULE BERNOULLIENNE

Allais s'emploie à réfuter la formule (9.9). D'abord du point de vue théorique, en partant d'une définition abstraite de la ***rationalité***. Selon lui, les seules implications de la rationalité sont au nombre de trois :

- un champ de choix ordonné
- le recours aux probabilités OBJECTIVES
- ce qu'il appelle l'*axiome de préférence absolue* disant que si pour toutes les éventualités, le gain de V_1 domine celui de V_2, il faut préférer V_1 à V_2.

Allais démontre mathématiquement (dans un autre article) qu'il est impossible de déduire l'équation (9.9) de ces trois postulats. Il démontre également que l'équation (9.9) est plus restrictive que l'équation (9.8). En fait, le formule bernoullienne est trop restrictive par rapport aux seules implications de la rationalité, ce qui est une forme d'irrationalité.

Un comportement très courant chez les entrepreneurs consiste à opérer un arbitrage entre le gain probable d'une opération et le risque de subir une grande perte. Ils ont alors un indicateur d'indifférence du type $S = f(M,q)$ où *M* est le gain probable et *q* la probabilité de cette perte. Pour conserver un niveau de satisfaction donné, il paraît rationnel que *M* doit augmenter beaucoup plus rapidement que *q*. Mais cette croissance plus que proportionnelle de *M* n'est pas compatible avec le caractère linéaire de la formule bernoullienne.

Allais s'attache également à réfuter l'*axiome d'indépendance*, notamment dans sa version de Samuelson, qu'il résume par l'équation :

$$\text{Si } P_1 < P_2, \text{ alors : } \alpha.P_1 + (1\text{-}\alpha).P_3 < \alpha.P_2 + (1\text{-}\alpha).P_3 \qquad (9.10)$$

Il écrit : « Ce point de vue est en réalité inacceptable, car il suppose le premier tirage correspondant aux probabilités [α;(1-α)] comme neutre, alors qu'il ne l'est pas, et il se place 'ex post' alors qu'il faut se placer 'ex ante' »[1]. Allais imagine une situation où P_2 et P3 sont des gains certains alors que P1 est un gain probable. Mêlé à P_3, P_2 perd le statut de gain certain qui

[1] Allais [7] p. 520

faisait son attrait ; il a donc plus à perdre que P_1 de la combinaison avec P_3. Voici un exemple chiffré :

> P2 donne100 millions certains, P1 donne 500 millions à 98/100 ou rien à 2/100. Les personnes prudentes mais néanmoins rationnelles préféreront P2 à P1. Introduisons maintenant la perspective P3 qui est la certitude d'un franc que l'on mêle à P1 et P2 avec une probabilité $1-\alpha = 99/100$. Ceux qui préféraient P2 à P1 pourraient maintenant opter pour 500 millions à 0,98/100 plutôt que 100 millions à 1/100. L'axiome de Samuelson est mis en échec.

L'erreur a une double cause :

- Si on se place EX POST comme Samuelson, on élimine inconsciemment l'essence du risque que constitue la forme de la distribution.
- « Notre psychologie est telle que nous préférons plus de sécurité au voisinage de la certitude qu'au voisinage de grands risques… »[1].

9.4.5. L'incertitude dans la *Théorie Générale* de Keynes

La conception de l'incertitude dans la *Théorie Générale*, que nous avons déjà évoquée au sous-chapitre 5.1.1, est cohérente avec le *Traité des probabilités*. Keynes et Knight sont dans le même camp. L'essentiel des problèmes rencontrés par les agents économiques relève de l'incertitude qui n'est pas probabilisable[2].

Je me contenterai de citer un passage très connu de l'article que Keynes avait écrit en 1937 pour commenter la Théorie générale où il définit l'incertitude :

> « The game of roulette is not subject, in this sense, to uncertainty (…) The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention or the position of private wealth owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know »[3].

Dans son article « Is Probability Theory relevant for Uncertainty? A Post Keynesian Perspective » (1991), **Davidson** cite également ce passage de Keynes. Il rappelle deux comportements que la Théorie Générale prêtait aux agents désorientés par l'incertitude : l'« animal spirit » et le repli dans une position inactive comme la détention de liquidités. Ces comportements contredisent fréquemment les prédictions de la théorie de l'utilité espérée.

Un comportement régulièrement rencontré dans la vie économique réelle est la conclusion de *contrats monétaires* : des contrats pour une relativement longue période où les prix sont spécifiés à l'origine et en unités monétaires. A priori il n'est pas rationnel de se corseter dans un prix dont on n'est pas sûr qu'il conviendra encore à plus lointaine échéance. Selon Davidson, il s'agit d'une réponse à la « true uncertainty ». Les contrats monétaires permettent aux agents de planifier leur trésorerie de façon à ne pas tomber en illiquidité, une situation très pénible pouvant avoir des suites judiciaires. La peur de l'illiquidité les pousse à détenir une réserve de monnaie, incompatible avec la maximisation de l'utilité espérée. « A shortage of money can

---

[1] Allais [7] p. 539

[2] Attention à la terminologie utilisée par les auteurs qui peut susciter des équivoques intempestives : le *Traité* de Keynes parle de probabilités qui ne peuvent pas être traitées mathématiquement ; l'objet est évidemment identique à l'incertitude dont Knight dit qu'elle ne peut pas être traitée par les probabilités.

[3] Keynes [183] p. 214

hold up the expansion of real output, despite of expected profits »[1]. La théorie néoclassique voit l'épargne comme un ajustement intertemporel des flux de dépenses. La thésaurisation pour fuir l'incertitude n'entre pas dans ce schéma.

N'étant pas fondés sur un calcul économique rationnel, parfois impossible, l'*animal spirit* et la *préférence pour la liquidité* peuvent aboutir à du chômage. Nul besoin de rigidité des salaires, d'information asymétrique… pour obtenir ce résultat. Il suffit de renoncer aux probabilités.

*

Pour conclure ce chapitre sur l'incertitude, on peut émettre l'hypothèse que si la variance des valeurs psychologiques des revenus possibles est inférieure à certain seuil, les comportements s'approcheront de la maximisation de l'utilité espérée, mais qu'au-dessus de ce seuil, les agents adopteront les comportements keynésiens.

[1] Davidson [68] p. 139

9.5- REVENU, CONSOMMATION ET EPARGNE

On appelle *fonction de consommation* la fonction dont la variable indépendante est le revenu et la variable dépendante la consommation, soit de l'individu (ou du ménage), soit de la société. Chronologiquement, la première fonction de consommation est celle de la *Théorie Générale* ; nous l'avons examinée au sous-chapitre 5.1.1. En résumé, la consommation y augmente avec le revenu, mais de façon moins prononcée que lui : la propension marginale à consommer est inférieure à la propension moyenne à consommer, ce que des études économétriques (régressions) travaillant sur des données à court terme semblent confirmer.

9.5.1. Les précurseurs, Kuznets et Samuelson

Dès avant la fin de la deuxième guerre mondiale, des économistes américains s'emparent de cet outil, à la fois théoriquement et empiriquement. Le premier d'entre eux est Kuznets, un des fondateurs de la comptabilité nationale dont l'immense travail empirique a éclairé toute cette génération d'économistes. Ses recherches aboutissent à une constatation paradoxale : sur le long terme, la consommation garde une proportion relativement constante du revenu.

Dans une contribution à un ouvrage collectif (1943), **Samuelson**, décidément omniprésent dans la science économique, se réfère aux statistiques de Kuznets montrant la relative stabilité du taux d'épargne sur le long terme. Pourtant, il croit à la fonction de consommation keynésienne ; il tente une synthèse permettant de marier les deux phénomènes : la propension marginale décroissante de la consommation à court terme et la stabilité à long terme.

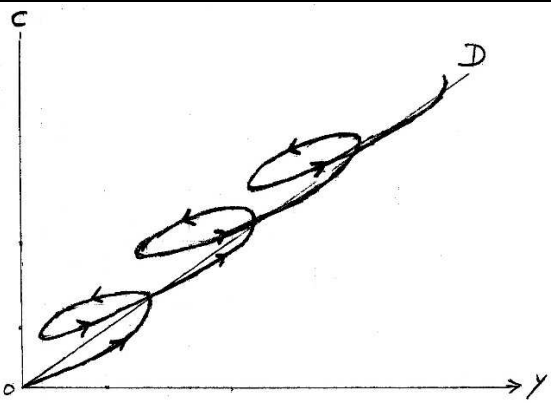
Sur le long terme, la consommation reste dans le sillage du revenu toujours croissant. Aucun phénomène de satiété ne se manifeste. « Since national income was increasing rapidly throughout this period, the most plausible explanation of this is to be found in the hypothesis that our enlarged scale of wants was causing an upward shift in the consumption function at about the same rate as improvements in our production potential, yielding a stable relation between percentage consumed out of national incomes corresponding to a given fraction of full employment income »[1]. Samuelson s'en félicite, car sans ce moteur, le plein emploi deviendrait de plus en plus difficilement accessible.

A court terme, la consommation suit le cycle conjoncturel. Mais les oscillations de la consommation sont de moindre amplitude que celles du revenu, et c'est encore plus vrai si dans la consommation, on n'inclut que l'utilisation réelle des biens durables.

La figure 9.6 expose la fonction de consommation qui associe la stabilité à long terme avec les fluctuations à court terme.

[1] Samuelson [312] p. 33

Graphique 9.6 : la fonction de consommation à court terme et à long terme



La variable temps intervient bien qu'elle ne figure ni en abscisse, ni en ordonnée. Les flèches indiquent le passage du temps.

L'extrémité droite de chaque boucle correspond au pic conjoncturel. A chaque pic, la courbe de consommation est proche de la droite OD qui représente un pourcentage constant du revenu, donné par sa pente ; c'est la stabilité à long terme.

9.5.2. Duesenberry et l'hypothèse du revenu relatif

En 1949, Duesenberry publie « Income, Saving and the Theory of Consumer Behavior », un livre aux idées sortant des chemins battus et bien documenté empiriquement.

L'auteur remonte à la théorie microéconomique de la consommation. La pierre angulaire de celle-ci est le système de préférence qui anime tout agent économique. Sur base de quelques paramètres, le revenu courant, les revenus futurs anticipés, les prix actuels, les prix futurs anticipés, le taux d'intérêt, la valeur des actifs à disposition, la solution au problème du consommateur est toujours déterminée. Le système des préférences semble une base solide pour expliquer les comportements. L'introspection et l'observation de notre entourage nous laissent penser que cette construction est plausible. Mais, selon Duesenberry, cette observation est superficielle.

En fait cette théorie dérive les préférences d'une entité exogène : les goûts. Leur origine extra-économique rend les préférences indépendantes des variables économiques, notamment des achats des autres agents. Selon Duesenberry, c'est ici que le bât blesse ; aucune fondation empirique ne conforte cette indépendance. En fait, constate-t-il, l'analyse des motivations du consommateur est absente de la théorie microéconomique. Il se fixe donc comme ambition de la reformuler en les y intégrant. « In short, we have to face up to the problems of the psychological bases of consumer choice (…) We shall make some definite commitments of a psychological and sociological nature »[1].

Les biens ne sont pas consommés pour eux-mêmes mais en tant que moyens pour réaliser tel objectif ou telle activité. Les différents biens offerts sur le marché qui concourent à un objectif donné n'y parviennent pas avec un niveau de satisfaction identique, à la fois pour des raisons objectives (efficacité, confort) et subjectives. Le consommateur classe ces biens par ordre de qualité. D'une façon générale, ce classement fait l'objet d'un consensus social. Malgré quelques divergences de goûts liées à des caractéristiques comme l'âge, les classements des différents

[1] Duesenberry [80] p. 19

individus sont fort semblables. Outre la qualité, la quantité et la variété affectent également le niveau de satisfaction.

L'individu préfère les biens de qualité supérieure qui sont plus chers, mais il souhaite se constituer une épargne. C'est à ce niveau qu'un arbitrage doit être opéré. Cependant, l'image courante chez les économistes, d'un consommateur planifiant à un moment donné tous ses achats, placements ou emprunts en un plan cohérent à long terme ne correspond pas à la réalité. Les décisions sont prises au fur et à mesure des besoins, souvent de façon impromptue. La rationalité du comportement vient de ce que le consommateur apprend par l'expérience. Il se construit progressivement son propre compromis entre consommation et épargne auquel les décisions qu'il sera amené à prendre devront se plier. Bien intégré mentalement, ce compromis est un rempart contre les tentations.

L'homme moderne est mû par une pulsion d'acheter toujours plus en quantité et en qualité. Les raisons objectives n'expliquent pas tout ; le modèle de consommation a une nature sociale : un niveau de vie croissant est un des buts majeurs de notre société. Il s'agit là d'un trait culturel marquant. Cette tendance précarise le compromis dont il était question au paragraphe précédent, même en l'absence de variation du revenu et des prix. Duesenberry explique ainsi ce qu'il appelle l'*effet de démonstration* : « For any particular family the frequency of contact with superior goods will increase primarily as the consumption expenditures of others increase (…) The result will be an increase in expenditures at the expense of saving »[1]. La fréquence de contact avec les biens supérieurs est l'un des éléments fondamentaux de la théorie de Duesenberry.

Dans notre société, l'accession à un niveau de vie supérieur est devenue un but en soi. La pulsion de consommer des biens supérieurs dépasse le besoin objectif que nous en avons. Le processus de socialisation instille cette envie dans l'esprit des individus. L'estime de soi en arrive à être conditionnée à un certain succès dans la satisfaction de ce désir. Dans une structure sociale différenciée, le succès se mesure par le statut social atteint. « Attainment of a given status also requires an ability to meet the behavioral standards of other members of high status groups »[2]. Le modèle de consommation du groupe social ayant le statut le plus convoité devient un critère pour juger du succès individuel dans la société. « Our social goal of a high standard of living, then, converts the drive for self-esteem into a drive to get high quality goods »[3]. La mobilité sociale et le non-cloisonnement entre les classes sociales, caractéristiques de notre société, augmentent les occasions où une personne est confrontée à des biens de qualité supérieure.

Duesenberry met l'équilibre de la consommation en équations. Comme l'indiquent les équations (9.11-A) et (9.11-B), l'individu *i* doit maximiser son utilité U_i, fonction des consommations et des actifs[4] détenus endéans l'horizon de *n* années ; ces variables sont chaque fois rapportées au paramètre R_i qui reflète la comparaison avec le niveau de vie d'autrui. Le paramètre R_i vaut la somme des consommations des individus *j* pondérées par le coefficient α_{ij} propre à chaque *j*. On a :

[1] Duesenberry [80] p. 27
[2] Duesenberry [80] p. 29
[3] Duesenberry [80] p. 31
[4] Cela peut sembler évident, mais tous les auteurs n'incluent pas les actifs dans la fonction d'utilité. Selon Fisher, seul comptait la consommation ; les actifs n'importaient que par la consommation future qu'ils rendaient possible.

$$U_i = U_i(C_{i1}/R_i,...C_{in}/R_i,A_{i1}/R_i,...A_{in}/R_i) \qquad (9.11\text{-A})$$
$$R_i = \sum \alpha_{ij}.C_j \qquad (9.11\text{-B})$$
$$C_i/R_i = f(Y_{i1}/R_i,...Y_{in}/R_i,A_i/R_i,r_1,...r_n) \qquad (9.11\text{-C})$$

L'équation (9.11-C) détermine la consommation courante ; elle est fonction des revenus présent et anticipés, du patrimoine présent, chacun rapporté à R_i ainsi que des taux d'intérêt présent et anticipés. S'il y a *m* individus, il y a *m* inconnues C_i et *m* équations ; le système a donc une solution.

Supposons que, partant de l'équilibre, toutes autres choses restant égales, les revenus présents et anticipés ainsi que les actifs des *m* individus soient multipliés par un même facteur *k*. Il s'ensuivrait que les C_i seraient également multipliés par *k*, car le numérateur et le dénominateur des ratios C_i/R_i, Y_i/R_i et A_i/R_i varient en proportion. La conclusion qu'en tire Duesenberry est celle-ci : « in equilibrium, consumption is proportionate to income and the savings ratio is independent of the absolute level of income »[1]. L'équilibre de la consommation est stable. Une augmentation du revenu moyen de la population pourrait créer l'illusion chez chaque individu que son statut s'est amélioré ; toutefois, il s'apercevra rapidement que le niveau de vie de ses concitoyens a également augmenté et l'éventuelle augmentation du taux d'épargne qui aurait prévalu un temps sera corrigée.

Le taux d'épargne macroéconomique est influencé par différents facteurs. Par exemple, la croissance démographique le pousse à la hausse car la part de la tranche d'âge qui épargne dans la population totale augmente par rapport à celle qui désépargne. L'effet de la croissance économique est plus complexe : elle tend également à augmenter la part de pouvoir d'achat détenue par la tranche d'âge qui épargne. L'effet final dépend du type d'actifs détenus par les retraités. La valeur des actions augmente avec les perspectives économiques, ce qui permet aux retraités d'augmenter leur consommation. A l'opposé, la détention de titres à revenu fixe réduit leur part dans le revenu global. La distribution des revenus, plus ou moins inégalitaire, est un autre facteur souvent cité comme exerçant une influence sur le taux d'épargne. Selon Duesenberry, cette problématique est complexe et « there is considerable doubt as to how saving would be affected by a redistribution »[2].

Comme Samuelson, Duesenberry constate que le ratio de la consommation sur le revenu est plus élevé dans les phases dépressives du cycle conjoncturel que dans les phases d'expansion. Les études empiriques montrent que pas mal de ménages désépargnent pendant les crises, surtout parmi les chômeurs. A revenu égal, les familles de chômeurs dépensent une part du revenu plus élevée que celles des employés à bas salaire. Trois raisons peuvent l'expliquer :

- Pendant la phase précédente, elles ont accumulé une réserve
- L'espoir de retrouver un travail lors de la reprise donne à penser qu'il y a juste une période difficile à passer
- Il est plus pénible de renoncer à ce qu'on a eu qu'à ce qu'on n'a jamais connu.

Duesenberry en tire cette conclusion : la consommation d'aujourd'hui dépend du revenu courant et des revenus futurs anticipés, mais également du revenu passé. Le taux d'épargne est faible lorsque le revenu présent est bas par rapport au revenu passé. Plus précisément, c'est le pic de revenu qu'a connu le ménage dans le passé qui continue à influencer sa consommation par la suite. « We can conclude then that the income or consumption of the last cyclical peak

[1] Duesenberry [80] p. 37
[2] Duesenberry [80] p. 45

will carry a special and very heavy weight in determining consumption at a given (lower) level of income during a depression »[1].

Duesenberry parle d'un *effet de cliquet*, produit par le caractère irréversible de la relation entre le revenu et la consommation. L'effet de cliquet se manifeste dans le rapport entre la croissance économique à long terme et les fluctuations conjoncturelles à court terme. Grâce à lui, en période de crise, l'économie ne régresse pas jusqu'à perdre le bénéfice du dernier boom : chaque cycle économique a donc son pic plus élevé que le précédent.

9.5.3. Friedman et l'hypothèse du revenu permanent

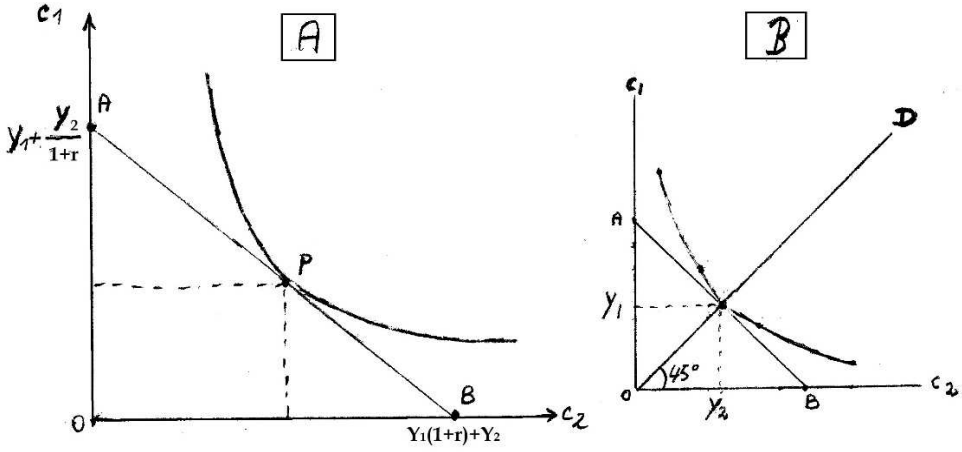
Dans les années quarante et cinquante, Milton Friedman s'est beaucoup préoccupé de la relation entre le revenu et la consommation. En 1957, il publie « A Theory of the Consumption Function », ouvrage dans lequel il formule sa fameuse *hypothèse du revenu permanent* (permanent income hypothseis : PIH).

Friedman veut enraciner sa fonction de consommation dans la théorie pure du comportement du consommateur, plus précisément dans sa version fishérienne. L'ouvrage de Friedman commence par réexposer le modèle graphique contenu dans « Theory of Interest »[2].

Le point de départ est une situation simplifiée à deux périodes, où l'agent reçoit respectivement un revenu Y_1 (l'année courante) et un revenu Y_2. Les prix restent constants. Chacune de ces deux années, il peut consommer un montant s'écartant du revenu annuel, grâce à la possibilité d'emprunter ou d'épargner au taux d'intérêt *r*. L'agent doit répartir optimalement sa consommation entre c_1 (consommation présente) et c_2 (consommation future).

Le graphique 9.7 montre comment s'établit l'équilibre dans le champ d'indifférence de l'agent entre c1 et c2.

Graphique 9.7 : répartition optimale de la consommation dans le temps



La contrainte budgétaire est donnée par le segment AB qui relie la consommation maximale de la première période, s'élevant à $Y_1+Y_2/(1+r)$ et la consommation maximale en deuxième période, valant $Y_2+Y_1.(1+r)$. L'optimum est évidemment donné par la tangence entre ce

[1] Duesenberry [80] p. 89

[2] Au sous-chapitre 3.2.8, j'avais omis l'exposé graphique de Fisher ; Friedman nous offre l'opportunité d'en prendre connaissance.

segment et une courbe d'indifférence. La forme des courbes d'indifférence reflète les préférences intertemporelles de l'agent. Leur convexité exprime la préférence pour une consommation équilibrée entre les deux années.

Apparemment, trois paramètres déterminent l'optimum : Y_1, Y_2 et *r*. Friedman les ramène à deux : comme on évalue habituellement un actif en actualisant ses revenus futurs, il crée une variable *richesse* (W) qui actualise le flux perpétuel de tous les revenus de l'agent. Dans l'exemple qui nous occupe, on a $W_1 = Y_1 + Y_2/(1+r)$. Des changements de Y_1 ou Y_2 n'affectent la consommation que pour autant qu'ils modifient W_1. On obtient comme fonction de consommation :

$$c_1 = f(W_1,r) \qquad (9.12)$$

W_1 détermine l'éloignement du segment AB par rapport à l'origine et *r* en détermine l'inclinaison. Y_1 ne fait plus partie des paramètres.

Friedman adhère à une conception largement répandue selon laquelle, si le revenu courant subit des variations importantes, la consommation restera plus ou moins stable et c'est l'épargne qui fluctuera en tant que facteur résiduel. Les concepts de CONSOMMATION et de REVENU au sens habituel de consommation COURANTE et revenu COURANT sont mal adaptés à l'étude du comportement du consommateur. Friedman met en avant les concepts de *consommation permanente* (c_p) et de *revenu permanent* (y_p), des variables englobant des périodes plus longues, ce qui leur procure une certaine stabilité. La consommation permanente s'oppose à la dépense en ce sens qu'une dépense de l'année *t* peut être consommée pendant *n* années qui suivent *t*. Le *revenu permanent* est défini comme le montant que l'agent pourrait consommer en laissant intacte sa richesse telle que définie ci-dessus. Friedman donne la formule suivante :

$$y_{pt} = r.W_t \qquad (9.13)$$

La relation qui importe est celle qui lie la consommation permanente avec le revenu permanent. Des équations (9.12) et (9.13), on peut déduire :

$$c_{pt} = f(y_{pt},r) \qquad (9.14)$$

La figure 9.7-B met en évidence le rôle du taux d'intérêt. La droite *D* à 45° comprend les combinaisons de consommation égales entre c_1 et c_2. Supposons qu'une même consommation rapporte une utilité identique en l'année 1 et l'année 2, c'est-à-dure qu'aucun facteur comme l'âge n'a altéré les préférences entretemps ; dans ce cas, les courbes d'indifférence sont symétriques par rapport à *D*. Si $Y_1 = Y_2$ et que le taux d'intérêt est nul (OA = OB), c_1 vaudra Y_1 ; il n'y aura ni épargne ni désépargne. Si le taux d'intérêt devient positif, la pente de AB diminue ; la tangence aura lieu plus à droite ; c_1 sera inférieure à Y_1 ; il y aura épargne l'année 1 et désépargne l'année 2.

Friedman tente de préciser la fonction *f* de l'équation (9.14) à l'aide de quelques hypothèses qui lui semblent plausibles. La principale est que la fonction d'utilité est homogène par rapport aux consommations annuelles ; le fait de multiplier la consommation de toutes les années par un même facteur ne modifie pas les préférences temporelles entre les années. L'argument est celui-ci : si on augmente la consommation c_1 par rapport à c_2, l'utilité marginale de c_2 augmente proportionnellement et diminue proportionnellement si on augmente c_2 ; le rapport doit donc rester plus ou moins constant lorsqu'on augmente les deux en parallèle. Friedman concède que cette hypothèse n'est qu'approximativement vraie. Nantie de cette hypothèse, l'équation (9.14) peut être reformulée comme suit :

$$c_{pt} = k(r,u).y_{pt} \qquad (9.15)$$

La consommation permanente est une fraction *k* du revenu permanent, fraction qui dépend exclusivement du taux d'intérêt *r* et d'un ensemble *u* de paramètres affectant la forme des courbes d'indifférences, tels que l'âge, l'évolution de la composition familiale, les préférences individuelles… Le niveau absolu du revenu permanent n'influe pas sur *k*.

Jusqu'ici, nous avons fait abstraction de l'incertitude ; l'agent connaît ses revenus futurs en totalité. La prise en compte de l'incertitude entraîne un double effet.

1- L'incertitude est un motif supplémentaire pour détenir de la *richesse non-humaine*[1]. Avec à un revenu permanent identique, l'individu plus fortuné pourra se permettre de consommer plus. Avec *w* pour symboliser le rapport de la richesse non humaine sur le revenu permanent, la fonction de consommation (9.15) devient :

$$c_p = k(r,w,u).y_p \qquad (9.16)$$

2- Les revenus futurs deviennent des distributions de probabilité. Comme l'attitude face au risque est elle-même question de préférence, « the sharp dichotomy between tastes and opportunities that is the central attraction of indifference analysis under certainty is shattered »[2]. L'enchevêtrement des influences se complique mais a priori, il n'est pas possible d'affirmer si cette complication fait hausser la proportion du revenu permanent qui serait consommée ou si elle la réduit. Friedman estime donc pouvoir conserver la fonction de consommation (9.15).

Tout en reconnaissant que cet exercice pose quelques problèmes, Friedman agrège les résultats de son analyse des comportements individuels. Il estime que l'équation (9.16) garde sa validité au niveau macroéconomique.

La majeure partie de l'ouvrage de Friedman est d'ordre économétrique. Un aspect déconcertant lui vient de l'interpénétration entre l'économétrie et la théorie. Selon Friedman, la PIH peut être résumée en trois équations :

$$c_p = k(r,w,u).y_p \qquad (9.17)$$
$$y = y_p + y_t$$
$$c = c_p + c_t$$

Nous retrouvons l'équation (9.16) agrégée, ainsi que la décomposition du *revenu mesuré* et de la *consommation mesurée* en une *composante permanente* et une *composante transitoire*[3]. A ces trois équations, Friedman ajoute une hypothèse à propos des composantes transitoires : la nullité des coefficients de corrélation entre composantes permanentes et transitoires, ainsi qu'entre les deux composantes transitoires. La nullité de ce dernier coefficient de corrélation est une hypothèse plus forte que celle des deux autres, mais elle est intuitivement plausible, puisque l'épargne est généralement considérée comme le facteur résiduel dans l'affectation du revenu.

Une régression entre le revenu mesuré et la consommation mesurée peut très bien confirmer le principe keynésien orthodoxe que le taux d'épargne augmente lorsque le revenu s'accroit mais cette constatation passe à côté du phénomène essentiel que la consommation permanente constitue une fraction constante du revenu permanent.

---

[1] Comme la richesse selon Friedman englobe l'actualisation de tous les revenus y compris ceux du travail, la précision « non-humaine » s'impose pour signifier qu'il s'agit du patrimoine mobilier et immobilier.

[2] Consumption [100] p. 15

[3] Friedman remarque qu'on aurait pu caractériser la décomposition de *c* et *y* comme multiplicative plutôt qu'additive, auquel cas il eut été sensé de travailler avec les logarithmes des variables.

Durant les soixante ans qui précèdent l'édition de l'ouvrage, la fraction *k* de l'équations (9.17) est restée plus ou moins constante à environ 0,88 aux Etats-Unis. Comme le note Friedman, cette constance n'est pas impliquée par la PIH. On constate une différence entre les salariés (*k* compris entre 0,90 et 0,95) et les indépendants (*k* compris entre 0,80 et 0,90). Friedman attribue cet écart à l'incertitude des revenus plus élevée chez les seconds.

Sur le plan social et politique, Friedman tire ces trois conclusions de la PIH :
- L'inégalité des revenus est parfois considérée comme favorable à l'épargne. La PIH ne conforte pas cette idée, car *k* est indépendant de la dispersion des revenus. Cet effet ne pourrait s'avérer que si l'inégalité est liée à la composante TRANSITOIRE du revenu.
- Inversement et pour la même raison, en diminuant l'inégalité, on ne stimule pas la consommation, contrairement à une autre idée courante.
- Le multiplicateur keynésien se trouve émoussé, car il dépend de la propension marginale à consommer que la PIH revoit à la baisse : le niveau de la consommation dépend du revenu permanent qui ne hausse que modérément lorsque le revenu mesuré s'accroît.

Cette critique du multiplicateur me paraît contestable. Elle s'appuie sur le principe indiscutable que la consommation dépend du revenu permanent et non du revenu courant, mais elle est incapable d'en apprécier les limites. Et le chômage en est une ; les études empiriques montrent une chute de la consommation lors de la perte d'un emploi et une hausse lors de la remise au travail[1]. Le chômage semble occuper une place particulière dans les comportements de consommation et d'épargne, peut-être parce qu'avant la perte de son emploi, l'individu garde l'espoir qu'il pourra y échapper et lorsqu'il y est tombé, la crainte de l'insolvabilité l'oblige à une certaine prudence. Lorsqu'une dépense autonome déclenche le multiplicateur, on attend d'elle qu'elle empêche des salariés de tomber dans le chômage ou qu'elle remette des chômeurs au travail.

L'ouvrage de Friedman est étonnamment peu disert sur le lien entre consommation et taux d'intérêt. Il n'y a pas d'estimation de l'élasticité de la consommation par rapport au taux d'intérêt. Plus les courbes d'indifférence du graphique 9.7 sont incurvées, moins l'influence du taux d'intérêt est sensible. Une courbure plus marquée signifie que les consommations de différentes années ne sont pas considérées comme substituables mais plutôt comme complémentaires. Mais précisément, la PIH attribue aux consommateurs un désir de consommation stable sur la durée de vie. Il me paraît logique d'en déduire une incurvation prononcée. Mon avis est que la PIH implique une influence faible du taux d'intérêt sur la consommation et l'épargne.

9.5.4. Modigliani et l'hypothèse du cycle de vie

L'*hypothèse du cycle de vie* (life cycle hypothesis : LCH) fut exposée la première fois en 1954 dans l'article « Utility Analysis and the Consumption Function : Interpretation of Cross Data » signé par Modigliani et Brumberg. Modigliani lui consacra par la suite de nombreux articles, seul ou avec des coauteurs. Nous résumerons d'abord l'article de 1954 et ensuite l'article « Life Cycle, Individual Thrift and the Wealth of Nations » (1986) dans lequel Modigliani synthétise ses travaux antérieurs.

Le principe de la LCH est que la consommation de l'année τ dépend, non des paramètres de cette année précisément, mais d'une manière rationnelle de planifier la consommation sur la

[1] Cf. par exemple Ganong et Noel [116].

longueur de la vie. Modigliani écrit : « Our purpose was to show that all the well-established empirical regularities could be accounted for in terms of rational, utility-maximizing, consumers allocating optimally their resources to consumption over their life, in the spirit of Irving Fisher (1930) »[1]. Algébriquement, le problème consiste à maximiser la fonction d'utilité :

$$U = U(c_t, c_{t+1}, \ldots c_L, a_{L+1}) \qquad (9.18)$$

En respectant la contrainte budgétaire intertemporelle :

$$a_t + \sum_{\tau=t}^{N} \frac{y_\tau}{(1+r)^{\tau+1-t}} = \frac{a_{L+1}}{(1+r)^{L+1-t}} + \sum_{\tau=t}^{L} \frac{c_\tau}{(1+r)^{\tau+1-t}} \qquad (9.19)$$

Dans le modèle, la vie de l'individu a une durée *L* qui se divise en un âge actif de durée *N* suivi d'une retraite de durée *M*[2]. L'indice *t* indique l'âge de l'individu. Les autres variables sont le revenu professionnel courant *y*, le revenu professionnel anticipé pour les années ultérieures y^e, le patrimoine *a*, la consommation *c* et l'épargne *s*, ainsi que le taux d'intérêt *r*. L_t est la durée restant à vivre à l'âge *t*. Le revenu *y* n'est gagné que pendant l'âge actif ; pour un âge t > N, le retraité désépargne sur le patrimoine qu'il s'est constitué pendant l'âge actif[3].

L'utilité intertemporelle dépend de la consommation aux différents âges et du patrimoine qu'on souhaite éventuellement léguer aux héritiers. La résolution des équation (9.18) et (9.19) par les multiplicateurs de Lagrange aboutit aux valeurs d'équilibre c_t*, une par année restant à vivre au nombre de L-t+1, et a_{L+1}*. Toutes les années où le revenu $y_\tau + r.a_\tau$ sera supérieur à c_τ*, il y aura épargne[4] ; dans le cas inverse, il y aura désépargne.

Tout ceci ressemble à la théorie de Friedman vue au sous-chapitre précédent. Selon Modigliani, la principale différence est que la LCH embrasse une durée de vie finie, tandis que la finitude de la vie du consommateur n'intervient pas dans la PIH : le revenu permanent est conçu comme une rente perpétuelle.

Pour analyser plus à fond les implications de leur théorie, Modigliani et Brumberg posent quatre hypothèses simplificatrices importantes qui, selon eux, n'orienteront pas le résultat final :

1- Le ménage représentatif ne laisse ni héritage ni dettes ; il n'hérite donc pas non plus.
2- La façon dont le ménage répartit sa consommation dans le temps est fonction exclusivement de ses goûts et ne dépend donc pas de ses ressources.
3- Le taux d'intérêt est nul. Le revenu professionnel est donc le seul revenu.
4- Le ménage répartit sa consommation entre les années de façon égale[5].

---

[1] Modigliani [261] p. 299

[2] Le modèle simplifie la réalité en ignorant l'enfance qui précède la phase professionnelle.

[3] L'assurance-vie entre dans ce schéma aussi bien que l'épargne directe.

[4] Modigliani et Brumberg distinguent quatre motifs d'épargne :
- le désir de laisser un héritage
- la non-coïncidence entre le cycle de revenu et le cycle des dépenses
- le souhait de disposer d'un actif de précaution contre l'incertitude
- la possibilité de réalise des achats coûteux sans emprunter.

[5] Pour les biens durables, il ne faut pas confondre la dépense et la consommation. Si vous achetez une voiture que vous conservez cinq ans, vous la consommez pendant ces cinq ans.

LCH ET FONCTION DE CONSOMMATION

Compte tenu des simplifications, la consommation individuelle pendant l'année courante est donnée par l'équation[1] :

$$c = c(y, y^e, a, t) = \frac{1}{L_t} y + \frac{(N-t)}{L_t} y^e + \frac{1}{L_t} a \qquad (9.20)$$

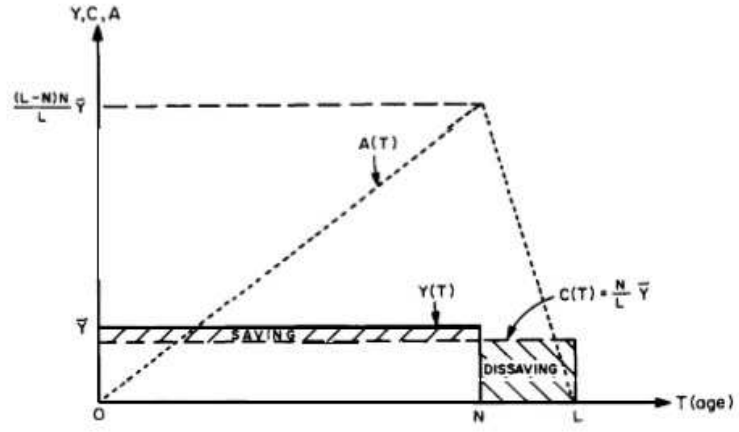
Cette consommation est donc une fonction linéaire et homogène du revenu courant, du revenu anticipé et de l'actif initial. Elle dépend de l'âge du consommateur. La *propension marginale à consommer* associée est donnée par la formule :

$$\frac{dc}{dy} = \frac{1}{L_t} + \frac{N-t}{L_t} \frac{dy^e}{dy} \qquad (9.21)$$

La formule tient compte de l'effet de la variation du revenu courant sur le revenu anticipé par la ratio dy^e/dy qui est une espèce d'élasticité d'anticipation. On constate que la propension marginale à consommer diminue avec l'âge ; la durée de la vie active allouable à la constitution d'un patrimoine suffisant se réduit. Elle est par contre indépendante du revenu courant, en contradiction avec la *Théorie Générale* de Keynes.

Le graphique 9.8 tiré de l'article de Modigliani de 1986 illustre l'équilibre individuel lors du cycle de vie moyennant les simplifications précitées et le cas particulier d'un revenu (courant et anticipé) constant. L'accumulation de l'épargne dans un premier temps et de la désépargne ensuite déterminent un sentier d'équilibre du patrimoine.

Graphique 9.8 : cycle de vie : consommation, épargne et patrimoine



La méthodologie de cette étude est paradoxale. Les formules ci-dessus sont les seules qui sont dérivées théoriquement. Toutes les autres et les considérations qui en découlent proviennent de la simulation d'analyses transversales sur des échantillons de ménages, alors que les données empiriques sont absentes.

L'épargne courante est donnée par l'équation :

[1] Les variables de l'année courante ne sont plus affectées d'un indice temporel.

$$s = \frac{M}{L} y^e + \frac{L-t}{L_t}(y - y^e) - \frac{1}{L_t}[a - a(y^e, t)] \qquad \textbf{(9.22)}$$

Comment évolue-t-elle si l'individu, après une période de revenu stationnaire connaît une augmentation imprévue de celui-ci ?

- Soit l'individu perçoit ce changement comme TRANSITOIRE. Dans ce cas, le troisième terme de (9.22) qui révèle le manque d'ajustement du capital à son chemin d'équilibre à long terme, est nul mais les deux autres sont positifs et l'épargne croîtra.
- Soit l'individu perçoit ce changement comme PERMANENT. Dans ce cas, ye s'ajuste à *y* et le deuxième terme du membre de droite s'annule[1]. Les deux autres sont positifs et l'épargne croîtra mais moins que dans le cas précédent, car l'accroissement du revenu permanent ye entraîne une augmentation de la consommation contrairement à l'augmentation du revenu non permanent.

En conséquence, les agents qui ont connu récemment une augmentation de leur revenu épargnent un part plus grande de celui-ci que ceux dont il est resté constant, qui eux-mêmes ont un taux d'épargne plus élevé que ceux dont le revenu a baissé. « According to the usual interpretation, these findings show that the proportion of income saved rises with the economic status of the household. According to our model, on the other hand, they only show that households whose income is above the level to which they are adjusted save an abnormally large proportion and those whose income is below this level save an abnormally low proportion, or even dissave »[2]. En conclusion, la relation entre revenu, épargne et consommation mise en évidence par Keynes est également expliquée par la LCH, mais avec une interprétation radicalement différente.

Modigliani et Brumberg montrent que leur modèle aboutit à la fonction de consommation semblable à la fonction keynésienne dessinée au graphique 7.3, mais ici encore, l'interprétation diverge. En l'absence de fluctuation, la droite de consommation partirait de l'origine et monterait à un taux élevé *N/L*. L'interception positive avec l'ordonnée et la pente inférieure sont la conséquence exclusivement des variations à court terme du revenu, pour la raison indiquée au paragraphe précédent et ces caractéristiques sont d'autant plus marquées que ces fluctuations sont importantes.

Une question largement débattue concerne l'influence que le patrimoine pourrait avoir sur la consommation[3]. Dans l'optique de la LCH, le patrimoine fait partie du plan de vie de l'individu au même titre que la consommation et l'épargne. Il n'est donc pas un déterminant de la consommation. Toutefois, les variations impromptues du patrimoine mettent son niveau en hiatus par rapport au plan, ce qui entraîne des ajustements qui affectent le rapport entre consommation et épargne.

IMPLICATIONS DU MODELE

De la LCH, Modigliani a tiré six déductions macroéconomiques qu'il a testées empiriquement, dont il reconnaît que certaines sont surprenantes et contre-intuitives.

1- Le taux d'épargne d'un pays est entièrement indépendant du revenu par tête.

---

[1] On suppose donc que l'élasticité d'anticipation vaut un.
[2] Modigliani & Brumberg [263] p. 24
[3] Cf. par exemple, l'*effet de richesse* de Pigou.

2- Des taux d'épargne nationaux divergents sont compatibles avec un même comportement LCH.
3- Entre pays ayant en commun le comportement LCH, un taux de croissance à long terme supérieur entraîne un taux d'épargne plus élevé.
4- Le ratio du patrimoine sur le revenu est une fonction décroissante du taux de croissance.
5- Une économie peut accumuler un patrimoine considérable, même compte non tenu de la richesse transmise par héritage.
6- Pour un taux de croissance donné, le taux d'épargne et le ratio du patrimoine sur le revenu dépendent principalement de la durée de la retraite.

Modigliani examine ensuite l'effet de la renonciation aux hypothèses simplificatrices.

1- **Taux d'intérêt positif**. Il a un double effet : EFFET DE REVENU par l'ajout d'un montant *r.a* au revenu ; EFFET DE SUBSTITUTION, en sens contraire, par le coût d'opportunité qui grève la consommation présente. Ce dernier dépend de l'élasticité de substitution entre les consommations aux différentes périodes. Tenant compte des données empiriques en sa possession, Modigliani conclut : « Unfortunately, despite a hot debate, no convincing general evidence either way has been produced, which leads me to the provisional view that *s* is largely independent of the interest rate »[1].
2- **Cycle de revenu et taille de la famille.** Le revenu professionnel tend à croître avec l'âge. Les dépenses de consommation dépendent du nombre d'enfants dans le ménage qui souvent s'accroît puis décroît ; il est possible que la désépargne soit courante chez des jeunes ménages. Ceci ne change pas fondamentalement la forme en bosse du patrimoine (figure 9.8).
3- **Contrainte de liquidité.** Celle-ci empêche les ménages d'emprunter autant qu'ils le souhaitent pour équilibrer leur consommation suivant leurs goûts. L'effet serait de reporter plus de consommation vers un âge plus avancé.
4- **La myopie des ménages.** Les ménages n'auraient pas le jugement ou la volonté suffisants pour épargner en vue des consommations futures. Modigliani déclare : « However, such myopia is not supported empirically »[2].
5- **Les durées respectives de la vie active et de la retraite.** Certains facteurs peuvent inciter à travailler moins longtemps. La croissance économique permet d'économiser autant en travaillant moins longtemps. La sécurité sociale procure un revenu qui n'est pas dépendant de l'épargne antérieure. La pointe de la bosse de la figure 9.8 est alors tirée vers la gauche.

Dans la LCH, les retraités sont censés désépargner ; à leur décès, leur patrimoine devrait approcher de zéro, comme sur la figure 9.8. Modigliani reconnaît que la réalité ne confirme pas ce scénario. Est-ce à dire que l'homme serait animé par un « motif d'héritage » ? Modigliani ne le pense pas. L'incertitude quant à la durée de vie crée un risque de pauvreté contre lequel la disposition permanente d'un patrimoine minimum permet de se prémunir. L'héritage est courant, non pas à la suite d'un éventuel *motif d'héritage* mais par le fait du *motif de précaution*. Modigliani admet qu'il s'agit là d'un amendement à apporter à la LCH. A la fin de son article, Modigliani rapporte toutefois des sondages selon lesquelles le motif d'héritage rencontrerait une certaine adhésion dans couches les plus prospères : « Thus the bequest motive seems to be limited to the highest economic classes »[3].

[1] Modigliani [261] p. 304
[2] Modigliani [261] p. 305
[3] Modigliani [261] p. 310

9.5.5. L’épargne des entreprises

Tous les auteurs étudiés au cours de ce chapitre prennent soin de préciser que leur analyse se limite à l’épargne des ménages. Mais bien-sûr, ajoutent-ils, l’épargne des entreprises (bénéfices non distribués) joue également un rôle important. A mon avis, il n’est pas correct de dissocier les deux aussi radicalement. L’épargne des entreprises nourrit indirectement le patrimoine des ménages. Les ménages détiennent les actions des entreprises dont les investissements autofinancés font croître la valeur. La plus-value des actions est une forme particulière de revenu et plus précisément de revenu non consommé. Il y a donc là indirectement épargne des ménages.

Avec la plus-value sur les participations, ne retrouve-t-on pas un taux d’épargne plus élevé suivant la classe de revenu du ménage ?

9.6- CAPITAL ET INVESTISSEMENT

Depuis le *Théorie Générale*, l'investissement occupe une place centrale dans la macroéconomie. La microéconomie l'a à peine abordé, à l'exception notable de Fisher. Nous avions terminé le sous-chapitre 3.2.8 avec la difficulté de réconcilier les deux concepts frères de CAPITAL et d'INVESTISSEMENT. Voyons dans le présent chapitre comment certains auteurs tenteront, à partir de cette réconciliation, d'élaborer une théorie de l'investissement alternative à celle de Keynes. Mais commençons par le développement de la thèse keynésienne.

9.6.1- Lerner et l'efficacité marginale de l'investissement

Dans son ouvrage « The Economics of Control » déjà connu du lecteur, Lerner s'intéresse au lien entre la MEI et la productivité marginale du capital.

L'entrepreneur peut faire varier son capital à souhait de façon à égaliser en permanence sa productivité marginale avec le taux d'intérêt. Mais ne pouvant emprunter qu'à elle-même, l'économie globale ne jouit pas de cette liberté à court terme[1]. Elle peut investir ou désinvestir, mais cela prend du temps. L'équilibre de courte période de l'économie globale se résume à l'égalisation de la MEI avec le taux d'intérêt. Selon Lerner, cette règle suffit à assurer l'usage optimal des ressources. L'équilibre de longue période implique quant à lui la double égalité : MEI = taux d'intérêt = productivité marginale du capital. Ceci suppose que le volume du capital ait atteint son niveau optimal ; l'investissement y est donc nul, en attendant une modification des paramètres qui relance la recherche d'un nouvel optimum.

Dans son article « On the Marginal Product of Capital and the Marginal Efficiency of Investment » (1953), Lerner va plus loin. L'idée même de l'égalisation de la productivité marginale du capital avec le taux d'intérêt est entachée d'une difficulté fondamentale. Il y a bien-sûr le problème bien connu de l'hétérogénéité du capital qui rend sa mesure physique impossible. Mais même s'il était homogène et mesurable physiquement, la difficulté persisterait. La productivité marginale du capital est donnée par la formule dY/dK où *Y* est la quantité produite et *K* le stock de capital. Mais *Y* et *K* sont hétérogènes l'un par rapport à l'autre puisque *Y* se compose d'une combinaison de biens de consommation et de biens d'équipement.

Avec l'exemple bien connu de la sylviculture, Lerner montre que la théorie autrichienne n'est pas plus capable d'élaborer un ratio à égaliser avec le taux d'intérêt. La quantité totale des arbres formant le capital à un moment donné dépend du rapport entre la plantation et la coupe des arbres, c'est-à-dire un facteur assimilable à l'investissement, qui est exogène par rapport à la période de production (durée de vie des arbres avant abattage).

Ce problème d'hétérogénéité ne se pose plus si la comparaison met en regard le sacrifice de consommation au temps t_0 et le surplus de consommation rendu possible au temps $t_0+\Delta t$. C'est cela l'essence de la MEI ; elle est toujours calculable et comparable avec le taux d'intérêt. Lerner écrit : « Is the marginal product of capital equal or not equal to the rate of interest in long run stationary equilibrium ? This question cannot be answered and should not be asked. It is essentially meaningless, because no problems about capital can ever arise. All the problems that look like capital problems turn out to be investment problems (…) So what we have is always a marginal efficiency of investment and never a marginal product of capital »[2].

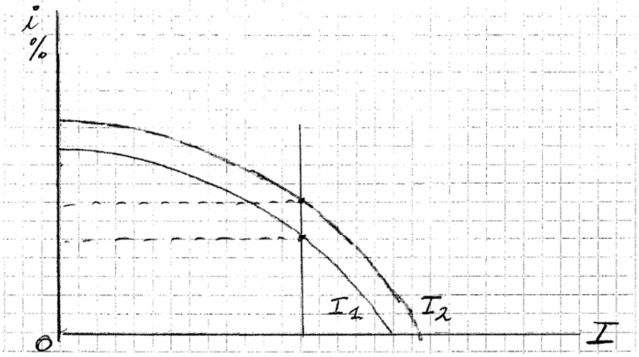
---

[1] Plus loin dans le chapitre, l'idée que l'entreprise individuelle peut adapter son capital instantanément sera remise en question.

[2] Lerner [226] p. 8

Pourquoi la courbe d'efficacité marginale de l'investissement (graphique 3.20) est-elle déclinante ? L'impossibilité d'augmenter la quantité de tous les inputs dans une proportion identique est une source de coût croissant. Soit les inputs meilleur marchés sont privilégiés, mais l'accroissement de leur proportion les rend moins efficaces ; soit les proportions optimales sont maintenues, mais les facteurs plus rares renchérissent.

Graphique 9.9 : MEI et accumulation du capital



La courbe I2 pourrait représenter un progrès technique par rapport à I1. Ou bien le passage de I2 à I1 signifierait une augmentation du stock de capital.

L'abscisse indique l'investissement, non le capital. Si à la suite de l'accumulation, le capital augmente, la courbe s'abaisse, tout en gardant sa forme. Par contre, le progrès technique tend à la déplacer vers le haut. Lerner craint toutefois qu'à très long terme, l'abaissement l'emporte. Avec une courbe MEI abaissée, un niveau d'investissement donné ne peut être maintenu que moyennant une diminution du taux d'intérêt. Outre ces facteurs à long terme, la courbe de MEI a son altitude principalement déterminée par les facteurs économiques qui rendent l'investissement plus ou moins souhaitable à court terme.

L'accroissement de l'investissement, c'es-à-dire la poussée vers la droite sur l'abscisse implique que la propension à épargner augmente pour compenser la baisse du taux d'intérêt. On sait que dans l'optique keynésienne, cette propension dépend du revenu ; en injectant des liquidités, l'autorité monétaire peut également favoriser l'épargne.

9.6.2- Taux interne de rentabilité et valeur actuelle nette

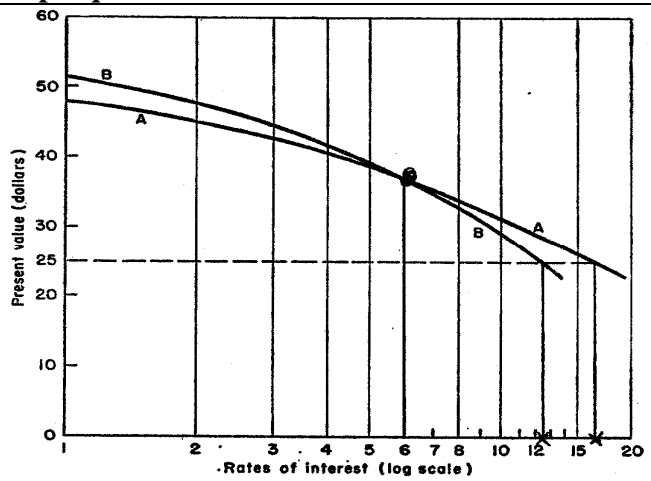
Dans les écoles de gestion, on enseigne que le *taux interne de rentabilité* et la *valeur actuelle nette* sont deux méthodes à la disposition des entreprises pour sélectionner les investissements à sélectionner.

- Le taux interne de rentabilité est le taux d'escompte qui, appliqué à toutes les recettes et dépenses présentes et anticipées, annule la valeur actuelle totale. Un investissement doit être effectué si ce taux est supérieur au taux d'intérêt en vigueur.
- La valeur actuelle nette est la somme des valeurs des recettes et des dépenses, actualisées au taux d'intérêt en vigueur. L'investissement doit être opéré si cette valeur est supérieure à zéro. On considère que la maximisation de la valeur actuelle nette de la firme devrait être l'objectif fondamental de sa direction.

La MEI keynésienne est clairement un taux interne de rentabilité. Keynes estimait simplement reprendre le concept de RROC élaboré par Fisher. C'est ce que conteste **Alchian** dans son article « The Rate of Interest, Fisher's Rate of Return over Costs and Keynes' Internal Rate of Return » (1955). Contrairement à la MEI, le RROC est un outil pour comparer deux flux de revenus et il hiérarchise les projets selon leur valeur actuelle nette. Alchian donne un exemple comparant deux flux de revenu *A* et *B* relatifs à une même dépense initiale de 25 mais distribués différemment dans le temps (*B* comportant une attente plus longue). La figure 9.10 donne leurs

valeurs actuelles nettes respectives pour différents taux d'intérêt en abscisse. Le RROC est le taux qui les égalise, ici 6%. En-deçà, *B* est plus avantageux, au-delà, *A* est préférable. Sur base des MEI, *A* (17%) semble absolument supérieur à *B* (12,5%), ce qui n'est pas un résultat correct.

Graphique 9.10 : taux interne de rentabilité et valeur actuelle nette



Certes, en définitive, la méthode de la MEI retient les deux investissements pour un taux d'intérêt inférieur à 12,5% et l'investissement *A* pour un taux entre 12,5 et 17%, ce qui n'est pas incorrect. Mais elle fait erreur lorsqu'il y a lieu de choisir entre deux opportunités d'investissement mutuellement exclusives. En conclusion, « A ranking of substitutable investments by the Keynesian internal rate of return is not consistent with the maximum of net present wealth »[1]. Comme fondement de la fonction de demande d'investissement, la méthode keynésienne de la MEI n'est donc qu'approximativement juste.

Dans son article « On the Theory of Optimal Investment decision », **Hirshleifer** teste la validité des deux méthodes, valeur actuelle nette et taux interne de rentabilité, dans différentes situations. Il part de l'analyse fishérienne mais il l'extrapole dans des situations plus complexes mais aussi plus réalistes : taux d'intérêt des emprunts supérieur à celui des prêts, taux d'intérêt croissant avec le montant emprunté ou rationnement du crédit.

Rappelons-nous que l'agent fishérien peut non seulement investir, mais également consommer, prêter et emprunter et donc éventuellement emprunter pour investir plus que ses moyens. Les variables sur lesquelles porte l'optimisation sont donc au nombre de quatre, ce que les deux méthodes tendent à occulter.

Après une analyse cas par cas, il s'avère, selon Hirshleifer, que la méthode du taux interne de rentabilité faut dans pas mal de situations. La valeur actualisée nette est plus fiable, sans être parfaite. La critique vient principalement du point de vue fishérien qui considère à tout moment la consommation du revenu comme une option. Les deux méthodes analysées se focalisent sur la meilleure sélection du projet comme si le montant à investir était une donnée intangible. Dans l'optique fishérienne, investir ou prêter reviennent à reporter la consommation à plus tard. Les courbes d'indifférence intertemporelles sont donc déterminantes.

*

Sans doute, Keynes eût-il été mieux avisé de mettre en avant la valeur actuelle nette plutôt que le taux interne de rentabilité, mais ne s'agit-il pas d'un détail par rapport à la globalité du

[1] Alchian [4] p. 940

problème macroéconomique ? Une idée de ce que serait la théorie keynésienne avec la valeur actuelle nette nous est donnée par le coefficient *q* de Tobin. Investir si ce coefficient est supérieur à un, revient à investir si la valeur actuelle nette est positive.

9.6.3- Une théorie néoclassique de l'investissement, est-elle possible ?

HAAVELMO POSE LA QUESTION

En 1960 paraît l'ouvrage « A Study in the Theory of Investment » de **Haavelmo**. Il s'agit d'une des premières études microéconomiques sur la question. Son but est de déduire l'investissement optimal de la maximisation du profit. La difficulté vient de ce la fonction de production a comme input les services du capital et non l'investissement ; le passage de l'un à l'autre n'est pas évident.

Prenons un cas simplifié où le capital est le seul facteur de production. On a :

$$x_t = F(K_t) \qquad (9.23)$$
$$\pi_t = p_t.x_t - r_t.q_t.K_t - \delta.q_t.K_t$$

où π est le profit, *x* est la quantité produite, *p* est le prix du produit, *K* est la quantité de capital, *q* est le prix des biens capitaux, *r* est le taux d'intérêt et δ est le taux de dépréciation du capital. La firme voudra maximiser :

$$\int_{t0}^{t0+\theta} e^{-r(t-t0)}.\pi_t.dt \qquad (9.24)$$

La lettre θ désigne l'horizon de la firme[1]. La condition pour la maximisation s'avère être :

$$p.[dF/dK] - q.(r+\delta) = 0 \qquad (9.25)$$

Haavelmo fait la constatation suivante : "if capital is perfectly mobile, it will be adjusted instantaneously at t_0, and kept constant over the whole horizon θ. There is absolutely no room for any non-zero dK/dt in this structure, (except a possible discontinuous jump in *K* at $t = t_0$). No finite change in the amount of capital is demanded for any finite period after $t = t_0$ unless there is a change in the data, i.e., in prices and interest rates"[2]. Bref, il est impossible de déduire une demande d'investissement[3] qui colle avec l'observation de la réalité montrant que la plupart des firmes opèrent des investissements chaque année.

Soit K* le capital optimal. Haavelmo montre que des écarts importants de *K* par rapport à K* sont susceptibles de se produire, sans que la firme s'écarte de son profit maximal. Il y a un paradoxe : alors que l'incitation à investir fait défaut la plupart du temps, "it is possible to argue that if there is the slightest degree of arbitrariness, or uncertainty about K*, then it is possible that dK/dt for short periods may be very large and fluctuate violently"[4]. L'auteur parle même d'une indétermination entre moins l'infini et plus l'infini. L'idée que le taux d'investissement serait infiniment volatile, si un élément n'intervient pas pour l'étaler, sera reprise dans nombre de contributions sur ce sujet. Elle a un côté absurde car elle se base sur deux éléments irréalistes : le temps continu (et non discret) et l'investissement instantané.

---

[1] Par hypothèse, l'entrepreneur élabore son plan en prévoyant que *p*, *q* et *r* restent constant sur l'horizon θ.
[2] Haavelmo [125] p. 163
[3] Il s'agit évidemment de l'investissement NET ; le remplacement aura toujours lieu si les conditions générales ne changent pas.
[4] Haavelmo [125] p. 164

Donc une fonction d'investissement CONTINUE n'a d'existence que dans le cas d'une évolution continue des paramètres de l'équation (9.25). Par exemple, si la fonction de production est sujette à un progrès technique continu[1]. Dans le cas contraire, l'investissement est nul, et ce, quel que soit le taux d'intérêt. Mais comme ce taux (r) est inclus dans l'équation (9.25), une modification de *r* affectera l'investissement ponctuellement, même alors que son niveau importe peu.

L'idée d'un ajustement du capital INSTANTANE, déjà rencontrée chez Lerner, est surprenante. Pourtant, Haavelmo reconnaît lui-même l'existence de délais de mise en œuvre des ajustements du capital. Il cite le temps de la prise de décision et de sa mise en application, la recherche du niveau optimal K* par essai et erreur, le coût d'ajustement d'autant plus élevé que celui-ci est rapide, les contraintes de crédit au niveau des prêts bancaires, la production d'équipements à la commande. Mais selon lui, ces facteurs ne suffisent pas à produire un flux d'investissement quasi-permanent. Etonnamment, il n'envisage pas que la conjonction de ces délais avec la modification des paramètres de 8.87 puisse suffire.

Haavelmo souhaite que la théorie permette de justifier un taux d'accumulation du capital positif et qu'elle puisse expliquer les variations à court terme du taux d'investissement. Mais elle n'y parviendra pas par le mode traditionnel d'une offre de capital qui s'égalise avec une demande correspondante sous l'effet régulateur d'un prix qui serait le taux d'intérêt. Haavelmo conclut : "What we should reject is the naive reasoning that there is a "demand schedule" for investment which could be derived from a classical scheme of producers' behavior in maximizing profit. The demand for investment cannot simply be derived from the demand for capital. Demand for a finite addition to the stock of capital can lead to any rate of investment, from almost zero to infinity, depending on the additional hypothesis we introduce regarding the speed of reaction of the capital-users"[2].

Haavelmo reconnaît qu'il pose plus de questions qu'il n'apporte de réponses. Il ouvre toutefois deux pistes de réflexion :

- le rapport entre *p* et *q* serait fondamental. Une valeur élevée de *q* suscitera l'élévation de la production de biens d'équipement.
- les perspectives de la consommation sont l'incitant ultime de l'investissement. A ce titre, le partage du revenu des ménages entre l'épargne et la consommation joue un rôle régulateur.

JORGENSON Y REPOND POSITIVEMENT

Faisant écho au livre de Haavelmo, l'article « Captial Theory and Investment Behavior » (1963) de Dale **Jorgenson** veut prouver qu'une théorie néoclassique de l'investissement est possible. "Stated baldly, the purpose of this paper is to present a theory of investment based on the neoclassical theory of optimal accumulation of capital"[3].

Jorgenson procède en deux étapes : d'abord déterminer la demande de capital et ensuite en déduire la demande d'investissement. Comme Haavelmo, Jorgenson raisonne dans le cadre de la concurrence parfaite.

---

[1] La présente théorie assigne comme fonction à l'investissement de toujours maintenir le capital à son niveau optimum. Haavelmo remarque que face à un progrès technique dans son secteur, la firme peut se lancer dans un investissement sans qu'intervienne cette motivation du capital optimum.

[2] Haavelmo [125] p. 216. Toutefois, une relation entre le taux d'intérêt et la demande d'investissement pourrait être établie dans la cadre de la statique comparative.

[3] Jorgenson [166] p. 248.

Suivant en cela Hirshleifer, Jorgensen considère que « Demand for capital is determined to maximize net worth »[1].

La valeur de l'entreprise équivaut à la somme des *cash flows* futurs escomptés[2]. On a :

$$R(t) = p(t).Q(t) - w(t).L(t) - q(t).I(t) \qquad (9.26)$$
$$W = \int_0^{\infty} e^{-rt}.R(t)\,dt$$

Où *R* est le cash flow de la période, *p* est le prix du produit, *Q* la quantité produite, *w* le salaire, *L* l'effectif employé, *q* le prix des biens d'investissement et *I* la quantité investie brute. *W* est la valeur de l'entreprise à maximiser et *r* le taux d'intérêt. Les contraintes sont :

$$dK(t)/dt = I(t) - \delta.K(t) \qquad (9.27)$$
$$F(Q,L,K) = 0 \qquad (9.28)$$

Où *K* indique les services du capital et δ le taux de remplacement dû à la dépréciation. L'équation (9.28) n'est autre que la fonction de production néoclassique bien connue.

Les conditions de la maximisation égalisent les productivités marginales des facteurs *K* et *L* avec leur rémunération : $\partial Q/\partial L = w/p$ et $\partial Q/\partial K = c/p$ où *c* « is the shadow price or implicit rental of one unit of capital service per period of time. We will call this price the user cost of capital ». *c* vaut :

$$c = q.(r+\delta) - dq/dt \qquad (9.29)$$

Jorgenson suppose que la fonction *F* est de type Cobb-Douglas et que les variations des gains en capital sont transitoires, ce qui permet de neutraliser dq/dt. Dans ces conditions, le capital optimal, noté K^* est donné par la formule :

$$K^* = \gamma.\frac{pQ}{c} \qquad (9.30)$$

où γ est l'élasticité de la production par rapport au capital.

Les mathématiques de ce modèle permettent de déterminer un taux d'intensité capitalistique K^*/L, mais si les rendements d'échelle sont constants, ils ne peuvent pas déterminer K^*, comme pourrait le faire un modèle macroéconomique où *L* devient une donnée exogène moyennant l'hypothèse du plein emploi. Pour sortir de cette indétermination, Jorgenson "assume(s) that output and employment on the one hand and capital stock on the other are determined by a kind of iterative process. In each period, production and employment are set at the level given by the first marginal productivity condition and the production function with capital stock fixed at its current level; demand for capital is set at the level given by the second marginal productivity condition, given output and employment. With stationary market conditions, such a process is easily seen to converge to the desired maximum of net worth"[3].

---

[1] Jorgenson [166] p. 248.

[2] Ce n'est pas le profit mais le *cash flow* qui entre dans la formule. Le coût d'acquisition des investissements est déduit, mais pas l'amortissement. La formule d'Haavelmo était donc erronée.

[3] Jorgenson [166] p. 249..Dans l'article « Old and New Formulations of the Neoclassical Theory of Aggregate Investment : A Critical Review », Girardi conteste que ce processus itératif résout le problème. D'abord, il est impossible de sélectionner un capital optimum en considérant comme fixes à la fois l'emploi et la production : c'est l'un ou l'autre. En situation de concurrence, la firme a d'ailleurs la capacité de les faire varier tous deux. Indépendamment de ce problème, en cas de rendements d'échelle constants, le processus itératif ne converge pas vers un capital optimum mais vers l'infini si le prix de vente est supérieur au coût de production et vers zéro s'il est inférieur. Face à ce dilemme, une partie des auteurs néoclassiques

Passons maintenant à la deuxième étape. Chaque période où les paramètres de l'optimisation ont changé (principalement les prix), des investissements sont décidés pour mener le capital réel au capital optimal. Le principe conduisant à un investissement quasi-permanent est tout simplement l'étalement de la concrétisation des investissements décidés.

Les investissements décidés en *t* sont notés I_t^N et les investissements exécutés en *t*, I_t^E. Jorgenson suppose une distribution fixe et identique de tous les investissements dont les paramètres sont *T* (durée) et les α_θ (proportion de l'exécution qui est réalisée en *t*). La somme des α vaut l'unité puisque la firme entend toujours combler le décalage entre K et K^*. La firme choisit d'ignorer que le stock de capital ne sera peut-être plus optimal au moment où il sera opérationnel.

$$I_t^E = \sum_{\theta=0}^{T} \alpha_{\theta.}\, I_{t-\theta}^N \qquad (9.31)$$

Cette formule nous donne l'investissement net. Pour obtenir l'investissement brut, il faut encore ajouter les investissements de remplacement que Jorgenson présume proportionnel au stock de capital dans un rapport δ, ce qui se justifie si le nombre d'équipements individuels est élevé. On a $I_t^R = \delta.K_t$.

Jorgenson teste ensuite sa théorie de l'investissement sur les données empiriques de l'industrie manufacturière américaine. Le résultat conforte la fonction d'investissement néoclassique, mais des critiques ont objecté que l'hypothèse de la fonction Cobb-Douglas avait une influence excessive sur ce résultat[1].

Nous avons donc une explication de l'investissement, mais pas encore la fameuse fonction d'investissement décroissante par rapport au taux d'intérêt. C'est dans un article publié en 1967 que Jorgenson construira cette fonction ; le modèle qui y est présenté a les mêmes bases que celui de 1963 mais il prend une direction totalement différente. Pour mettre en avant le taux d'intérêt, Jorgenson donne à l'équation (9.29) une place centrale. Sa démonstration de la relation entre *I* et *r* est peu convaincante[2]. Quoi qu'il en soit, la postérité a gardé la mémoire de l'article de 1963, devenu un classique de la littérature néoclassique sur l'investissement, mais a réservé peu d'intérêt à celui de 1967. Jorgenson concluait pourtant celui-ci par un satisfecit : "We conclude that it is possible to derive the demand for investment goods as a function of the rate of interest on the basis of purely neoclassical considerations. However, the demand for investment goods depends on the rate of interest through a comparison of alternative paths of capital accumulation, each continuous and each depending on a time path of the rate of interest"[3].

*

mettent l'accent sur les rendements d'échelle décroissants. D'autres renoncent à cette caractéristique de la concurrence parfaite que les firmes peuvent vendre une quantité illimitée au prix du marché.

[1] Girardi indique que d'une majorité d'études empiriques se dégage une faible élasticité de l'investissement par rapport au coût d'usage du capital, donc une élasticité de substitution basse, en contradiction avec les caractéristiques de la fonction Cobb-Douglas. La faible élasticité-intérêt de l'investissement est également mise en évidence par Sharpe et Suarez.

[2] Le cheminement de la démonstration impose à un moment la constance du coût d'usage *c*. Ceci implique qu'en cas de variation de *r*, il y ait compensation exacte entre Δq.r et dq/dt. Jorgenson, estime pouvoir assurer que cette égalité est normalement satisfaite, une affirmation que contestera Tobin. Les gains en capital, neutralisés dans l'article de 1963, jouent ici un rôle déterminant.

[3] Jorgenson [167] p. 150.

Ce résultat théorique trouve peu d'écho empirique. Une multitude de sondages auprès d'entrepreneurs révèlent que le taux d'intérêt est un critère d'importance tout-à-fait secondaire lors de la prise de décision en matière d'investissement. La première de ces études fut le fait de l'OERG, dans le sillage de celle qui mit en avant le *full cost pricing*. D'autres études dans d'autres pays confirmèrent le résultat. Bien-sûr, la méthodologie de ce genre d'études donne toujours lieu à débat. Ce fut encore le cas ici.

9.6.4- Le coût d'ajustement convexe : solution ultime ?

Jorgensen a écrit ou coécrit plusieurs articles sur la dérivation de l'investissement à partir du capital optimum et il n'est pas le seul à avoir adopté ce point de vue. Dans l'article "Adjustment Costs in the Theory of Investment of the Firm" (1968), **Gould** écrit à ce propos : « The standard approach in these studies has been to derive desired capital stock K^*, from comparative static profit maximization considerations and then to use this K^* as the "long-run" desired capital stock in some auxiliary adjustment mechanism to determinate the firm's investment path. An example of such an auxiliary adjustment mechanism is the geometrically declining distributed lag which may be stated in continuous form by:

$$dK(t)/dt = \gamma.(K^* - K) \qquad (9.32) \text{ »}[1]$$

Gould impute des lacunes à cette approche: « One resulting difficulty is that many of the variables used to define K^* such as sales or profits are in fact affected by (9.32) and hence do not reflect the 'true' desired capital stock at any point of the investment path before full equilibrium has been achieved »[2]. L'équation (9.32) devrait d'une manière ou l'autre être intégrée dans le problème de maximisation, plutôt que de l'accompagner en externe.

Gould développe mathématiquement sa solution. Son système reprend les équations (9.27) et (9.28) de Jorgenson. Mais l'équation (9.26) est modifiée en :

$$R(t) = p(t).Q(t) - w(t).L(t) - C(I) \qquad (9.33)$$

Ce n'est plus l'investissement *I* qui figure dans l'équation, mais une fonction C(I) qui signifie l'investissement « gonflé » par un ensemble de facteurs qui en aggravent le coût[3]. C'est ce qu'on appelle le *coût d'ajustement de l'investissement*. La notion, qui s'est imposée dans la théorie moderne de l'investissement, est assez large et même un peu vague. On distingue parfois les coûts d'ajustement internes aux entreprises et les coûts d'ajustements externes aux entreprises. La première catégorie comporte toutes les dépenses annexes à l'investissement ainsi que les coûts indirects comme le détournement d'une partie du personnel de ses tâches productives normales. La seconde catégorie comprend essentiellement la hausse du prix des biens d'équipement due à l'accroissement de leur demande, notamment lorsque les producteurs de ce type de biens travaillent à coûts croissants.

Un aspect essentiel de ce concept est la convexité du coût d'ajustement. Par rapport à *I*, C(I) augmente plus que proportionnellement. On a :

---

[1] Gould [121] p. 47. Les modèles représentés par l'équation (9.32) sont parfois appelés modèles d'*accélérateur flexible*. Comme K^* dépend de la production anticipée, on peut dresser un parallèle entre cette formule et les modèles étudiés au sous-chapitre 8.3.1 qui faisaient dépendre l'investissement des variations du revenu. L'effet des variations du revenu pourra être amplifié au niveau de l'investissement.

[2] Gould [121] p. 47

[3] Les façons de modéliser varient : dans l'équation (9.33), C(I) inclut l'investissement proprement dit ; d'autres auteurs le présentent comme une addition à l'investissement.

$$C(I) > 0 \; ; \; C(0) = 0 \; ; \; C'(I) > 0 \; ; \; C''(I) > 0 \qquad (9.34)$$

L'investissement *I* est un flux lié à une période ; la concentration de dépenses d'investissement sur une période provoquerait un C(I) trop élevé sur cette période, ce qui représente une incitation à les étaler sur plusieurs périodes. Grâce au C(I) convexe, le modèle peut déboucher sur une demande d'investissement généralement positive.

Gould construit un tel modèle, qui aboutit à une fonction d'investissement. Les équations sont trop complexes pour être reprises ici. Gould constate que le sentier de l'investissement I(t) est dépendant du sentier entier de *p*, *w* et *r*. Si on postule la constance de ces trois prix, on retombe sur l'équation (9.32). Mais en dehors de cette hypothèse peu réaliste, il semble impossible d'énoncer une règle concernant le sentier d'investissement, qui serait généralement valable.

Sans les coûts d'ajustement, la firme peut maximiser le profit instantanément à chaque période. Avec ces coûts, elle doit à tout moment considérer le problème dans sa dimension intertemporelle.

Gould n'est pas le premier à mettre en avant les coûts d'ajustement. Eisner et Strotz (1963) et Lucas (1967) en avaient déjà émis l'idée. Voyons comment **Lucas** l'intègre dans son article « Adjustment Costs and the Theory of Supply ». L'article part de cette constatation : la théorie néoclassique de la firme à la mode Marshall-Viner est mise en défaut par les études empiriques ; celles-ci montrent que la croissance de l'offre d'une industrie est généralement satisfaite par la croissance des firmes existantes plus que par celle du nombre de firmes ; il apparaît également que des firmes de tailles variées croissent ensemble sur de nombreux marchés. Nombre d'auteurs cherchent l'explication dans les rendements d'échelle constants, ce qui disqualifie les courbes de coût à long terme en *U*. Lucas propose une explication alternative, où intervient le coût d'ajustement de l'investissement.

Le coût d'ajustement a ici la même signification que chez Gould, mais sa présence dans les équations diffère. La maximisation de la valeur de l'entreprise intervient comme chez Gould et Jorgensen ; les équations (9.26) et (9.27) de ce dernier restent le point de départ. L'originalité se situe dans l'inclusion de l'investissement brut dans la fonction de production :

$$Q(t) = F[L(t),K(t),I(t)] \qquad (9.35)$$
$$F_L > 0 \; ; \; F_K > 0 \; ; \; F_I < 0 \; ; \; F_{LL} < 0 \; ; \; F_{KK} < 0 \; ; \; F_{II} < 0$$

Comme l'indiquent les dérivées par rapport à *I*, un investissement plus élevé au cours d'une même période entraîne une réduction de la production, elle-même de plus en plus sensible. Cela peut s'expliquer par une diversion des ressources.

Lucas ajoute quelques hypothèses qui ont pour but de faciliter le traitement mathématique : la fonction *F* est homogène de degré un[1] ; elle est concave ; elle est décomposable en une fonction F(K,L) et une fonction F(K,I), dont les dérivées partielles doivent répondre à certaines conditions.

Un optimum existe, à partir duquel il est possible de déduire la fonction d'offre Q(t) et les fonctions de demande L(t) et I(t). On a :

[1] Ce qui ne signifie pas que les rendements d'échelle sont constants, vu la présence de *I* parmi les variables indépendantes.

$$L(t) = K(t).D_1(w/p) \qquad (9.36)$$
$$I(t) = K(t).D_2(r,\delta,w/p,q/p) \qquad (9.37)$$

Les fonctions (9.36) et (9.37) sont à court terme, c'-à-d pour *K* donné. D_1 est une fonction décroissante de w/p. D_2 est une fonction décroissante de w/p, de q/p, de *r* et de δ. Q(t) est fonction à la fois de D_1 et de D_2. L'effet d'une variation de *p* sur *Q* est contradictoire : d'une part, l'accroissement de *p* relativement à *q* et *w* est le principal incitant à produire plus ; d'autre part, l'augmentation de *I* qui s'ensuit entrave *Q*. Selon Lucas, le premier effet devrait prédominer.

L'équation (9.38) est déduite de (9.37), avec cette fois la prise en compte de la variation de *K* pour établir l'équilibre de longue période.

$$(dK/dt)/K = D_2(r,\delta,w/p,q/p) - \delta \qquad (9.38)$$

Révélation importante, avec des prix (*p,w,r,q*) constants, le stock de capital augmentera ou baissera à taux constant, un taux qui est indépendant de la dimension du capital (et donc de l'entreprise). L'investissement brut, l'emploi et la production varieront ensemble à un taux valant D_2-δ.

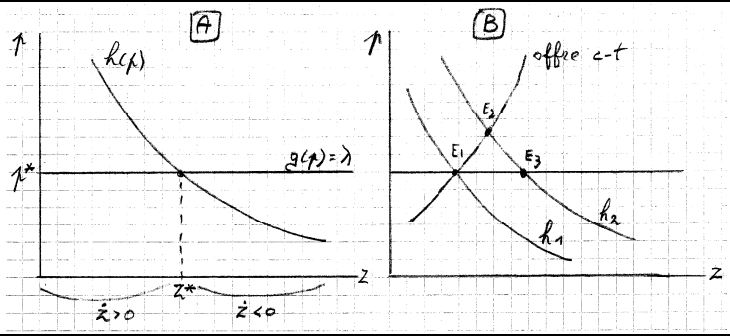
Voyons maintenant l'équilibre au niveau du marché, par hypothèse en concurrence parfaite. Supposons que la demande connaît une croissance continue de taux λ, expliquée par exemple par l'augmentation du revenu national à long terme. L'équation de cette demande est :

$$Q(t).e^{-\lambda t} = z(t) = h(p(t)) \qquad (9.39)$$

Où *z* est la quantité échangée compte tenu de la situation propre à la période considérée (où le taux de croissance λ n'intervient donc pas) et où h(p) est la relation décroissante bien connue de la demande par rapport au prix.

Les firmes ont toutes la même fonction de production ; elles ne diffèrent que par leur valeur de K(0). Lucas émet l'hypothèse que l'accroissement de l'offre se réalise par l'augmentation des productions individuelles des firmes existantes[1]. Le taux de croissance de l'offre globale est, suivant l'hypothèse néoclassique du prix paramétrique, une fonction croissante du prix, notée g(p). Il y a un prix p* qui assure l'égalité entre g(p) et λ, significative de l'équilibre de longue période. La quantité échangée relativement à la dimension du marché prend alors la valeur d'équilibre, z*, correspond au prix p*. Lorsque le prix du marché (donné par la courbe *h*) est supérieur à p* (à gauche de z* sur la figure 9.11-A), g(p) >λ et *z* augmente ; lorsqu'il lui est inférieur, g(p)< λ et *z* diminue. L'équilibre z* est donc stable.

Graphique 9.11 : ajustement de l'offre à la croissance de la demande



[1] Les deux autres voies, l'accroissement du nombre de firmes et les fusions-acquisitions sont juste esquissées.

La figure 9.11-B montre l'effet d'un déplacement (augmentation) de la courbe de demande (pour toute raison qui rend le produit plus désirable pour le consommateur)[1]. Un nouvel équilibre provisoire E2 s'établit à l'intersection de la nouvelle demande avec l'offre de courte période. Mais g(p) est maintenant supérieur à λ. L'offre va s'accroître jusqu'à l'égalisation entre ces deux variables en E3. Lucas constate que la dynamique de ce rééquilibrage est comparable à celle qui caractériserait une industrie qui fonctionne à coût constant.

Si le taux de croissance de la demande λ avait augmenté, la droite p* se serait élevée ; le nouvel équilibre requerrait un prix plus élevé et une quantité échangée moindre.

En agrégeant (9.37), on peut obtenir la fonction d'investissement du secteur industriel. Ici, la variable *p* devient endogène et elle dépend de λ. « The industry investment demand function so obtained will be a first order differential (or difference) equation in I(t)/K(t) with the levels and rates of change in factor prices *w*, *q* and *r* as 'forcing variables' and with the rate of demand shift, λ, as an additional variable (…) The presence of λ in this function was, of course, a feature of the accelerator hypothesis »[2].

9.6.5- Le *q* de Tobin

Les trois dernières théories étudiées (datant des années soixante) ont en commun de baser l'investissement optimum sur la maximisation de la valeur de l'entreprise. Dans les années soixante-dix, une nouvelle conception vient concurrencer celle-là : la théorie du coefficient *q*, initiée par Tobin, que nous avions analysée au sous-chapitre 7.1.4. Ce coefficient, rappelons-le, est le ratio de la valeur actualisée des revenus escomptés d'un incrément d'une unité de capital sur le prix de livraison de cette unité de capital. La firme a intérêt à investir tant que q >1.

La conception du *q* peut être accommodée du coût d'ajustement au même titre que celle de la maximisation de la valeur. Dans ce cas, *q* vaudra à 1 lorsque l'accroissement des revenus actualisés sera égal à la dépense totale générée par l'investissement en ce compris le coût d'ajustement.

Dans son article « Tobin's Marginal *q* and Average *q*: A Neoclassical Interpretation » (1982), **Hayashi** démontre que les deux conceptions reviennent au même : « It is increasingly recognized that Tobin's conjecture that investment is a function of marginal *q* is equivalent to the firm's optimal capital accumulation problem with adjustment costs »[3]. Il montre comment *q* peut être dérivé des variables intervenant dans la maximisation de la valeur de l'entreprise.

Le coefficient *q* est une variable marginale : c'est l'investissement d'une unité de capital qui est envisagée. « Unfortunately, however, marginal *q* is not directly observable. What we can (in principle) observe is average *q* ». Le *q* moyen égale la valeur de l'entreprise divisée par le coût de remplacement de la totalité de son capital. Hayashi démontre que "marginal *q* and average *q* are essentially the same in the special yet important case where the firm is a price-taker and the production function and the installation function are homogenous"[4].

---

[1] L'accroissement naturel de la demande au taux λ n'entraîne pas le déplacement de cette courbe.
[2] Lucas [236] p. 331.
[3] Hayashi [137] p. 213.
[4] Hayashi [137] p. 218.

9.6.6. Le coût du capital selon Modigliani et Miller

Une marche arrière chronologique nous ramène à 1958. Modigliani et Miller publient l'article "The Cost of Capital, Corporation Finance and the Theory of Investment", qui répond à la question "What is the 'cost of capital' to a firm in a world in which funds are used to acquire assets whose yields are uncertain; and in which capital can be obtained by many different media… ?"[1]. L'article s'articule autour de trois propositions axées sur la structure du capital de l'entreprise (répartition du capital entre fonds propres et fonds empruntés). Les auteurs supposent la concurrence parfaite, l'efficience des marchés financiers, notamment que tous les agents (entreprises et particuliers) jouissent d'une égale facilité d'emprunter ; les obligations des différentes firmes sont considérées comme de parfaits substituts par les épargnants. Les entreprises (indice *j*) sont rangées dans des classes de risque (indice *k*) distinguées par leur taux de capitalisation des revenus sur le marché des actions ρ_k.

1- "the market value V_j of any firm is independent of its capital structure and is given by capitalizing its expected return X_j at the rate ρ_k appropriate to its class"[2]. Avec S_j pour la capitalisation boursière des actions et D_j pour la valeur de marché des emprunts, à l'équilibre, on a :

$$V_j \equiv (S_j + D_j) = X_j/\rho_k \quad \forall j \in k \qquad (9.40)$$

Comme corollaire, le *coût moyen du capital* X_j/V_j est également indépendant de la structure du capital :

$$X_j/(S_j + D_j) \equiv X_j/V_j = \rho_k \qquad \forall j \in k \qquad (9.41)$$

Les auteurs prouvent cette proposition en montrant que si deux firmes 1 et 2, identiques en tous points sauf la structure du capital, n'avaient pas $V_1 = V_2$, leurs actionnaires pourraient profiter de possibilités d'arbitrage.

2- Dans l'hypothèse normale où le capital investi rapporte plus que le taux d'intérêt des obligations *r*, l'entreprise *j* de la classe *k* qui se finance partiellement par la dette pourra offrir à ses actionnaires un return supérieur à ρ_k. La règle est celle-ci : le return attendu *i* sera une fonction linéaire du taux d'endettement. C'est ce qu'on appelle l'*effet de levier* de l'endettement.

$$i_j = \rho_k + (\rho_k - r).D_j/S_j \qquad (9.42)$$

3- "If a firm in class *k* is acting in the best interest of the stockholders at the time of the decision, it will exploit an investment opportunity if and only if the rate of return on the investment, say ρ^*, is as large as or larger than ρ_k"[3]. Cette règle est valable, quel que soit le mode de financement de l'investissement: autofinancement, emprunt ou augmentation du capital propre.

L'exposé est ici simplifié. Modigliani et Miller analysent les effets de la fiscalité qui habituellement permet la déduction des intérêts mais pas des dividendes.

Le modèle considère ρ_k et *r* comme des données exogènes. ρ_k ne couvre que le risque commercial et industriel. L'écart entre i_j et ρ_k est censé couvrir le risque financier. Modigliani et Miller reconnaissent que si D_j/S_j atteint un niveau inquiétant, *r* qui devient une fonction de l'endettement s'en trouvera accru ; de par (9.42), $(i_j - \rho_k)$ diminue alors avec l'endettement. La prime couvrant le risque financier se réduit donc lorsqu'il s'accroît. Cette conclusion me paraît illogique.

---

[1] Modigliani & Miller [262] p. 261. L'optique de l'article est microéconomique.
[2] Modigliani & Miller [262] p. 268. J'ai intercalé Vj et Xj.
[3] Modigliani & Miller [262] p. 288

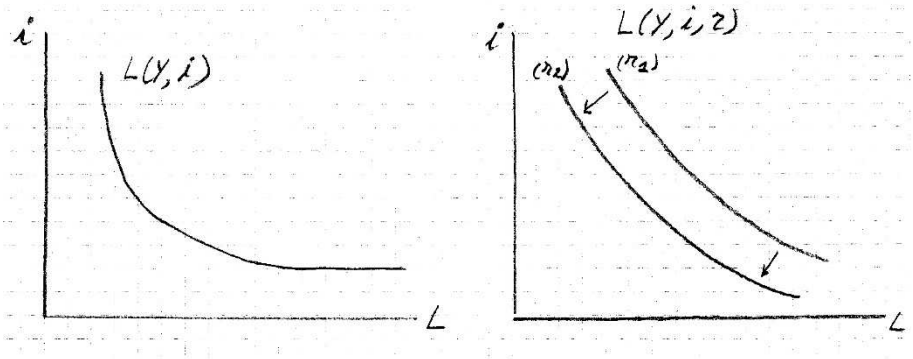
9.7- ON KEYNESIAN ECONOMICS AND THE ECONOMICS OF KEYNES

Au sous-chapitre 7.5.2, nous avons évoqué la défense de la théorie du déséquilibre contenue dans cet ouvrage de **Leijonhufvud**. Mais le cadre de cet ouvrage est bien plus large. C'est une analyse de l'œuvre de Keynes à la fois perspicace et compétente. C'est également un excellent ouvrage de macroéconomie. Caractéristique étonnante, il ne comporte aucune équation et compte un seul et unique graphique. L'objectif affirmé est la redécouverte du vrai Keynes, souvent mal interprété, particulièrement par les néo-keynésiens. Leijonhufvud pense que la *Théorie Générale* marque moins une rupture par rapport à l'œuvre antérieure de Keynes qu'on a bien voulu le croire. Globalement Leijonhufvud endosse la conception économique de Keynes mais n'en souligne pas moins certaines faiblesses.

Une idée courante parmi les économistes est que la théorie keynésienne du chômage repose essentiellement sur la rigidité du salaire nominal, ce que contestent les keynésiens. Leijonhufvud juge cette affirmation peu crédible eu égard à la déflation salariale majeure qui a marqué la grande crise des années trente, dont on sait qu'elle servit de contexte à l'écriture de la Théorie Générale. Sous la plume des détracteurs de Keynes, le terme rigidité des salaires est assorti du sous-entendu que le comportement anti-concurrentiel serait le nœud de la question. Il s'agit là d'une conception étriquée de la rigidité.

Les néo-keynésiens ont défiguré l'explication de l'intérêt, de l'épargne et de l'investissement. Notamment leur postulat selon lequel tant l'épargne que l'investissement sont inélastiques par rapport au taux d'intérêt n'a pas de fondement chez Keynes. Le sous-emploi prévaut, non pas du fait que l'investissement serait insensible à la baisse du taux, mais parce que, précisément, il est si malaisé d'abaisser ce taux. En comprendre le pourquoi constitue l'un des nœuds de la pensée keynésienne ; la thèse néo-keynésienne de la *trappe des liquidités* en rend compte très imparfaitement. Keynes fait intervenir le taux d'intérêt sous DEUX variables distinctes : le taux d'intérêt proprement dit (variable *i*) et le niveau que l'opinion considère comme stable en longue période (variable *r*)[1]. La formule de la demande de monnaie est en fait L = L(Y,i,r) et non pas L = L(Y,i) comme dans le modèle IS-LM. Le graphique 9.12 met côte à côte la version néo-keynésienne (à gauche) où la demande de monnaie devient parfaitement élastique lorsque le taux est bas et la version de Keynes (à droite) où l'on constate un recul global de la demande de monnaie lorsque le *taux normal* est revu à la baisse.

Graphique 9.12 : la résistance à la baisse du taux d'intérêt



Chez Keynes, lorsque le taux « normal » baisse de r_1 à r_2, la courbe de demande de monnaie se déplace vers le bas.

[1] Il s'agit du taux auquel l'épargnant peut placer ses avoirs sans trop devoir craindre une perte en capital.

On comprend que dans l'optique de Keynes, le salut réside dans la baisse de *r*. A court terme, la capacité de la banque centrale d'affecter ce taux est limitée, mais à long terme, elle peut y arriver. La banque centrale doit avoir pour objectif de garder *r* au niveau de ce que Wicksell appelait le *taux naturel*, c'-à-d la productivité du capital.

Selon Leijonhufvud, les commentateurs de Keynes tendent à sous-estimer son intérêt pour un mécanisme qu'il avait décrit sous l'appellation *effet d'aubaine*. Il consiste en un EFFET DE RICHESSE dont bénéficie le détenteur de titres en cas de baisse du taux d'intérêt. On connaît le doute exprimé par Keynes quant au fait qu'une variation du taux d'intérêt affecterait la répartition du revenu entre la consommation et l'épargne. D'après Leijonhufvud, ces doutes portaient exclusivement sur l'effet de SUBSTITUTION entre ces deux grandeurs. Une baisse des taux accroît la valeur du portefeuille ; en conséquence, le détenteur de titres se sent plus riche et de ce fait augmente sa propension à consommer comme dans l'*effet Pigou*[1].

Keynes a simplifié le monde financier en l'emboitant dans un modèle à deux marchés, ceux de la monnaie et des obligations, et un seul prix, le taux d'intérêt des obligations. L'offre et la demande de monnaie suffisent à déterminer ce taux. Mais si on intègre dans le modèle, la multiplicité des marchés financiers du monde réel, cette explication simpliste ne tient plus. Les agents choisissent le marché sur lequel ils prêtent ou empruntent en fonction des caractéristiques propres des différents actifs et des écarts de taux d'intérêt. Il apparaît alors que les forces de l'épargne et de la productivité, chères aux classiques, ont été évacuées trop hâtivement par Keynes. La théorie doit expliquer à la fois le niveau des taux d'intérêt et la dispersion entre eux. Mais la *préférence pour la liquidité* n'éclaire que la dispersion. Elle n'est qu'un élément parmi d'autres dans une théorie complète de l'intérêt.

Leijonhufvud est encore plus sévère contre la théorie néoclassique de l'intérêt. Elle reste la grosse lacune de la microéconomie, le trou à combler. Cette théorie ne voit dans l'épargne qu'une consommation différée, négligeant le mobile de la possession. C'est avec l'ascendant de l'école néowalrassienne que cette conception a triomphé. Parmi les grands précurseurs, Pareto, Knight, Cassel, Pigou ont reconnu l'existence du motif de possession.

Leijonhufvud consacre de nombreuses pages à la relation des taux d'intérêt avec la *structure des termes des actifs*. Il rappelle les trois grandes théories qui ont tenté d'expliquer les écarts entre les taux portant sur des termes différents :

- la ***théorie des anticipations*** (Meiselman) : le marché est dominé par des traders qui spéculent sur les futurs taux à court terme. Le taux présent à long terme reflète le consensus entre eux sur l'évolution future des taux à court terme.
- la ***préférence pour la liquidité*** (Keynes, Hicks) : les taux à plus long terme sont plus élevés, car ils incluent une prime de liquidité
- la ***segmentation du marché*** : les entreprises désirent conserver un certain parallélisme entre la structure de l'actif de leur bilan et celle du passif.

Meiselman a vertement critiqué la théorie de la préférence pour la liquidité, pointant son invalidation empirique : il n'est pas rare que le taux à court terme dépasse le taux à long terme. Leijonhufvud montre que Keynes était loin de planer au-dessus de ces questions. Dans les années vingt, le taux à long terme avait enfin décliné après une longue période où il était historiquement élevé. Il s'ensuivit tout normalement un boom des valeurs boursières, face auquel l'autorité monétaire britannique, craignant la spéculation, mena une politique monétaire

[1] Personnellement, je doute qu'un effet de richesse causé par des facteurs à court terme et réversibles comme les variations du taux d'intérêt ou du niveau général des prix puisse affecter la propension à consommer.

restrictive. Contrairement à la majorité des économistes, Keynes prônait une politique laxiste pour permettre ce nécessaire ajustement et il attribua le taux à court terme trop élevé à cette politique inappropriée.

Un facteur influençant le niveau du taux d'intérêt est le risque encouru par l'investisseur. Lequel de l'investissement à court terme ou de celui à long terme est le plus risqué ? En fait, il faut distinguer deux types de risques :

- le ***risque de perte de revenu*** : si j'investis à court terme, j'ignore à quelle condition je réinvestirai le capital à l'échéance
- le ***risque de perte de capital*** : si j'investis à long terme et que j'ai besoin de dépenser la somme avant l'échéance, je ne sais pas à quel prix je pourrai revendre l'actif.

Si le premier risque domine, le taux à court terme doit être plus élevé, sinon le taux à long terme doit l'emporter. Il est extrêmement difficile de dégager une règle en cette matière. Ceci suffit à Meiselman pour condamner la théorie de la préférence pour la liquidité.

Keynes avait abordé la segmentation du système financier réel. Les entreprises doivent investir à long terme pour profiter des rendements supérieurs des méthodes de production plus détournées. En même temps, les ménages souhaitent éviter le risque des placements à long terme. Le rôle des banques et des autres organismes de crédit est de financer les actifs à long terme des entreprises par des placements à court terme des ménages. Historiquement, constate Keynes, la scission est de plus en plus marquée entre la décision d'épargner et celle d'investir. En rendant plus efficace la canalisation du capital, la société a donc accru l'instabilité.

Quelle est finalement la cause du sous-emploi d'après Keynes ? Leijonhufvud rejoint les autres théoriciens du déséquilibre : le vecteur des prix en vigueur n'est pas équilibrant. Mais le prix à blâmer n'est pas le salaire sur lequel s'acharnent les classiques. C'est le prix des actifs financiers qui est trop bas. L'incitation à investir est moindre que la propension à épargner. La faute est imputable aux spéculateurs qui thésaurisent la recette des ventes de titres au lieu de la réinvestir. Leijonhufvud estime que ce jugement de Keynes pèche par excès de généralisation.

Le prix des actifs peut être trop bas, soit parce que l'efficacité marginale du capital est anémique, soit parce que le taux d'intérêt à long terme est excessif. Selon Keynes, cette deuxième éventualité est la cause la plus courante. Le taux d'intérêt à long terme est au centre du problème.

Le déséquilibre se fait principalement sentir sur le marché du travail (chômage) ; pourtant, sa cause est un disfonctionnement, pas tant du marché du travail que de l'équilibre général du système. La réduction du salaire prônée par les classiques n'est donc pas la solution. Non pas que la vision de Keynes conclut à l'inexistence d'un vecteur d'équilibre ; simplement la baisse du salaire n'y mènera pas. Keynes ne ferme toutefois pas la porte à un effet remédiant de la déflation salaire-prix ; elle peut être avantageuse à la condition qu'elle accroît la valeur des actifs, par exemple via l'*effet Keynes*[1].

La politique monétaire serait assez inefficace, contrairement à la politique budgétaire. Telle est la sentence généralement attribuée à Keynes, notamment par les néo-keynésiens. A tort, d'après Leijonhufvud, selon qui Keynes n'a jamais renié la politique monétaire. Dans l'optique de Keynes, la politique monétaire serait plus efficace pour empêcher le retournement conjoncturel de se produire, pour prévenir la dégradation de la production et de l'emploi. Par contre, la politique budgétaire est à préférer lorsque le marasme s'est installé et qu'il faut relancer une économie déjà fort affaiblie. Keynes s'oppose à la hausse du taux d'intérêt en période de

[1] Cf. sous-chapitre 5.1.1.

tension conjoncturelle. Cette position ne peut que surprendre et heurter les tenants de la théorie du surinvestissement, pour qui cette réduction participe à la nécessaire répression de la demande en ces circonstances. Keynes défend au contraire des taux d'intérêt peu élevés en permanence.

Il semble que sur ce point, Leijonhufvud interprète Keynes assez librement. Comme nous l'avons vu au chapitre 5.1, les citations de la Théorie générale ne manquent pas, qui mettent l'accent sur l'efficacité marginale du capital et la politique budgétaire. Toutefois, il est bien certain que la politique monétaire est restée en grâce auprès de Keynes.

La politique monétaire qui incite efficacement les entreprises à investir est celle qui leur garantit que le capital investi vaudra plus que le coût d'acquisition des équipements. Idéalement, l'autorité monétaire doit veiller à accroître la valeur des actifs qui sont les substituts les plus proches du capital physique. Keynes recommande qu'elle achète des titres à long terme. Dans beaucoup de pays, l'autorité monétaire est réticente à l'acquisition de ces titres, leur préférant des effets à court ou moyen terme. Lorsque l'autorité monétaire se concentre sur les titres à court terme, la substitution entre actifs de termes différents opérée par les offreurs et les demandeurs exercera un effet de dominos, communiquant la baisse des taux vers des actifs de plus en plus durables. La politique monétaire risque alors d'atteindre son but trop lentement. En outre la simple formule mathématique déterminant la valeur des titres montre clairement que la sensibilité de cette valeur aux variations du taux d'intérêt est une fonction croissante de leur terme.

Leijonhufvud termine l'ouvrage en expliquant pourquoi les théories néoclassique et keynésienne sont irréductibles l'une à l'autre. Les néoclassiques et les keynésiens proclament tous deux que leur théorie est plus générale que celle de l'adversaire. Selon Leijonhufvud, c'est l'absence du commissaire-priseur chez Keynes qui permet de considérer sa théorie comme plus générale.

Dans le paradigme néoclassique, le système de prix a une double fonction :
- fournir les stimulants pour que les agents réalisent les ajustements souhaitables
- fournir aux agents l'information nécessaire pour la coordination des activités

Keynes retient la première de ces deux fonctions mais ne croit pas à la seconde. Il conserve également du paradigme néoclassique l'hypothèse que les agents économiques sont animés d'un comportement optimisateur.

En conclusion, Leijonhufvud dépeint le contraste entre une théorie néoclassique désincarnant l'économie réelle dans un modèle au formalisme prodigieux et une théorie keynésienne dont l'appareil analytique brouillon n'est pas à la hauteur de sa conception économique très inspirée.

9.8- POURQUOI Y A-T-IL DES ENTREPRISES ?

LA NATURE DE LA FIRME

« The Nature of the Firm » est un article publié en 1937 par Ronald **Coase** où il pose la question qui, suivant l'optique, peut sembler la plus grave ou la plus légère de la science économique : pourquoi y a-t-il des firmes ? Pourquoi la firme existe-t-elle ? Ce qui motive cette question, c'est que la théorie néoclassique a démontré l'extrême efficacité du marché. Alors, pourquoi celui-ci n'est-il le lieu de passage que d'une partie des relations coopératives tissées par les hommes pour produire ? Les relations coopératives à l'intérieur de la firme sont en effet d'une nature différente de celles établies sur les marchés qu'on dit si efficaces. Pourquoi toutes les utilisations de ressources, ne font-elles pas l'objet d'un échange sur le marché ?

La raison réside dans les *coûts de transaction* : coûts de recherche d'un contractant, qui plus est, du meilleur contractant. Coûts à renouveler sans cesse, car sans la firme, il n'y a pas de contrats stables. La stabilité des contrats permet d'économiser des coûts de transaction. Or, l'essence de la firme, ce sont précisément des contrats à long terme par lesquels, dans les limites prévues par la loi, les détenteurs des facteurs de production se soumettent à l'autorité de l'entrepreneur.

La firme a une taille plus ou moins grande selon que plus ou moins de transactions sont gérées par l'entrepreneur plutôt qu'échangées sur le marché. Qu'est-ce qui détermine la taille de l'entreprise ? En fait, la fonction d'entrepreneur a elle-même des rendements décroissants. Jusqu'à une certaine taille, l'économie des frais de transaction dépasse les avantages du marché ; au-delà, le rapport s'inverse. « The question always is, will it pay to bring an extra exchange transaction under the organising authority ? At the margin, the cost of organising within the firm will be equal either to the cost of organising in another firm or to the cost involved in leaving the transaction be 'organised' by the price mechanism »[1].

L'ENTREPRENEUR SELON ALCHIAN ET DESMETZ

L'explication de Coase n'eut pas l'heur de convaincre Alchian et Desmetz. Ils reposèrent donc la même question dans un article publié en 1972 et y apportèrent leur propre réponse.

Ils s'opposent à l'idée que le pouvoir autoritaire de l'entrepreneur serait l'essence de la firme. La seule sanction qu'il peut appliquer à ses collaborateurs est de les licencier. Exactement comme quand sur le marché, il change de fournisseur. La durée des contrats n'est pas plus le nœud de la question : il est possible d'établir des contrats cadres à long terme sur le marché, tout comme il est possible d'employer du personnel à la journée.

La spécificité de la firme découle de la conjonction de deux facteurs : « it is in a team use of inputs and a centralized position of one party in the contractual arrangements of all other inputs »[2].

La production sera plus efficace si chaque agent est rémunéré suivant sa productivité. S'ensuit la nécessité de mesurer l'apport de chacun et sa rétribution. Le fait est que ce mesurage n'est pas réalisable à coût nul. Le problème vient de ce que dans la production coopérative (team production), le résultat n'est pas séparable. L'observation du produit ne permet pas de mesurer l'apport propre de chaque intervenant. Renoncer au mesurage reviendrait à donner l'avantage

[1] Coase [60] p. 404
[2] Alchian & Desmetz [5] p. 778

aux tire au flanc[1], mais consacrer plus de ressources au mesurage qu'il n'apporte de bénéfice serait irrationnel. La question devient donc : comment organiser un team pour optimiser l'efficacité ?

Dans le cadre d'une production coopérative, les transactions sur le marché sont impuissantes contre le *shirking*[2]. L'organisation la plus efficace est celle qu'Alchian et Desmetz appellent « the classical firm ». Elle consiste à attribuer la lutte contre le shirking à un agent spécialisé. Mais comment éviter le shirking de la part de cet agent ? En le rétribuant avec la recette résiduelle après paiement des autres inputs. La fonction de l'entrepreneur est avant tout l'anti-shirking, mais elle inclut les responsabilités connexes que sont l'embauche des autres facteurs, la distribution des tâches, la révision et la fin des contrats. Est également incluse dans cette fonction, le pouvoir de vendre les droits précités.

La fonction d'entrepreneur peut être partagée par plusieurs intervenants[3] (par exemple dans une coopérative), mais alors le risque de shirking des 'moniteurs' s'accroît avec leur nombre. Cette forme d'entreprise convient donc mieux pour les petits groupes.

Comme Coase, Alchian et Desmetz réfutent la théorie de l'entreprise développée par Knight. Comme nous l'avons vu (cf. chapitre 9.4), celui-ci fondait la fonction d'entrepreneur sur la prise de risque : il y a des entrepreneurs parce que certains, plus téméraires, doivent prendre la responsabilité de faire vivre les entreprises. Nos auteurs doutent que la distribution de l'entreprenariat et celle de l'aversion pour le risque coïncident à ce point.

Parmi les inputs, certains sont possédés par l'entrepreneur, d'autres sont loués. Vu l'interdiction de l'esclavage, les facteurs humains sont toujours loués. La tendance est à posséder les inputs matériels lorsqu'il s'agit de biens durables dont la durée de vie dépend du soin avec lesquels ils seront maniés. Le moniteur central, observateur permanent des processus internes de l'entreprise, pourra mieux s'assurer du bon usage de ce matériel qu'un propriétaire bailleur absent.

L'exercice de sa fonction par l'entrepreneur fait de lui un agent superinformé des caractéristiques des inputs internes à l'entreprise. Ceci lui permet de les combiner plus efficacement que ne le pourrait un acteur qui les acquiert tous par le marché.

On peut reprocher à l'analyse d'Alchian et Desmetz qu'elle postule la supériorité de la firme classique sur la marché en matière de lutte contre le shirking, plus qu'elle ne la démontre.

ALCHIAN, DESMETZ ET LES DROITS DE PROPRIETE

L'année suivante, les mêmes auteurs publient un article intitulé « The Property Right Paradigm » qui répond à la question : qu'est-ce que la propriété ? « What is owned are rights to use resources, including one's body and mind »[4]. Il existe une multitude de droits qui ont leurs

---

[1] Si tout le monde tire au flanc, la situation est désavantageuse pour le bien-être global, car le taux marginal de substitution RÉEL entre le loisir et le travail qui s'égalise avec leur rapport de prix ne serait pas égal au taux marginal de substitution SOUHAITÉ et les agents travailleront moins que le niveau qui maximise leur utilité.

[2] Le lecteur me pardonnera cet anglicisme. La langue française n'a pas de substantif pour l'action d'être un tire-au-flanc. Le terme *shirking* revient très régulièrement dans l'article.

[3] Ce que les auteurs appellent « profit sharing ».

[4] Alchian & Desmetz [6] p. 17

limites légales. « What are owned are socially recognized rights of action ». Des droits différents relatifs à une même ressource peuvent être détenus par des agents différents.

Les auteurs distinguent trois situations alternatives :
- les droits communaux
- la propriété publique (étatique ou locale)
- la propriété privée.

Les droits communaux sont ceux qui n'empêchent pas les autres personnes d'exercer les mêmes droits. Contrairement aux deux formes de propriété, ils ne comportent aucune exclusivité[1]. Ce serait par exemple le droit de chasse ouvert à tous dans un bois communal.

Une grande partie de l'article a précisément pour but de critiquer ce type de situation. Les droits y sont exercés suivant le principe « first come, first serve ». Il n'y a donc aucune incitation à économiser les ressources, puisque les bénéficiaires n'en supportent pas le coût ». Dans l'exemple ci-dessus, la préservation du gibier serait mise en danger. Un autre exemple est celui des embouteillages sur les routes gratuites. Les auteurs ne cachent pas leur préférence pour des routes payantes. « The communal right system raises transaction costs by creating a free rider problem »[2].

Une loi déclarant que les animaux abattus sur le bois communal appartiendraient à la communauté et non au chasseur qui les a tués résoudrait le problème de l'« overhunting » mais pourrait susciter le problème inverse de l' « underhunting ». La société ne pourrait y remédier que par un surplus de régulation ou d'endoctrinement.

La conclusion d'Alchian et Desmetz est celle-ci : « Contrary to some popular notions, it can be seen that *private* rights can be socially useful precisely because they encourage persons to take account of *social* costs »[3].

[1] Les auteurs remarquent que couramment, « communal ownership is technically associated with state ownership, as in the case of public parks, wherein the state technically has the capability of excluding persons from using his property » (p.19)
[2] Alchian & Desmetz [6] p. 21
[3] Alchian & Desmetz [6] p. 24

TITRE X- CONCLUSIONS

L'introduction se terminait par la question du statut de l'économie politique, présente également dans le titre de l'ouvrage. La conclusion tentera d'esquisser une réponse. Mais en guise de préparation à cet exercice complexe, j'offre maintenant à mon esprit critique un espace de liberté totale, je le lâche face aux théories qui sont exposées du titre II au titre IX. Dans ces pages, je m'étais permis çà et là de les commenter. Ici, je voudrais mettre en question des hypothèses fondamentales partagées par la majorité des économistes néoclassiques. C'est l'objet des chapitres 10.1 et 10.2 consacrés respectivement à la microéconomie et à la macroéconomie[1].

Les économistes déduisent parfaitement des conclusions correctes de leurs hypothèses. C'est l'avantage du formalisme mathématique pour lequel ils ont opté. Encore faut-il que les hypothèses soient saines. Pour les évaluer, il est impératif de distinguer :

- Les **hypothèses de base** : exemples : un revenu présent est préféré à un revenu futur ; les firmes maximisent le profit…
- Les **simplifications** : exemples : le nombre de ménages est constant, il n'y a qu'un type de travail, tous les consommateurs ont les mêmes goûts…

Ce sont les hypothèses de base qui nous intéressent. Des simplifications, on peut tolérer certains écarts par rapport à la réalité économique, car ces écarts ont une incidence limitée et clairement déterminée. Par contre, les hypothèses de base sont censées représenter les grandes caractéristiques du monde économique réel. Un manque de réalisme aurait ici des conséquences graves. J'insiste sur cette distinction, car les économistes néoclassiques tendent à considérer que tous les aspects de l'économie réelle sont susceptibles de faire l'objet de « simplifications »[2].

Les hypothèses de base ont trait à des phénomènes complexes, multidimensionnels ; elles sont des théories en soi. Il est donc difficile de les qualifier simplement de « vraies » ou « fausses », comme on pourrait le faire à propos d'un élément ponctuel d'une théorie. Le cas échéant, il est toutefois possible de révéler le manque de solidité de certaines hypothèses.

Le manque de solidité d'une hypothèse de base se manifeste par la négligence d'aspects importants de la vie économique. Bien-sûr, l'ampleur de la littérature économique est telle que tous les sujets y ont déjà été traités. Mais l'existence de quelques articles traitant de tel objet ne suffit pas à démentir que la théorie le néglige. Une contribution relative à une hypothèse de base ne s'intègre dans l'édifice de la théorie économique que si la plupart des modèles ultérieurs traitant du même objet ne s'autorisent plus à l'ignorer[3].

---

[1] Certaines de ces critiques sont des prises de position relativement à des débats existants. D'autres émanent directement de ma réflexion. Vu l'étendue de la littérature économique dont je n'ai pas connaissance, il est peu probable que je sois précurseur dans l'une ou l'autre.

[2] Qu'on se souvienne de l'argument dont Friedman s'était servi pour disqualifier toute critique contre le réalisme des hypothèses (sous-chapitre 5.5.6) : sans hypothèses, théoriser deviendrait impossible. Cet argument simpliste ne s'accorde pas avec notre distinction entre les deux types d'hypothèses.

[3] Donnons un exemple : en 1972, Hayne Leland a produit un article très intéressant sur le comportement des producteurs faisant face à une demande aléatoire. Depuis sa parution, l'immense majorité de la littérature économique traitant de l'équilibre du producteur ignore ses conclusions. Cet article ne fait pas partie du corpus de la théorie économique. Il n'existe que comme satellite de ce corpus.

10.1- MICROECONOMIE

LES FIRMES

Avant de mettre en question telle ou telle hypothèses, il me faut déplorer la confusion qui entoure la conception générale de l'entreprise, des ses objectifs ainsi que de l'entrepreneur. Des éléments aussi fondamentaux devraient être clarifiés avec une vue globale, mais l'observateur a l'impression de se retrouver devant un puzzle.

La théorie est laconique au sujet des entreprises. On connaît leur but qui semble unique[1] : maximiser ce qu'on appelait auparavant le profit et qui est devenu la maximisation de leur valeur conçue comme l'actualisation des cash flows futurs. On peut sans doute imputer au faible intérêt pour les questions de financement que cette valeur n'est que rarement spécifiée être celle des actionnaires ou propriétaires. Les firmes de la théorie n'ont ni âge ni histoire ; leur rapport avec l'entrepreneur n'est pas pris en considération. A priori, cela semble acceptable puisque la théorie doit se préoccuper du général, non du particulier ; mais trop d'éléments ne sont-ils pas rejetés dans le particulier simplement parce qu'ils sont plus difficiles à modéliser ? La théorie devrait expliquer comment une entreprise dont la taille est limitée par la volonté de contrôle des propriétaires peut être florissante, comment des firmes rentables de tailles diverses coexistent sur un marché. La détermination théorique de la taille est purement mécanique et fait intervenir trop peu de facteurs.

Le risque semble absent de l'environnement des entreprises. La théorie ne nous explique pas comment la firme choisira entre des investissements qui diffèrent à la fois par la rentabilité et le risque. Une fonction d'utilité mêlant profit et risque n'apparaît que trop rarement. Certains économistes transfèrent entièrement la gestion du risque des épaules de l'entrepreneur vers le marché financier. Les actionnaires sélectionneraient leur portefeuille en fonction du risque associé aux différentes entreprises ; il suffit alors aux managers d'écouter le marché, ordonnateur des prix, pour prendre les bonnes décisions. Cette conception néglige certains faits : toutes les entreprises ne sont pas cotées en bourse ; les actionnaires qui contrôlent les entreprises cotées n'ont pas la même mobilité que les simples épargnants. En outre, les managers disposent d'une information plus fournie que les actionnaires.

L'entrepreneur est aussi fantomatique que son entreprise. On sait qu'il se rémunère avec le profit résiduel, ce qui est interloquant puisque sur le long terme ce profit est censé tendre vers zéro. On peut en inférer que l'entrepreneur doit donc être le détenteur de l'un des facteurs de production, mais ce n'est nulle-part écrit noir sur blanc. La théorie s'est préoccupée du rôle de l'entrepreneur mais de son rôle uniquement. Par exemple, le mystère plane quant à savoir lequel de ses organes est l'entrepreneur d'une société par actions.

Un des grands paradoxes est que l'économie politique décrit la fonction de l'entrepreneur de telle façon qu'il pourrait être remplacé par un ordinateur : élaborer un plan de production à partir d'un système de prix. Alors qu'en société, les qualités le plus saluées sont l'initiative et l'intuition.

[1] L'instinct de survie face au risque de faillite n'est même pas mentionné

LA CONCURRENCE

Les économistes définissent des *structures de marché*. Les principales sont : la concurrence parfaite, le monopole, l'oligopole et la concurrence monopolistique. Les critères pour les départager sont :

- Le nombre d'offreurs
- L'homogénéité ou l'hétérogénéité du produit
- L'absence ou la présence de collusion
- L'absence ou de la présence de barrières à l'entrée.

La *concurrence parfaite* capte une part prépondérante de l'attention des économistes[1]. Pourtant, dans l'économie réelle, les quatre critères ci-dessus sont rarement combinés de la façon qui y mène, ce qui fait d'elle un cas plutôt rare. Mais, comme son nom l'indique, la concurrence parfaite est érigée en une perfection théorique vers laquelle le secteur concurrentiel de l'économie est censé tendre. A ce titre, elle en est une représentation sublimée, sans quoi l'effort d'analyse qui la concerne serait disproportionné.

La concurrence parfaite, est-elle réellement la représentation simplifiée d'une vaste réalité ? Certaines de ses propriétés en font douter. Un exemple : à l'équilibre de longue période, la firme de la théorie produit au minimum de son coût moyen, en conséquence de quoi, un accroissement de la demande sur le marché ne peut être rencontré que par une variation du nombre de firmes. Est-ce là une représentation même simplifiée de l'économie réelle ?

A mon avis, la concurrence parfaite est représentative d'un type de marchés très spécifique : les marchés organisés, tels ceux de certaines matières premières. Les autres situations s'en écartent. Idéalement, l'appellation « concurrence parfaite » devrait même être remplacée par celle de « marchés organisés ». En fait, ce qui est « parfait » dans cette structure de marché, ce n'est pas l'ardeur compétitive des firmes protagonistes mais l'adaptation aux besoins de la modélisation économique. La concurrence parfaite est « modélogénique ».

En fait, un type de structure de marché est ignoré des économistes. Celle où un petit nombre d'entreprises se livrent à une concurrence acharnée. Elle est certainement plus courante que la concurrence dite parfaite. Comme nous l'avons vu au sous-chapitre 5.5.4, dans le cadre de la *controverse marginaliste*, Harrod regrettait que cette structure n'était même pas nommée ; il proposait l'appellation « libre concurrence avec marché imparfait ». Dans les pages qui suivent, je l'appellerai « concurrence oligopolistique ». Les producteurs à la demande non horizontale sont suspicieusement regardés comme au moins partiellement protégés de la concurrence. Pourtant, on connaît des situations où des milliers d'artisans boulangers appliquent les tarifs préconisés par leur association professionnelle et vendent le pain plus cher que des grandes surfaces sensiblement moins nombreuses mais au comportement plus compétitif.

Le piédestal de la concurrence parfaite s'accorde avec la prééminence du premier critère, le nombre d'offreurs. Que des facteurs aussi importants que la liberté d'accès ou la pratique de la collusion ne soient pas discriminants heurte le sens commun. Leur tort aux yeux des économistes est d'être moins mesurables[2].

---

[1] L'équilibre général d'Arrow-Debreu qui sert de socle à la microéconomie moderne postule la concurrence parfaite. La grande majorité des modèles macroéconomiques posent également cette hypothèse.

[2] « Mesurables » n'est pas le terme le plus exact. L'utilité n'est pas mesurable non plus ; elle est pourtant omniprésente dans l'économie. Les économistes apprécient les concepts qui sont des quantités ou sonnent comme telles, afin d'en faire des variables dans leurs équations.

Quel est l'objet ou quels sont les objets de la concurrence ? Selon Chamberlin, les firmes manient trois paramètres pour se faire une place sur le marché :

- le prix,
- la qualité du produit,
- les dépenses promotionnelles[1]

Toujours comme suite de la préférence des économistes pour le quantifiable, la première de ces formes de concurrence est infiniment plus étudiée que les deux autres. Mais elle l'est par le petit bout de la lorgnette. Le prix comporte deux éléments : le coût de production et la marge bénéficiaire. La concurrence peut porter sur chacun de ces deux éléments. Tout observateur sensé admettra que la concurrence sur les coûts est l'une des caractéristiques essentielles de l'économie réelle. La place que lui réserve la théorie économique est anecdotique. Comme si la concurrence sur les prix était avant tout une concurrence sur les marges. Mais la malléabilité des marges par la concurrence est forcément limitée[2] : elles doivent fournir un profit suffisant pour satisfaire les marchés boursiers. Un fossé sépare la concurrence réelle et l'image qu'en donnent les ouvrages de microéconomie.

Les formes de concurrences portant sur les coûts, sur la qualité des produits et sur la préférence de marque sont une course à la performance. Il s'agit d'être plus productif que les concurrents, d'offrir de meilleurs produits ou de se créer une image plus attirante. Elles offrent aux entreprises l'opportunité de se distinguer alors que la concurrence sur les marges est uniformisante.

L'EQUILIBRE DU PRODUCTEUR

Le monde de l'entreprise comporte certainement plus de *faiseurs de prix* que de *preneurs de prix*. Peu d'entrepreneur se reconnaîtront dans la description walrassienne : l'homme à l'écoute des prix criés par le commissaire priseur et qui sait exactement combien produire à chaque prix. L'entreprise réelle recherche la combinaison prix-quantité idéale et cette quête est complexe, car interviennent de multiples paramètres. Une relation lie généralement la qualité du produit d'une part avec le coût et le prix d'autre part. Ces variables sont au centre d'un choix. Le rapport qualité-prix n'étant pas quantifiable, il intéresse peu les économistes. Autre choix, les dépenses R&D influent sur le rapport entre la qualité et le coût. Les dépenses de marketing importent également, tout comme la publicité gratuite qu'offre un volume de ventes élevé. La fixation du prix ne peut s'expliquer qu'au vu de tous ces paramètres.

Dans l'équilibre général d'Arrow-Debreu, l'ensemble des firmes comprend aussi celles qui n'existent pas encore ; de même l'ensemble des produits comporte ceux qui restent à créer. Globaliser les produits existants et futurs est une négation de la concurrence, car la mise sur le marché de nouveaux produits est justement l'un de ses objets fondamentaux ; c'est l'occasion par excellence de faire mieux que ses concurrents ; les nouveaux produits ne sont pas un fruit qu'on cueille mais un arbre qu'on plante.

Revenons à l'idée selon laquelle le plan de production peut être décidé sur base de la seule connaissance du vecteur des prix présents et futurs. Si l'on se réfère à des idées émises par Muth et Radner (cf. supra), l'anticipation des prix futurs pourrait même se limiter au décodage des prix présents et passés. La toute-puissance du système de prix anime le modèle de Lucas ; tout ne serait qu'harmonie sans l'inaptitude potentielle du producteur à distinguer si le prix a varié à cause de l'inflation ou d'un déplacement de la demande. Dans le même ordre d'idées,

[1] Les deux dernières formes sont parfois regroupées sous l'appellation « nonprice competition ».
[2] Je considère ici la malléabilité volontaire, reflétant les décisions stratégiques de l'entreprise.

Hayek écrit que “production is generally guided not by any knowledge of the actual size of the total demand, but by the price to be obtained in the market (… The entrepreneur) is not in the least concerned with the amount by which, in a given case, the total amount demanded will alter; he only looks at the price he can expect to get after the change in question has taken place”[1].

Peu d’entrepreneurs se reconnaîtront dans cette description. Ils tentent au contraire de s’informer à toutes les sources possibles à propos de tous les aspects de leur commerce : fiscalité, technologie, politique publique, goûts du public, actions des concurrents, évolution conjoncturelle… Les grandes entreprises et les associations professionnelles disposent souvent d’un bureau d’étude ; l’information commerciale est devenue l’objet d’un commerce où opèrent des firmes spécialisées. Reconstituer le puzzle du futur à partir de toutes ces pièces d’information est un jeu où certains entrepreneurs se révèlent plus inspirés que d’autres. L’affirmation de Lucas (cf. supra) selon laquelle les entrepreneurs se soucient peu de l’évolution conjoncturelle et ont du mal à se repérer dans le cycle paraît erronée. C’est l’une des lacunes de la microéconomie de se croire dispensée d’analyser l’effet du cycle conjoncturel sur le comportement des agents. Les économistes se sont préoccupés d’assurer les MICROFONDATIONS de la macroéconomie mais ont oublié les MACROFONDATIONS de la microéconomie, comme si le contexte dans lequel opèrent les agents n’influençait pas leurs actions.

Grâce au coût unitaire croissant (à partir d’un certain niveau de production), les courbes de recette marginale et de coût marginal se rejoignent, ce qui permet de déterminer un optimum de la firme, tant à court terme qu’à long terme. Pourtant, les coûts croissants sont loin de faire l’unanimité et les études empiriques ne tendent pas à les confirmer. En *concurrence oligopolistique*, la courbe de demande de la firme est décroissante ; un équilibre existera donc bien, même si le coût unitaire est constant ou légèrement décroissant. Mais cet équilibre, vient-il vraiment de l’intersection entre la recette marginale et le coût marginal ?

Dans la théorie orthodoxe, en égalisant recette marginale et coût marginal, les firmes obtiennent normalement un profit net ; celui-ci attire de nouveaux arrivants. La demande de chaque firme s’en trouve progressivement réduite jusqu’à ce que ce profit disparaisse. En voulant maximiser le profit à court terme, les entreprises ont fait disparaître le profit à long terme et ont vu fondre leur part de marché et leur taille. Il est peu probable que des firmes gérées par des managers rationnels se comportent de la sorte. Comme expliqué au sous-chapitre 5.5.4, Harrod avait parfaitement résolu ce problème. Les firmes savent le prix qu’elles peuvent demander sans stimuler l’accroissement de l’offre. En l’absence de collusion et de barrières à l’entrée, ce prix ne peut évidemment pas comporter de surprofits importants.

Puisque la marge bénéficiaire ne doit assurer que la rémunération normale du capital comprise dans le coût, il est parfaitement rationnel pour une entreprise de produire au-delà de l’égalisation de la recette marginale avec le coût marginal. Obtenir un profit supérieur reste possible ; la clé est d’être plus performant que les concurrents : avoir un rapport qualité-coût supérieur permet de gagner soit sur la quantité soit sur la marge.

L’erreur des marginalistes est de croire que l’entreprise maximise son profit à long terme en maximisant son profit instantané à tout moment. L’intuition des full costers fut de comprendre que les profits aux différents instants ne sont pas des variables indépendantes et que la maximisation du profit à long terme, forcément le véritable objectif de la firme, doit comporter une règle propre. Malheureusement, les full costers ont exprimé leur intuition maladroitement

[1] Hayek [143] p. 33

en discréditant la maximisation du profit. Trop heureux de l'erreur de leurs adversaires, les marginalistes en ont profité pour clore la controverse prématurément.

CAPITAL, INTERET, EPARGNE ET PREFERENCE TEMPORELLE

La théorie fisherienne demeure le fond de la pensée néoclassique en cette matière. Elle a été intégrée dans l'équilibre général néo-walrassien. La préférence temporelle y joue un rôle fondamental.

Avant d'aborder la théorie de l'intérêt, examinons plus en détail la maximisation de l'utilité intertemporelle du consommateur. Dans plusieurs modèles analysés tout au long du présent ouvrage apparaissent des variantes de l'équation générique (présentée ici en temps discret) :

$$V_t = \sum_{k=0}^{T-t} \beta^k . u(C_{t+k}) \quad \text{où } \beta = \left(\frac{1}{1+\rho}\right) \qquad (10.1)$$

L'utilité intertemporelle au temps *t* notée V_t est la somme des utilités instantanées *u* de la consommation aux différentes époques t+k, actualisées au taux de préférence temporelle ρ. L'initiateur de cette équation n'est autre que Samuelson (1937), s'inspirant de Rae, Böhm Bawerk et Fisher. Mais Samuelson, sceptique à l'égard de sa créature, ne revendique pour elle ni valeur normative, ni valeur descriptive. Cette réticence n'a pas empêché l'économie politique moderne d'en faire l'un de ses piliers.

Les dernières décennies du XXe siècle virent une importante littérature analyser le bien-fondé de l'escompte des satisfactions futures, aux niveaux tant de l'individu que du gouvernement appelé à ménager les intérêts discordants de la génération présente et des générations à venir. Cette réflexion recouvre un double questionnement : est-il rationnel et est-il éthiquement défendable ? Le débat fut occupé par les philosophes plus que par les économistes.

Pour répondre adéquatement à la question, il faut sérier deux aspects apparemment semblables mais indépendants :

- la *préférence temporelle pure* (le terme β de 10.1) par laquelle une satisfaction aujourd'hui est préférable à la même satisfaction dans le futur ;
- le fait que divers facteurs tendent à réduire l'utilité instantanée *u* qu'une consommation *C* produira dans le futur par rapport à la même consommation présentement.

Mais l'effet de dévalorisation de la consommation future est identique et cumulable.

Un argument avancé en faveur de la préférence temporelle pure est que l'être humain n'a pas une personnalité continue. Il est au cours des phases de sa vie une succession de personnalités différentes. Que la personnalité présente ne se soucie que modérément du sort de ses « successeurs » n'a rien d'irrationnel.

En faveur du déclin de l'utilité instantanée future, il est argué notamment qu'avec la croissance économique, les sujets anticipent généralement une hausse du revenu futur, donc de leur consommation dont l'utilité marginale est ainsi appelée à décroître. Cet argument me paraît peu pertinent ; comme l'a montré Duesenberry, l'individu s'attache surtout au rapport entre sa consommation et celle du reste de la société ; en moyenne, ce rapport ne varie forcément pas ; l'utilité marginale de la consommation ne subit donc pas de biais systématique de ce type.

L'incertitude est un autre argument souvent invoqué. L'individu se préoccupe moins de ses intérêts futurs, car il ne sait pas s'il sera encore en vie, ni quel seront son état de santé, ses goûts etc.

En dépit de tous les arguments exprimés dans le cadre de ce débat et antérieurement, Trostel et Taylor arrivent en 2001 à cette conclusion partagée par d'autres auteurs : « Despite the almost universal assumption of time preference in models of intertemporal choice, an adequate explanation of why people discount the future has not been provided »[1].

En l'absence d'une préférence systématique pour la consommation présente, la prime au report de la consommation serait rendue inutile. Il faudrait alors rechercher l'explication de l'intérêt dans d'autres types de considérations. Déterminer la forme préférée de la répartition temporelle de la consommation s'avère donc critique pour l'ensemble de la problématique.

Les théories que nous avons étudiées au chapitre 9.5 mettent en avant le souhait de lisser la consommation, d'en faire l'élément constant par opposition à l'épargne qui est l'élément résiduel. Ceci ne suffit pas à exclure une préférence pour un trend croissant ou décroissant sur la vie. Les expériences « psychologiques » révèlent que les sujets mis face à cette alternative préfèrent généralement un trend croissant à un trend décroissant[2], ce qui n'est pas étonnant. Il me semble en outre que terminer sa vie dans le dénuement est pour l'homme moyen une perspective effrayante et qu'il est motivé à prendre les dispositions nécessaires pour l'éviter.

Knight, un des rares économistes sceptiques à l'égard de la préférence pour le présent, pose correctement la problématique : « If a man were given his entire income for a year in a lump-sum payment on January first, how would he distribute its expenditure through the year? (…) A zero-time preference obviously means a uniform distribution in time (…) As to the facts of human nature, it is safe to assume that different individuals would give the most varied forms of distribution. »[3]. L'individu se construit une image de l'évolution de ses besoins au cours de sa vie, fonction de sa situation familiale, de sa santé, de ses projets personnels et professionnels... Il tentera de faire adhérer sa courbe de consommation à cette courbe de besoins tout en respectant sa contrainte budgétaire[4]. Cette tendance pourrait être appelée « préférence pour une consommation équilibrée » ; elle semble plus plausible que la préférence pour le présent.

Une facette de la conception néoclassique connexe au principe de l'impatience consiste à ne voir en l'épargne qu'une consommation différée. Fisher argumente que seule la consommation apporte la jouissance et que l'épargne n'a donc aucune justification si la consommation n'est pas au bout du chemin. En est-il vraiment ainsi ? Des grands économistes comme Pigou ont reconnu le désir de posséder comme motif à l'épargne. Il arrive que des auteurs concèdent cette motivation aux très riches, mais elle semble généralement déniée aux autres classes sociales. Admettons que les très pauvres ne sont pas tentés par l'épargne, car elle mettrait leur vie en danger. Mais les membres de la classe moyenne semblent trouver une satisfaction à posséder un capital. L'économie politique, peut-elle réellement répondre de façon scientifique à ces questions : comment les différentes catégories sociales épargneraient-elles en l'absence d'intérêt ? Et si l'intérêt devait être une agent motivant indispensable, quelle gêne servirait-il à

[1] Trostel & Taylor [367] p. 379. Dans cet article, les auteurs proposent une « nouvelle » justification de l'escompte. Ils évacuent toute *préférence temporelle pure*, mais considèrent qu'en vieillissant, l'individu voit baisser sa capacité de profiter de la consommation, ce qui abaisse l'utilité instantanée future. L'anticipation de ce déclin l'amène à escompter le futur.

[2] Cf. Frederick & al [96] p. 363

[3] Knight [189] § 2.4.51. Dès 1921, il écrivait : « The assumption of a general preference in human nature for present over future goods is so commonly and so confidently made that some courage is required to call in question the foundations of the entire body of doctrine on the subject ; yet it must be done ».

[4] Cette contrainte est d'ailleurs souvent définie erronément dans les modèles, car elle néglige la réticence et les obstacles à l'emprunt.

compenser : l'impatience, le simple sentiment d'être séparé de son argent, la peur de le perdre[1] ? A moins que gagner de l'intérêt ne devienne un but en soi ?

Chez Fisher, les écarts entre le taux de préférence pour le présent et le taux d'intérêt génèrent des mouvements de capitaux qui tendent à ramener le taux d'intérêt vers le taux d'impatience. Mais les études empiriques semblent indiquer une élasticité très faible de l'épargne par rapport au taux d'intérêt. Keynes avait d'ailleurs intégré ce trait dans sa Théorie Générale. Dans ces conditions, il est douteux que le taux d'intérêt puisse être guidé par la préférence pour le présent, même dans l'hypothèse implausible où celle-ci est avérée.

Marx avait élaboré le concept de *plus-value* pour montrer que tout l'intérêt provient de l'exploitation du travail. Nous avons vu au sous-chapitre 2.4.1 que ce concept inobservable est impropre à l'analyse scientifique. Böhm Bawerk vînt avec le concept de *préférence pour le présent*, tout aussi inobservable et lui aussi plus adapté à la spéculation qu'à l'analyse scientifique. La vérité est qu'au XXIème siècle, l'économie politique est incapable d'expliquer l'intérêt.

L'EQUILIBRE ET LE FACTEUR TEMPS

Considérons une économie de marché assez semblable à nos économies réelles. Admettons que si on pouvait en écrire les équations, elles aboutiraient à l'existence d'un équilibre général. En analyse statique, il serait parfaitement raisonnable de présumer que l'état de cette économie correspond à son équilibre général, sauf à supposer qu'il est mathématiquement démontré que cet équilibre est instable, une hypothèse que nous écartons. L'analyse dynamique prend en compte une multitude de chocs qui affectent cette économie, ce qui a pour effet de faire diverger son état et son équilibre. De ces chocs, écartons les désordres monétaires et le cycle conjoncturel, de façon à ne retenir que les chocs microéconomiques. Inévitablement, sur une majorité des marchés de cette économie, la demande excédentaire restera positive ou négative, parfois faiblement, parfois plus fort selon la brutalité des chocs. Le déséquilibre est donc la situation habituelle de toute économie et il serait présomptueux de prétendre que ces demandes excédentaires sont généralement de faible proportion. Si les agents étaient effectivement des homo economicus tels que les décrit l'économie politique, la dépression nerveuse les guetterait. Heureusement, les agents réels n'ont généralement qu'une idée assez vague de ce qu'est leur optimum ; de plus, le facteur temps tempère leur perception du déséquilibre : ce qu'ils n'ont pu acheter ou vendre aujourd'hui alimente l'espoir de transactions futures.

L'économie néoclassique a inventé deux processus qui doivent mener cette économie immanquablement à l'équilibre : le *tâtonnement* de Walras et le *re-contrat* d'Edgeworth. Totalement irréalistes, il est peu probable que l'un ou l'autre soient le reflet imagé de forces plus obscures à l'œuvre dans l'économie réelle. Toutes deux ont le tort de négliger le facteur temps. Sauf sur les marchés financiers et les autres marchés organisés, les réajustements menant à l'équilibre prennent au mieux quelques mois, au pire de nombreuses années lorsque des investissements importants sont nécessaires. Certains croient balayer cette critique en considérant que c'est la normalité qu'une économie court toujours derrière l'équilibre sans jamais l'atteindre. L'équilibre resterait malgré tout le moteur du fonctionnement. Mais ladite critique se place à un niveau plus fondamental que le simple écoulement du temps. Deux facteurs, l'insertion de l'économie dans son contexte socio-politique et le dynamique propre à

[1] L'expression « le risque zéro n'existe pas » est entrée dans le langage courant pour une multitude de questions non économiques. En économie, le terme « actifs sans risque » intervient couramment. Est-on sûr que certains placements ne rapportent pas une prime de risque ? Que cette prime de risque n'englobe pas toute la rémunération du capital ?

l'économie (le progrès) la soumettent à une succession de stimuli qui altèrent son équilibre à un rythme effréné. Il y a une disproportion entre les deux échelles de temps : celle des modifications de l'équilibre et celle du retour vers l'équilibre.

L'une des lois battues en brèche par cette difficulté est celle qui égalise la rémunération des facteurs avec leur productivité marginale. Faisons abstraction des causes statiques qui entravent cette égalité, comme les fonctions de production non homogènes, le salaire d'efficience... Clark et après lui Pigou et d'autres insistèrent sur ce qu'il faut entendre par une unité supplémentaire de capital : non pas ajouter un outil identique en plus mais réorganiser le capital à partir d'outils plus performants et plus coûteux (cf. supra). Les chocs technologiques, la variation du taux d'épargne, l'évolution démographique impliquent des réajustements tels que la substitution entre les facteurs ou des déplacements de main d'œuvre, sans lesquels l'égalisation entre la rémunération et la productivité marginale n'aura pas lieu. Mais ces réajustements nécessitent des années. Un temps de latence est ainsi offert aux facteurs de production pendant lequel règne une certaine indétermination des rémunérations et où triomphent les rapports de force. Lorsque l'ampleur des délais atteint un certain seuil, l'égalité entre la rémunération et la productivité marginale cesse d'être une description pertinente de la réalité.

10.2- MACROECONOMIE

LES MICROFONDATIONS

Un leitmotiv des macroéconomistes modernes est de fonder la macroéconomie sur un socle microéconomique, c'est-à-dire qu'ils récusent toute thèse dont on ne pourrait garantir la cohérence avec la doctrine microéconomique. Tout naturellement, ils se tournent vers la microéconomie néoclassique orthodoxe. Nous avons vu au chapitre précédent que la solidité de cette théorie est discutable. Il n'est donc pas certain que la macroéconomie soit bénéficiaire de ces fondations qu'on veut lui assurer.

Si les macroéconomistes avaient daigné s'inspirer de la microéconomie non orthodoxe, ils se serait épargné beaucoup d'efforts pour analyser la *rigidité des prix*. Si on considère les entreprises comme des *faiseurs de prix* déterminant en situation d'incertitude un rapport qualité-prix et un rapport qualité-coût optimaux, la rigidité des prix devient une évidence ainsi que son corolaire, la réaction-quantité aux variations de la demande. Si les entreprises sont convaincues de l'optimalité de cette attitude, le déséquilibre entre l'offre et la demande sera persistant.

LES MODELES MACROECONOMIQUES

Les hypothèses rencontrées le plus fréquemment sont :

- Fonction de production de Cobb-Douglas à rendements d'échelle constants. Même si à première vue, l'hypothèse des coûts constants paraît raisonnable dans le cadre de l'économie globale, elle est généralement posée sans argumentation. De plus, la déduction de l'épuisement du produit, est-elle valide si la production ne manifeste des rendements d'échelle constants qu'en moyenne ?
- Concurrence parfaite sur les marchés des facteurs et les marchés des produits.
- Un produit unique sert à la fois pour la consommation et l'investissement. La *controverse du capital* avait pourtant relevé que l'abandon de cette hypothèse pouvait réserver des surprises. Le capital agrégé continue à être mesuré en unités de ce produit unique, comme si la controverse n'avait pas eu lieu[1].
- L'équilibre du marché du travail détermine non pas le nombre d'emplois, mais le nombre d'heures travaillées par chaque salarié. Il n'y a donc pas de chômeurs mais des heures chômées[2].
- Soit l'incertitude est absente, soit les agents connaissent les distributions probabilistes des variables et tentent de maximiser ou minimiser l'espérance mathématique de l'une d'elles.

Chacune de ces hypothèses et bien d'autres encore amènent un biais. Quand tous ces biais sont combinés, le résultat est exponentiellement biaisé, d'autant plus que tous vont dans le même sens : celui qui convient le mieux à l'équilibre.

POURQUOI Y A-T-IL DU CHOMAGE ?

Le caractère statique des arguments échangés dans le débat entre les monétaristes et les keynésiens est surprenant. Déjà la Théorie générale accorde trop peu d'importance aux facteurs

[1] La microéconomie néowalrassienne a désagrégé le capital ce qui l'immunise partiellement de la critique cambridgienne, mais les macroéconomistes sont restés fidèles aux fonctions agrégées.

[2] On devrait même dire simplement : des heures non consacrées au travail, puisque certaines théories en font le résultat d'un choix.

dynamiques, comme l'avait souligné Leontief. Le débat entre les néo-keynésiens et les monétaristes est resté dans la même veine. Pour l'observateur, il est évident qu'il y a du chômage (non frictionnel) parce qu'il y a un cycle conjoncturel. Mais le cycle conjoncturel a été étudié comme un objet particulier, un peu déconnecté. La réputation de Hicks lui vient plus du très statique modèle IS-LM que de son modèle du multiplicateur-accélérateur, infiniment plus éclairant.

Si l'économie réelle était statique, il est probable que le plein emploi prévaudrait. En dernier ressort, c'est l'emploi qui rapporte le profit tant convoité. Pour que l'économie s'en écarte durablement, elle doit combiner ces deux facteurs :

- il faut que le déséquilibre soit possible. L'analyse statique montre où est la faille.
- Il faut qu'opèrent l'un ou l'autre processus cumulatifs auto-alimentés. Ceux-ci ne se conçoivent que dans le cadre d'une analyse dynamique.

Le caractère cumulatif du processus dynamique permet au déséquilibre de toujours garder une longueur d'avance sur les tendances au rééquilibrage[1]. Le déséquilibre change de sens sans que les forces d'équilibre aient pu stabiliser l'économie. Les théoriciens du cycle ont déployé des trésors d'imagination pour générer des oscillations non amorties et non explosives. Les oscillations sont un phénomène bien connu des physiciens et certains économistes s'en sont inspiré. Il n'est toutefois pas certain que ce soit la manière idéale d'appréhender le cycle économique. Peut-être faut-il simplement rechercher dans chacune des quatre phases les germes de la suivante.

LA MONNAIE

L'attachement à une conception purement exogène de la monnaie a de quoi étonner. On la trouve aussi bien chez certains keynésiens, dont Keynes, que chez les monétaristes. La monnaie naît du crédit ; la demande de crédit, tant des entreprises que des particuliers, varie avec le niveau d'activité économique. Bien-sûr, en manipulant les taux d'intérêt, l'autorité monétaire peut influencer le volume du crédit et donc celui de la monnaie. La création de monnaie endogène subit une modulation exogène. Dans ce cas, la baisse du taux d'intérêt n'est pas une conséquence de l'élargissement monétaire mais une condition nécessaire à sa survenance. L'*open market policy* permet aux banques commerciales d'adapter leurs moyens à la demande des emprunteurs ; en achetant les titres à bon prix ou en les vendant à bas prix, la banque centrale peut les influencer dans un sens ou dans l'autre. Mais la variation finale de la masse monétaire dépendra de l'élasticité-intérêt de la demande de crédit[2]. Le domaine de la monnaie purement exogène se limite au crédit direct de la banque centrale aux pouvoirs publics. Cette forme de crédit est décriée et en perte de vitesse.

Que la création monétaire soit endogène ou exogène n'est pas anodin du point de vue de l'équilibre macroéconomique. Par exemple, ce qu'on a appelé l'*effet Keynes* (cf. 5.1.1) n'opère que si la monnaie est exogène.

La théorie nous explique que lorsqu'il dispose d'une encaisse accrue, le consommateur augmente sa demande de biens, poussant leurs prix à la hausse. Mais cette hypothèse est mal posée, car elle ne raconte pas l'histoire à partir du début. D'où vient cet élargissement

[1] Si la macroéconomie fait abstraction du cycle, elle ne pourra jamais comprendre la persistance d'un déséquilibre.

[2] Il est frappant de voir que les très amples variations de la *base monétaire* dans la zone Euro depuis la crise de 2008 se sont traduites en variations des *réserves bancaires*, avec une *masse monétaire* croissant sagement.

monétaire ? Dans l'« Optimum Quantity of Money »[1], Friedman imagine un hélicoptère larguant des billets de banque. Voyons quelques causes potentielles plus réalistes :

- ***les purs transferts d'actif monétaire entre les agents, entreprises ou particuliers***. Ils ne nous concernent pas puisque la hausse de l'encaisse de l'un est compensée par la baisse de celle de l'autre.
- ***L'épargne par thésaurisation*** permet à un agent d'augmenter son encaisse, mais alors il s'agit d'un choix, pas d'un fait exogène et il serait absurde d'en déduire un accroissement de la demande de biens.
- Reste la cause la plus naturelle de l'élargissement monétaire, ***l'accroissement du revenu nominal***. Mais dans ce cas, la demande accrue de biens découle de la hausse du revenu, pas de celle de l'encaisse ; cette dernière ne fait que suivre le revenu.

La hausse du revenu nominal reflète soit celle de la production soit celle des prix. Comme l'avait déjà remarqué Tooke il y a deux siècles, ces deux phénomènes opèrent souvent en amont de la monnaie.

LES ANTICIPATIONS RATIONNELLES

Le principe des *anticipations rationnelles* est des plus convaincants. Les agents ne se contentent certainement pas d'extrapoler le passé lorsqu'ils tentent de prévoir les évolutions futures. Ils tentent, au contraire, de rassembler le maximum d'informations et sont parfaitement capables de les interpréter correctement.

Il est toutefois un phénomène qui échappe aux anticipations rationnelles : la spéculation. Y interviennent des éléments de la psychologie de groupe. L'optimisme et le pessimisme sont communicatifs et les humeurs communiquées se renforcent mutuellement. C'est beaucoup demander aux anticipations rationnelles de prévoir le moment exact où se produira le retournement d'une spirale de prix. Et chaque intervenant veut éviter d'être le premier à subir les effets du retournement de la courbe du prix. Ainsi adviennent les *bulles spéculatives*. C'est sur les marchés spéculatifs que les anticipations rationnelles ont le plus de chances d'être battues en brèche. Les bulles spéculatives ne sont pas une exception anodine ; on les constate régulièrement avant un retournement conjoncturel.

Sur les autres marchés et tout particulièrement le marché du travail, les anticipations rationnelles me semblent la règle.

N'est-ce pas étonnant qu'il ait fallu attendre les *nouveaux classiques* pour qu'apparaisse cette hypothèse raisonnable ? Ils n'en sont pas peu fiers et semblent l'avoir sacralisée. Pourtant, elle ne joue qu'un rôle tout-à-fait accessoire dans leur théorie. Très justement, Muth écrit : « I should like to suggest that expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory ». Le tout est de savoir quelle est la théorie adéquate. Pour les nouveaux classiques, pas de doutes : c'est la théorie classique, avec notamment l'*équation des échanges* fisherienne.

Considérons un salarié rationnel qui croit que la théorie keynésienne est correcte. Supposons qu'il assiste à un accroissement de la masse monétaire décidé par l'autorité qui veut relancer une économie fatiguée. Comme Keynes, il sait que ce ΔM va se répartir en un ΔT et un ΔP selon une proportion qui est fonction du taux de sous-emploi. Il prendra soin de se prémunir

[1] Friedman [102] p. 4.

contre le ΔP[1]. Les précautions des agents voulant assurer l'intégrité de leur revenu réel n'absorberont qu'une part de la monnaie nouvellement crée ; le solde servira à accroître la demande et le revenu réels.

CAPITAL, INVESTISSEMENT ET PROGRES TECHNIQUE

Le capital est un fluide homogène qui coule dans la tuyauterie économique. L'objet des théories de l'investissement est de déterminer l'ajout de fluide nécessaire pour que celui-ci atteigne sa quantité optimale.

Dans l'économie réelle, l'hétérogénéité du capital est essentielle. Le plus souvent, un équipement arrivé en fin de vie est remplacé partiellement ou totalement par un équipement aux caractéristiques différentes, normalement plus performant ; l'investissement d'extension et celui de remplacement se confondent. Dire à quel taux la quantité de capital a augmenté au cours de l'opération est impossible. Si la firme atteint son objectif, le capital sera devenu qualitativement plus optimal. Mais qu'en est-il quantitativement ?

La variable la plus importante économiquement est la CAPACITE PRODUCTIVE du capital installé[2]. Elle combine les aspects quantitatif et qualitatif du capital. Elle manque toutefois d'une unité de mesure adéquate tout autant que la quantité et la qualité du capital. La seule variable quantifiable dont on dispose est la VALEUR du capital[3]. Pas plus que sa quantité physique, elle n'est représentative de la capacité productive. Sur la longue période, l'augmentation de la valeur installée (hors inflation) sous-estime l'augmentation de la capacité productive puisque celle-ci réduit le coût de production unitaire des nouveaux équipements. Ce domaine de la science économique est en proie à un problème de quantification fondamental. Les modèles économétriques sont obligés de quantifier physiquement le capital à partir de statistiques relatives à sa valeur. Ne serait-il dès lors pas plus honnête de la part de la théorie de travailler avec des valeurs ?

C'est tout à l'honneur de Solow d'avoir introduit le concept de *progrès technique incorporé* et l'approche *vintage* dans la théorie de la croissance et du progrès technique. Mais il n'est pas allé au bout de sa logique, car il s'obstine à distinguer les parts respectives de la croissance qui sont dues au progrès technique et à l'accumulation. Considérons un investissement qui incorpore un progrès technique ; la croissance qui s'ensuit ne peut être imputable ni au progrès technique ou à l'accumulation exclusivement, ni divisée en deux parts distinctes. L'accumulation n'est pas indépendante du progrès technique ; celui-ci élargit les opportunités d'investir. Rien n'indique qu'autant de capital serait investi si le progrès technique était inférieur. Il me paraît douteux que la seule substitution entre le travail et le capital à technologie égale puisse soutenir un niveau d'investissement comme celui qui a caractérisé les économies industrielles ces dernières décennies. De ce point de vue, on peut se demander si les théories de l'investissement étudiées au chapitre 9.6 ne traitent pas d'un aspect très secondaire de l'économie réelle.

---

[1] Dans beaucoup de cas, sa rationalité ne sera même pas nécessaire, car le contrat de travail ou la convention collective prévoient l'indexation automatique.

[2] On suppose que le capital installé est le résultat de l'optimisation de la combinaison du capital avec la main d'œuvre.

[3] Pour autant qu'on ait choisi une méthode de valorisation.

10.3- CONCLUSION GENERALE

L'énumération de faiblesses de l'économie politique dans les chapitres 10.1 et 10.2 n'est certainement pas exhaustive. Elle noircit toutefois le tableau à l'excès précisément parce qu'il s'agit d'une liste de points faibles.

Je suis bien conscient qu'à chacune des déficiences mentionnées ci-dessus peuvent être opposées des objections et sans doute des contre-objections à ces objections et ainsi de suite. C'est précisément la raison pour laquelle j'envisage l'économie politique comme une discipline de type philosophique plutôt que scientifique. Il est vrai que la science se construit aussi avec des débats ; mais toute discussion scientifique doit déboucher sur une conclusion ; c'est ce qui justifie l'existence du débat. Certaines oppositions entre économistes me semblent avoir un caractère irréductible. Les controverses entre keynésiens et anti-keynésiens, entre néoclassiques et néo-ricardiens ou entre marginalistes et antimarginalistes, diffèrent-elles par leur nature de la dispute perpétuelle entre empiristes et rationalistes ?

Au cours du vingtième siècle, la philosophie des sciences s'est énormément développée. Des philosophes comme Popper, Kuhn, Lakatos et bien d'autres ont réfléchi sur ce qu'est la science, sur ce qui permet à une théorie scientifique ou à un savoir de prétendre à la validité. L'unanimité ne règne pas plus dans ce domaine qu'en économie politique. Envisageons la question : « l'économie politique est-elle une science ? » ; et la question connexe : « la théorie économique dominante, satisfait-elle aux critères de validité du savoir scientifique ? ». L'ampleur de ces questionnements et de la production philosophique existante à exploiter pour y répondre est telle que la présente conclusion se trouverait bien étroite. C'est un deuxième volume qu'il faudrait écrire et je serais bien incapable d'en être l'auteur.

L'ambition de cette conclusion se limitera donc à ceci : je relèverai quelques difficultés de l'économie politique qui semblent épargnées aux sciences dites exactes et en conséquence desquelles ses « lois » me semblent mériter un statut proche de celui des doctrines philosophiques[1]. Ces critiques méthodologiques ne sont évidemment pas originales, tant le sujet a déjà été débattu.

Le problème fondamental est comment valider ou invalider des lois dont les objets n'ont pas d'existence matérielle et ne peuvent donc pas être observées directement ou faire l'objet d'expériences en laboratoire. Prenons l'exemple du chômage : on peut certes observer la manifestation concrète du phénomène, en ce sens qu'on sait que des personnes sont chômeuses. Il s'agit alors d'une observation subjective totalement différente de l'observation d'une étoile au télescope ou d'une cellule au microscope. Pour observer le chômage en tant que tel, il doit être conceptualisé ; son observation sera alors elle aussi conceptuelle et non directe, avec tous les risques de biais que cela engendre. L'économie politique tente de contourner cette difficulté en ayant un maximum de ses objets qui sont des grandeurs ; on parlera plus du taux de chômage que du chômage. Si la grandeur est mesurable, les mesures peuvent effectivement servir de base au processus de validation. A ce que la mesure d'une grandeur économique apporte une validation fiable, deux types de difficultés font obstacle :

1. Les mesures empiriques se trouvent dans les statistiques. Celles-ci jouent donc un rôle fondamental en économétrie. Mais les statistiques reflètent le monde réel avec toutes ses interactions. Le résultat des forces économiques y est mêlé avec les éléments de

[1] Loin de moi l'idée que la philosophie est inférieure à la science. Elles sont complémentaires. La science explique et la philosophie éclaire.

politique intérieure, de politique internationale, les éléments culturels, les détournements de flux par les frontières, les luttes sociales, la fiscalité… Pour valider des théorèmes qui isolent certaines variables, on se réfère à des statistiques ou tout est mêlé.

2. L'économie politique manipule nombre de grandeurs non mesurables et pas des moindres : l'utilité, l'escompte du temps sont des exemples notables. Dans le même registre, l'offre et la demande sont des courbes dont on ne peut observer qu'un seul point à la fois ; les observations successives sont généralement insuffisantes et n'excluent pas les interférences parasites. La réalité est qu'en microéconomie, peu de mesures peuvent être réalisées.

Les statistiques brutes étant trop biaisées, les économistes ont l'habitude de les retravailler pour neutraliser certains biais qu'ils entrevoient. Premier écueil, l'analyste doit être capable et désireux de percevoir tous les biais importants, dont certains sont bien cachés. Deuxième écueil, le risque existe de remplacer un biais par un autre.

L'économétrie regorge de régressions, de calculs de corrélation et de covariance. J'avoue avoir été très impressionné par le travail statistique notamment dans le débat entre partisans et opposants de la théorie du *cycle réel*. Mais impressionner ne suffit pas pour convaincre. Les régressions offrent trop de liberté à l'analyste pour assurer que son équation est la seule possible :

- Le choix du type de fonction, a priori très vaste, même si ce sont souvent les mêmes qui reviennent.
- Le choix des variables, parfois fort nombreuses
- Les délais. Telle variable sera comptée *n* périodes en retard en vertu du délai d'adaptation présumé ou même en avance si on estime qu'elle est suffisamment anticipée. Parfois une variable et ses déclinaisons retardées apparaissent ensemble dans une même équation.

Le travail empirique devient presque aussi abstrait que le travail théorique et il a ses propres hypothèses. En principe, cela se justifie, mais ces validations ne sont plus que semi-empiriques et perdent donc une part de leur crédibilité.

Les concepts qui ne sont pas observables directement peuvent être appréhendés indirectement par des expériences « psychologiques », des interviews d'acteurs ou par l'introspection. Par exemple, l'expérimentateur intéressé par l'effet de l'incertitude sur le comportement posera à un panel de cobayes des questions du type : « préférez-vous 100 euros avec certitude ou un choix aléatoire entre 60 euros à 50% et 150 euros à 50% ». Mais si la personne questionnée n'est pas amenée à ressentir psychiquement le sentiment d'incertitude quant à sa situation, peut-on déduire des conclusions fiables de sa réponse ? L'introspection, quant à elle, est généralement méprisée par les économistes. Peut-être à tort. L'introspection, aurait-elle laissé dire que le taux d'épargne s'élève lorsque le taux d'intérêt augmente ?

Le plus souvent, la théorie ne se contente pas de constater une corrélation entre deux variables, elle établit une relation causale entre elles. L'une est placée au rang de cause, l'autre à celui d'effet. La corrélation statistique est impuissante à établir cette distinction. Vu l'importance des anticipations en économie, la chronologie des changements ne peut même pas être un critère déterminant. La subjectivité de l'analyste intervient nécessairement.

Dans les questions de causalité, le problème le plus grave est celui des déterminations cachées, lorsque la cause est une variable extérieure au modèle. Imaginons un modèle analysant la corrélation avérée entre deux variables *X* et *Y* et concluant sur base d'arguments rationnels que *X* est causal. Il se peut que *X* et *Y* son tous deux l'effet d'une troisième variable *Z*. Le modèle

aura mal interprété la corrélation entre *X* et *Y*. Un exemple déjà évoqué illustre ce type de méprise : les économistes considèrent généralement les variations de l'encaisse comme une cause des variations de la demande de biens, mais peut-être les variations de l'encaisse et celles de la demande de biens découlent-elles toutes deux des variations du revenu nominal.

Outre le caractère causal d'une variable par rapport à une autre, les économistes s'interrogent parfois sur un type de cause qui mériterait un C majuscule : la cause d'un phénomène important, tel que l'intérêt, l'entreprise... Comme nous l'avons vu (chapitre 9.8), Alchian et Desmetz voient en la lutte contre le shirking la cause originelle de l'entreprise. Leur causalité semble signifier une condition à la fois nécessaire et suffisante. Pour juger leur explication, il faudrait procéder à l'exercice mental d'imaginer une société en tous points identique à la nôtre sauf qu'elle ne connaîtrait pas la mentalité de tire-au-flanc. Y existerait-il des entreprises ? Une réponse absolument sûre est évidemment impossible. Mais le bon sens fait douter que ce monde ignorerait les entreprises. Pour des phénomènes sociaux aussi fondamentaux, il existe probablement une multitude de conditions SUFFISANTES et peut-être aucune condition NECESSAIRE significative. Des causes diverses opèrent, les unes de nature sociale, les autres de nature économique. Comme chacune est suffisante, ses partisans ont l'illusion de détenir la vérité. Nous avions terminé le chapitre 10.1 en constatant l'impuissance de l'économie politique face au phénomène de l'intérêt. Cette impuissance s'explique par la déconnexion de l'économie politique des sciences sociales et de l'histoire. Trop d'abstraction tue la connaissance. L'observation de la société découvre de multiples causes SUFFISANTES à l'intérêt. Mais les économistes s'échinent à trouver LA cause à la fois nécessaire et suffisante, qui, peut-être, n'existe tout simplement pas.

Revenons encore aux hypothèses. Leur rôle est stratégique, car d'elles dépend la démontrabilité de telle ou telle idées. Friedman n'admettait pas qu'elles pussent être critiquées sur leur réalisme (cf. 5.5.6). Selon lui, l'unique critère de validité de la théorie est la confirmation empirique. Les considérations qui précèdent font douter que ce critère suffise. La critique du caractère réaliste des hypothèses me semble un acte d'hygiène scientifique. Le choix des hypothèses dépend de la subjectivité de l'analyste. Une voie qui s'offre à lui consiste à choisir des hypothèses qui mènent immanquablement à la thèse. On retrouve cette attitude chez les *nouveaux classiques*, comme l'admet candidement l'un d'eux : "Rational expectations models, however, contain an additional element that has little to do with the formation of expectations : the assumption of equilibrium. In other words, supply is assumed to equal demand in all markets at all times"[1]. Cette hypothèse est sous-jacente dans tous leurs modèles et porte littéralement leurs conclusions.

Fréquemment, l'économiste choisit l'une ou l'autre hypothèses en fonction des besoins mathématiques de la démonstration. Cette démarche est acceptable si l'hypothèse ainsi posée ne bouleverse pas la mécanique du modèle, Une hypothèse comme la durée de vie infinie des agents semble innocente en apparence. Pourtant, il est évident qu'un individu éternel n'aurait pas les mêmes comportements qu'un humain, essentiellement en matière d'utilité intertemporelle. L'individu éternel ne peut pas optimiser ses choix sans escompter le futur : son cas ne peut évidemment pas être généralisé aux mortels. L'infini a des propriétés mathématiques intéressantes qui séduisent les mathématiciens et donc également certains économistes. Dans l'économie réelle, il est dénué de toute signification.

*

[1] Holland [158] p. 5

Par rapport à ses rivales existantes ou potentielles, l'orthodoxie néoclassique jouit d'un avantage énorme. De par le monde, des dizaines de milliers d'économistes travaillent à l'enrichir et la consolider, parfois ou souvent dans le cadre de programmes de recherche structurés et organisés. Ils sont parfaitement entraînés à l'usage d'un appareil mathématique avancé. Après un siècle et demi, il en résulte une construction impressionnante. Mais paradoxalement, il s'agit d'une construction en hauteur sur une assise relativement restreinte. Considérant les traitements mathématiques sophistiqués d'hypothèses souvent discutables, Paul Davidson ironise : « For Keynes as for post keynesians, the guiding motto is 'it is better to be roughly right than precisely wrong!' »[1]. Je cite cette sentence parce que mes lectures m'ont laissé exactement la même impression : le contraste entre le perfectionnement mathématique et le simplisme de certaines hypothèses, comme si l'apparence scientifique comptait plus que le fond. L'apparence scientifique exerce une réelle force de séduction. Auréolée du prestige des mathématiques, l'économie « mainstream » a plus ou moins fait le vide autour d'elle.

*

Le scepticisme affiché dans cette conclusion ne signifie pas que l'économie politique s'est rendue antipathique à mes yeux. Au contraire, sa découverte fut une aventure intellectuelle passionnante. Dès le départ, j'étais plus à la recherche d'idées stimulantes que d'une vérité révélée ; c'est effectivement ce que j'ai trouvé et j'en sais gré aux économistes que j'ai lus. J'écrivais plus haut : la science explique, la philosophie éclaire. Ce sont les débats qui se sont révélés le plus éclairant. Mais pas plus qu'une bonne connaissance des théories philosophiques n'apporte la compréhension du sens de la vie, bien connaître l'économie politique ne permet de comprendre le fonctionnement de l'économie réelle.

[1] Davidson [66] p. 574. Le commentaire de Knight, pourtant néoclassique, va dans le même sens : « The mathematical economists have commonly been mathematicians first and economists afterword, disposed to oversimplify the data and underestimate the divergence between their premises and the facts of life » [190] p. 589.

LES GRANDS ECONOMISTES

Nom	Pays	Dates	class	néo-class	keynes
1- La sainte trinité					
Ricardo David	Grande Bretagne	1772-1823	X		
Walras Léon	France	1834-1910		X	
Keynes John Manyard	Grande Bretagne	1883-1946			X
2- Les figures historiques de la science économique					
Smith Adam	Grande Bretagne	1723-1790	X		
Malthus Thomas Robert	Grande Bretagne	1766-1834			
Mill John Stuart	Grande Bretagne	1806-1873	X		
Marx Karl	Allemagne	1818-1883	X		
Jevons William Stanley	Grande Bretagne	1835-1882		X	
Menger Carl	Autriche	1841-1921		X	
Marshall Alfred	Grande Bretagne	1842-1924		X	
Edgeworth Francis Ysidro	Grande Bretagne	1845-1926		X	
Clark John Bates	Etats-Unis	1847-1948		X	
Pareto Vilfredo	Italie	1848-1923		X	
Böhm Bawerk (von) Eugen	Autriche	1851-1914		X	
Wicksell Knut	Suède	1851-1926		X	
Wicksteed Philip	Grande Bretagne	1851-1927		X	
Fisher Irving	Etats-Unis	1867-1947		X	
Pigou Arthur Cecil	Grande Bretagne	1877-1959		X	
Knight Frank	Etats-Unis	1885-1972		X	
Hayek (von) Friedrich	Autriche	1889-1992		X	
Sraffa Piero	Italie	1898-1983	X		
Kalecki Michal	Pologne	1899-1970	X		X
Harrod Roy	Grande Bretagne	1900-1978			X
Neumann (von) John	Etats-Unis	1903-1957	X	X	
Lerner Abba	Etats-Unis	1903-1982		X	X
Robinson Joan	Grande Bretagne	1903-1983	X		X
Lange Oskar	Pologne	1904-1965		X	X
Hicks John Richard	Grande Bretagne	1904-1989		X	X
Kaldor Nicholas	Grande Bretagne	1908-1986	X		X
Friedman Milton	Etats-Unis	1912-2006		X	
Samuelson Paul	Etats-Unis	1915-2009		X	X
Arrow Kenneth	Etats-Unis	1921-2017		X	
Debreu Gérard	Etats-Unis	1921-2004		X	
Solow Robert	Etats-Unis	1924-		X	X
Lucas Robert	Etats-Unis	1937-		X	
Stiglitz Joseph	Etats-Unis	1943-		X	X
3- Les autres grands économistes					
Cantillon Richard	France-Irlande	1680-1734			
Quesnay François	France	1694-1774			

Hume David	Grande Bretagne	1711-1776			
Turgot Jacques	France	1727-1781			
Bentham Jeremy	Grande-Bretagne	1748-1832			
Thornton Henry	Grande-Bretagne	1760-1815			
Say Jean-Baptiste	France	1767-1832		X	
Tooke Thomas	Grande Bretagne	1774-1858	X		
Thünen (von) Johann Heinrich	Prusse	1780-1870		X	
Cournot Antoine Augustin	France	1801-1877		X	
Wieser (von) Friedrich	Autriche	1851-1926		X	
Hobson John Atkinson	Grande Bretagne	1858-1940			
Barone Enrico	Italie	1859-1924		X	
Cassel Gustav	Suède	1866-1945		X	
Bowley Arthur	Grande Bretagne	1869-1957		X	
Warming Jens	Danemark	1873-1939			X
Aftalion Albert	France	1874-1956			X
Slutsky Eugen	Russie	1880-1948		X	
Mises (von) Ludwig	Autriche	1881-1973		X	
Schumpeter Joseph	Autriche	1883-1950		X	
Clark John Maurice	Etats-Unis	1884-1963			
Robertson Dennis	Grande Bretagne	1890-1963		X	
Viner Jacob	Etats-Unis	1892-1970		X	
Douglas Paul	Etats-Unis	1892-1976		X	
Hotelling Harold	Etats-Unis	1895-1973		X	
Frisch Ragnar	Norvège	1895-1973		X	
Robinson Austin	Grande Bretagne	1897-1993		X	X
Chamberlin Edward	Etats-Unis	1899-1967		X	
Ohlin Bertil	Suède	1899-1979		X	
Kuznets Simon	Etats-Unis	1901-1985			X
Leontief Vassily	Etats-Unis	1901-1999	X		
Wald Abraham	Autriche	1902-1950		X	
Machlup Fritz	Autriche	1902-1983		X	
Menger Karl	Autriche	1902-1985		X	
Tinbergen Jan	Pays Bas	1903-1994			X
Ramsey Frank	Grande Bretagne	1903-1930		X	
Kahn Richard	Grande Bretagne	1905-1989			X
Coase Ronald	Grande Bretagne	1910-2013		X	
Koopmans Tjalling	Etats-Unis	1910-1986		X	
Scitovsky Tibor	Etats-Unis	1910-2002		X	X
Haavelmo Trygve	Norvège	1911-1999		X	X
Allais Maurice	France	1911-2010		X	
Stigler George	Etats-Unis	1911-1991		X	
Metzler Lloyd	Etats-Unis	1913-1980		X	X
Andrews Philip	Grande Bretagne	1914-1971			
Alchian Armen	Etats-Unis	1914-2013		X	
Phillips Alban William	Grande Bretagne	1914-1975			X
Bergson Abram	Etats-Unis	1914-2003		X	
Brunner Karl	Suisse	1916-1989		X	
Savage Leonard	Etats-Unis	1917-1971			
Modigliani Franco	Etats-Unis	1918-2003		X	X
Tobin James	Etats-Unis	1918-2000			X
Duesenberry James	Etats-Unis	1918-2009			X
Buchanan James	Etats-Unis	1919-2013		X	
Sylos-Labini Paolo	Italie	1920-2005			
Patinkin Don	Etats-Unis	1922-1995		X	
Baumol William	Etats-Unis	1922-2017			X

Malinvaud Edmond	France	1923-2015		X	X
Hahn Frank	Grande Bretagne	1925-2013		X	
Clower Robert	Etats-Unis	1926-2011		X	
Archibald Georges	Grande Bretagne	1926-1996		X	
Markowitz Harry	Etats-Unis	1927-		X	
Radner Roy	Etats-Unis	1927-		X	
Lipsey Richard	Canada	1928-		X	X
Meltzer Allen	Etats-Unis	1928-		X	
Drèze Jacques	Belgique	1929-			
Scarf Herbert	Etats-Unis	1930-2015		X	
Pasinetti Luigi	Italie	1930-	X		
Muth John	Etats-Unis	1930-2005		X	
Davidson Paul	Etats-Unis	1930-			X
Phelps Edmund	Etats-Unis	1933-		X	
Sen Amartya	Inde	1933-		X	
Leijonhufvud Axel	Suède	1933-		X	X
Wallace Neil	Etats-Unis	1938-		X	
Prescott Edward	Etats-Unis	1940-		X	
Akerlof George	Etats-Unis	1940-			X
Diamond Peter	Etats-Unis	1940-			
Sargent Thomas	Etats-Unis	1943-		X	
Kydland Finn	Norvège	1943-		X	
Barro Robert	Etats-Unis	1944-		X	
Blinder Alan	Etats-Unis	1945-			X
Starr Ross	Etats-Unis	1945-		X	
Benassy Jean-Pascal	France	1948-			X
Blanchard Olivier	France	1948-			X
Plosser Charles	Etats-Unis	1948-		X	
Hart Oliver	G-B, E-U	1948-			
Christiano Lawrence	Etats-Unis	1952-			
Romer Paul	Etats-Unis	1955		X	
Mankiw Gregory	Etats-Unis	1958-			X
Romer David	Etats-Unis	1958-			X
Gali Jordi	Espagne	1961-			X

BIBLIOGRAPHIE

[1] Aftalion Albert (1913) *Les crises périodiques de surproduction*, Marcel Rivière et cie, Paris

[2] Akerlof George A, Janet L. Yellen (1985) «A Near-Rational Model of the Business Cycle, With Wage and Price Inertia» in *The Quarterly Journal of Economics*, vol 100 suppl, pp. 823-838.

[3] Alchian Armen A. (1950) «Uncertainty, Evolution and Economic Theory» in *Journal of Political Economy*, vol 58 n°3, pp. 211-221.

[4] Alchian Armen A. (1955) «The Rate of interest, Fisher's Rate of Return over Cost and Keynes' Internal Rate of Return» in *American Economic Review*, vol 45 n°5, pp. 938-943.

[5] Alchian Armen A., Harold Desmetz (1972) «Production, Information Costs, and Economic Organization» in *American Economic Review*, vol 62 n°5, pp. 777-795.

[6] Alchian Armen A., Harold Desmetz (1973) «The property Rights Paradigm» in *Journal of Economic History*, vol 33 n°1, pp. 16-27.

[7] Allais Maurice (1953) «Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Américaine » in *Econometrica,* vol 21 n°4, pp. 503-546.

[8] Allais Maurice (1994) *Traité d'économie Pure*, Clément Juglar Editions juridiques et économiques, Paris

[9] Andrews P.W.S. (1959) *Manufacturing Business*, Macmillan & co ltd, London

[10] Archibald G.C., R.G. Lipsey (1958) «Monetary and Value Theory: a Critique of Lange and Patinkin» in *The Review of Economic Studies*, vol 26 n°1, pp. 1-22.

[11] Arrow Kenneth J. (1951) «Alternative Approaches to the Theory of Choice in Risk-taking situations» in *Econometrica,* vol 19 n°4, pp. 404-437.

[12] Arrow Kenneth J. (1951) *Social Choice and Individual Values*, John Wiley & sons, New York

[13] Arrow Kenneth J. (1953) «Le rôle des valeurs boursières pour la répartition la meilleure des risques», Reprinted from the International Colloquium on Econometrics 1952 by the *Centre National de la recherche Scientifique*, Paris, Cowles Foundation Paper n°77.

[14] Arrow Kenneth J. (1962) «The Economic Implications of Learning by Doing» in *The Review of Economic Studies*, vol 29 n°3, pp. 155-173.

[15] Arrow Kenneth J. (1974) «General Economic Equilibrium: Purpose, Analytic Technics, collective Choice» in *American Economic Review*, vol 64 n°3, pp. 253-272.

[16] Arrow Kenneth J., F.H. Hahn (1971) *General Competitive Analysis*, Holden Day Inc, San Francisco

[17] Arrow Kenneth J., Gérard Debreu (1954) «Existence of an Equilibrium for a Competitive Economy» in *Econometrica,* vol 22 n°3, pp. 265-290.

[18] Auerbach Paul, Dimitris P. Sotiropoulos (2012) «Revisiting the Socialist Calculation Debate: the Role of Markets and Finance in Hayek's response to Lange's Challenge», Kingston University London, Economics Discussion Paper 2012-6.

[19] Backhouse Roger, Mauro Boianovsky (2005) «Disequilibrium Macroeconomics: an episode in the Transformation of Modern Macroeconomics», ANPEC, Proceedings of the 33rd Brasilian Economic Meeting 012.

[20] Ball Laurence, Gregory N.Mankiw, David Romer (1988) «The New Keynesian Economics and the Output-Inflation Trade-off» in *Brookings Papers on Economic Activity*, n°16, pp. 1-82.

[21] Bardhan Pranab (1991) «Risk-taking, Capital Markets and Market Socialism», Berkeley Department of Economics. http://www.escholarship.org/uc/item/5vg1v6kd

[22] Bardhan Pranab, John E. Roemer (1992) «Market Socialism: A Case for Rejuvenation» in *Journal of Economic Perspectives*, vol 6 n°3, pp. 101-116.

[23] Bardhan Pranab, John E. Roemer (1994) «On the Workability of Market Socialism» in *Journal of Economic Perspectives*, vol 8 n°2, pp. 177-181.

[24] Barone E. (2009) «The Ministry of Production in the Collectivist State» in *Collectivist Economic Planning*, F.A. Hayek (éd), The Ludwig von Mises Institite, Auburn Alhabama, pp. 245-290.

[25] Barro Robert J. (1974) «Are Government Bonds net Wealth ?» in J*ournal of Political Economy*, vol 82 n°6, pp. 1095-1117.

[26] Baumol William J. (1952) «The Transaction Demand for Cash: an Inventory Theoretic Approach» in *The Quarterly Journal of Economics*, vol 65 n°4, pp. 545-556.

[27] Baumol William J. (1958) «On the Theory of Oligopoly» in *Economica,* vol 25 n°99, pp. 187-198.

[28] Baumol William J. (1982) «Contestable Markets: an Uprising in the Theory of Industry Structure» in *American Economic Review*, vol 72 n°1, pp. 1-15.

[29] Bauvert Joanna (2007) «La théorie pure de Léon Walras: économie réelle ou économie monétaire ? Discussion à propos d'une controverse fondamentale» in *Cahiers du Ceras*. Fifth Conference of the International Walras Association (Lausanne), 5e hors série, pp. 87-101.

[30] Benassy Jean-Pascal (1990) «Non Walrasian Equilibria, Money and Macroeconomics» in *Handebook of Monetary Economics*, B.M. Friedman et F.H. Hahn (éd), Elsevier Science Publishers, vol 1, pp. 103-169.

[31] Bergson Abram (1938) «A Reformulation of certain Aspects of Welfare Economics» in *The Quarterly Journal of Economics*, vol 52 n°2, pp. 310-334.

[32] Bertrand J. (1883) «Revue de Walras (théorie mathématique de la richesse Sociale) et revue de Cournot » in *Journal des savants*, septembre, pp. 499-508.

[33] Bhagwati Jagdish N. (1970) «Oligopoly Theory, Entry Prevention, and Growth» in *Oxford Economic Papers*, vol 22 n°3, pp. 297-310.

[34] Blanchard Olivier J., Lawrence H.Summers (1987) «Hysteresis in Unemployment» in *European Economic Review*, vol 31 n°3, pp. 288-295.

[35] Blaug Mark (1986) *La pensée économique. origine et développemen*t, Economica , Paris

[36] Blinder Alan S. (1994) «On Sticky Prices: Academic Theories Meet the Real World» in *Monetary Policy*, Gregory Mankiw (éd), The University of Chicago Press, pp. 117-154.

[37] Blinder Alan S., Robert M. Solow (1973) «Does Fiscal Policy Matter ?» in *Journal of Public Economics*, vol 2, pp. 319-337.

[38] Bliss C.J. (1975) *Capital Theory and the Distribution of Income,* North-Holland Publishing Company, Amsterdam Oxford

[39] Bradley Michael E., Manuela Mosta (2010) «Enrico Barone's Ministry of Production': Content and Context», Universita di Lecce, Department of Economics, Mathematics and Statistics, vol 144 n°51. https://papers.ssrn.com/sol13/papers.cfm?abstract_id=1734325

[40] Brainard William C., James Tobin (1968) «Pitfalls in Financial Models Building» in *American Economic Review*, vol 58 n°2, pp. 99-122.

[41] Briggs Vernon M. (1998) «Thomas Robert Malthus: The Economist» in *The Social Contract,* vol 8 n°3, pp. 206-216.

[42] Brunner Karl, Allan H. Meltzer (1971) «The Uses of Money: Money in the Theory of an Exchange Economy» in *American Economic Review*, vol 61 n°5, pp. 784-805.

[43] Buchanan James M (1976) «Barro and the Ricardian Equivalence Theorem» in *Journal of Political Economy*, vol 84 n°2, pp. 337-342.

[44] Buchanan James M (1999) *Public Principles of Public Debt: a Defense and Restatement,* Liberty Fund, Indianapolis. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Buchanan/buchCv2.html

[45] Cantillon Richard (1755) *Essai sur la nature du Commerce en général*, Chez Fletcher Gyles dans Holborn, Londres

[46] Cassel Gustave (1929) *Traité d'économie politique*, Marcel Giard, Paris

[47] Chamberlin Edward Hastings (1962) *Theory of Monopolistic Competition*, Harvard University Press, Cambridge Massachussets

[48] Choi Hak (2014) «A Detailed Rebuttal Report on Barro's Net Wealth Theory». https://papers.ssrn.com/sol13/papers.cfm?abstract_id=2490358

[49] Christiano Lawrence J., Martin Eichenbaum (1992) «Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations» in *American Economic Review,* vol 82 n°325, pp. 430-450.

[50] Christiano Lawrence J., Martin Eichenbaum, Robert Vigfusson (2003) «What Happens After a Technology Shock ?», *National Bureau of Economic research*, NBER Working Paper n° 9819.

[51] Clark John Bates (1891) «Distribution as Determined by a Law of Rent» in *The Quarterly Journal of Economics*, vol 5 n°3, pp. 289-318.

[52] Clark John Bates (1895) «Real Issues concerning Interest» in *The Quarterly Journal of Economics*, vol 10 n°1, pp. 98-102.

[53] Clark John Bates (1895) «The Origin of Interest» in *The Quarterly Journal of Economics*, vol 9 n°3, pp. 257-278.

[54] Clark John Bates (1908) *The Distribution of Wealth: a Theory of Wages, Interest and Profits*, Macmillan, New York. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Clark/clkDW.html

[55] Clark John Maurice (1917) «Business Acceleration and the Law of Demand: A Technical Factor in Economic Cycles» in *Journal of Political Economy*, vol 25 n°3, pp. 217-235.

[56] Clark John Maurice (1931) «Capital Production and Consumer-Taking: A Reply» in *Journal of Political Economy*, vol 39 n°6, pp. 814-816.

[57] Clower Robert W. (1967) «A Reconsideration of the Microfoundations of Monetary Theory» in *Western Economic Journal*, vol 6, pp. 1-9.

[58] Clower Robert W. (1984) «The Keynesian Counter-Revolution: a Theoretical Appraisal» in *Money and Markets: Essays by Robert Clower*, Donald A. Walker (éd), Cambridge University Press, Cambridge, pp. 34-58.

[59] Clower Robert, Axel Leijonhufvud (1975) «The Coordination of Economic Activities: A Keynesian Perspective» in *American Economic Review*, vol 65 n°2, pp. 182-188.

[60] Coase Ronald H. (1937) «The Nature of the Firm» in *Economica,* vol 4 n°16, pp. 386-405.

[61] Coase Ronald H. (1960) «The Problem of Social Cost» in *Journal of Law and Economics*, vol 3, pp. 1-44.

[62] Cobb Charles W., Paul H. Douglas (1928) «A Theory of Production» in *American Economic Review*, vol 18 n°1 (supplément), pp. 139-165.

[63] Cohen Avi J., G.C. Harcourt (2003) «Whatever Happened to the Cambridge Capital Theory Controversies» in *Journal of Economic Perspectives*, vol 17 n°1, pp. 199-201.

[64] Cournot A. A. (1980) *Recherches sur les principes mathématiques de la théorie des richesses*, Librairie philosophique J.Vrin, Paris

[65] Davidson Paul (1972) «Money and the Real World» in *The Economic Journal,* vol 82 n°325, pp. 101-115.

[66] Davidson Paul (1984) «Reviving Keynes's Revolution» in *Journal f Post Keynesian Economics*, vol 6 n°4, pp. 561-576.

[67] Davidson Paul (1988) «Endogenous Money, the Production Process and Inflation Analysis» in *Economie appliquée*, tome 41 n°1, pp. 151-169.

[68] Davidson Paul (1991) «Is Probability Theory Relevant for Uncertainty ? A Post Keynesian Perspective» in *Journal of Economic Perspectives*, vol 5 n°1, pp. 129-143.

[69] Davidson Paul, Sidney Weintraub (1977) «Money s Cause ad Effect» in *The Economic Journal,* vol 83 n°332, pp. 1117-1132.

[70] de Mourgues Michelle (1993) *La monnaie. Système Financier et théorie monétaire*, Economica , Paris

[71] De vroey Michel (2004) «Théorie du déséquilibre et chômage involontaire» in *Revue économique*, vol 55 n°4, pp. 647-668.

[72] Debreu Gérard (2001) *Théorie de la valeur*, Dunod, Paris

[73] Debreu Gérard, Herbert Scarf (1963) «A Limit Theorem on the Core of an Economy» in *International Economic Review*, vol 4 n°3, pp. 235-246.

[74] Debreu Gérard, Herbert Scarf (1972) «The Limit of the Core of an Economy» in *Decision and Organization*, C.B. Mc Guire et Roy Radner (éd), North-Holland, pp. 283-295.

[75] Diamond Peter A. (1967) «The Role of the Stock Market in a General Equilibrium Model with Technological Uncertainty» in *American Economic Review*, vol 57 n°4, pp. 759-776.

[76] Dickinson H.D. (1933) «Price Formation in a Socialist Economy» in *Economica,* vol 43 n°170, pp. 237-250.

[77] Domar Evsey D. (1946) «Capital Expansion, Rate of Growth, and Employment» in *Econometrica,* vol 14 n°2, pp. 137-147.

[78] Domar Evsey D. (1947) «Expansion and Employment» in *American Economic Review*, vol 37 n°1, pp. 34-55.

[79] Drèze Jacques H. (1985) «(Uncertainty and) The Firm in General Equilibrium Theory» in *The Economic Journal*, vol 95 suppl (conference papers), pp. 1-20.

[80] Duesenberry James S. (1949) *Income, Saving and the Theory of Consumer Behavior,* Harvard University Press, Cambridge Massachussets

[81] Edgeworth Francis Ysidro (1925) «On some Theories due to Professor Pigou» in *Papers relating to Political Economy*, Macmillan and co, London, vol 2, pp. 429-449.

[82] Edgeworth Francis Ysidro (1925) «The Law of Increasing and Diminishing Returns» in *Papers relating to Political Economy*, Macmillan and co, London, vol 1, pp. 61-99.

[83] Edgeworth Francis Ysidro (1925) «The Pure Theory of Monopoly» in *Papers relating to Political Economy*, Macmillan and co, London, vol 1, pp. 111-142.

[84] Edgeworth Francis Ysidro (1925) «The Theory of Distribution» in *Papers relating to Political Economy*, Macmillan and co, London, vol 1, pp. 13-60.

[85] Edgeworth Francis Ysidro (1967) *Mathematical Psychics*, Augustus M.Kelley , New York

[86] Edwards H.R. (1955) «Price Formation in Manufacturing Industry and Excess Capacity» in *Oxford Economic Papers*, vol 7 n°1, pp. 94-118.

[87] Eiteman Wilford J. (1947) «Factors Determining the Location of the least Cost point» in *American Economic Review*, vol 37 (dec), pp. 910-918.

[88] Eiteman Wilford J., Glenn E. Guthrie (1952) «The Shape of the Average Cost Curve» in *American Economic Review*, vol 42 (dec), pp. 832-838.

[89] Farm Ante (2014) «Competitive Price Leadership», Swedish Institute for Social Research, Stockholm University. www2.sofi.su.se/ -afa/docs/Competitive_Price_Leadership.pdf

[90] Fischer Stanley (1977) «Long Term Contracts, Rational Expectations, and the Optimal Money Supply Rule» in *Journal of Political Economy*, vol 85 n°1, pp. 191-205.

[91] Fisher Irving (1997) The Works of Irving Fisher volume 3, *The rate of Interest*, Pickering & Chatto, London

[92] Fisher Irving (1997) The Works of Irving Fisher volume 4, *The Purchasing Power of Money*, Pickering & Chatto, London

[93] Fisher Irving (1997) The Works of Irving Fisher volume 9, *The Theory of Interes*t, Pickering & Chatto, London

[94] Flateau Paul (2001) «Some Reflection about the Pigou-Robinson Theory of Exploitation» in *History of Economics Review,* vol 33 n°1, pp. 1-16.

[95] Frederick Shane (1999) «Discounting, Time Preference, and Identity», Carnegie Mellon University, Department of Social and Decision Sciences. https://researchgate.net/publication/318850872_Time_preference_and_personal_identity

[96] Frederick Shane, George Loewenstein, Ted O'Donoghue (2002) «Time Discounting and Time Preference: A Critical Review» in *Journal of Economic Literature*, vol 40 (june), pp. 351-401.

[97] Friedman Milton (1951) «A Monetary and Fiscal Framework for Economic Stability» in *Readings in Monetary Economics*, American Economic Association (éd), George Allen and Unwin Ltd, London, pp. 369-393.

[98] Friedman Milton (1953) *Essays in Positive Economics*, University of Chicago Press, Chicago

[99] Friedman Milton (1956) «The Quantity Theory of Money- a Restatement» in *Studies in the Quantity Theory of Money*, Milton Friedman (éd), The University of Chicago Press, Chicago, pp. 3-21.

[100] Friedman Milton (1957) *A Theory of the Consumption Function*, Princeton University Press, Princeton

[101] Friedman Milton (1968) «The Role of Monetary Policy» in *American Economic Review*, vol 58 n°1, pp. 1-17.

[102] Friedman Milton (1969) *The Optimum Quantity of Money and Other Essays*, Macmillan, New York

[103] Friedman Milton (1970) «A Theoretical Framework for Monetary Analysis» in J*ournal of Political Economy*, vol 78 n°2, pp. 193-238.

[104] Friedman Milton (1971) «A Monetary Theory of Nominal Income» in *Journal of Political Economy*, vol 79 n°2, pp. 323-337.

[105] Friedman Milton (1974) «Using Escalators to Help Fight Inflation» in *Fortune,* jul, pp. 94-97, 174-176.

[106] Friedman Milton (1977) «Inflation and Unemployment» in *Journal of Political Economy*, vol 85 n°3, pp. 451-472.

[107] Friedman Milton (1984) «Currency Competition: A Skeptical View». https://miltonfriedman.hoover.org/friedman_images/Collections/2016c21/Hague_1984.pdf

[108] Friedman Milton, Anna Jacobson Schwartz (1963) *A Monetary History of the US (1867-1960)*, Princeton University Press, Princeton

[109] Friedman Milton, Anna Jacobson Schwartz (1987) «Has Government any Role in Money» in *Money in Historical Perspective*, Anna J. Shwartz (éd), University of Chicago Press, Chicago, pp. 289-314.

[110] Friedman Milton, L.J. Savage (1948) «The Utility Analysis of Choices Involving Risk» in *Journal of Political Economy*, vol 56 n°4, pp. 279-304.

[111] Friedman Milton, L.J. Savage (1952) «The Expected Utility Hypothesis and the Measurability of Utility» in *Journal of Political Economy*, vol 60 n°6, pp. 463-474.

[112] Frisch Ragnar (1931) «The Interrelation Between Capital Production and Consumer-Taking» in *Journal of Political Economy,* vol 39 n°5, pp. 646-654.

[113] Frisch Ragnar (1933) «Propagation Problems and Impulse Problems in Dynamic Economics» in *Economic Essays in Honour of Gustav Cassel*, George Allen and Unwin Ltd, London, pp. 171-205.

[114] Gali Jordi (1999) «Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations ?» in *American Economic Review*, vol 89 n°1, pp. 249-271.

[115] Gali Jordi, Pau Rabanal (2004) «Technology Shocks and Aggregate Fluactuations: How Well Does the RBC Model Fit Postwar US Data ?», *National Bureau of Economic research,* Working Paper n°10636. http://www.nber.org/papers/w10636

[116] Ganong Peter, Pascal Noel (2015) «How Does Unemployment Affect Consumer Spending ?». https://scholar.harvard.edu/files/ganong/files/ganong_jmp_unemployment_spending.pdf

[117] Garcia-Molina Mario (2005) «Capital Theory and the Origin of the Elasticity of Substitution» in *Cambridge Journal of Economics*, vol 29 n°3, pp. 423-437.

[118] Gillies Donald (2003) «Probability and Uncertainty in Keynes's General Theory» in *The Philosophy of Keynes's Economics: Probability, Uncertainty and Convention*, Sohei Mizuhara, Jochen Runde (éd), Routledge, London and New York, pp. 111-129.

[119] Girardi Daniele (2017) «Old and New Formulations of the Neoclassical Theory of Aggregate Investment: A Critical Review», *University of Massachusetts Amherst, Economic Department Working Papers* Series 2019. https://scholarworks.umass.edu/econ_workingpaper/219

[120] Godley W.A.H., W.D. Nordhaus (1972) «Pricing in the Trade Cycke» in *The Economic Journa*l, vol 82 (sept), pp. 853-882.

[121] Gould J.P. (1968) «Adjustment Costs in the Theory of Investment of the Firm» in *The Review of Economic Studies*, vol 35 n°1, pp. 47-56.

[122] Greenwald Bruce C., Joseph E. Stiglitz (1989) «Towards a Theory of Rigidities» in *American Economic Review,* vol 79 n°2, pp. 364-369.

[123] Greenwald Bruce C., Joseph E. Stiglitz (1990) «Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior» in *American Economic Review*, vol 80 n°2, pp. 160-165.

[124] Gurley John G., Edward S. Shaw (1973) *La monnaie dans une théorie des actifs financiers*, Editions Cujas, Paris

[125] Haavelmo Trygve (1960) *A Study int the Theory of Investment*, The University of Chicago Press, Chicago

[126] Hahn F.H., R.C.O. Matthews (1964) «The Theory of Economic Growth: A Survey» in *The Economic Journal*, vol 74 n°296, pp. 779-902.

[127] Hahn Frank (1984) *Equilibrium and Macroeconomics*, Basil Blackwell, Oxford

[128] Hall R.L., C.J. Hitch (1939) «Price Theory and Business Behaviour» in *Oxford Economic Papers*, vol 2, pp. 12-45.

[129] Hall Robert E., Olivier jean Blanchard, R. Glenn Hubbard (1986) «Market Structure and Macroeconomic Fluctuations» in *Brookings Papers on Economic Activity*, n°2, pp. 285-338.

[130] Hansen Gary D. (1985) «Indivisible Labor and the Business Cycle» in *Journal of Monetary Economics*, n°16, pp. 309-327.

[131] Harrod R.F. (1936) *The Trade Cycle: An Essay*, Oxford at the Clarendon Press, Oxford

[132] Harrod R.F. (1939) «an Essay in Dynamic Theory» in *The Economic Journal*, vol 49 n°193, pp. 14-33.

[133] Harrod R.F. (1948) *Towards a Dynamic Economics*, Macmillan & co ltd, London

[134] Harrod R.F. (1952) *Economic Essays*, Harcourt, Brace and Company, New York

[135] Hart Olivier D. (1975) «On the Optimality of equilibrium when the Market Structure is Incomplete» in *Journal of Economic Theory*, vol 11 n°3, pp. 418-443.

[136] Hausman Daniel M. (1981) *Capital, Profit and Prices*, Columbia University Press, New York

[137] Hayashi Fumio (1982) «Tobin's Marginal q and Average q : A Neoclassical Interpretation» in *Econometrica,* vol 50 n°1, pp. 213-224.

[138] Hayek (von) Friedrich A. (1936) «The Mythology of Capital» in *The Quarterly Journal of Economics*, vol 50 n°2, pp. 199-228.

[139] Hayek (von) Friedrich A. (1948) «The Competitive Solution» in *Individualism & Economic Order*, The University of Chicago Press, Chicago, pp. 181-208.

[140] Hayek (von) Friedrich A. (1948) «The Use of Knowledge in Society» in *Individualism & Economic Order*, The University of Chicago Press, Chicago, pp. 77-91.

[141] Hayek (von) Friedrich A. (1975) «Profits, Interest and Investment» in *Profits, Interest and Investment and Other Essays on the Theory of Industrial Fluctuations*, Augustus Kelley , Clifton New Jersey, pp. 3-71.

[142] Hayek (von) Friedrich A. (1990) *Denationalisation of Money- The Argument Refined*, The Institute of Economic Affairs, London

[143] Hayek (von) Friedrich A. (2008) «Monetary Theory and the Trade Cycle» in *Prices and Production and Other Works: F.A. Hayek on Money, the Business Cycle, and the Gold Standard.*, The Ludwig von Mises Institut, Auburn Alhabama, pp. 1-130.

[144] Hayek (von) Friedrich A. (2008) «Prices and Production» in *Prices and Production and Other Works: F.A. Hayek on Money, the Business Cycle, and the Gold Standard.*, The Ludwig von Mises Institut, Auburn Alhabama, pp. 189-329.

[145] Hayek (von) Friedrich A. (2009) «The Nature and History of the Problem» in *Collectivist Economic Planning*, F.A. Hayek (éd), The Ludwig von Mises Institite, Auburn Alhabama, pp. 1-40.

[146] Hayek (von) Friedrich A. (2009) «The Present State of the Debate» in *Collectivist Economic Planning*, F.A. Hayek (éd), The Ludwig von Mises Institite, Auburn Alhabama, pp. 201-243.

[147] Hayek (von) Friedrich A. (2009) *The Pure Theory of Capital*, The Ludwig von Mises Institute, Auburn Alabama

[148] Heflebower Richard B. (1955) «Full Cost, Cost Changes and Prices» in *Business Concentration and Price Policy*, National Bureau of Economic Research (éd), Princeton University Press, pp. 359-394. http://www.nber.org/books/univ55-1

[149] Hicks J.R. (1934) «A Reconsideration of the Theory of Value. Part I» in *Economica,* vol 1 n°1, pp. 52-76.

[150] Hicks J.R. (1935) «A Suggestion for Simplifying the Theory of Money» in *Economica,* vol 2 n°5, pp. 1-19.

[151] Hicks J.R. (1935) *Theory of Wages*, Macmillan & co ltd, London

[152] Hicks J.R. (1937) «Mr. Keynes and the Classics: a Suggested Interpretation» in *Econometrica,* vol 5 n°2, pp. 147-159.

[153] Hicks J.R. (1951) *A Contribution to the Theory of the Trade Cycle*, Oxford at the Clarendon Press, Oxford

[154] Hicks J.R. (1956) *Valeur et capital*, Dunod, Paris

[155] Hicks J.R. (1963) *Theory of Wages*, Macmillan & co ltd, London

[156] Hirschleifer J. (1958) «On the Theory of Optimal Investment Decision» in *Journal of Political Economy*, vol 66 n°4, pp. 329-352.

[157] Hobson J.A. (1910) *The Industrial System*, Charles Scribner's sons, New York

[158] Holland Steven A. (1985) «Rational Expectations and the Effects of Monetary Policy: a Guide for the Uninitiated» in *Ferderal Bank of Saint Louis*, may, pp. 5-11.

[159] Hoover Kevin D. (2008) «Was Harrod Right ?», CNRS- GREDEG, GREDEG WP n°2013-02. http://www.gredeg.cnrs.fr/working-papers/GREDEG-WP-2013-02.pdf

[160] Hotelling Harold (1932) «Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions» in *Journal of Political Economy*, vol 40 n°5, pp. 577-616.

[161] Hotelling Harold (1938) «The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates» in *Econometrica,* vol 6 n°3, pp. 242-269.

[162] Hume David (1987) *Essays Moral, Political, and Literary*, Liberty Fund Inc, Indianapolis. Site Internet Library of Economics and Liberty http://www.econlib.org/library/LFBooks/Hume/hmMPL.html

[163] Humphrey Thomas M. (1991) «Noneutrality of Money in Classical Monetary Thought» in *Federal Reserve Bank of Richmond Economic Review*, mar/apr, pp. 3-15.

[164] Humphrey Thomas M. (1997) «Fisher and Wicksell on the Quantity Theory» in *Federal Reserve Bank of Richmond* Quaterly, vol 83 n°4, pp. 71-90.

[165] Jevons W. Stanley (1888) *Theory of Political Economy*, Macmillan, London. Site Internet Library of Economics and Liberty http://www.econlib.org/library/YPDBooks/Jevons/jvnPE.html

[166] Jorgenson Dale W. (1963) «Capital Theory and Investment Behavior» in *American Economic Review*, vol 53 n°2, pp. 247-259.

[167] Jorgenson Dale W. (1967) «The Theory of Investment Behavior» in *Determinants of Investment Behavior*, Robert Ferber (éd), National Bureau of Economic Research, pp. 129-175. http://www.nber.org/books/ferb67-1

[168] Kahn R.F. (1931) «The Relation of Home Investment to Unemployment» in *The Economic Journal*, vol 41 n°162, pp. 173-198.

[169] Kahn R.F. (1952) «Oxford Studies in the Price Mechanism» in *The Economic Journa*l, vol 62 (mar), pp. 119-130.

[170] Kaldor Nicholas (1937) «Annual Survey of Economic Theory: The Recent Controversy on the Theory of Capital» in *Econometrica,* vol 5 n°3, pp. 201-233.

[171] Kaldor Nicholas (1939) «Capital Intensity and the Trade Cycle» in *Economica,* vol 6 n°21, pp. 40-66.

[172] Kaldor Nicholas (1942) «Professor Hayek and the Concertina-Effect» in *Economica,* vol 9 n°36, pp. 359-382.

[173] Kaldor Nicholas (1960) «A Model of the Trade Cycle» in *Essays on Economic Stability and Growth*, Gerald Duckworth, London, pp. 177-192.

[174] Kaldor Nicholas (1961) «Capital Accumulation and Economic Growth» in *The Theory of Capital*, F.A. Lutz, Hague Douglas (éd), Macmillan & co, pp. 177-222.

[175] Kaldor Nicholas (1964) «Economic Growth and the Problem of Inflation» in *Essays on Economic Policy*, Gerald Duckworth, London, vol 1, pp. 166-199.

[176] Kaldor Nicholas, James Trevithick (1981) «A Keynesian Perspective on Money» in *Lloyds Bank Review*, 139 (jan), pp. 1-19.

[177] Kalecki Michal (1937) «A Theory of the Business Cycle» in *The Review of Economic Studies*, vol 4 n°2, pp. 77-97.

[178] Kalecki Michal (1938) «The Determinants of Distribution of the National Income» in *Econometrica,* vol 6 n°21, pp. 97-112.

[179] Kalecki Michal (1944) «Professor Pigou on "the Classical Stationary State" A Comment» in *The Economic Journal*, vol 54 (apr), pp. 131-132.

[180] Kalecki Michal (1971) «Costs and Prices» in *Selected Essays on the Dynamics of the Capitalist Economy*, Cambridge University Press, Cambridge, pp. 43-61.

[181] Keynes J.M. (1937) «Alternative Theories of the Rate of Interest» in *The Economic Journal*, vol 47, pp. 241-252.

[182] Keynes J.M. (1937) «The Ex-ante Theory of the Rate of Interest» in *The Economic Journal*, vol 47 (dec), pp. 663-669.

[183] Keynes J.M. (1937) «The General Theory of Employment» in *The Quarterly Journal of Economics*, vol 51 n°2, pp. 209-223.

[184] Keynes J.M. (1939) «Relative Movements of Real Wage and Output» in *The Economic Journal*, vol 49 n°193, pp. 34-51.

[185] Keynes J.M. (1940) *How to Pay for the War*, Macmillan & co ltd, London

[186] Keynes J.M. (1971) *Théorie générale de l'emploi*, de l'intérêt , de la monnaie, Petite bibliothèque Payot, Paris

[187] Keynes J.M. (1973) *A Treatise on Probability*, Macmillan for the Royal Economic Society, London

[188] King Robert G., Charles I. Plosser (1984) «Money,Credit, and Prices in a Real Business Cycle» in *American Economic Review*, vol 74 n°3, pp. 363-380.

[189] Knight Frank H. (1921) *Risk, Uncertainty and Profits*, Hart, Schaffner & Marx, Boston. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Knight/knRUP.html

[190] Knight Frank H. (1923) «The Ethics of Competition» in *The Quarterly Journal of Economics*, vol 38 n°3, pp. 579-624.

[191] Knight Frank H. (1934) «Capital; Time and the Interest Rate» in *Economica,* vol 1 n°3, pp. 257-286.

[192] Koopmans Tjalling C. (1951) «Analysis of Production as an Efficient combination of activities» in *Activity Analysis of Production and Allocation*, Tjalling Koopmans (éd), Cowles Commission for Research in Economics. John Wiley & Sons, New York, pp. 33-97.

[193] Koopmans Tjalling C. (1951) «Efficient Allocation of Resources» in *Econometrica,* vol 19 n°4, pp. 455-465.

[194] Koopmans Tjalling C. (1960) «Stationary Ordinal Utility and Impatience» in *Econometrica,* vol 28 n°2, pp. 287-309.

[195] Koopmans Tjalling C. (1965) «On the Concept of Optimal Economic Growth» in *Academia Scientiarum Scripta Varia*, vol 28 n°1, pp. 225-300.

[196] Koopmans Tjalling C. (1967) «Objectives, Constraints and Outcomes in Optimal Growth Models» in *Econometrica,* vol 35 n°1, pp. 1-15.

[197] Koopmans Tjalling C., Diamond Peter A., Richard E. Williamson (1964) «Stationary Utility and Time Perspective» in *Econometrica,* vol 32 n°1-2, pp. 82-100.

[198] Kornai Janos (1986) «The Soft Budget Constraint» in *Kyklos,* Fasc 1, pp. 3-30.

[199] Kurz Heinz D. (2000) «Wicksell and the Problem of the "Missing Equation"» in *History of Political Economy*, vol 32 n°4, pp. 765-788.

[200] Kurz Heinz D., Neri Salvadori (1993) «von Neumann's Growth Model and the Classical System» in *European Journal of History of Economic Thought*, vol 1 n°1, pp. 129-160.

[201] Kurz Heinz D., Neri Salvadori (1995) *Theory of Production, A Long-Period Analysis*, Cambrige University Press, Cambrige

[202] Kurz Heinz D., Neri Salvadori (2000) «Piero Sraffa's Contributions to Economics: a Brief Survey» in *Critical Essays on Piero Sraffa's Legacy in Economics*, Heinz D. Kurz (éd), Cambridge University Press, Cambridge, pp. 3-24.

[203] Kurz Heinz D., Neri Salvadori (2002) «A Single Theory of Two ? Walras's Critique of Ricardo» in *History of Political Economy*, vol 34 n°4, pp. 365-398.

[204] Kydland Finn E., Edward C. Prescott (1977) «Rules rather than Discretion: the Inconsistency of Optimal Plans» in *Journal of Political Economy*, vol 85 n°3, pp. 473-492.

[205] Kydland Finn E., Edward C. Prescott (1982) «Time to Build and Aggregate Fluctuations» in *Econometrica,* vol 50 n°6, pp. 1345-1370.

[206] Lange Oskar (1936) «On the Economic Theory of Socialism: part one» in *The Review of Economic Studies*, vol 4 n°1, pp. 53-71.

[207] Lange Oskar (1937) «On the Economic Theory of Socialism: part Two» in *The Review of Economic Studies*, vol 4 n°2, pp. 123-142.

[208] Lange Oskar (1942) «Say's Law: a Restatement and Criticism» in *Studies in Mathematical Economics and Econometrics,* O. Lange, F. Mcintyre, T.O. Yntema (éd), University of Chicago Press, Chicago, pp. 49-68.

[209] Lange Oskar (1967) «The Computer and the Market» in *Socialism, Capitalism and Economic Growrh: Essays Presented to Maurice Dobb*, C. Feinstein (éd), Cambridge University Press, Cambridge, pp. 158-161.

[210] Lange Oskar (1973) «Mr. Lerner's Note on Socialist Economics» in *Readings in Welfare Economics*, M.J. Farrell (éd), Macmillan, pp. 39-43.

[211] Lanzillotti Robert F. (1957) «Competitive Price Leadership- A Critique of Price Leadership Models» in *The Review of Economics and Statistics*, vol 39, pp. 55-64.

[212] Lanzillotti Robert F. (1958) «Pricing Objectives in Large Companies» in *American Economic Review*, vol 48 n°5, pp. 921-940.

[213] Lee F.S., J.Irving Lessmann (1992) «The Fate of an Errant Hypothesis: the Doctrines of Normal Cost-Price» in *History of Political Economy*, vol 24 n°2, pp. 273-309.

[214] Lee Frederic S. (1999) *Post Keynesian Price Theory*, Cambrige University Press, Cambridge

[215] Leijonhufvud Axel (1968) *On Keynesian Economics and the Economics of Keynes*, Oxford University Press, New York, London, Toronto

[216] Leijonhufvud Axel (1973) «Effective Demand Failures» in *The Swedsh Journal of Economics,* vol 75 n°1, pp. 27-48.

[217] Leijonhufvud Axel (1974) «The Varieties of Price Theory: What Microfoundations for Macro Theory», UCLA, Discussion Paper n°44. http://www.econ.ucla.edu/workingpapers/wp044.pdf

[218] Leland Hayne E. (1972) «Theory of the Firm Facing Uncertain Demand» in *American Economic Review*, vol 62 n°3, pp. 278-291.

[219] Leontief Wassily W. (1937) «The Fundamental Assumptions of Mister Keynes' Theory of Unemployment» in *The Quarterly Journal of Economics*, vol 51 n°1, pp. 192-197.

[220] Leontief Wassily W. (1958) *La Structure de l'économie américaine 1919-1939*, Editions Génin, Paris

[221] Lerner Abba P. (1937) «Statics and Dynamics in Socialist Economics» in *The Economic Journal,* vol 47 n°186, pp. 253-270.

[222] Lerner Abba P. (1943) «Functional Finance and the Federal Debt» in *Social Research*, vol 10 n°1, pp. 38-51.

[223] Lerner Abba P. (1946) *The Economics of Control,* The Macmillan Company, New York

[224] Lerner Abba P. (1949) «The Inflation Process: some Theoretical Aspects» in *The Review of Economics and Statistics,* vol 31 n°3, pp. 193-200.

[225] Lerner Abba P. (1951) «Fighting Inflation» in *The Review of Economics and Statistics*, vol 33 n°3, pp. 194-196.

[226] Lerner Abba P. (1953) «On the Marginal Product of Capital and the Marginal Efficiency of Investment» in *Journal of Political Economy*, vol 61 n°1, pp. 1-14.

[227] Lerner Abba P. (1973) «A Note on Socialist Economics» in *Readings in Welfare Economics*, M.J. Farrell (éd), Macmillan, pp. 44-45.

[228] Lerner Abba P. (1973) «The Concept of Monopoly and the Measurement of Monopoly Power» in *Readings in Welfare Economics*, M.J. Farrell (éd), Macmillan, pp. 1-19.

[229] Lerner Abba P. (2002) «The Burden of the National Debt» in *The Economics of Budget Deficits*, Charles Rowley, William Shughart, Robert Tollison (éd), Edward Elgar Publishing, Cheltenham UK, vol 1, pp. 255-275.

[230] Lester Richard A. (1946) «Shortcomings of Marginal Analysis for Wage-Employment Problems» in *American Economic Review*, vol 36 n°1, pp. 63-82.

[231] Lester Richard A. (1947) «Marginalism, Minimum Wages and Labor Markets» in *American Economic Review*, vol 37 n°1, pp. 135-148.

[232] Lipsey Richard G. (1960) «The Relation between Unemployment and the Rate of Change of Money Wages in the United kingdom, 1861-1957: a further Analysis» in *Economica,* vol 27 n°105, pp. 1-31.

[233] Long John B., Charles I. Plosser (1983) «Real Business Cycles» in *Journal of Political Economy*, vol 91 n°1, pp. 39-69.

[234] Longuet Stéphane (2006) «Hayek et Knight (1933-1936), La controverse sur le capital» in *Cahiers d'économie politique*, fasc 2 n°51, pp. 177-195.

[235] Lucas M.R. (2003) «Pricing decision and the neoclassical theory of the firm» in *Management Accounting Research*, n°14, pp. 201-217.

[236] Lucas Robert E. Jr (1967) «Adjustment Cost and the Theory of Supply» in *Journal of Political Economy*, vol 75 n°4, pp. 321-334.

[237] Lucas Robert E. Jr (1972) «Expectations and the Neutrality of Money» in *Journal of Economic Theory*, vol 4 n°2, pp. 103-124.

[238] Lucas Robert E. Jr (1976) «Econometric Policy Evaluation : a Critique» in *The Phillips Curve and Labor Markets*, Karl Brunner, Allan H. Meltzer (éd), Carnegie Rochester. North Holland Publishing Company, Amsterdam- New York- Oxford, pp. 19-46.

[239] Lucas Robert E. Jr (1977) «Understanding Business Cycles» in *Carnegie-Rochester Conference Series on Public Policy*, vol 5, pp. 7-29.

[240] Lucas Robert E. Jr, Leonard A. Rapping (1969) «Real Wages, Employment and Inflation» in *Journal of Political Economy*, vol 77 n°5, pp. 721-752.

[241] Machlup Fritz (1935) «Professor Knight and the Period of Production» in *Journal of Political Economy*, vol 43 n°5, pp. 577-624.

[242] Machlup Fritz (1946) «Marginal Analysis and Empirical research» in *American Economic Review*, vol 36 n°4, pp. 519-554.

[243] Machlup Fritz (1947) «Rejoinder to a Anti-Marginalist» in *American Economic Review*, vol 37 n°1, pp. 148-154.

[244] Malinvaud Edmond (1953) «Capital Accumulation and Efficient Allocation of Resources» in *Econometrica,* vol 21 n°2, pp. 233-266.

[245] Malinvaud Edmond (1980) *Réexamen de la théorie du chômage*, Calmann-Lévy, France

[246] Malinvaud Edmond (1999) *Leçons de Microéconomie*, Dunod, Paris

[247] Malthus Thomas Robert (1810) «Depreciation of Paper Currency» in *The Edimburgh Review or Critical Journal*, vol 17 n°1, pp. 340-372.

[248] Malthus Thomas Robert (1922) *Principles of Political Economy*, Augustus M.Kelley , New York

[249] Mandel Ernest (1962) *Traité d'économie marxiste*, René Julliard, Paris

[250] Mankiw N. Gregory (1985) «Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly» in *The Quarterly Journal of Economics*, vol 100 n°2, pp. 529-538.

[251] Mankiw N. Gregory (1989) «Real Business Cycles: a New Keynesian Perspective» in *Journal of Economic Perspectives*, vol 3 n°3, pp. 79-90.

[252] Markowitz Harry (1952) «The Utility of Wealth» in *Journal of Political Economy*, vol 60 n°2, pp. 151-158.

[253] Marshall Alfred (1920) Principles of Economics, Macmillan, London. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Marshall/marP.html

[254] Marx Karl (1975) *Le Capital (livres I, II et III)*, Editions sociales, Paris

[255] Menger Carl (1892) «On the Origins of Money» in *The Economic Journal*, vol 2 n°6, pp. 239-255.

[256] Menger Carl (1981) *Principles of Economics*, New York University Press, New York and London

[257] Metzler Lloyd A. (1951) «Wealth, Saving and the Rate of Interest» in *Journal of Political Economy*, vol 59 n°2, pp. 93-116.

[258] Mill John Stuart (1909) *Principles of Political Economy with some of their Applications to Social Philosophy* (2 volumes), Longmans, Green and Co, London. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Mill/mlP.html

[259] Modigliani Franco (1944) «Liquidity Preference and the Theory of Interest and Money» in *Econometrica,* vol 12 n°1, pp. 45-88.

[260] Modigliani Franco (1958) «New Developments on the Oligopoly Front» in *Journal of Political Economy*, vol 66 n°3, pp. 2015-232.

[261] Modigliani Franco (1986) «Life Cycle, Individual Thrift and the Wealth of Nations» in *American Economic Review*, vol 76 n°3, pp. 297-313.

[262] Modigliani Franco, Merton H. Miller (1958) «The Cost of Capital, Corporation Finance and the Theory of Investment» in *American Economic Review*, vol 48 n°3, pp. 261-297.

[263] Modigliani Franco, Richard Brumberg (2005) «Utility Analysis and the Consumption Function : Interpretation of Cross Data» in *The Collected Papers of Franco Modigliani*, The MIT Press, Cambridge Massachusetts, vol 6, pp. 3-46.

[264] Mongin Philippe (1992) «The Full Cost Controversy of the 1940s and the 1950s: A Methodological Assessment» in *History of Political Economy*, vol 24 n°2, pp. 311-356.

[265] Muth John F. (1961) «Rational Expectations and the Theory of Price Movements» in *Econometrica,* vol 29, pp. 315-335.

[266] Nelson Charles R, Charles I. Plosser (1982) «Trends and Random Walks in Macroeconomic Times Series» in *Journal of Monetary Economics*, vol 10 n°2, pp. 139-168.

[267] Niehans Jürg (1971) «Money and Barter in General Equilibrium with Transaction Costs» in *American Economic Review*, vol 56 n°5, pp. 773-783.

[268] Nitzan Jonathan (1990) «Price Behaviour and Business Behaviour», Mc Gill University Montreal, Department of Economics Discussion Papers. http://bnarchives.yoku.ca/160

[269] Nubbemeyer Elmar (2010) *A Reconsideration of Full-Cost Pricing*, Ludwig-Maximilians-Universität München, Inaugural-Dissertation zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.). www.edoc.ub.uni-muenchen.de/12313/2/nubbemeyer_Elmar.pdf.

[270] Ohlin Bertil (1937) «Alternative Theories of the Rate of Interest: Rejoinder» in *The Economic Journa*l, vol 47 (sept), pp. 423-427.

[271] Ohlin Bertil (1937) «Some Notes on the Stockholm Theory of Savings and Investment (I+II)» in *The Economic Journal*, vol 47 (mars et juin), pp. 53-69 et 221-240.

[272] Oliver Henry M. Jr (1947) «Marginal Theory and Business Behaviour» in *American Economic Review*, vol 37 n°3, pp. 375-383.

[273] Pareto Vilfredo (1927) *Manuel d'économie Politique*, Marcel Giard, Paris

[274] Pasinetti Luigi L. (1962) «Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth» in *The Review of Economic Studies*, vol 29 n°4, pp. 267-279.

[275] Pasinetti Luigi L. (1978) «Wicksell Effects and Reswitching of Techniques in Capital Theory» in *Scandinavian Journal of Economics*, vol 80 n°2, pp. 181-189.

[276] Patinkin Don (1989) *Money, Interest and Price*, MIT Press, Cambridge Massachussets

[277] Patinkin Don (2002) «Price Flexibility and Full Employment» in *The Origins of Macroeconomics*, Robert W. Dimand (éd), Routledge, London and New York, vol 3, pp. 173-204.

[278] Petri Fabio (2000) «On the Likelihood and Relevance of Reverse Capital Deepening», University of Sienna, Department of Economics, Working Paper n°279. https://dx.doi.org/10.2139/ssrn.234880

[279] Phelps Edmund (1961) «The Golden Rule of Accumulation: A Fable for Growthmen» in *American Economic Review*, vol 51 n°4, pp. 638-643.

[280] Phelps Edmund (1965) «Second Essay on the Golden Rule of Accumulation» in *American Economic Review*, vol 55 n°4, pp. 793-814.

[281] Phelps Edmund (1967) «Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time» in *Economica,* vol 34 n°135, pp. 254-281.

[282] Phillips A. W. (1958) «The Relation between Unemployment and the Rate of Change of Money Wages in the United kingdom, 1861-1958» in *Economica,* vol 25 n°100, pp. 283-299.

[283] Pigou A.C. (1932) *The Economics of Welfare*, Macmillan and Co Ltd, London.

[284] Pigou A.C. (1943) «The Classical Stationary State» in *The Economic Journal,* vol 53 (dec), pp. 343-353.

[285] Pigou A.C. (1947) «Economic Progress in a Stable Environment» in *Economica,* vol 14 (aug), pp. 180-188.

[286] Pigou A.C. (1951) «The Value of Money» in *Readings in Monetary Economics*, American Economic Association (éd), George Allen and Unwin Ltd, London, pp. 162-183.

[287] Plosser Charles I. (1989) «Understanding Real Business Cycles» in *Journal of Economic Perspectives*, vol 3 n°3, pp. 51-77.

[288] Prescott Edwarsd C. (1986) «Theory Ahead of Business Cycle Measurement» in *Federal Reserve Bank of Minneapolis Quarterly Review*, vol 10 n°4, pp. 9-22.

[289] Pressman Steven (1999) *Fifty Major Economists*, Routledge, London

[290] Quesnay François (1766) «Analyse de la formule arithmétique du tableau économique de la distribution des dépenses annuelles d'une Nation agricole» in *Journal de l'agriculture, du commerce & des finances*, tome 2, 3e partie, pp. 11-41.

[291] Radner Roy (1972) «Existence of Equilibrium of Plans, Prices and Prices Expectations in a Sequence of Markets» in *Econometrica,* vol 40 n°2, pp. 289-303.

[292] Radner Roy (1979) «Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices» in *Econometrica,* vol 47 n°, pp. 655-678.

[293] Ramsey F.P. (1928) «A Mathematical Theory of Savings» in *The Economic Journal*, vol38 n°152, pp. 543-559.

[294] Ramsey F.P. (1931) «Truth and Probability» in *The Foundations of Mathematics and other Logical Essays*, R.B. Braitwaite (éd), Kegan, Paul Trench, Trubner & Co, London, pp. 156-198.

[295] Rancan Antonella (2015) «The Origin of the Sylos Postulate: Modigliani's and Sylos Labini's Contributions to Oligopoly Theory » in *Journal of History of Economic Thought*, vol 37 n°3, pp. 431-448.

[296] Rebelo Sergio (2005) «Real Business Cycle Models: Past, Present and Future» in *Scandinavian Journal of Economics*, vol 107 n°2, pp. 217-238.

[297] Ricardo David (1888) *The Works of David Ricardo*, John Murray, London

[298] Robertson D.H. (1940) *Essays in Monetary Theory*, P,S, King & Son Ldt, London

[299] Robinson Austin (1950) «The Pricing of Economic Products» in *The Economic Journal*, vol 60 (dec), pp. 771-780.

[300] Robinson Joan (1936) «Review of "The Trade Cycle" by R.F.Harrod» in *The Economic Journa*l, vol 46 n°184, pp. 691-693.

[301] Robinson Joan (1946) *The Economics of Imperfect Competition*, Macmillan and co ldt, London

[302] Robinson Joan (1953) «Imperfect Competition Revisited» in *The Economic Journal*, vol 63 n°251, pp. 581-593.

[303] Robinson Joan (1954) «The Production Function and the Theory of Capital » in *The Review of Economic Studies*, vol 21 n°2, pp. 81-106.

[304] Robinson Joan (1962) *Essays on the Theory of Economic Growth*, Macmillan St Martin's Press, London

[305] Robinson Joan (1972) *L'accumulation du capital*, Dunod, Paris

[306] Romer David (1993) «The New Keynesian Synthesis» in *Journal of Economic Perspectives*, vol 7 n°1, pp. 5-22.

[307] Romer David (1997) *Macroéconomie approfondie*, Ediscience international, Paris

[308] Romer Paul M. (1990) «Endogenous Technical Change» in *Journal of Economic Perspective*s, vol 98 n°5, pp. S71-S102.

[309] Samuelson Paul A. (1938) «A Note on the Pure Theory of Consumer's Behavior» in *Economica,* vol 5 n°17, pp. 61-71.

[310] Samuelson Paul A. (1939) «A synthesis of the Principle of Acceleration and the Multiplier» in *Journal of Political Economy*, vol 47 n°6, pp. 786-797.

[311] Samuelson Paul A. (1939) «Interactions Between the Multiplier Analysis and the Principle of Acceleration» in *The Review of Economics and Statistics*, vol 4, pp. 75-78.

[312] Samuelson Paul A. (1943) «Full Employment After the War» in *Postwar Economic Problems*, Seymour E. Harris (éd), Mac Graw Hill Book Company, New York, pp. 27-54.

[313] Samuelson Paul A. (1948) «Consumption Theory in Terms of Revealed Preference» in *Economica,* vil 15 n°60, pp. 243-253.

[314] Samuelson Paul A. (1952) «Probability, Utility and the Independence Axiom» in *Econometrica,* vol 20 n°4, pp. 670-678.

[315] Samuelson Paul A. (1954) «The Pure Theory of Public Expenditure» in The *Review of Economics and Statistics*, vol 36 n°4, pp. 387-389.

[316] Samuelson Paul A. (1955) «Diagrammatic Exposition of a Theory of Public Expenditure» in *The Review of Economics and Statistics*, vol 37 n°4, pp. 350-356.

[317] Samuelson Paul A. (1956) «Social Indifference Curves» in *The Quarterly Journal of Economics*, vol 70 n°1, pp. 1-22.

[318] Samuelson Paul A. (1958) «Aspects of Public Expenditure Theories» in *The Review of Economics and Statistics*, vol 40 n°4, pp. 332-338.

[319] Samuelson Paul A. (1962) «Parable and Realism in Capital Theory: the Surrogate Production Function» in *The Review of Economic Studies*, vol 29 n°80, pp. 193-259.

[320] Samuelson Paul A. (1966) «A summing up» in *The Quarterly Journal of Economics*, vol 80 n°4, pp. 568-583.

[321] Samuelson Paul A. (1966) *Foundations of Economic Analysis*, Harvard University Press, Cambridge

[322] Samuelson Paul A. (1998) «Abstract of a Theorem Concerning Substituability in Open Leontief Systems» in *Input-Output Analysis*, Heinz Kurz, Erik Dietzenbacher, Christian Lager (éd), Edward Elgar Publishing, Cheltenham UK, vol 3, pp. 183-187.

[323] Samuelson Paul A. (2002) «The Simple Mathematics of Income Determination» in *The Origins of Macroeconomics*, Robert W. Dimand (éd), Routledge, London and New York, vol 4, pp. 17-39.

[324] Samuelson Paul A., Robert M. Solow (1960) «Problem of Achieving and Maintaining a Stable Price Level, Analytical Aspects of Anti-Inflation Policy» in *American Economic Review*, vol 50 n°2, pp. 177-194.

[325] Sanches Daniel (2012) «The Optimum Quantity of Money» in *Federal Reserve Bank of Philadelphia Quarterly Review*, Q4, pp. 8-15.

[326] Sargent Thomas J (1973) «Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment» in *Brookings Papers on Economic Activity*, n°2, pp. 429- 480.

[327] Sargent Thomas J., Neil Wallace (1976) «Rational Expectations and the Theory of Economic Policy» in *Journal of Monetary Economics*, vol 2 n°2, pp. 169-183.

[328] Say Jean-Baptiste (1966 (éd, 1841)) *Traité d'économie politique ou simple exposé de la manière dont se forment, se distribuent et se consomment les richesses*, Otto Zeller, Osnabrück

[329] Schumpeter Joseph A. (1976) *Capitalism, Socialism and Democracy*, Routledge, London

[330] Scitovsky (de) T. (1973) «Reconsideration of the Theory of Tariffs» in *Readings in Welfare Economics*, M.J. Farrell (éd), Macmillan, pp. 58-79.

[331] Shapiro Carl, Joseph E. Stiglitz (1984) «Unemployment Equilibrium as a Worker discipline device» in *American Economic Review*, vol 74 n°3, pp. 433-444.

[332] Sharpe Steve A., Gustavo A. Suarez (2015) «Why isn't Investment More Sensitive to Interest Rates: Evidence from Surveys», Federal Reserve Board, Washington DC, Finance and Economics Discussion Series n°2014-002. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2667352

[333] Shleifer Andrei, Robert W. Vishny (1994) «The Politics of Market Socialism» in *Journal of Economic Perspectives*, vol 8 n°2, pp. 165-176.

[334] Simon Herbert A. (1978) «Rational Decision-making in Business Organizations» in *Nobel Prize in Economics. Documents*, 1978-1, pp. 343-371.

[335] Slutsky Eugen E. (1998) «On the Theory of the Budget of the Consumer» in *Consumer Theory*, Kelvin J. Lancaster (éd), Edward Elgar Publishing, Cheltenham UK, pp. 3-32.

[336] Smith Adam (1904) *An Inquiry into the Nature and Causes of the Wealth of Nations* (2 volumes), Edwin Cannan. Methuen & Co, London. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Smith/smWN.html

[337] Solow Robert M (1955) «The Production Function and the Theory of Capital » in *The Review of Economic Studies*, vol 23 n°2, pp. 101-108.

[338] Solow Robert M (1956) «A Contribution to the Theory of Economic Growth» in *The Quarterly Journal of Economics*, vol 70 n°1, pp. 65-94.

[339] Solow Robert M (1957) «Technical Change and the Aggregate Production Function» in *The Review of Economics and Statistics*, vol 39 n°3, pp. 312-320.

[340] Solow Robert M (1959) «Investment and Technical Progress» in *Mathematical Methods in the Social Sciences*, Kenneth Arrow, Samuel Karlin, Patrick Suppes (éd), Stanford University Press, Stanford, pp. 89-104.

[341] Solow Robert M (1962) «Substitution and Fixed Proportions in the Theory of Capital» in *The Review of Economic Studies*, vol 29 n°3, pp. 207-218.

[342] Solow Robert M (1962) «Technical Progress, Capital Formation, and Economic Growth» in *American Economic Review*, vol 52 n°2, pp. 76-86.

[343] Solow Robert M., Joseph E. Stiglitz (1968) «Output, Employment and Wages in the Short Run» in *The Quarterly Journal of Economics*, vol 82 n°4, pp. 537-560.

[344] Sraffa Piero (1925) «On the Relations between Cost and Quantity Produced», G. Langer et F. Lee, Chicago, Translation of "Sulle relazioni fra costo e quantita prodotta" by A. Roncaglia and J. Eatwell.. http://cas2.umkc.edu/ECON/economics/faculty/lee/courses/502/reading/yt1.pdf

[345] Sraffa Piero (1926) «The Laws of Return under Competitive Conditions» in *The Economic Journal,* vol 36 n°144, pp. 535-550.

[346] Sraffa Piero (1977) *La production des marchandises par les marchandises*, Bordas, Paris

[347] Starr Ross M. (1972) «The Structure of Exchange in Barter and Monetary Economies» in *The Quarterly Journal of Economics*, vol 86 n°2, pp. 290-302.

[348] Stigler George J. (1968) *The Organization of Industry*, Richard Irwin inc, Homewood Illinois

[349] Stiglitz Joseph E. (1999) «Towards a General Theory of Wage & Price Rigidities and Economic Fluctuations» in *American Economic Review*, vol 89 n°2, pp. 75-80.

[350] Stiglitz Joseph E. (2001) «Information and the Change in the Paradigm in Economics» in *Nobel Prize in Economics documents*, Nobel Prize Comittee, 2001-8, pp. 472-540. https://nobelprize.org/nobelprizes/economic-sciences/laureates/2001/stiglitz-lecture.pdf

[351] Suzumura Kotaro (2004) «An Interview with Paul Samuelson: Welfare Economics, 'Old' and 'New', and Social Choice Theory», COE/RES, Discussion Paper n°37. http://www1.econ.ut-u.ac.jp/coe-res/dp_doc/no_37_dp.pdf

[352] Swan T.W. (1956) «Economic Growth and Capital Accumulation» in *The Economic Record*, vol 32 n°2, pp. 334-361.

[353] Sylos-Labini Paolo (1962) *Oligopoly and Technical Progress,* Harvard University Press, Cambridge Massachussets

[354] Sylos-Labini Paolo (1971) «La théorie des prix en régime d'Oligopole et la théorie du développement» in *Revue d'économie politique*, vol 81 n°2, pp. 244-272.

[355] Sylos-Labini Paolo (2007) «Oligopoly: Static and Dynamic Analysis» in *Revue d'économie industrielle* (en ligne:http://rei.revues.org/1793), 118 (2e trim) , pp. 91-107.

[356] Taylor Fred M. (1929) «The Guidance of Production in a Socialist State» in *American Economic Review*, vol 19 n°1, pp. 1-8.

[357] The Royal Swedish Academy of Sciences (2004) «Finn Kydland and Edward Prescott's Contribution to Dynamic Macroeconomics: The Time Consistency of Economic Policy and the Driving Forces Behind Business Cycles», The Royal Swedish Academy of Sciences, Advanced information on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel. https://www.nobelprize.org/nobel_prizes/Economique/laureates/2004/advanced-economicsciences2004.pdf

[358] Thornton Henry (1965) *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*, Augustus M.Kelley , New York

[359] Tinbergen J. (1938) «Statistical Evidence on the Acceleration Principle» in *Economica,* vol 5 n°18, pp. 164-176.

[360] Tobin James (1956) «The Interest Elasticity of Transactions Demand for Cash» in *The Review of Economics and Statistics*, vol 38 n°3, pp. 241-24.

[361] Tobin James (1958) «Liquidity Preference as Behavior towards Risk» in *The Review of Economic Studies*, vol 25 n°2, pp. 65-86.

[362] Tobin James (1969) «A General Equilibrium Approach to Monetary Theory» in *Journal of Money, Credit and Banking*, vol 1 n°1, pp. 15-29.

[363] Tobin James (1972) «Friedman's Theoretical Framework» in *Journal of Political Economy*, vol 80 n°5, pp. 852-863.

[364] Tobin James (1972) «Inflation and Unemployment» in *American Economic Review*, vol 62 n°1, pp. 1-18.

[365] Tobin James, Willem Buiter (1976) «Long Run Effects of Fiscal and Monetary Policy on Aggregate Demand» in *Monetarism: Studies in Monetary Economics*, North-Holland, pp. 273-309.

[366] Tooke Thomas (1844) *An Inquiry into the Currency Principle*, Longman, Brown, Green and Longmans, London

[367] Trostel Philip A., Grant A. Taylor (2001) «A Theory of Time Preference» in *Econmic Inquiry*, vol 39 n°3, pp. 379-395.

[368] Turgot Jacques (1966) «Observations sur le Mémoire de M. de Saint-Peravy» in *Œuvres de Turgot*, Osnabrück Otto Zeller, vol 1, pp. 418-433.

[369] Turgot Jacques (1966) «Réflexions sur la Formation et la Distribution des richesses» in *Œuvres de Turgot*, Osnabrück Otto Zeller, vol 1, pp. 1-67.

[370] Turgot Jacques (1966) «Valeurs et monnaies» in *Œuvres de Turgot,* Osnabrück Otto Zeller, vol 1, pp. 75-93.

[371] Vercelli Alessandro, Serena Sordi (2009) «Genesis and Evolution of the Multiplier-Accelerator Model in the Years of High Theory», Universita di Sienna, Workshop "Alternative Approaches to the Builiding of Economic Dynamics in the Years of High Theory". https://storep.org/workshpdynamics/papers.htm

[372] Viner Jacob (1932) «Cost Curves and Supply Curves» in *Zeitschrift für Nationalökonomie*, Bd 3 hft 1, pp. 23-46.

[373] von Böhm Bawerk Eugen (1895) «The Origin of Interest » in *The Quarterly Journal of Economics*, vol 9 n°4, pp. 380-387.

[374] von Böhm Bawerk Eugen (1959) *Positive Theory of Capital*, Libertarian Press, South Holland Illinois

[375] von Mises Ludwig (2009) «The Problem of Value in the Socialist Society» in *Collectivist Economic Planning*, F.A. Hayek (éd), The Ludwig von Mises Institite, Auburn Alhabama, pp. 87-130.

[376] von Neumann J. (1998) «A Model of General Economic Equilibrium» in *Input-Output Analysis*, Heinz Kurz, Erik Dietzenbacher, Christian Lager (éd), Edward Elgar Publishing, Cheltenham UK, vol 1, pp. 133-141.

[377] von Neumann John, Oskar Morgenstern (1964) *Theory of Games and Economic Behavior*, John Wiley & sons, New York

[378] Wald Abraham (1951) «On some Systems of Equations of Mathematical Economics» in *Econometrica,* vol 19 n°4, pp. 368-403.

[379] Walras Léon (1898) *Etudes d'économie appliquée*, F. Rouge, Lausanne

[380] Walras Léon (1988) *Œuvres économiques complètes d'Auguste et Léon Walras*: volume VIII: Eléments d'économie Politique Pure, Economica , Paris

[381] Warming Jens (1932) «International Difficulties Arising out of the Financing of Public Works during Depression» in *The Economic Journal*, vol 42 n°166, pp. 211-224.

[382] Weintraub E. Roy (1985) *General Equilibrium Analysis*, Cambrige University Press, Cambridge

[383] White William H (1956) «Interest Inelasticity of Investment Demand - The Case from Business Attitude Surveys Re-examined.» in *American Economic Review*, vol 46 n°4, pp. 565-587.

[384] Wicksell Knut (1907) The Influence of the Rate of Interest on Prices, *The Economic Journal*, vol 17 pp 213-220. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Essays/wcksInt1.html

[385] Wicksell Knut (1934 (I) et 1935 (II)) *Lectures on Political Economy* volumes I + II, Routledge & Kegan Paul Ltd, London

[386] Wicksell Knut (1954) *Value, Capital and Rent*, George Allen & Unwin ltd, London

[387] Wicksteed Philip H. (1910) *The Common sense of Political Economy*, Macmillan and Co Ltd, London. Site Internet Library of Economics and Liberty http://www.econlib.org/library/Wicksteed/wkCS.html

[388] Wicksteed Philip H. (1914) «The Scope and Method of Political Economy» in *The Economic Journal*, vol 24 n°93, pp. 1-23.

[389] Wicksteed Philip H. (1932) *An Essay on the Co-ordination of Laws of Distribution*, London School of Economics, Scarce Tracts of Economic and Political Science: Reprint n°12.

[390] Yunker James A. (1995) «Post-Lange Market Socialism: an Evaluation of Profit-Oriented Proposals» in *Journal of Economic Issues*, vol 29 n°3, pp. 683-717.

Je ne voudrais pas clore cette bibliographie sans mentionner l'excellent site internet « History of Economic thought », qui est un outil de vulgarisation haut de gamme. L'URL est : http://hetwebsite.net/het/home.htm

LISTE DES GRAPHIQUES

Graphique 2.1 : le Tableau Economique de Quesnay 11
Graphique 2.2 : la loi de la demande selon Cournot 47
Graphique 2.3 : l'équilibre du duopole 48
Graphique 3.1 : l'utilité marginale et l'utilité totale 60
Graphique 3.2 : l'offre de travail selon Jevons 62
Graphique 3.3 : le montant de l'investissement de capital selon Jevons 63
Graphique 3.4 : l'indétermination des contrats 79
Graphique 3.5 : le rétrécissement de la courbe de contrat à quatre intervenants 80
Graphique 3.6 : la productivité marginale 84
Graphique 3.7 : la rémunération des facteurs suivant la productivité marginale 86
Graphique 3.8 : les courbes d'offre et de demande 94
Graphique 3.9 : le surplus du consommateur 96
Graphique 3.10 : l'élasticité de la demande 97
Graphique 3.11 : effets d'une hausse de l'offre ou la demande en longue période 99
Graphique 3.12 : l'offre et la demande suivant Wicksteed 104
Graphique 3.13 : la structure du capital selon Wicksell 110
Graphique 3.14 : le modèle du vin 111
Graphique 3.15 : les courbes d'indifférences de Pareto 116
Graphique 3.16 : l'optimum du consommateur : deux biens 117
Graphique 3.17 : l'équilibre de l'échange 118
Graphique 3.18 : L'optimum du consommateur : de nombreux biens 119
Graphique 3.19 : la forme du revenu 121
Graphique 3.20 : les opportunités d'investissement et le RROC 123
Graphique 3.21 : la productivité marginale du capital et le taux d'intérêt 128
Graphique 4.1: les courbes de productivité et de coût 139
Graphique 4.2: le coût moyen et ses composants 140
Graphique 4.3 : le coût moyen à long terme et les rendements d'échelle 140
Graphique 4.4 : la courbe de coût moyen à long terme en tant que courbe enveloppe 141
Graphique 4.5: la maximisation du profit à court terme 143
Graphique 4.6 : l'équilibre du monopoleur 147
Graphique 4.7 : l'équilibre à long terme 151
Graphique 4.8 : le taux marginal de substitution 159
Graphique 4.9 : la courbe de dépense 160
Graphique 4.10 : la courbe de demande de Hicks 161
Graphique 4.11 : les isoquantes 163
Graphique 4.12 : l'élasticité de substitution 166
Graphique 4.13 : L'équilibre du producteur et le chemin d'expansion 168
Graphique 4.14 : Coupe verticale de la colline de production 168
Graphique 4.15 : la boite d'Edgeworth-Bowley 171
Graphique 4.16 : La boite d'Edgeworth à la production 172
Graphique 4.17 : L'adéquation de la production à la consommation 173
Graphique 4.18: rendement social d'un état économique 175
Graphique 4.19 : prix s'écartant du coût marginal 176
Graphique 4.20 : maximisation de la fonction de bien-être social 179
Graphique 4.21: la valeur de la monnaie en fonction de sa quantité 188
Graphique 4.22 : l'offre et la demande de monnaie chez Pigou 191
Graphique 4.23 : la théorie des fonds prêtables 196
Graphique 5.1 : la détermination de l'emploi par la demande effective 203
Graphique 5.2 : l'équilibre monétaire et le taux d'intérêt selon Keynes 210
Graphique 5.3 : le prix selon que le coût soit croissant ou constant 220
Graphique 5.4 : la relation linéaire entre le salaire et le taux de profit 225
Graphique 5.5: la vitesse de dépréciation du capital fixe 226
Graphique 5.6: comparaison de deux méthodes de production d'un bien 227
Graphique 5.7: valeur des termes d'âge différent suivant le taux d'intérêt 228
Graphique 5.8: effet des variations du taux d'intérêt sur les prix relatifs 229
Graphique 5.9 : la fonction de production intensive dans sa version néo-classique 231
Graphique 5.10 : la fonction de production alternative de Robinson 233
Graphique 5.11 : la frontière du prix des facteurs 236
Graphique 5.12: la frontière du prix des facteurs subrogée 237

Graphique 5.13 : le double aiguillage et la marche arrière du capital 238
Graphique 5.14: la demande de capital et le double aiguillage 240
Graphique 5.15 : la courbe de demande brisée 247
Graphique 5.16 : les éléments du coût moyen à long terme selon Andrews 253
Graphique 6.1 : la stabilité de l'équilibre sur un marché 279
Graphique 6.2 : activités dans l'espace vectoriel 287
Graphique 6.3 : la frontière d'efficience 288
Graphique 6.4 : la convexité des ensembles de production faisable 289
Graphique 6.5 : la maximisation du profit 290
Graphique 6.6 : l'optimum du consommateur 292
Graphique 6.7 : non-convexités et équilibre 296
Graphique 6.8 : l'arbre des événements 301
Graphique 6.9: la production dans le temps selon Malinvaud 305
Graphique 6.10: demande de monnaie et équilibre monétaire 318
Graphique 6.11: l'échange avec coûts de transaction 328
Graphique 7.1 : le modèle IS-LM 336
Graphique 7.2 : l'influence du revenu sur l'épargne et l'investissement. 337
Graphique 7.3 : le diagramme en croix des keynésiens 341
Graphique 7.4 : la préférence pour la liquidité comme attitude face au risque. 345
Graphique 7.5 : l'équilibre IS-LM chez les monétaristes et les keynésiens 357
Graphique 7.6 : l'utilité des encaisses et le taux d'intérêt 360
Graphique 7.7 : l'épargne, l'investissement et l'effet de richesse 361
Graphique 7.8 : Blinder, Solow et l'effet d'éviction 368
Graphique 7.9 : la courbe de Phillips 372
Graphique 7.10 : les courbes d'ajustement du salaire à la demande excédentaire 373
Graphique 7.11 : la courbe de Phillips à court et à long terme 376
Graphique 7.12 : distribution probabiliste du niveau général des prix et des prix relatifs 387
Graphique 7.13 : arbitrage optimal entre l'inflation et le chômage 392
Graphique 7.14: chômage involontaire et déséquilibre suivant Patinkin 394
Graphique 7.15 : offre excédentaire et demande excédentaire chez Solow et Stiglitz 398
Graphique 7.16 : l'équilibre de l'emploi et du salaire réel 400
Graphique 7.17 : les domaines des trois types d'équilibre non walrassien 404
Graphique 7.18 : l'élasticité de l'offre de monnaie 409
Graphique 7.19 : l'effet de la politique monétaire dans le cas horizontaliste 410
Graphique 7.20 : le marché du travail avec salaire d'efficience 414
Graphique 7.21 : l'effet des coûts de menu sur le bien-être 416
Graphique 7.22 : l'incitation à adapter les prix selon Romer 417
Graphique 8.1 : l'épargne optimale selon Ramsey 426
Graphique 8.2 : la croissance équilibrée selon Solow 434
Graphique 8.3 : croissance équilibrée : rôle de la rémunération des facteurs 435
Graphique 8.4: la croissance équilibrée selon Swan 437
Graphique 8.5: le taux d'accumulation désiré 444
Graphique 8.6 : Kaldor et le progrès technique 449
Graphique 8.7 : la règle d'or de l'accumulation 451
Graphique 8.8 : La demande dérivée : influence des deux composantes 461
Graphique 8.9 : la demande dérivée : aspect temporel 462
Graphique 8.10; évolution de revenu national en fonction de s et v 465
Graphique 8.11 : le cycle conjoncturel selon Hicks 467
Graphique 8.12 : la décision d'investir selon Kalecki 470
Graphique 8.13 : le cycle selon Kalecki 471
Graphique 8.14 : fonctions d'épargne et d'investissement linéaires 472
Graphique 8.15 : fonctions d'épargne et d'investissement non linéaires 472
Graphique 8.16 : le cycle conjoncturel selon Kaldor 473
Graphique 8.17 : le triangle hayekien : production plus détournée 475
Graphique 9.1: la fonction d'input 513
Graphique 9.2 : la valeur du capital 514
Graphique 9.3 : l'épargne, l'investissement et le taux d'intérêt chez Hayek 515
Graphique 9.4 : l'utilité d'une perspective aléatoire 547
Graphique 9.5 : la fonction d'utilité stochastique chez Friedman-Savage et Markowitz 548
Graphique 9.6 : la fonction de consommation à court terme et à long terme 554
Graphique 9.7 : répartition optimale de la consommation dans le temps 557
Graphique 9.8 : cycle de vie : consommation, épargne et patrimoine 562
Graphique 9.9 : MEI et accumulation du capital 567

Graphique 9.10 : taux interne de rentabilité et valeur actuelle nette _____ *568*
Graphique 9.11 : ajustement de l'offre à la croissance de la demande _____ *575*
Graphique 9.12 : la résistance à la baisse du taux d'intérêt _____ *579*

TABLE DES MATIERES

Avertissement concernant la notation 1

TITRE I- Introduction 2

TITRE II- L'économie politique pré-moderne 6

2.1- les précurseurs 7
2.1.1. Les mercantilistes 7
Cantillon et la circulation monétaire 8
2.1.2. Les physiocrates 10
Le tableau économique de François Quesnay 11

2.2- Libéralisme et pensée économique 14
2.2.1. L'économie politique à la fin du siècle des lumières 14
2.2.2. Smith et la « main invisible » du marché 17

2.3- Le classicisme ricardien et son temps 22
2.3.1. Ricardo : le coût de production détermine la valeur 23
2.3.2. Mill, la synthèse parfaite de l'économie classique 29
2.3.3. Say et la loi des débouchés 33
2.3.4. Malthus et la controverse sur la surproduction générale 36
2.3.5. Les débats monétaires à l'époque classique 40

2.4- Deux directions opposées de l'économie politique 46
2.4.1. Cournot : l'introduction du calcul différentiel en économie 46
2.4.2. Marx : de la plus-value au profit 50

TITRE III- L'école néoclassique 57

3.1- Naissance d'une nouvelle école 58
3.1.1. Jevons : de l'utilitarisme à l'utilité marginale 58
3.1.2. Menger et la satisfaction des besoins 64
3.1.3. Walras : l'équilibre général 68
L'échange pur 70
L'échange et la production 72
Intégration de l'épargne et des biens capitaux. 75
Dynamique de l'équilibre : le tâtonnement 76
Implications sociales et politiques 76
3.1.4. Edgeworth et l'indétermination de l'échange 77

3.2- La consolidation 82
3.2.1. Clark et la productivité marginale 83
3.2.2. Böhm Bawerk et la théorie autrichienne du capital 87
3.2.3. Marshall : l'offre et la demande 94
Les courbes de l'offre et de la demande 94
L'utilité et la demande 95
La production et l'équilibre du marché 97
Les facteurs de production 99
3.2.4. Wicksteed et la fonction de production 100
La fonction de production 101
L'offre, la demande et le coût d'opportunité 103
3.2.5. Wicksell : le capital en long et en large 105
Théorie de la valeur et gain de l'échange 105
La production sans le capital 106
La production avec le capital 107
Raffinements 112
1- Application des facteurs primaires en continu 112
2- Analyse de longue période 113
3- La prise en compte du capital fixe 113
3.2.6. Pareto : l'équilibre et l'optimum 114
Les goûts : la consommation 115
L'équilibre général 118

3.2.7. Fisher, l'épargne et l'investissement ... 120
La préférence pour le présent ... 120
L'opportunité d'investir ... 122
L'équilibre individuel ... 123
L'équilibre du marché ... 123
Taux d'intérêt monétaire et taux réel ... 125
3.2.8. Différentes conceptions du capital ... 125
Capital physique ou capital valeur ? ... 125
Capital temps ou capital stock ? ... 126
Capital ou investissement ? ... 128

3.3- Les néoclassiques face aux classiques ... 130

3.4- Une voix discordante ... 133
Hobson contre la théorie marginaliste de la rémunération des facteurs ... 133

TITRE IV- Développement de l'économie néoclassique ... 136

4.1- La firme et le marché : coûts, prix et profit ... 137
4.1.1. Productivité et coût ... 137
Analyse à court terme ... 137
Analyse à long terme ... 140
4.1.2. Equilibre du producteur et du marché en concurrence parfaite ... 142
Analyse à court terme ... 142
Analyse à long terme ... 144
4.1.3. Le producteur et le marché hors concurrence parfaite ... 146
Le monopole ... 146
Robinson et la concurrence imparfaite ... 149
Chamberlin et l'oligopole ... 153
La concurrence monopolistique ... 154
La différenciation du produit ... 154
Le coût des ventes ... 155
L'équilibre de la firme ... 155
La concurrence monopolistique et la théorie de la rétribution des facteurs ... 156

4.2- Le renouveau parétien de l'entre-deux guerres ... 158
4.2.1. Le comportement du consommateur individuel ... 158
Le taux marginal de substitution ... 158
La courbe de demande ... 160
4.2.2. Le comportement du producteur ... 162
La substitution entre les facteurs et la substitution entre les produits ... 162
L'élasticité de substitution ... 165
L'échelle de la production ... 167
Les indivisibilités ... 169
4.2.3. L'optimum ... 170
L'optimum de la consommation ... 170
L'optimum de la production ... 172
L'adéquation de la production à la consommation ... 173
Equilibre concurrentiel et optimum ... 175
Optimum et répartition du bien-être ... 177
Les biens publics ... 181
4.2.4. Nouvelles mathématiques de l'optimisation ... 182

4.3- La macroéconomie avant Keynes ... 185
4.3.1. Irving Fisher et l' « équation de l'échange » ... 185
4.3.2. Wicksell et le taux d'intérêt naturel ... 187
4.3.3. Demande d'encaisse et théorie quantitative ... 190
Le coefficient *k* des Cambridgiens ... 190
L'encaisse désirée suivant Walras ... 192
4.3.4. Caractéristiques essentielles de la macroéconomie pré-keynésienne ... 194
La dichotomie et la neutralité de la monnaie ... 194
L'impossibilité du chômage involontaire ... 195
La théorie des fonds prêtables ... 195

4.4- Les fonctions de production agrégées ... 197

TITRE V- Les critiques contre la théorie néoclassique 199

5.1- La révolution keynésienne 200
Keynes et la *Théorie Générale* 200
5.1.1. Les mécanismes keynésiens 200
1- Le marché de l'emploi et la rigidité des salaires nominaux 200
2- La demande effective et la fonction d'emploi 202
3- La consommation et l'épargne 203
4- L'incitation à investir 204
5- L'ajustement par le revenu et le multiplicateur 205
6- L'intérêt et la préférence pour la liquidité 209
7- Retour sur la question du salaire nominal 211
8- L'équation quantitative 212
5.1.2. Le statut de la monnaie 213
5.1.3. Est-il possible de combattre le chômage et comment ? 214
5.1.4. La controverse Ohlin - Keynes : *ex-ante* ou *ex-post* 216
5.1.5. Le salaire et la marge face au cycle conjoncturel 218

5.2- Sraffa remet Ricardo à l'honneur 220
5.2.1. Le parcours de Sraffa 220
5.2.2. La production de marchandises par les marchandises 221
L'équilibre général ricardien 221
Profit, salaire et étalon invariant 223
Cas particuliers 225
Le changement des méthodes de production 226
La réduction à des quantités de travail de périodes différentes 227

5.3- La controverse du capital 230
5.3.1. La critique des fonctions de production agrégées 230
Modèle simplifié avec le capital et le travail 231
La réplique de Solow 234
5.3.2. La *frontière des prix des facteurs de production* 235
5.3.3. Le double aiguillage et la marche arrière du capital 237

5.4- Néo-ricardiens et néoclassiques 241
5.4.1. La réponse néoclassique 241
5.4.2. L'économie néo-ricardienne 242
5.4.3. Les théorèmes de non-substitution 243

5.5- Les débats sur le marginalisme 246
5.5.1. Le « Full cost pricing » 246
5.5.2. La productivité marginale, l'emploi et le salaire 248
5.5.3. La contre-attaque marginaliste 250
5.5.4. La polémique reprend de plus belle … et puis s'éteint 252
Andrews et le *normal cost* 252
Harrod et la prévention de l'entrée 254
Heflebower et la fin de la controverse 256
5.5.5. Approfondissement du *full cost pricing* 257
La prévention de l'entrée 257
Le taux de profit cible 259
Le prix du marché 260
5.5.6. Théorie évolutionniste de la firme 261
5.5.7. Le *mark-up* de Kalecki 263

TITRE VI- Les néo-walrassiens 266

6.1- Walras : un nouveau départ 268
6.1.1. Le modèle de walras-cassel 268
Cassel, contre les théories de la valeur 268
Le « colloque viennois » et le modèle de Wald 271
Le modèle de Wald de 1936 272
6.1.2. von Neumann : le modèle de production linéaire 274

6.2- Hicks : stabilité et équilibre temporaire 279
La stabilité de l'équilibre général 279
Equilibre général et temporalité 281

6.3- Emergence de la théorie actuelle de l'équilibre général 286
6.3.1. Progrès mathématiques et la « commission Cowles » 286
Convexité et théorie du producteur 286
Convexité et optimum du consommateur 290
6.3.2. L'équilibre général selon Arrow et Debreu 292
La preuve de l'existence par Arrow et Debreu 292
Equilibre d'Arrow-Debreu : caractéristiques et cas particuliers 294
Les non-convexités 295
Unicité de l'équilibre 296
Stabilité (dynamique) de l'équilibre général 297
La conjecture d'Edgeworth remise à l'honneur 299
L'équilibre intertemporel et l'incertitude 299

6.4- Deux dimensions supplémentaires : le temps et l'aléa 303
6.4.1. Malinvaud, le capital et l'intérêt 303
Le taux d'intérêt néo-walrassien 303
La théorie du capital et de l'intérêt de Malinvaud 304
6.4.2. L'incertitude et les anticipations 308
Arrow et les valeurs mobilières 308
Radner et le modèle d'Arrow-Debreu en économie séquentielle 308
Equilibre de Radner et anticipations rationnelles 311
L'économie des marchés incomplets 312

6.5- Une place pour la monnaie dans l'équilibre général ? 314
6.5.1. La filière Walras – Hicks – Lange – Patinkin 314
Lange : la loi de Walras et la loi de Say 314
Patinkin et l'effet d'encaisse réelle 316
Les « Patinkin controversies » 321
6.5.2. Nouvelles ouvertures 322
Hahn contre la monnaie « inessentielle » 322
Clower et la contrainte financière 323
Starr et la coïncidence des besoins 325
Niehans et les coûts de transaction 327
Brunner et Meltzer et le coût d'acquisition de l'information 329
Conclusion 331

TITRE VII- La macroéconomie après la théorie générale 332

7.1- la synthèse néoclassique-keynésienne 334
Introduction 334
7.1.1. Hicks et le modèle IS-LM 334
Le modèle IS-LM 334
La stabilité de l'équilibre temporaire 337
7.1.2. Modigliani, la rigidité des salaires et la préférence pour la liquidité 338
7.1.3. Samuelson : variables autonomes et variables induites 340
7.1.4. Tobin et la théorie du portefeuille 343
La préférence pour la liquidité 343
Le coefficient « q » 346

7.2- Le monétarisme 348
7.2.1. La théorie quantitative remise à l'honneur 348
La théorie quantitative 348
L'équation friedmanienne de la demande de monnaie 348
La monnaie, le revenu nominal, le revenu réel et les prix 351
7.2.2. La politique monétaire et budgétaire 352
7.2.3. « The Opimal Quantity of Money » 353
7.2.4. Comment Friedman perçoit sa controverse avec les keynésiens 355
7.2.5. La contre-révolution monétariste : acte deux 357

7.3- Quelques grands débats 359
7.3.1. Baisse des prix et effet de richesse 359
Pigou introduit l'effet de richesse 359
Critique de l'effet Pigou selon Patinkin 361
Monnaie interne et monnaie externe 362
7.3.2. Les déficits publics 364

Le fardeau de la dette ... 364
L'effet d'éviction ... 367
Barro et l'équivalence ricardienne ... 369
7.3.3. La courbe de Phillips et le taux de chômage naturel ... 371
Origine et développement de la courbe de Phillips ... 371
La critique de Phelps et Friedman ... 374
La défense de Tobin ... 377
L'inflation vue par les keynésiens ... 377

7.4- Les nouveaux classiques ... 382
7.4.1. Les anticipations rationnelles ... 382
7.4.2. Le modèle d'information imparfaite de Lucas ... 384
7.4.3. Vision de la politique économique ... 387
Taux d'intérêt réel et politique monétaire ... 387
Sargent et Wallace : la politique discrétionnaire impuissante ... 389
Kydland & Prescott : politique optimale ou cohérente ... 390

7.5- La macroéconomie du déséquilibre ... 393
7.5.1. Patinkin, le précurseur ... 393
7.5.2. L'hypothèse de décision duale ... 395
Clower introduit l'hypothèse de décision duale ... 395
Leijonhufvud soutient le keynésianisme clowerien ... 396
7.5.3. Solow et Stiglitz : emploi et salaire ... 397
7.5.4. L'école française du déséquilibre ... 400
Benassy, le rationnement et l'équilibre « fixprice » ... 400
Malinvaud réexamine la théorie du chômage ... 402
7.5.5. Conclusions quant à l'école du déséquilibre ... 405
Les critiques ... 405
Les suites ... 405

7.6- Les post keynésiens ... 407

7.7- Les nouveaux keynésiens ... 412
Fischer : les contrats de travail à long terme ... 412
Shapiro et Stiglitz : les salaires d'efficience ... 413
Impact de premier ordre pour des frictions nominales de deuxième ordre ... 415
L'effet de la presque-rationalité ... 415
Les coûts de menu ... 416
Romer et l'incitation à adapter les prix ... 417
Blanchard & Summers et l'hystérèse ... 418
Stiglitz et la rigidité des prix ... 421

TITRE VIII- La macroéconomie dynamique ... 423

8.1- L'épargne optimale suivant Ramsey ... 425

8.2- La croissance et le progrès technique ... 429
8.2.1. Instabilité de la croissance dans le modèle Harrod-Domar ... 429
Domar : croissance et plein emploi ... 431
8.2.2. La réponse néoclassique : stabilité ... 432
Solow : accumulation et substitution entre les facteurs ... 432
Le résidu de Solow ... 435
Swan : croissance de la population, du capital et du revenu ... 436
8.2.3. Progrès technique et fonction de production ... 437
Influence du progrès technique sur la répartition du revenu ... 437
Progrès technique incorporé ... 440
8.2.4. La croissance vue par les Cambridgiens ... 442
Croissance et accumulation ... 442
Le progrès technique ... 445
Pasinetti complète le modèle ... 446
Kaldor et la fonction de progrès technique ... 448
8.2.5. La règle d'or et le retour de Ramsey ... 450
Phelps et la règle d'or ... 450
Cass, Koopmans et Ramsey ... 451
8.2.6. La croissance endogène ... 453

Arrow et l'apprentissage par la pratique ... 453
Paul Romer et la recherche et développement ... 454

8.3- Le cycle conjoncturel ... 458
8.3.1. Le cycle causé par le surinvestissement ... 458
Le principe d'accélération : Aftalion et Clark ... 458
Les critiques contre l'accélérateur ... 462
Modèles combinant le multiplicateur et l'accélérateur ... 464
D'autres théories du surinvestissement ... 468
8.3.2. Le cycle causé par les phénomènes monétaires ... 474
Hayek : crédit et détours de la production ... 474
La théorie monétaire de Hayek ... 477
La controverse entre Hayek et Kaldor ... 478
La dénationalisation de la monnaie ... 480
8.3.3. Le « cycle économique réel » ... 481
Lucas et le cycle des affaires ... 481
Lucas et le marché du travail ... 481
La nature du cycle conjoncturel ... 483
Les modèles RBC ... 484
Kydland et Prescott : le modèle de base ... 485
Modèles améliorant le réalisme ... 489
Les critiques contre les modèles RBC ... 496
Critiques générales ... 496
Controverse relative aux effets des chocs technologiques sur l'emploi ... 497

TITRE IX- Quelques théories et quelques débats importants ... 501

9.1- L'économie du bien-être selon Pigou ... 502
9.1.1. Introduction : le dividende national ... 502
9.1.2. La distribution et la redistribution des revenus ... 503
9.1.3. Les externalités ... 505
La critique de Coase ... 507

9.2- Faut-il croire en la théorie autrichienne ? ... 510
9.2.1. Knight contre Hayek à propos du capital ... 510
La période de production ... 510
L'attente, le capital permanent ... 511
9.2.2. « The Pure Theory of Capital » ... 512

9.3- Le débat sur le calcul socialiste ... 517
9.3.1. Prélude ... 517
9.3.2. La controverse entre les deux Guerres ... 520
von Mises et le lancement de la controverse ... 520
Taylor et Dickinson : la possibilité d'un calcul socialiste ... 521
Hayek fait le point ... 524
Lange : l'équilibre général ... 526
Lerner : l'égalisation du prix avec le coût marginal ... 528
Hayek : au-delà du calcul ... 529
Premières conclusions ... 531
9.3.3. Une vision moderne du socialisme de marché ... 533
Socialisme de marché ? Mais encore ... 533
Bardhan et Roemer : comment remplacer le marché financier ... 534
La « contrainte budgétaire douce » ... 535

9.4- Le risque et l'incertitude ... 538
9.4.1. La dispute philosophique ... 538
Keynes : le degré de croyance non mesurable ... 538
Ramsey: le degré de croyance mesurable ... 539
9.4.2. Knight : distinction entre le risque et l'incertitude ... 540
9.4.3. Théorie de l'utilité espérée ... 543
Le paradoxe de Saint-Pétersbourg ... 543
von Neumann et Morgenstern ... 543
L'axiome d'indépendance ... 545
La courbe d'utilité d'un revenu incertain ... 546
9.4.4. La critique d'Allais ... 549

Réfutation de la formule bernoullienne ... 550
9.4.5. L'incertitude dans la *Théorie Générale* de Keynes ... 551

9.5- Revenu, consommation et épargne ... 553
9.5.1. Les précurseurs, Kuznets et Samuelson ... 553
9.5.2. Duesenberry et l'hypothèse du revenu relatif ... 554
9.5.3. Friedman et l'hypothèse du revenu permanent ... 557
9.5.4. Modigliani et l'hypothèse du cycle de vie ... 560
LCH et fonction de consommation ... 562
Implications du modèle ... 563
9.5.5. L'épargne des entreprises ... 565

9.6- Capital et investissement ... 566
9.6.1- Lerner et l'efficacité marginale de l'investissement ... 566
9.6.2- Taux interne de rentabilité et valeur actuelle nette ... 567
9.6.3- Une théorie néoclassique de l'investissement, est-elle possible ? ... 569
Haavelmo pose la question ... 569
Jorgenson y répond positivement ... 570
9.6.4- Le coût d'ajustement convexe : solution ultime ? ... 573
9.6.5- Le *q* de Tobin ... 576
9.6.6. Le coût du capital selon Modigliani et Miller ... 577

9.7- On Keynesian economics and the economics of Keynes ... 579

9.8- Pourquoi y a-t-il des entreprises ? ... 583
La nature de la firme ... 583
L'entrepreneur selon Alchian et Desmetz ... 583
Alchian, Desmetz et les droits de propriété ... 584

Titre X- Conclusions ... 587

10.1- Microéconomie ... 589
Les firmes ... 589
La concurrence ... 590
L'équilibre du producteur ... 591
Capital, intérêt, épargne et préférence temporelle ... 593
L'équilibre et le facteur temps ... 595

10.2- Macroéconomie ... 597
Les microfondations ... 597
Les modèles macroéconomiques ... 597
Pourquoi y a-t-il du chômage ? ... 597
La monnaie ... 598
Les anticipations rationnelles ... 599
Capital, investissement et progrès technique ... 600

10.3- Conclusion générale ... 601