

Université libre de Bruxelles

Introduction à la macroéconomie

Pierre-Guillaume Méon

Avec la collaboration de :

Julien Ravet, Marie Montigny, Geoffrey Minne, Gwenael Tondreau, Thibaut Dort, Heng-Sheng Quach

Ont également contribué à ce syllabus :

Oscar Bernal, Imane Chaara, Naïm Cordemans, Benoit Crutzen, Quentin David, Hafsatou Diallo, Alice Duhaut, Iason Foscolos, Renaud Foucart, Denis Herbaux, Gisèle Hittes, Alice Mc Cathie, Arnaud Lemaire, Elena Mattevi, Marie-Christine Moreels, Fulvio Mulatero, John Neve De Mevergnies, Sile O'Dorchai, Benjamin Rausch, Idesbald Van Der Gracht et Catherine Van Den Steen.

Année académique 2010-2011

Table des matières

CHAPITRE 1 : QU'EST-CE QUE LA MACROECONOMIE ?	1
INTRODUCTION : UNE DEFINITION	1
SECTION 1 : LA MACROECONOMIE : UNE DISCIPLINE SCIENTIFIQUE	2
I- Evolution de la macroéconomie en tant que discipline.....	2
A- Une discipline jeune.....	3
B- Une discipline en débat.....	5
II- Questions méthodologiques.....	10
A- Ce qu'on attend de la macroéconomie.....	11
B- L'importance de la modélisation	12
C- La relation entre micro et macroéconomie	13
SECTION 2 : LES PRINCIPAUX AGREGATS	16
I- Flux et stocks	16
II- La mesure du revenu	18
A- Le PIB nominal	18
B- La comparaison des PIB dans le temps et l'espace.....	25
III- L'inflation.....	34
A- L'indice des prix à la consommation.....	35
B- La mesure de l'inflation grâce au déflateur du PIB	39
IV- Le chômage.....	41
A- La distinction entre chômage, inactivité et emploi	41
B- La difficulté de mesurer le chômage.....	41
CONCLUSION.....	43
REFERENCES	43
PREMIERE PARTIE : LA CROISSANCE	45
CHAPITRE 2 : LA CROISSANCE PAR ACCUMULATION DE RESSOURCES	50
INTRODUCTION : LES ORIGINES : DAVID RICARDO	50
SECTION 1 : LES HYPOTHESES DU MODELE DE SOLOW.....	51
I- La fonction de production agrégée	51
A- Une représentation simplifiée de la production	51
B- Forme de la fonction de production	52
II- L'accumulation de capital	54
A- La répartition de la production entre investissement et consommation	55
B- La variation du stock de capital	55
SECTION 2 : LA CROISSANCE DANS LE MODELE DE SOLOW	58
I- L'état stationnaire.....	58
II- La convergence vers l'état stationnaire.....	60
SECTION 3 : LA RELATION ENTRE CROISSANCE ET EPARGNE	62
I- L'absence de relation entre épargne et croissance à l'état stationnaire	62
II- Une relation croissante entre épargne et croissance hors de l'état stationnaire	64
CONCLUSION : UN RETOUR SUR LA CONVERGENCE.....	66
REFERENCES	68
CHAPITRE 3 : LE ROLE DU PROGRES TECHNIQUE DANS LA CROISSANCE	69
INTRODUCTION : LES ORIGINES : ADAM SMITH	69
SECTION 1 : UNE MESURE DE LA PART DU PROGRES DANS LA CROISSANCE.....	70
I- Construction du résidu de Solow	70
II- La part du progrès technique dans la croissance.....	74
SECTION 2 : LE PROGRES TECHNIQUE DANS LE MODELE DE SOLOW	76
I- Effet d'une croissance ponctuelle de la productivité	76
II- La prise en compte d'une augmentation tendancielle de la productivité	79
SECTION 3 : LES DETERMINANTS DU PROGRES TECHNIQUE	81
I- L'investissement dans la recherche	82
A- Motivation des investissements en R&D.....	83
B- L'appropriabilité des résultats de la R&D	83

Les notes du cours ne sont transmises qu'à titre indicatif.

II- L'efficacité des efforts de recherche.....	85
A- Facteurs institutionnels.....	85
B- Le caractère cumulatif des connaissances.....	88
CONCLUSION.....	90
REFERENCES	90
ANNEXES	91
DEUXIEME PARTIE : L'ECONOMIE DANS LE COURT TERME : LES FLUCTUATIONS.....	92
CHAPITRE 4 : LE MARCHE DES BIENS	97
INTRODUCTION : L'IMPORTANCE DE LA DEMANDE DE BIENS	97
SECTION 1 : LES COMPOSANTES DE LA DEMANDE GLOBALE	97
I- La consommation.....	98
A- La consommation : une fonction du revenu courant.....	98
B- La consommation : une fonction des revenus futurs.....	101
II- L'investissement	104
A- L'investissement : une fonction du taux d'intérêt	105
B- L'investissement : une fonction de la demande et des anticipations.....	106
III- Les dépenses de l'Etat.....	106
SECTION 2 : L'EQUILIBRE SUR LE MARCHE DES BIENS.....	108
I- Le diagramme à 45 degrés.....	108
II- Le multiplicateur	112
CONCLUSION.....	118
REFERENCES	119
ANNEXES	119
CHAPITRE 5 : LE MARCHE DE LA MONNAIE ET LES MARCHES FINANCIERS	120
INTRODUCTION : CE QU'EST LA MONNAIE ET CE QU'ELLE N'EST PAS	120
SECTION 1 : L'OFFRE DE MONNAIE.....	121
I- Fonctions et formes de la monnaie	121
A- Les trois fonctions de la monnaie	121
B- L'évolution des formes de la monnaie.....	124
C- La mesure de l'offre de monnaie	125
II- La création monétaire	126
A- La création monétaire en l'absence de banques commerciales.....	127
B- La création monétaire en présence de banques commerciales.....	128
III- Les moyens de contrôle de l'offre de monnaie	133
SECTION 2 : LA DEMANDE DE MONNAIE	135
I- La théorie quantitative de la monnaie	136
A- La quantité de monnaie nécessaire aux transactions.....	136
B- La relation entre la masse monétaire et les prix.....	138
II- La préférence pour la liquidité.....	140
A- Le motif de transaction.....	141
B- Le motif de précaution.....	141
C- Le motif de spéculation	142
D- La fonction de demande de liquidité	145
SECTION 3 : LA DETERMINATION DU TAUX D'INTERET	147
I- L'équilibre sur le marché de la monnaie.....	147
II- L'impact d'une augmentation de l'offre de monnaie.....	148
III- L'impact d'une augmentation du revenu.....	150
CONCLUSION.....	151
REFERENCES	151
ANNEXE	152
CHAPITRE 6 : LE MODELE IS-LM.....	153
INTRODUCTION : 70 ANS ET TOUTES SES DENTS	153
SECTION 1 : CONSTRUCTION DU DIAGRAMME IS-LM	154
I- La courbe IS.....	154
A- Pourquoi IS est décroissante.....	154

Les notes du cours ne sont transmises qu'à titre indicatif.

B- Forme et position de la courbe IS	156
II- La courbe LM	160
A- Pourquoi LM est croissante	160
B- Forme et position de la courbe LM	161
III- L'équilibre global	165
SECTION 2 : LA POLITIQUE ECONOMIQUE DANS LE MODELE IS-LM	167
I- La politique monétaire	167
A- Principe de la politique monétaire	168
B- Efficacité de la politique monétaire	169
II- La politique budgétaire	171
A- Principe de la politique budgétaire	172
B- Efficacité de la politique budgétaire	174
III- Le policy-mix	178
A- Politique budgétaire expansionniste accompagnée d'une politique monétaire restrictive	178
B- Politique budgétaire expansionniste accompagnée d'une politique monétaire expansionniste	179
CONCLUSION	180
REFERENCES	181
CHAPITRE 7 : LE MARCHE DU TRAVAIL	182
INTRODUCTION : LES DIFFERENCES INTERNATIONALES DE TAUX DE CHOMAGE	182
SECTION 1 : DESCRIPTION D'UN MARCHE DU TRAVAIL CONCURRENTIEL	183
I- L'offre de travail	184
II- La demande de travail	186
III- L'équilibre du marché du travail	187
SECTION 2 : LA POSSIBILITE DU SOUS-EMPLOI	190
I- Les rigidités spécifiques au marché du travail	190
A- Les rigidités exogènes au marché du travail	190
B- Rigidités créées du côté de l'offre de travail	192
C- Rigidités créées du côté de la demande de travail	193
II- Le chômage involontaire	195
A- Conséquence d'une rigidité à la baisse du salaire réel	195
B- La distinction entre chômage structurel et chômage conjoncturel	197
CONCLUSION	198
REFERENCES	199
ANNEXE	200
CHAPITRE 8 : L'OFFRE AGREGEE ET LE MODELE DE LA SYNTHESE	201
INTRODUCTION : POURQUOI IL FAUT COMPLETER LE MODELE IS-LM	201
SECTION 1 : CONSTRUCTION ET EQUILIBRE DANS LE MODELE DEMANDE AGREGEE-OFFRE AGREGEE	202
I- L'offre agrégée	202
A- La rigidité des salaires comme fondement de l'offre agrégée	203
B- Trois autres fondements de l'offre agrégée	209
II- La demande agrégée	217
A- Construction de la courbe de demande agrégée	218
B- Position de la courbe de demande agrégée	219
III- L'équilibre global	221
SECTION 2 : LA POLITIQUE ECONOMIQUE DANS LE MODELE DEMANDE AGREGEE-OFFRE AGREGEE	227
I- La politique budgétaire	227
II- La politique monétaire	230
III- Que faire face à un choc de demande ?	233
IV- Que faire face à un choc d'offre ?	236
CONCLUSION	241
REFERENCES	242
ANNEXE	242
TROISIEME PARTIE : L'ECONOMIE OUVERTE	243
INTRODUCTION : L'OUVERTURE DES ECONOMIES	243
CHAPITRE 9 : LA COMPTABILITE DES TRANSACTIONS AVEC LE RESTE DU MONDE	245

Les notes du cours ne sont transmises qu'à titre indicatif.

INTRODUCTION : UN DOCUMENT COMPTABLE.....	245
SECTION 1 : CONSTRUCTION DE LA BALANCE DES PAIEMENTS.....	245
I- Définition de la balance des paiements.....	245
II- L'enregistrement des transactions avec le reste du monde.....	246
SECTION 2 : INTERPRETATION DE LA BALANCE DES PAIEMENTS.....	248
CONCLUSION.....	251
REFERENCES.....	251
ANNEXE.....	252
CHAPITRE 10 : LE MARCHÉ DES BIENS EN ÉCONOMIE OUVERTE.....	253
INTRODUCTION : LA CONTRAINTE EXTERIEURE.....	253
SECTION 1 : LES DÉTERMINANTS DES EXPORTATIONS NETTES.....	254
I- Les revenus.....	254
II- Le taux de change.....	255
SECTION 2 : L'ÉQUILIBRE DU MARCHÉ DES BIENS EN ÉCONOMIE OUVERTE.....	259
I- Le revenu d'équilibre.....	259
II- L'effet d'une augmentation exogène de la demande globale.....	263
III- Une dépréciation.....	265
CONCLUSION.....	268
REFERENCES.....	268
CHAPITRE 11 : LE MARCHÉ DES CAPITAUX ET DE LA MONNAIE DANS UNE ÉCONOMIE OUVERTE.....	269
INTRODUCTION : PRESQUE AUTANT DE MONNAIES QUE DE PAYS.....	269
SECTION 1 : LES MOUVEMENTS INTERNATIONAUX DE CAPITAUX.....	270
SECTION 2 : LA GESTION DE LA MONNAIE DANS UNE ÉCONOMIE OUVERTE.....	274
I- Les contreparties de la masse monétaire en économie ouverte.....	274
II- Les régimes de change.....	276
CONCLUSION.....	279
REFERENCES.....	280
ANNEXE.....	280
CHAPITRE 12 : LA POLITIQUE MACROÉCONOMIQUE DANS UNE ÉCONOMIE OUVERTE : LE MODELE DE MUNDELL-FLEMING.....	282
INTRODUCTION : UNE INNOVATION THÉORIQUE AUX IMPLICATIONS PRATIQUES IMMÉDIATES.....	282
SECTION 1 : CONSTRUCTION DU MODELE DE MUNDELL-FLEMING.....	283
I- Le marché des biens : la courbe IS.....	283
II- Le marché de la monnaie et les marchés financiers : la courbe LM et la droite d'intégration financière.....	285
A- La courbe LM.....	285
B- La droite d'intégration financière.....	286
III- L'équilibre dans le modèle de Mundell-Fleming.....	287
SECTION 2 : LA POLITIQUE ÉCONOMIQUE DANS LE MODELE DE MUNDELL-FLEMING.....	288
I- La politique budgétaire.....	288
A- Politique budgétaire en changes fixes.....	289
B- Politique budgétaire en changes flexibles.....	290
II- La politique monétaire.....	292
A- Politique monétaire en changes fixes.....	293
B- Politique monétaire en changes flexibles.....	294
III- La dévaluation.....	296
CONCLUSION.....	298
REFERENCES.....	299

Les notes du cours ne sont transmises qu'à titre indicatif.

Chapitre 1 : Qu'est-ce que la macroéconomie ?

Introduction : une définition

Bien qu'une discipline ne se limite pas à une définition d'une seule phrase, celle de la macroéconomie s'avère assez facile à trouver.

Macroéconomie ou macro-économie : discipline qui a pour objet d'analyser l'économie d'un pays d'un point de vue global.

Cette définition simple appelle au moins deux commentaires :

- d'abord, il faut insister sur le caractère global de l'analyse. On ne s'intéressera par définition pas au fonctionnement du marché du blé, de l'automobile ou de la confiture de framboise, mais à toute l'économie à la fois.

Les variables d'intérêt seront donc ce qu'on appelle des *agrégats*. Il s'agira par exemple de la production totale, du niveau général des prix ou encore de l'emploi global. Tout au plus pourra-t-on s'intéresser à un marché s'il joue un rôle important dans l'évolution globale de l'économie (ex. marché du travail).

La perspective est donc d'emblée celle d'un état, au moins. On pourra de plus s'intéresser aux relations entre états. C'est l'objet de la macroéconomie internationale.

- ensuite, la macroéconomie se veut une discipline scientifique. L'objectif de scientificité impose aux macroéconomistes des contraintes de rigueur et de confrontation aux faits de leurs analyses.

Ces contraintes tiennent à la fois aux principes généraux d'une démarche scientifique, mais aussi aux particularités de l'objet de la macroéconomie.

Dans ce chapitre, nous allons tenter de préciser la méthode et l'objet de la macroéconomie. Nous allons pour ce faire procéder en deux étapes, qui correspondent à deux sections.

-1^{ère} section : nous allons décrire la macroéconomie en tant que discipline scientifique en insistant sur sa méthode et son évolution. Il s'agira donc, pour parler pompeusement, d'une section historique, méthodologique et épistémologique.

-2^{ème} section : La deuxième section sera consacrée aux variables centrales qui font l'objet d'une analyse macroéconomique. Il s'agira de définir ce qu'on appelle des agrégats. Nous nous intéresserons aux principaux d'entre eux, qui seront utilisés dans la suite du cours.

Section 1 : La macroéconomie : une discipline scientifique

Bien que les phénomènes auxquels la macroéconomie s'intéresse soient vieux comme le monde, la discipline ne s'est constituée que peu à peu. Elle ne s'est en particulier distinguée du reste de l'analyse économique qu'avec retard.

C'est en ayant à l'esprit son évolution qu'on pourra mieux saisir sa méthode et les débats qu'elle suscite.

C'est pourquoi cette section est divisée en deux sous-sections. La première décrit l'histoire de la discipline (I), la seconde s'intéresse à sa méthode (II).

I- Evolution de la macroéconomie en tant que discipline

L'existence d'une discipline appelée macroéconomie nous semble aujourd'hui aller de soi puisque dans presque toutes les universités du monde on l'enseigne en tant que telle dès les cours d'introduction. Pourtant, son autonomie ne va pas de soi.

Après tout, il n'est pas a priori évident que les phénomènes économiques globaux doivent faire l'objet d'une discipline à part puisque la macroéconomie reste avant tout de l'économie. Cette idée ne s'est d'ailleurs imposée que tardivement au sein de la science économique. Par exemple, la première édition du manuel de Paul Samuelson, parue en 1948, ne reposait pas sur la distinction entre micro et macroéconomie.

C'est pourquoi la macroéconomie reste encore une discipline jeune à l'échelle de l'analyse économique moderne (A). Elle s'est constituée peu à peu et continue d'évoluer à partir des débats qui l'animent (B).

Cette section brosse à grands traits l'évolution de la macroéconomie depuis les années trente. Cela vous permettra de comprendre d'où viennent les connaissances que nous allons voir ensemble. Pour ce faire, nous allons malheureusement devoir mentionner des théories que nous ne verrons que plus tard dans le cours. C'est pourquoi vous avez intérêt à relire cette section à la fin du quadrimestre, voire en cours de quadrimestre. Vous la verrez sous un autre jour et comprendrez plus facilement l'articulation des éléments du reste du cours.

A- Une discipline jeune

On considère en général que la macroéconomie est née avec la révolution keynésienne, dans les années trente. C'est John Maynard Keynes (1883-1946) qui sert de référence, et surtout la publication de son livre, la *Théorie générale de l'emploi, de l'intérêt et de la monnaie*, en 1936.

Pourtant les économistes s'intéressaient depuis longtemps aux phénomènes globaux en les appréhendant directement à l'échelle des nations. Adam Smith avait consacré son ouvrage de référence à la *Richesse des nations* (1776). Keynes lui-même rendait hommage à François Quesnay, le chef de file des physiocrates, pour son *Tableau économique*, ou zigzag, qui modélisait les interdépendances au sein d'une économie dès 1758.

Le terme de macroéconomie est même fondé dès 1933 par Ragnar Frisch, un économiste norvégien qui obtiendra le premier prix Nobel d'économie en 1969.

Pour bien comprendre la rupture de l'analyse keynésienne, il faut rappeler l'état de la réflexion économique avant les années trente. La base de la réflexion s'inscrivait dans le courant néo-classique qui considérait que les marchés aboutissaient spontanément à l'optimum. Il en résultait des prescriptions de laisser-faire.

On percevait bien les fluctuations conjoncturelles. On avait ainsi identifié plusieurs familles de cycles imbriquées les unes dans les autres, des « cycles mineurs » de Joseph Kitchin aux « cycles longs » de Nicholas Kondratieff. On considérait cependant que ces cycles constituaient une réaction normale de l'économie aux chocs que pouvaient représenter des inventions nouvelles, de mauvaises récoltes, des changements de préférences des consommateurs etc.

Il convenait donc de laisser l'économie s'adapter à ces changements. Les plus interventionnistes préconisaient des politiques de soutien sectoriel, mais on ne concevait pas de politique économique conjoncturelle à l'échelle de tout un pays.

A la même époque, les économistes marxistes pensaient que l'économie de marché était vouée à l'effondrement. Ils s'intéressaient aussi à la meilleure façon de gérer une économie planifiée, puisque l'URSS venait d'opter pour ce système.

Toutes ces analyses se sont avérées incapables de proposer une explication et des remèdes convaincants à la crise de 1929. Cette crise semble au départ habituelle. On observe une baisse de la production et des prix comme on l'avait déjà fait par le passé, comme en 1921. On pense alors qu'elle se résorbera au bout de quelques mois.

Ce qui frappe, et qui fait que la crise de 1929 reste une référence encore aujourd'hui, ce sont à la fois son ampleur et sa durée.

La production industrielle mondiale hors URSS diminue de plus d'un tiers en quatre ans. Elle est presque divisée par deux entre 1929 et 1932 aux Etats-Unis et reste inférieure à son niveau de 1929 jusqu'à la guerre.

En opposition aux analyses en vogue à l'époque, la théorie keynésienne va proposer une triple rupture :

- en rupture avec l'analyse néoclassique, Keynes s'intéresse à des agrégats et à des *comportements globaux* : produit national, consommation, investissement etc. Il substitue aussi à un raisonnement qui repose sur des marchés où les prix ajustent l'offre et la demande, un raisonnement où l'équilibre se définit en termes de flux.

Cette perspective peut être critiquée car elle ne repose pas sur le comportement observable d'individus. Elle a en revanche le grand avantage de se prêter à la quantification. Il n'est pas étonnant que la comptabilité nationale ait commencé à se développer à peu près au même moment.

- Keynes se concentre sur un nouvel objet d'étude : le *niveau d'activité*. Il montre que le plein emploi n'est pas l'état normal de l'économie de marché. Par conséquent, il faut étudier ce qui détermine le niveau de l'emploi.

- Une autre conséquence de ce changement de perspective est qu'il devient possible de réfléchir aux politiques économiques susceptibles d'influencer le niveau d'activité. Keynes va suggérer qu'une politique bien menée peut atténuer les fluctuations conjoncturelles.

Entre le laisser-faire et la planification, il propose donc une troisième solution qui consiste à confier à l'Etat la gestion de l'activité globale dans une économie de marché. L'instrument privilégié est le budget de l'Etat qui peut relancer l'activité en augmentant ses dépenses. On parle de *politique budgétaire*.

Par conséquent la théorie keynésienne propose des solutions à un problème majeur dont les théories précédentes s'avèrent incapables de rendre compte. Cela explique sans aucun doute son succès scientifique et politique.

Encadré 1 : La personnalité de John Maynard Keynes

Pour être juste, il faut souligner que d'autres économistes de premier plan proposaient au même moment des théories qui reposaient sur des idées très proches de Keynes. C'est le cas par exemple de Gunmar Myrdal, suédois et prix Nobel en 1974, ou de Michael Kalecki, polonais. Les idées de Keynes étaient donc dans l'air.

On peut alors se demander ce qui explique la prééminence de Keynes dans l'histoire de la macroéconomie. On peut y voir le résultat du prestige des économistes britanniques (Keynes enseignait à Cambridge).

La position scientifique, administrative et intellectuelle de Keynes a aussi dû jouer un rôle : il avait en effet mené une triple carrière d'enseignant (à Cambridge), de haut fonctionnaire (il conseillait le Trésor britannique et la Banque d'Angleterre. Il avait par exemple participé aux négociations de paix de 1919), et de financier avisé (il avait fait fortune en bourse et dirigeait une société d'assurance). Il était devenu en 1911 le rédacteur en chef de l'*Economic Journal*, l'une des plus prestigieuses revues scientifiques en économie.

Mais son aura dépassait le cercle des économistes professionnels. Il écrivait régulièrement des chroniques dans la presse grand public. Enfin, il s'intéressait aussi aux arts. Il collectionnait les œuvres d'art contemporain et les livres rares, finançait une troupe de théâtre et épousa une danseuse russe.

Keynes était donc un personnage hors du commun. Cela explique peut-être pourquoi il est le seul économiste avec Marx à avoir donné son nom à un courant.

Si Keynes a influencé la science économique en imposant ses idées, il a aussi provoqué des réactions. En un sens, l'histoire de la macroéconomie depuis cinquante ans est celle de la remise en cause et de la défense des idées keynésiennes.

B- Une discipline en débat

Au lendemain de la deuxième Guerre Mondiale, la théorie keynésienne tient une position quasiment hégémonique sur la macroéconomie. Elle va déboucher sur un consensus qu'on appellera la *synthèse néoclassique* (1). Pourtant, des critiques vont émerger, ce qui va relancer le débat jusqu'à aujourd'hui. Ces critiques vont d'abord porter sur l'application des conclusions keynésiennes à la politique économique (2). Elles vont ensuite remettre également en cause la méthode et l'apport théorique keynésien (3).

1) La synthèse néoclassique

En 1955, Paul Samuelson, qui est l'une des figures de proue de la synthèse néoclassique, écrivait dans son manuel que le résultat de la synthèse néoclassique était « accepté par tous, sauf 5% d'économistes d'extrême droite ou d'extrême gauche. »

Ce consensus scientifique reposait fondamentalement sur les idées keynésiennes traduites et formalisées. La théorie était complétée par des relations issues de la théorie néoclassique et de travaux statistiques qui permettaient d'intégrer les mouvements de prix.

Dans tous les cas on considérait qu'une politique économique judicieuse permettait de réduire efficacement les fluctuations conjoncturelles. Par conséquent, la recherche se concentrait sur l'amélioration des théories et des modèles de la synthèse.

Encadré 2 : La recherche en macroéconomie au temps de la synthèse

Même s'il existait un consensus autour de la validité des idées keynésiennes, la recherche ne s'était pas pour autant arrêtée. Trois champs de recherche étaient particulièrement actifs, et ont posé les bases de ce qui constitue aujourd'hui la macroéconomie :

- le développement du modèle IS-LM, que nous étudierons dans le chapitre 6 du cours. Ce modèle permet d'analyser les effets des politiques macroéconomiques (budgétaire, monétaire). Il s'agissait de déterminer au mieux les paramètres de ce modèle, la forme des relations, ainsi que les variables qui devaient être intégrées, pour le rendre plus réaliste et précis dans ses prévisions. On a par exemple intégré les échanges commerciaux avec l'étranger (Robert Mundell, prix Nobel 1999).

Ces travaux ont été accompagnés du développement de la comptabilité nationale (Simon Kuznets, prix Nobel 1971 et Richard Stone, prix Nobel 1984) et des modèles macroéconométriques qui permettent les simulations et les prévisions.

- le deuxième champ de recherche particulièrement actif s'est alors avéré être l'amélioration de la description des comportements des principaux agrégats (investissement, consommation...). On a ainsi amélioré la théorie de la consommation pour tenir compte de sa dimension intertemporelle (Franco Modigliani, prix Nobel 1985, et Milton Friedman, prix Nobel 1976). On s'est de la même façon intéressé aux déterminants de l'investissement (James Tobin, prix Nobel 1981). Nous évoquerons les résultats de ces travaux dans le chapitre 4 du cours.

- parallèlement, on a proposé une théorie de la croissance économique afin d'expliquer l'augmentation du niveau de vie. C'est à Robert Solow (prix Nobel 1987) qu'on doit la principale avancée sur ce plan. Nous étudierons son modèle dans le prochain chapitre du cours.

Cependant, les conclusions de la théorie keynésienne finirent par susciter des critiques de plus en plus vives.

2) La critique monétariste

Dans les années soixante, les premières attaques viennent du *courant monétariste* mené par Milton Friedman. Le débat keynésiens-monétaristes allait structurer le débat jusqu'aux années septante.

Les critiques monétaristes de la théorie keynésienne portaient de trois arguments principaux :

- les monétaristes ont souligné l'efficacité de la *politique monétaire* comme instrument de politique économique. La politique monétaire consiste à jouer sur la quantité de monnaie en circulation et les taux d'intérêt pour influencer l'activité. La théorie keynésienne privilégiait au contraire la politique budgétaire.¹

- Dans les années cinquante et soixante, les économistes de la synthèse considéraient que grâce à la politique macroéconomique, il était possible d'arbitrer en permanence entre chômage et inflation. Cette conviction s'appuyait sur la *courbe de Phillips*, issue des travaux statistiques d'Alban W. Phillips.²

Les monétaristes, Milton Friedman en tête, se sont opposés à l'idée qu'une telle relation puisse exister. Pour eux, elle était incompatible avec la théorie économique parce qu'elle supposait une forme d'illusion monétaire permanente.

Selon les monétaristes, toute tentative d'exploiter cette relation ne pouvait de plus déboucher que sur de l'inflation, sans effet sur le chômage. L'apparition de la stagflation (chômage et inflation) dans les années septante participa à populariser leurs arguments.

- Enfin, pour les monétaristes, toute tentative de limiter les fluctuations par la politique économique était vouée à l'échec. En effet, ils considéraient que nos connaissances du fonctionnement de l'économie sont trop imparfaites pour prétendre gérer efficacement le cycle économique. De plus, la réaction des autorités se fait forcément avec retard, et a alors toutes les chances de produire ses effets à contretemps.

¹ Dans un ouvrage célèbre intitulé *Histoire Monétaire des Etats-Unis, 1867-1960*, Milton Friedman et Anna Schwartz (1963), ont défendu l'idée que les fluctuations de la quantité de monnaie en circulation (la *masse monétaire*) permettaient de prévoir les fluctuations de la production. En particulier, la crise de 1929 pouvait s'expliquer, selon ces auteurs, par la contraction de la masse monétaire provoquée par les faillites bancaires qui s'étaient produites suite au krach boursier.

² A.W. Phillips "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957", *Economica*, 1958.

Par conséquent, il vaut mieux avoir recours à des *règles* de politique économique. Milton Friedman proposait par exemple que la politique monétaire ait pour principe de maintenir constant le taux de croissance de la masse monétaire, quelles que soient les circonstances.

Malgré tout, les monétaristes acceptaient dans une large mesure la théorie keynésienne. Leur critique portait plus sur sa capacité à formuler des conseils de politique économique que sur sa pertinence théorique. Le cadre conceptuel des deux écoles était largement commun. Dans les années septante en revanche, une critique radicale des analyses keynésiennes allait apparaître et remettre en cause ce cadre commun.

3) Les nouveaux classiques

On appelle *nouveaux classiques* les économistes qui, à la suite notamment de Robert Lucas (prix Nobel 1995), ont remis en cause la théorie keynésienne à partir des années 70, en s'en prenant à ses fondements théoriques. L'expression de *nouveaux classiques*, est due au fait que les conclusions de cette école rappellent celles des classiques, en s'opposant notamment à l'interventionnisme keynésien.

Les nouveaux classiques critiquent notamment le manque de fondements microéconomiques de la théorie keynésienne. Ils accusent en particulier la théorie keynésienne de ne pas correctement prendre en compte les anticipations des agents et leurs effets sur leurs décisions présentes. Ils proposent au contraire de bâtir les analyses macroéconomiques sur l'hypothèse *d'anticipations rationnelles*. C'est pourquoi on parle parfois du *courant des anticipations rationnelles*.

L'hypothèse des anticipations rationnelles consiste à supposer que les agents économiques fondent leurs anticipations de l'évolution des variables économiques qui les concernent aussi rationnellement que possible, en tenant compte de toute l'information disponible. Cette hypothèse va amener à remettre en cause les principales conclusions keynésiennes en termes d'efficacité des politiques économiques. Puisque les agents anticipent la politique économique, ils s'y adaptent instantanément. La politique ne peut produire des effets que si elle surprend le secteur privé.

Robert Lucas remet même en cause l'utilisation des modèles macroéconométriques. Selon la *critique de Lucas*, puisque les agents s'adaptent aux politiques mises en œuvre, les paramètres estimés par le passé ne sont plus valables au moment où la politique produit ses effets. Les prédictions des modèles ne sont donc pas fiables.

La prise en compte des anticipations rationnelles a donc remis en cause profondément le consensus des années cinquante et soixante. C'est toujours l'hypothèse de travail de l'immense majorité des théories macroéconomiques. Cependant, cette hypothèse a été digérée dans les années 80 et 90 pour aboutir à nouveau à une situation de débat entre macroéconomistes :

- les théoriciens de la *nouvelle économie classique* ont cherché à proposer des explications satisfaisantes des cycles économiques, sans avoir recours aux analyses keynésiennes. C'est ce qu'on appelle la *théorie des cycles réels* ou « *real business cycles* » (RBC). Finn Kydland et Edward Prescott (prix Nobel 2004) ont largement contribué à ce mouvement.

- d'un autre côté, des économistes ont montré que les anticipations rationnelles étaient compatibles avec certaines conclusions keynésiennes, en leur donnant des fondements microéconomiques. C'est pourquoi on parle de *nouveaux keynésiens* ou de *nouvelle économie keynésienne*. Ces auteurs sont encore relativement jeunes mais leurs noms sont connus parce qu'ils ont rédigé des manuels conseillés pour ce cours : Gregory Mankiw ou Olivier Blanchard.

A l'heure actuelle, les recherches se poursuivent. On peut remarquer que les méthodes utilisées par les nouveaux keynésiens et les nouveaux classiques sont les mêmes. Cela facilite donc le débat et la confrontation des hypothèses. Certains avancent même l'idée d'une nouvelle synthèse.

On le voit, la macroéconomie est une discipline en débat permanent, ce qui explique pourquoi elle évolue encore. Ce constat appelle trois remarques importantes qu'il faudra garder à l'esprit dans la suite du cours :

- il faut rester modeste. L'histoire de la pensée montre que ce qui semble évident à un moment donné peut être remis en cause quelques années plus tard. C'est le propre d'une discipline scientifique.

- cependant, malgré les débats, la macroéconomie a réussi à constituer un socle de connaissances communes qui permet d'organiser le débat. Par exemple, aucun économiste n'accepterait l'idée qu'un décuplement de la masse monétaire peut se faire sans inflation.

- en quelque sorte, le fil rouge de la macroéconomie est le débat sur l'efficacité des politiques macroéconomiques. On aura donc implicitement affaire à des arguments en faveur de plus ou moins d'intervention de l'Etat. Ce débat a des implications politiques évidentes.

Par exemple, les idées keynésiennes ont donné une base intellectuelle au développement de l'Etat providence défendu par des partis situés plutôt à gauche de l'échiquier politique. A l'inverse, les arguments des monétaristes ou des nouveaux classiques peuvent amener à des propositions plus libérales.

Cependant, en tant qu'économiste, il n'est pas question de disqualifier ces théories à partir d'arguments dogmatiques et d'a priori idéologiques.

D'abord, on ne peut atteindre ces objectifs qu'en ayant fait un diagnostic correct de l'état et du fonctionnement de l'économie. Les théories qui se sont succédées ont permis d'affiner ce diagnostic.

De plus, les chefs de file des courants qui ont marqué le corpus central de la macroéconomie ne peuvent pas être considérés comme des idéologues stupides. On a vu par exemple que le prix Nobel a récompensé des chercheurs appartenant à des courants opposés. L'apport des uns et des autres a pu s'avérer inconfortable pour telle ou telle idéologie, mais il a toujours alimenté le débat de façon pertinente.

Enfin, l'assimilation des keynésiens à la gauche et des autres à la droite peut provoquer des surprises. Kenneth Arrow qui a reçu le prix Nobel en 1972 pour sa contribution à la théorie de l'équilibre général était socialiste. Plus près de nous, Gregory Mankiw, qui a développé l'économie néo-keynésienne, est devenu conseiller de Georges W. Bush qui n'est pas connu pour être un gauchiste convaincu.

La méthode que nous allons suivre dans le cours découle de ces remarques. Il ne s'agit en aucun cas d'imposer un credo. Au contraire, il s'agit de vous donner les outils et des éléments de réflexion de base pour vous préparer à la suite de vos études d'économie et vous permettre de vous faire une opinion personnelle.

Pour ce faire, il est nécessaire de préciser la méthode utilisée en macroéconomie. C'est l'objet de la prochaine section.

II- Questions méthodologiques

La méthode en macroéconomie n'est pas fondamentalement différente de celle du reste de la discipline, même s'il existe des spécificités. Avant de discuter de ces questions de méthode, il est nécessaire de définir ce qu'on attend de la macroéconomie (A). On verra alors

l'importance de la formalisation (B), ce qui amènera à discuter des relations entre la macroéconomie et la microéconomie (C).

A- Ce qu'on attend de la macroéconomie

L'économie est une science sociale qui s'est constituée au cours des siècles pour conseiller les décisions des gouvernements. C'est pourquoi on parle encore d'*économie politique*. La macroéconomie étant une branche de l'économie politique, elle n'échappe pas à cette fonction.

Ce que le public et les gouvernements attendent donc de la macroéconomie est de formuler des jugements et de proposer des politiques à mettre en œuvre.

Comme un médecin doit avoir compris l'origine des maladies et le fonctionnement du corps humain pour soigner ses malades, la macroéconomie ne peut émettre directement des recommandations de politique économique. Elle doit au contraire passer par une première étape au cours de laquelle elle va avoir pour objet d'expliquer les phénomènes économiques.

Cette première étape relève de ce qu'on appelle une *démarche positive*. Elle consiste à établir des relations entre certains faits, donc à fournir des explications, des théories. L'étape qui consiste à formuler des recommandations relève d'une *démarche normative*.³

Pour être convaincante, l'économie normative doit s'appuyer sur des théories, donc sur l'économie positive.

Pourtant, il existe une différence fondamentale entre une démarche positive et une démarche normative : une démarche normative repose forcément sur un *jugement, ou norme, de valeur*. Ce sont les valeurs qui permettent de choisir des objectifs et de les hiérarchiser afin de définir la politique que l'on recommande. Elles sont par définition subjectives.

Par analogie avec la microéconomie, on peut dire que l'analyse positive a pour objet de déterminer la contrainte à laquelle on est confronté. Les valeurs déterminent les préférences qui vont permettre de définir un optimum.

On voit donc que la démarche normative est inséparable de la démarche positive. C'est ce qui fait la différence entre les économistes et les habitués du café du commerce.

Il y a par conséquent deux raisons pour lesquelles les macroéconomistes peuvent défendre des politiques différentes. Ils peuvent croire à des théories différentes, ce qui relève de la démarche positive. Ils peuvent adhérer à des valeurs différentes qui les amènent à

³ Cette distinction est due au philosophe anglais du XVIII^{ème} siècle David Hume.

poursuivre des objectifs différents. Tel économiste verra dans le chômage l'objectif prioritaire de la politique économique alors que tel autre pourra être beaucoup plus sensible à l'inflation.

Idéalement, un économiste qui propose de mettre en œuvre une politique donnée (démarche normative) devrait préciser la vision du fonctionnement de l'économie sur laquelle il s'appuie (démarche positive) et préciser l'objectif qu'il poursuit.

Malheureusement, les valeurs et les objectifs restent trop souvent implicites : cherche-t-on à améliorer le revenu moyen, faire baisser le chômage, rendre plus équitable la distribution des revenus ? Cela ne facilite pas le débat.

Dans le cours, nous insisterons sur les aspects positifs de la macroéconomie. En posant des questions telles que : « qu'est-ce qui explique les différences de niveau de vie entre les pays ? », « quel est l'effet d'une augmentation des dépenses publiques ? », ou encore « quelle est l'origine de l'inflation ? » etc.

Ce faisant, nous pourrions proposer telle ou telle politique pour améliorer les niveaux de vie, ou telle autre pour réduire l'inflation. Nous pourrions le faire parce qu'implicitement nous considérerons que l'amélioration des niveaux de vie est souhaitable alors que l'inflation ne l'est pas.

Dans tous les cas, notre démarche sera guidée par l'utilisation de modèles. Nous allons voir à présent en quoi ils consistent.

B- L'importance de la modélisation

La théorie macroéconomique repose largement sur des modèles. Un modèle est l'expression d'une théorie sous la forme d'un ensemble d'hypothèses qui relie entre elles des variables jugées pertinentes.

La plupart du temps, les modèles macroéconomiques sont formalisés. On y distingue alors deux types de variables :

- les *variables endogènes* sont celles que le modèle va expliquer ;
- les *variables exogènes* sont les variables qui sont considérées comme « données ».

Le but du jeu est d'expliquer les variations des variables endogènes par celles de certaines variables exogènes. Les autres variables exogènes sont appelées des *paramètres*.

Comme toute théorie, un modèle est forcément une simplification de la réalité. Si on pouvait comprendre le monde tel quel, on n'aurait plus besoin de théorie, mais il est trop

complexe pour cela. On doit par conséquent choisir les simplifications pertinentes pour se focaliser sur les relations importantes.

ex : Il n'est probablement pas nécessaire de tenir compte de la couleur de la cravate du gouverneur de la Banque centrale européenne pour comprendre l'effet des mesures qu'il annonce. Pourtant, tant la couleur de la cravate que les mesures annoncées font partie de la « réalité ».

Par ailleurs, une simplification utile pour comprendre certains phénomènes se révélera abusive pour en expliquer d'autres. Par conséquent, il n'est pas choquant d'avoir recours à des modèles différents en fonction des problèmes étudiés. Les macroéconomistes ont donc recours à une variété de modèles qui peuvent être complémentaires.

Cependant, tous les modèles macroéconomiques sont confrontés à la même *contrainte de bouclage*. Cette contrainte impose de tenir compte du fonctionnement et des interactions de tous les marchés qui constituent le modèle (biens, marchés financiers, marché du travail etc.). Il n'est pas possible de s'intéresser à un marché sans tenir compte de ses effets sur les autres marchés.

On doit donc toujours raisonner en termes d'*équilibre global* de l'économie. Certaines *égalités comptables*, que nous verrons dans la suite de ce chapitre, doivent en particulier être respectées sans quoi le modèle sera incohérent.

C- La relation entre micro et macroéconomie

La plupart des remarques que nous venons de passer en revue pourraient s'appliquer autant à la microéconomie qu'à la macroéconomie. Partant, il n'est pas étonnant que la distinction entre micro et macroéconomie soit l'objet d'un débat entre économistes.

En effet, il n'y a *a priori* pas de raison de distinguer les deux disciplines. Les variables agrégées ne tombent pas du ciel mais sont la somme de variables qui sont le reflet des décisions individuelles et des interactions entre individus. L'offre de travail est bien la somme de toutes les offres individuelles de travail.

De même, l'impératif de bouclage des modèles macroéconomiques s'impose aussi à la théorie de l'équilibre général. C'est d'ailleurs grâce à des raisonnements de nature microéconomique qu'on traitait les problèmes confiés à la macroéconomie à partir des années trente ou quarante.

Alors pourquoi cette séparation ?

Si on a dû développer des analyses spécifiques pour la macroéconomie, c'est au moins au départ parce que la microéconomie butait sur deux problèmes qui l'empêchaient de rendre correctement compte des phénomènes agrégés.

- D'abord, comme Keynes l'a souligné, les comportements collectifs résultent certes de décisions individuelles, mais ces dernières sont prises dans un contexte de grande incertitude. Les agents ignorent l'évolution future de beaucoup de variables importantes ainsi que les décisions qui vont être prises par leurs congénères, ce qui pose des problèmes de *coordination*.

La microéconomie de la première moitié du XX^{ème} siècle était mal préparée pour traiter de ces questions. Il était alors nécessaire de développer une approche alternative.

Cependant, les progrès de la microéconomie, grâce à la théorie des jeux, et le développement de la théorie des anticipations ont largement résorbé ces lacunes.

- La deuxième limite de la microéconomie est plus fondamentale et reste problématique. Il s'agit du problème de l'*agrégation* des comportements individuels. On s'est rendu compte qu'il était très difficile voir impossible de déduire des fonctions de comportement globales utilisables (consommation, investissement etc.) à partir des comportements individuels.

Certains en ont tiré l'argument du « *no bridge* » selon lequel il n'existerait pas de « pont » entre les phénomènes individuels et globaux.

C'est pour ces deux raisons qu'on a eu au départ recours à des raisonnements purement agrégés qui seuls permettaient de rendre compte de façon satisfaisante les problèmes de sous-emploi. La distinction entre microéconomie et macroéconomie s'opérait donc autour d'un double clivage :

- différence de méthode : la macroéconomie utilisait des raisonnements *holistes* (qui reposaient sur des considérations globales) alors que la microéconomie relevait de l'*individualisme méthodologique*.

- différence d'objet d'étude : la macroéconomie s'intéressait aux phénomènes de sous-emploi et d'excès de capacité alors que la microéconomie était consacrée aux questions d'allocation optimale des ressources.

Il faut souligner ici qu'une démarche agrégée ou holiste n'est ni plus ni moins scientifique qu'une démarche purement microéconomique. Le critère de scientificité est un critère de réfutabilité par confrontation aux observations. Or, la démarche agrégée satisfait tout à fait ce critère.

Personne ne reproche aux hydrologues qui étudient un cours d'eau de ne pas modéliser la trajectoire des molécules d'eau une à une.

De même, les raisonnements microéconomiques font eux aussi des hypothèses arbitraires lorsqu'ils reposent sur une forme particulière de fonction d'utilité ou de production.

Pourtant la méthode holiste restait sujette à critique :

- Il y avait des fondements microéconomiques dans les modèles agrégés mais ces fondements restaient implicites. Il valait mieux exposer explicitement les hypothèses sous-jacentes aux modèles.

- Au fur et à mesure que l'analyse se raffinait et portait sur des comportements de plus en plus précis, il devenait difficile de ne pas en spécifier les fondements microéconomiques, sous peine de faire des hypothèses arbitraires ou incohérentes, voire de ne plus être capable d'interpréter ses résultats.

- Enfin, les modèles agrégés interdisent par construction les comparaisons en termes de bien-être puisqu'ils ne reposent pas sur la maximisation de fonctions d'utilité.

Pour essayer de répondre à ces critiques, les chercheurs ont trouvé un compromis. Ils ont cherché à donner des *fondements microéconomiques* aux théories macroéconomiques. Mais ces fondements microéconomiques sont rendus gérables par l'hypothèse simplificatrice de l'*agent représentatif*.

Au lieu de modéliser une infinité de consommateurs, on suppose qu'il n'y en a qu'un qui représente tous les consommateurs. C'est lui qui va permettre de déterminer les choix de consommation et d'offre de travail. De même, on ne considère qu'un producteur représentatif etc.

Cette démarche s'est avérée fructueuse et à permis de mieux décrire des phénomènes importants, comme la rigidité des prix ou les comportements intertemporels.

On voit donc qu'après s'être séparée de la microéconomie, la macroéconomie a fini par s'en rapprocher. La principale différence reste probablement l'intérêt de la macroéconomie pour les agrégats. Ce sont ces agrégats que nous allons étudier à présent.

Section 2 : Les principaux agrégats

Comme la macroéconomie étudie les phénomènes économiques d'un point de vue agrégé, elle doit utiliser des données elles aussi agrégées. C'est ce qu'on appelle les agrégats et c'est la *comptabilité nationale* qui permet de les construire.⁴

Le nom des principaux agrégats est connu parce que ces derniers font l'objet de commentaires réguliers dans les médias. Pourtant leur définition exacte est souvent ignorée, ce qui donne lieu à des erreurs d'interprétation.

L'objet de cette section est donc de définir les principaux agrégats que nous serons amenés à utiliser dans le cours afin de bien savoir de quoi nous allons parler.

Cependant, il existe une pléthore d'agrégats qui mériteraient d'être commentés. Il faut donc faire un choix. Dans cette section, nous allons nous concentrer sur les agrégats qui mesurent les principaux objectifs de la politique économique : le revenu (II), le niveau des prix (III), et le chômage (IV).

Ces trois agrégats correspondent à trois des côtés de ce qu'on appelle le *carré magique* de la politique économique.⁵ Le carré magique est composé de la croissance, du plein-emploi, la stabilité des prix et l'équilibre externe. Le dernier objectif ne sera pas traité ici, mais dans la partie du cours qui portera sur l'économie ouverte.

Cependant, avant de se tourner vers les trois agrégats, nous allons commencer par définir une distinction essentielle en macroéconomie, celle qui oppose les variables de flux aux variables de stock (I).

I- Flux et stocks

La distinction entre flux et stocks peut être définie simplement :

Flux : variable qui se mesure sur un intervalle de temps.

Stock : variable qui se mesure à un instant donné.

La distinction entre flux et stocks est ce qui permet de traiter les problèmes de robinet dans les manuels de calcul. La quantité d'eau qui s'échappe du robinet au cours de chaque

⁴ Vous pouvez d'ailleurs être fiers d'être inscrits dans l'université qui a mis au point la comptabilité nationale belge dans les années cinquante.

⁵ L'expression est due à l'économiste Nicholas Kaldor en 1971.

minute est un flux. La quantité d'eau recueillie dans la baignoire à un moment donné est un stock.

En macroéconomie, la distinction entre flux et stock se retrouve en permanence. Le nombre de licenciements et d'embauches sont des flux. Le nombre de chômeurs et le nombre de personnes qui ont un emploi sont des stocks. De même, l'investissement est un flux mais le capital est un stock. Le déficit est un flux, mais la dette est un stock.

On comprend dès lors qu'il existe une relation mécanique entre flux et stock. Un stock peut être vu comme une accumulation de flux.

Par exemple, la dette publique (stock) est constituée des déficits passés (flux).⁶ Pour s'en convaincre, il suffit de partir de la dernière dette en date (t), qui est égale à la dette de la période précédente à laquelle s'ajoute le déficit de l'année :

$$D_{pub_t} = D_{pub_{t-1}} + \text{déficit}_t$$

Mais on peut écrire la même chose pour la période précédente ($t-1$) :

$$D_{pub_{t-1}} = D_{pub_{t-2}} + \text{déficit}_{t-1}$$

$$\Rightarrow D_{pub_t} = \text{déficit}_t + \text{déficit}_{t-1} + D_{pub_{t-2}}$$

En remontant le temps de la même façon (raisonnement par récurrence), on obtient l'égalité suivante :

$$D_{pub_t} = \text{déficit}_t + \text{déficit}_{t-1} + \dots + \text{déficit}_{t-n+1} + D_{pub_{t-n}}$$

En remontant suffisamment loin dans le temps, on trouvera bien une période $t-n$ où la dette publique était nulle. La dette est donc bien la somme des déficits passés.

On peut à l'inverse exprimer les flux en fonction des stocks. Si on connaît le montant de la dette de cette année et de l'année dernière, on pourra déduire le montant du déficit en calculant la différence entre les deux.

$$\text{déficit}_t = D_{pub_t} - D_{pub_{t-1}}$$

Plus généralement, on peut écrire que les flux correspondent aux variations des stocks :

⁶ Vous pouvez trouver le déficit et la dette des pays membres de l'Union Européenne sur le site d'Eurostat : http://epp.eurostat.ec.europa.eu/portal/page/portal/government_finance_statistics/data/database.

$$\Delta \text{STOCK} = \text{FLUX}$$

II- La mesure du revenu

Lors de ses déclarations d'impôts un contribuable a parfois du mal à calculer son revenu. On imagine que lorsqu'il s'agit de calculer le revenu de toute une économie, les choses sont encore plus compliquées. C'est pourquoi il n'existe pas une mesure unique du revenu, mais plusieurs.

La plus utilisée en macroéconomie est le *produit intérieur brut* (PIB). C'est au calcul de cette mesure que nous nous intéresserons en priorité (A). Nous verrons alors que pour l'utiliser de façon pertinente, il est nécessaire d'utiliser plusieurs variantes du PIB, notamment pour étudier son évolution ou faire des comparaisons internationales (B). Cependant, il existe d'autres mesures du revenu que nous verrons dans une dernière sous-section (C).

A- Le PIB nominal

On va d'abord définir le PIB (1) avant de passer en revue les trois façons de le calculer (2). On déduira alors une égalité comptable importante (3).

1) Définition du PIB

Le PIB a une définition unique mais peut être calculé de trois façons différentes, cohérentes entre elles.

Produit intérieur brut ou *PIB* : valeur monétaire de l'ensemble des biens et services finaux produits sur le territoire d'un pays pendant une période donnée.

Tous les termes sont importants :

- ensemble des biens et services : le PIB est censé mesurer tout ce qui est produit comme richesse dans l'économie. Cela inclut donc tous les biens et tous les services.

Certaines activités sont cependant exclues du PIB à cause de la difficulté rencontrée pour les mesurer. C'est le cas des activités illicites et des activités domestiques.⁷

⁷ Cela implique que lorsqu'on épouse son coiffeur ou sa coiffeuse, on diminue le PIB.

- pendant une période donnée : le PIB est donc un flux. Ce n'est pas une mesure de la richesse totale mais de la richesse produite au cours d'une période donnée. Les données sont en général disponibles à une fréquence trimestrielle ou annuelle.

- biens et services finaux : on exclut les consommations intermédiaires pour éviter de les compter deux fois.

Par exemple, si on n'exclut pas les pneus du PIB, on va les compter lorsqu'ils seront vendus à un constructeur automobile par le producteur de pneus et une seconde fois lors de la vente des véhicules.

Cette règle admet cependant une exception : si les pneus sont stockés, ils sont considérés comme une consommation finale pour la période courante. Ils seront déduits du PIB au moment où ils seront déstockés.

- produits sur le territoire : c'est la production des *unités résidentes* qui compte. Un belge qui travaille au Luxembourg ne fait pas augmenter le PIB belge mais le PIB luxembourgeois. En revanche, un luxembourgeois qui travaille en Belgique fait augmenter le PIB belge.⁸

- valeur monétaire : comme le PIB doit agréger des biens et services très différents, il faut les exprimer dans une unité commune. Cette unité commune est une unité monétaire (dollar, euro etc.) et elle permet d'additionner des pommes, des poires, des machines outils etc.

On mesure donc la valeur des biens grâce à leur prix. C'est pourquoi on parle de *PIB nominal*. En général si on parle de PIB, on a en tête le PIB nominal. Il existe pourtant d'autres PIB, comme nous le verrons dans la suite de cette section.

Le PIB mesure la production totale d'un pays. Il dépend donc de sa taille, ce qui ne facilite pas les interprétations. C'est pourquoi on utilise souvent le *PIB par habitant*, appelé aussi *PIB par tête*, ou *PIB per capita*. Il s'agit simplement du PIB divisé par le nombre d'habitants.

⁸ Ceci reste vrai même si chacun réside dans son pays d'origine car c'est le lieu de versement du salaire qui compte.

2) Les trois modes de calcul du PIB

Il existe trois façons de calculer le PIB qui aboutissent toutes au même nombre. Elles reposent chacune sur une autre optique de la mesure de la richesse créée pendant une période donnée.

1^{ère} optique : optique de la production

C'est l'application directe de la définition du PIB que nous venons de commenter. On cherche à mesurer la richesse produite dans l'économie.

Dans ce cas, le PIB est simplement la somme des valeurs ajoutées :

$$PIB = \sum VA$$

La *valeur ajoutée* d'un bien correspond à la différence entre la valeur du bien ou du service et la valeur des biens intermédiaires (inputs matériels) qui ont été nécessaires pour le réaliser.

ex : si un livre vaut 10€ et qu'il a fallu 1,50€ de papier et 50 centimes d'encre pour l'imprimer, la valeur ajoutée de ce livre vaudra $10 - 1,50 - 0,50 = 8€$

Calculer le PIB à partir de la somme de valeurs ajoutées permet d'éviter de compter plusieurs fois les produits intermédiaires. Dans l'exemple, on aurait surestimé la richesse créée en additionnant les prix du livre, du papier et de l'encre, puisque le prix du livre inclut ceux du papier et de l'encre.

2^{ème} optique : optique de la dépense

Le PIB mesure le revenu d'une année, mais ce revenu va aussi être dépensé. Or, au niveau du pays dans son ensemble, la dépense va être égale au revenu. On peut donc mesurer le PIB indirectement en mesurant les dépenses.

Une autre façon de présenter la même idée consiste à dire qu'on va mesurer la richesse au moment où elle va être utilisée.

Dans les deux cas on peut dire que le PIB va être égal à la somme des dépenses ou des utilisations. On dit que le PIB est égal à la somme des *demandes finales*.

$$PIB = \sum Demandes\ finales$$

Encore faut-il définir ce que sont les demandes finales. Il s'agit de la consommation, des investissements, des dépenses publiques, et des exportations auxquelles on soustrait les importations.

Pour s'en convaincre, il suffit de se demander ce que pourrait faire de sa production une économie productrice uniquement de véhicules automobiles. Les véhicules pourraient être achetés par des consommateurs (consommation), des entreprises (investissement), le gouvernement (dépenses publiques), ou bien être exportés (exportations).

Cependant, on peut aussi imaginer que certains véhicules utilisés proviennent de l'étranger (importations). Ces véhicules n'auraient pas été fabriqués dans le pays et ne feraient donc pas partie de la richesse créée dans le pays. C'est pourquoi on soustrait les importations des demandes finales.

Enfin, certains véhicules peuvent avoir été stockés pour être vendus à l'avenir. Si on ne les incluait pas dans les demandes finales, on oublierait une partie de la richesse produite.

A l'inverse, certains véhicules vendus pendant l'année courante peuvent avoir été produits au cours de l'année précédente. Si on les incluait dans les demandes finales, on surestimerait la production de l'année courante. On doit donc les retrancher des demandes finales.

Par conséquent, on ajoute la variation des stocks aux demandes finales. Une autre façon de le justifier est de considérer l'accumulation des stocks comme un investissement et la diminution des stocks comme un désinvestissement.

Le PIB calculé par l'optique de la dépense est donc égal à la somme de la consommation finale, de l'investissement (ou formation brute de capital fixe), de la variation des stocks et des exportations nettes.

3^{ème} optique : optique des revenus

Lorsque la richesse a été produite, elle doit être distribuée. Une troisième façon de la mesurer consiste donc à calculer la somme des revenus des facteurs de production (salaires, revenus du capital...).

$$PIB = \Sigma \text{revenus des facteurs}$$

Plus précisément le PIB sera la somme des rémunérations des salariés versées par les unités résidentes (revenus du travail) et des excédents bruts d'exploitation des unités résidentes (revenus du capital), auxquels on ajoutera les impôts liés à la production et on soustraira les subventions d'exploitation (revenus nets de l'Etat liés à la production).

NB : Les trois mesures doivent donner exactement le même résultat.

Si on reprend l'exemple du livre en supposant que c'est la seule production de l'économie pendant la journée, on peut dire que la somme des valeurs ajoutées est égale à celle du livre (8€), du papier (1,5€) et de l'encre (0,50€). La somme des valeurs ajoutées est donc de 10€

Il n'y a qu'une seule consommation finale, le livre (10€).

Enfin, le libraire aura gagné $10 - 1,50 - 0,50 = 8€$, le marchand de papier 1,5€ et le marchand d'encre 0,50€. En tout, les revenus des facteurs s'élèveront à 10€

Les trois méthodes donnent donc bien le même résultat : 340 milliards d'euros pour la Belgique en 2010.

Le PIB belge en 2010 (en millions d'euros)

Optique de la production		Optique de la dépense		Optique du revenu	
VA brute	303366	Consommation privée (ménages et ISBLSM)	177789	Rémunération des salariés	179275
+ Impôts sur les produits	37894.5	Consommation finale des administrations publiques	83681	Excédent d'exploitation et revenu mixte	124034
- Subventions sur les produits	-2098.5	FBCF	72269	Impôts sur la production et les importations	44119.8
		- Variation des stocks	-3868	-Subventions	-8266.8
		Exportations nettes	9291		
PIB	339162	PIB	339162	PIB	339162

Source : Eurostat.

Ce que nous venons de calculer et qui apparaît dans le tableau ci-dessus est ce qu'on appelle le *PIB au prix du marché*. En effet, nous avons utilisé les prix de vente des biens sur le marché pour calculer les valeurs ajoutées.

On calcule aussi un *PIB au coût des facteurs*. Pour l'obtenir, on soustrait les impôts et on réintègre les subventions au *PIB au prix du marché*. Ainsi :

$$\text{PIBcf} = \text{PIBpm} - \text{impôts indirects} + \text{subventions}$$

Le PIBcf est une meilleure mesure de la production que le PIBpm parce qu'il est insensible aux variations des prélèvements et subventions publics. En augmentant les impôts

indirects, l'Etat fait augmenter le prix des biens, ce qui augmente le PIBpm sans augmenter la quantité de biens produite. Le PIBcf reste quant à lui constant dans les mêmes circonstances.

En 2010, il s'élevait à 303,366 milliards d'euros (source Banque Mondiale).

3) L'identité comptable fondamentale

A partir de la définition du PIB, on peut déduire une *identité comptable*, c'est-à-dire une égalité qui sera par construction toujours vraie. Cette identité est d'une importance colossale en macroéconomie car elle doit être respectée par tous les modèles macroéconomiques lors du bouclage. Tout modèle qui ne la respecterait pas serait forcément incohérent. C'est pourquoi on l'appelle l'*identité comptable fondamentale*.

Nous allons commencer par étudier l'identité comptable fondamentale d'une économie fermée et sans Etat. Il n'existe qu'une seule économie réellement fermée, c'est l'économie mondiale, et il n'y a pas de pays sans Etat. Pourtant cette première étape est utile parce qu'elle est pédagogique et parce que nous serons amenés à utiliser un certain nombre de modèles d'économie fermée et sans Etat (ex : modèles de croissance).

Partons de la définition du PIB par la dépense. Dans une économie fermée et sans Etat, le PIB est utilisé soit pour la consommation soit pour l'investissement.

Dorénavant, nous noterons Y le PIB, C la consommation agrégée et I l'investissement. On peut alors écrire :

$$Y \equiv C + I \quad (1)$$

Cette égalité est une définition, d'où le signe « \equiv ». Elle est donc toujours vraie.

On va compléter cette première égalité par une autre qui définira l'épargne. Par analogie avec le cas d'un consommateur individuel, l'épargne (S) est définie par la différence entre le revenu et la consommation (C). Or on sait que le PIB (Y) est aussi égal à la somme des revenus. Par conséquent :

$$S \equiv Y - C \quad (2)$$

Comme il s'agit aussi d'une définition, cette deuxième égalité est, elle aussi, toujours vraie. Par conséquent, si on combine les deux définitions, on obtiendra une expression toujours vraie.

On remplace alors Y par sa valeur dans (2) :

$$S = (C + I) - C$$

$$\Rightarrow \boxed{S = I} \quad (3)$$

On obtient alors l'identité comptable fondamentale en économie fermée sans Etat. Comme elle résulte de la combinaison de deux égalités toujours vraies, elle est aussi toujours vraie.

Elle traduit le fait que la production est répartie entre les consommateurs et les entreprises. Ce qui n'est pas consommé, donc épargné, peut être investi et vice versa.

Une fois cette identité comptable comprise, on peut par le même raisonnement l'étendre au cas d'une économie ouverte avec un Etat, qui est beaucoup plus réaliste.

On part de la définition du PIB par la dépense qui donne à présent :

$$Y \equiv C + I + G + NX \quad (4)$$

Où G représente les dépenses publiques et NX les exportations nettes des importations, c'est-à-dire les exportations moins les importations.

On définit toujours l'épargne comme la différence entre le revenu et la consommation. Cependant, il faut à présent soustraire un prélèvement important sur le revenu brut avant d'obtenir l'épargne : les impôts. Si on note les impôts T , on peut écrire :

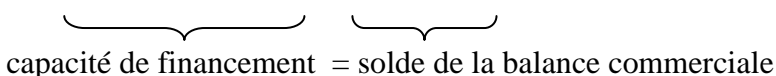
$$S = Y - C - T \quad (5)$$

Il n'y a plus qu'à remplacer le revenu par sa valeur dans l'expression de l'épargne :

$$S = (C + I + G + NX) - C - T \quad (6)$$

En simplifiant et en réarrangeant, on obtient une identité comptable intéressante :

$$\boxed{(S - I) + (T - G) = NX} \quad (7)$$



On peut interpréter cette identité comme traduisant le fait que les productions qui ne sont pas utilisées à l'intérieur des frontières du pays peuvent être envoyées à l'étranger. En d'autres termes, l'excédent commercial est égal à la somme de l'épargne privée et de l'épargne publique.

A l'inverse, si l'excédent budgétaire est négatif, c'est-à-dire si on observe un déficit budgétaire, et si l'épargne privée ne le compense pas, on observera aussi un déficit commercial. Il est donc fort probable qu'un déficit budgétaire soit accompagné d'un déficit commercial et réciproquement.

C'est pourquoi on parle des *déficits jumeaux* ou « twin deficits ». L'expression avait connu un certain succès dans les années quatre-vingts lorsque les Etats-Unis avaient vu augmenter parallèlement leur déficit budgétaire et leur déficit commercial.

La relation n'est cependant pas systématique puisqu'il faut tenir compte de l'épargne privée. Par exemple, dans les années quatre-vingts la Belgique concédait un déficit public très élevé. Elle ne connaissait pourtant pas de déficit extérieur parce que son épargne privée était suffisamment élevée.

Il faut se garder pour l'instant d'expliquer un déficit par un autre. L'identité comptable fondamentale n'est qu'une relation comptable. Elle est toujours vraie mais ne dit rien sur l'origine des déficits ni sur la causalité qui les relie. Tout ce que nous pouvons dire est que l'identité sera toujours respectée *ex post*.

Si on veut expliquer l'évolution des différentes composantes de l'identité, on doit disposer d'une théorie. Pour l'instant nous n'en disposons pas, mais les prochains chapitres nous permettront d'y remédier. Toutes les théories que nous verrons respecteront l'identité comptable fondamentale. En revanche, elles pourront fournir des explications contradictoires sur les variables qui vont s'ajuster pour y parvenir.

B- La comparaison des PIB dans le temps et l'espace

Jusqu'à présent, nous avons calculé le PIB pour une période donnée et un pays donné. Cela nous a permis d'observer comment il était utilisé et réparti. On pourrait souhaiter aller plus loin dans l'analyse. En particulier, il serait instructif de pouvoir décrire son évolution au cours du temps, pour mesurer la croissance. Il serait aussi intéressant de pouvoir comparer les PIB de différents pays, ne serait-ce que pour avoir une idée des écarts entre la Belgique et ses partenaires européens ou entre pays développés et en développement ;

Ces deux comparaisons, dans le temps et dans l'espace, sont impossibles avec le PIB nominal que nous avons défini dans le (A), et pour la même raison. Le PIB agrège des quantités de biens différents en les pondérant par leur prix. Or les prix évoluent au cours du temps. De même, ils diffèrent d'un pays à l'autre.

Par conséquent, lorsqu'on compare des PIB nominaux, on ignore si les différences observées sont dues à des différences de quantités ou de prix. Ceci est vrai pour les comparaisons dans le temps et dans l'espace. Nous allons voir à présent les deux astuces qui permettent de contourner ces difficultés, en nous concentrant sur l'aspect intertemporel (1) avant de passer plus succinctement sur l'aspect géographique (2).

1) Le PIB réel

Lorsque l'on souhaite comparer les PIB de deux années différentes, on souhaite comparer la richesse produite pendant ces deux années. Or entre ces deux années, les prix ont forcément changé. Cela peut être dû à l'inflation et/ou à des variations de prix relatifs, mais une tonne de blé reste une tonne de blé quel que soit son prix.

Il faut donc trouver un moyen de neutraliser les mouvements de prix. Pour ce faire, le plus simple est de partir d'un exemple simple.

Supposons qu'une économie produise uniquement du blé et de l'acier. Le tableau suivant décrit l'évolution des quantités et des prix entre 2009 et 2010 :

	$P_{\text{blé}}$	$Q_{\text{blé}}$	P_{acier}	Q_{acier}
2009	200 €	2 t	100 €	3 t
2010	300 €	1 t	250 €	4 t

On calcule le PIB nominal pour les deux années :

$$\text{PIB nominal}_{2009} = 200 \times 2 + 100 \times 3 = 700$$

$$\text{PIB nominal}_{2010} = 300 \times 1 + 250 \times 4 = 1300$$

Le PIB nominal a donc augmenté d'une année sur l'autre. Il a même presque doublé. Pourtant cette augmentation est trompeuse. En effet, la production de blé a été divisée par deux et la production d'acier n'a augmenté que d'un tiers.

L'augmentation du PIB nominal est donc largement due à l'augmentation très importante des prix.

Pour se faire une véritable idée de l'évolution des quantités produites, on doit neutraliser l'inflation. On va pour cela définir le *PIB réel*.

La méthode est simple, on choisit une année de référence, l'*année de base*, et on utilise les prix de cette année pour les appliquer aux quantités produites pendant les autres années. On obtient ainsi le PIB réel de chaque année au prix de l'année de base.

PIB réel, en volume, ou à prix constants : production de biens et services valorisée aux prix de l'année de base.

Les PIB réels des différentes années sont à présent comparables puisque les différences entre eux ne sont plus dues qu'à l'évolution des quantités produites. L'astuce consistait à fixer les prix à une date arbitraire.

Calculer le PIB réel pour les deux années :

$$\text{PIB réel}_{2009} = 200 \times 2 + 100 \times 3 = 700$$

$$\text{PIB réel}_{2010} = 200 \times 1 + 100 \times 4 = 600$$

On constate que le PIB réel de 2009 est égal au PIB nominal de la même année. Ce résultat est toujours vérifié pour l'année de base puisqu'on utilise les mêmes prix pour le PIB nominal et le PIB réel.

On remarque par ailleurs que le PIB réel de 2010 est bien inférieur au PIB nominal de 2009. Il est surtout inférieur au PIB réel de 2009. La production a donc diminué. On voit que l'évolution du PIB nominal surestime l'évolution de la production, à cause de l'évolution des prix.

Plus généralement, l'augmentation du PIB nominal surestime toujours celle du PIB réel en période d'inflation. Elle la sous-estime en période de déflation.

Pour calculer le taux de croissance entre deux années, on doit donc absolument toujours utiliser le PIB réel. Par commodité, on se contente de parler du taux de croissance du PIB, mais il faut entendre taux de croissance du PIB réel.

Dans l'exemple, le taux de croissance du PIB est donné par :

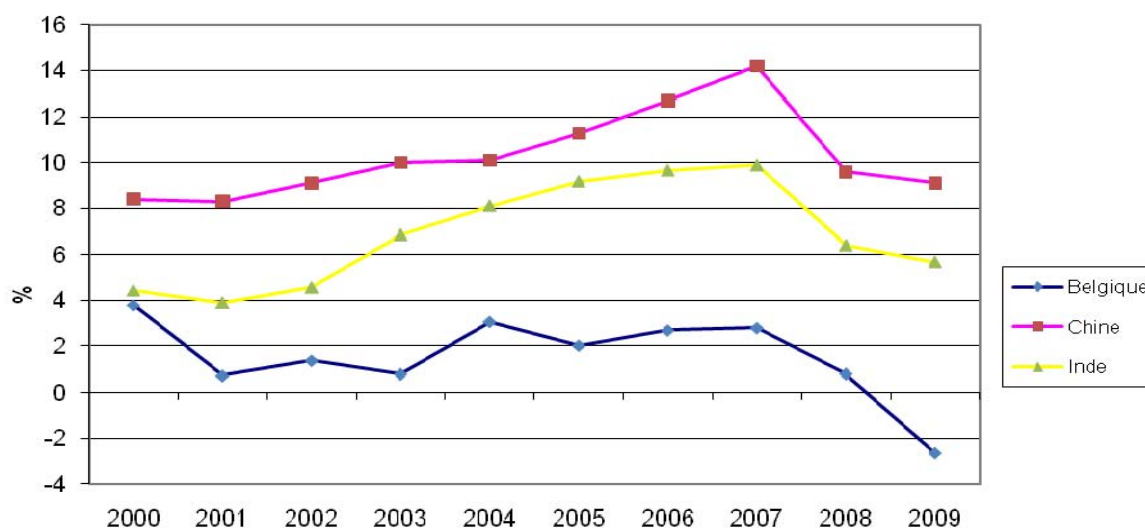
$$\begin{aligned} \text{Taux de croissance} &= 100 \times (\text{PIB réel}_{2010} - \text{PIB réel}_{2009}) / \text{PIB réel}_{2009} \\ &\cong -14,29 \% \end{aligned}$$

Le taux de croissance est négatif, ce qui traduit une diminution de la production du pays en termes réels.

En Belgique, ces dix dernières années (2000-2009), le taux de croissance (du PIB aux prix de 2000) a fluctué entre 3,80% en 2000 et -2,65% en 2009.

On trouve les taux de croissance les plus élevés dans certains pays en développement qui sont en train de décoller, comme la Chine ou l'Inde. Le graphique ci-dessous illustre l'évolution du taux de croissance dans ces trois pays. Bien qu'ils fléchissent tous les trois en fin de période en raison de la récession mondiale, on constate qu'ils sont toujours plus élevés en Chine et en Inde qu'en Belgique.

Taux de croissance du PIB réel : 2000-2010



Source : Banque Mondiale.

Lorsque l'on a calculé le PIB réel et le PIB nominal, on peut en déduire le *déflateur du PIB*.

$$\text{déflateur du PIB} = \frac{\text{PIB nominal}}{\text{PIB réel}}$$

Comme le PIB nominal mesure la production aux prix courants alors que le PIB réel utilise les prix de l'année de base, le déflateur mesure l'évolution des prix entre les deux années. Le déflateur est donc une mesure de l'inflation.

La définition du déflateur implique que le PIB nominal est égal au produit du déflateur et du PIB réel. On a donc réussi à décomposer l'évolution du PIB nominal entre une composante qui reflète l'évolution des prix et une autre qui reflète l'évolution des quantités produites.

On peut de la même façon obtenir le PIB réel en divisant le PIB nominal par le déflateur.

Dans l'exemple, le déflateur pour 2010 est égal à : $1300 / 600 \cong 2,167$

Pour l'année de base (2009), il est égal à un par construction.

Nous illustrerons l'évolution du déflateur en Belgique quand nous verrons comment on peut mesurer l'inflation, dans le paragraphe III de cette section.

2) Le PIB à parité des pouvoirs d'achat

Dans une large mesure, les problèmes qui se posent pour comparer les PIB dans le temps se posent également lors des comparaisons dans l'espace. En effet, les prix peuvent différer d'un pays à l'autre.

ex : l'immobilier est plus cher à Paris qu'à Bruxelles.

De même, une coupe de cheveux est beaucoup plus chère à Bruxelles qu'à New Delhi.

Dans le cas des comparaisons entre des pays qui n'utilisent pas la même monnaie, une complication supplémentaire apparaît. Pour convertir les deux PIB dans une unité commune, on utilise un *taux de change*, qui est simplement le prix d'une monnaie exprimé dans une autre.

Or les taux de change fluctuent de façon importante et sans qu'on sache toujours pourquoi. Cela peut venir fausser de façon significative les comparaisons internationales de revenu.

ex : Depuis son lancement en janvier 1999, l'euro a fluctué entre une valeur de 0.8324 USD/euro le 27 octobre 2000 à 1.5940 USD/euro le 23 avril 2008, ce qui équivaut à une appréciation de plus de 90 pourcents. Cette variation est sans commune mesure avec les écarts de croissance entre l'Europe et les Etats-Unis qui sont de l'ordre du point de pourcentage.⁹

Par conséquent, en comparant les PIB nominaux, on obtient une idée faussée des différences de revenu d'un pays à l'autre parce que les différences observées proviennent à la fois de différences dans les quantités produites, dans les prix, et dans les taux de change.

Pour résoudre ce problème, on doit encore faire preuve d'astuce. Une solution est d'utiliser la même série de prix pour calculer tous les PIB. La solution la plus simple serait d'utiliser les prix américains ou européens, mais cela accorderait trop d'importance à la structure des prix relatifs d'un seul pays. La solution retenue est de faire la même chose, mais à partir de moyennes internationales de prix.

Cette solution revient à calculer le taux de change d'une monnaie fictive qui respecte la *parité des pouvoirs d'achats*. Le taux de change de cette monnaie fictive est tel qu'elle permette d'acheter la même quantité de biens dans tous les pays. On obtient ainsi un dollar PPA ou un euro PPA.

⁹ L'évolution du taux de change de l'euro vis-à-vis de nombreuses autres devises est disponible sur le site de la BCE, à l'adresse suivante : <http://www.ecb.int/stats/exchange/eurofxref/html/eurofxref-graph-usd.en.html>.

Ce travail est réalisé par différents organismes internationaux avec des méthodes et des résultats légèrement différents. Cependant, les ordres de grandeur sont respectés.¹⁰

On a parfois des surprises.

En 2009, le PIB nominal par habitant, s'élevait à 45934 dollars aux Etats-Unis et à 39740 dollars au Japon (données FMI).

En revanche, le PIB PPA par habitant s'élevait à 45934 dollars aux Etats-Unis, par construction le même chiffre que précédemment, et à 32554 dollar au Japon (données FMI).

L'utilisation du PIB nominal donnait donc l'impression que le revenu japonais était plus proche du revenu américain qu'il ne l'était en réalité : 86% contre 71% du PIB par tête américain. Outre le cours des deux monnaies, la différence est aussi due au fait que les prix sont plus élevés au Japon qu'aux Etats-Unis.¹¹

Il faut garder à l'esprit que des écarts de PIB très faibles ne sont en général pas significatifs. Il y a trop d'approximations dans les calculs et d'imprécisions dans les données. En revanche, les ordres de grandeur sont parlants lorsque l'on compare des pays très différents :

PIB par habitant nominal et PPA rapportés à celui des Etats-Unis (2009)

	PIB par habitant en dollars courants		PIB par habitant en dollars PPA	
	En niveau	Par rapport aux USA (en %)	En niveau	Par rapport aux USA (en %)
Etats-Unis	45989	100	45989	100
Belgique	43671	94.96	36249	78.82
Mexique	8143	17.71	14335	31.17
Corée	17078	37.14	27168	59.08
Ethiopie	344	0.75	934	2.03

Source : Banque Mondiale.

Compte tenu de l'imprécision des données, les organisations internationales préfèrent présenter leurs résultats sous forme de classement. L'OCDE par exemple distingue ainsi les pays à revenu élevé (Etats-Unis), élevé-moyen (Belgique), moyen-bas (Corée), et bas (Russie ou Mexique).

¹⁰ On trouvera un commentaire plus détaillé de la construction des PIB PPA dans le chapitre 3 de Blanchard (2004), à la page 40.

¹¹ L'impression peut être faussée davantage encore. Ainsi, en 1999, le PIB nominal par habitant s'élevait à 34700 \$ au Japon et 31900 \$ au Etats-Unis. En revanche, le PIB PPA par habitant s'élevait à 23800 \$ au Japon et à 30900 \$ aux Etats-Unis (données Banque Mondiale). Le PIB nominal donnait donc à tort l'impression que le Japon était plus riche que les Etats-Unis.

L'un des objectifs que nous poursuivrons dans la deuxième partie du cours sera d'expliquer ces écarts colossaux de revenu par habitant.

3) D'autres mesures du revenu

Au lieu du PIB, on entend souvent parler du PNB ou Produit National Brut.¹² Les deux agrégats sont très proches l'un de l'autre et sont parfois utilisés indifféremment dans la discussion courante. Il existe cependant une différence de taille entre les deux qui apparaît en comparant les deux définitions :

Produit national brut ou *PNB* : valeur monétaire de l'ensemble des biens et services finaux produits par les facteurs de production nationaux pendant une période donnée.

Ici, ce qui compte est la nationalité des détenteurs des facteurs de production et non le lieu de leur activité. Un belge qui travaille au Luxembourg fait augmenter le PNB belge mais pas le PNB luxembourgeois. De même, un luxembourgeois qui travaille en Belgique ne contribue qu'au PNB luxembourgeois.

Depuis l'adoption des nouvelles normes de comptabilité nationale dans l'Union Européenne (SEC95), le *Revenu National Brut* a remplacé le *Produit National Brut*. En toute rigueur, on devrait donc aujourd'hui parler de *Revenu National Brut*. Le changement de nom souligne que quelques conventions ont changé mais ces changements ne modifient ni le principe général de construction du PNB, ni sa relation avec les principaux autres agrégats. Surtout, les vieilles habitudes sont difficiles à perdre. On continue donc de parler à la fois de PNB et de RNB, ces deux termes étant presque synonymes.¹³

On obtient facilement le PNB (RNB) à partir du PIB. Il suffit d'y ajouter les revenus perçus dans le reste du monde par des facteurs de production nationaux et d'y soustraire les revenus versés aux facteurs de production du reste du monde.

$$\begin{aligned} \text{PNB (RNB)} &= \text{PIB} + \text{revenus des facteurs reçus du reste du monde} \\ &\quad - \text{revenus des facteurs versés au reste du monde} \end{aligned}$$

¹² On trouvera un commentaire plus détaillé de la distinction entre PNB et PIB dans le chapitre 15 de Blanchard (2004), à la page 327.

¹³ Le même changement s'est opéré en dehors de l'Union Européenne. Dans les statistiques internationales le Gross National Product (PNB) a été remplacé par le Gross National Income (RNB). Le Gross Domestic Product (PIB) existe toujours.

La relation entre le PIB et le RNB en Belgique en 2006
(en millions d'euros)

Produit intérieur brut		316621.8	316621.8
Revenus perçus du reste du monde	Rémunération des salariés reçue du reste du monde	5434.5	56145.3
	Subsides reçus du reste du monde	792.4	
	Revenus de la propriété reçus du reste du monde	49918.4	
Revenus payés au reste du monde	Rémunération des salariés payée au reste du monde	1474.4	52276
	Impôts sur la production et les importations payés au reste du monde	2078.0	
	Revenus de la propriété payés au reste du monde	48723.6	
Revenu national brut		320491.1	320491.1

Source : INS 2007.

On constate que la différence entre le PIB et le PNB belge est très faible, autour de 1% (1.222). Cela vient du fait que les revenus perçus du reste du monde et versés au reste du monde s'équilibrent.

Ce n'est pas le cas pour tous les pays : les pays de forte émigration tendent à avoir un PNB supérieur à leur PIB parce qu'une grande partie de leurs nationaux travaillent à l'étranger (ex : Etats-Unis, Japon). De même les pays qui possèdent un nombre important d'entreprises à l'étranger auront un PNB supérieur à leur PIB (ex : Etats-Unis, Japon).

A l'inverse, les pays sur le territoire desquels opèrent des entreprises détenues par des étrangers auront un PIB supérieur à leur PNB. (ex : Pakistan).

Différence entre le PNB et le PIB en 2009

	PIB en millions de dollars courants	PNB en millions de dollars courants	(PIB – PNB)/PNB
Etats-Unis	14 119 000	14 011 000	0.77%
Japon	5 068 996	5 228 304	-3.05%
Pakistan	161 989	166 370	-2.63%

Source : Banque Mondiale World Development Indicators.

On peut par ailleurs avancer que le PIB surestime la richesse réellement disponible pendant une année. En effet, une partie du stock de capital s'est usée et a vieilli. Une partie de

Les notes du cours ne sont transmises qu'à titre indicatif.

la production totale a donc dû être consacrée au remplacement du capital devenu hors d'usage. C'est ce que mesurent les *amortissements*. En retranchant les amortissements du PIB, on obtient le *Produit Intérieur Net* :

$$\text{Produit Intérieur Net} = \text{PIB} - \text{Amortissements}$$

Le produit intérieur net mesure donc la richesse créée nette de celle qui a été détruite au cours de la production. C'est donc une meilleure mesure du revenu d'un pays. C'est pourquoi on l'appelle également *revenu national*.

Il faut se méfier de l'utilisation du PIB ou du PNB par habitant pour mesurer le bien-être d'un pays. Le PIB présente en effet plusieurs défauts qui amènent à l'interpréter avec circonspection :

- le PIB par habitant est une moyenne et peut cacher des disparités importantes au sein de la population.

- le PIB ne prend pas en compte la pollution. Si la production d'un bien polluant augmente (ex : essence), le PIB va augmenter non seulement à cause de cette production mais parce qu'elle va engendrer une série de dépenses provoquées par les effets de la pollution. (soins médicaux, accidents, dépollution des plages etc.).

- Le PIB ne tient pas compte de l'augmentation du temps de loisir et la diminution du temps de travail. A PIB par habitant égal, le bien-être sera pourtant supérieur dans un pays dont les habitants disposent de plus de temps libre.

- Le PIB ne tient pas compte de l'utilisation des ressources naturelles qui peuvent pourtant s'épuiser. (mines, forêts, biodiversité etc.).

Pour pallier ces lacunes, on a développé des mesures alternatives :

L'*Indice de Développement Humain* (IDH)¹⁴ est un indice composite qui dépend certes du revenu par habitant mais tient compte de la décroissance de l'utilité marginale du revenu, de l'espérance de vie, du taux d'alphabétisation et du taux de scolarisation.

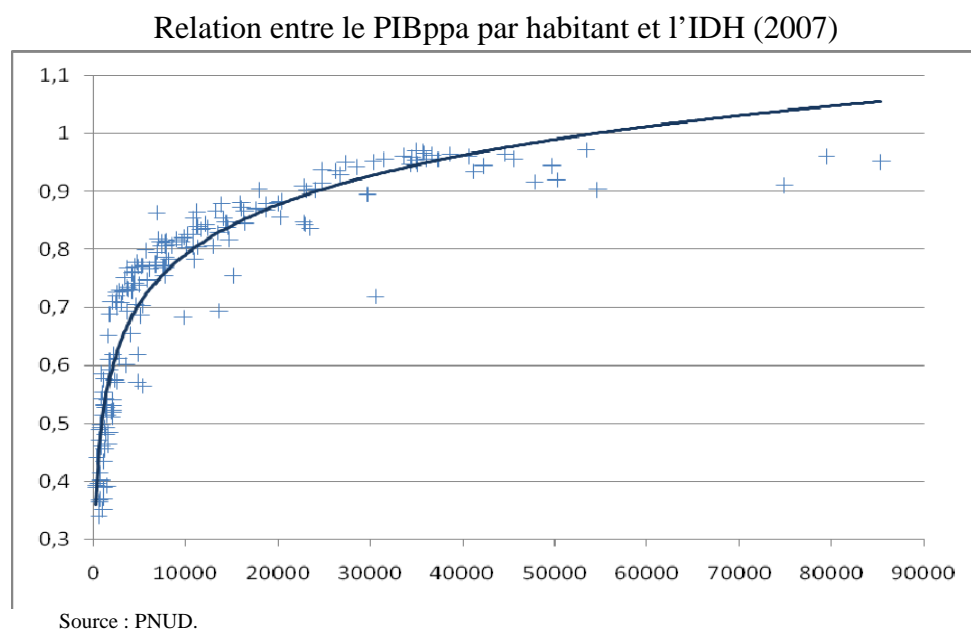
Bien que le classement des pays en fonction de leur IDH soit globalement proche de celui obtenu à partir de leur PIB par tête, on observe des différences.

¹⁴ Cet indice a été notamment développé par l'économiste indien Amartya Sen (prix Nobel 1998).

ex : Le sultanat d'Oman a un PIBppa par habitant qui est presque quatre fois plus élevé que celui de Cuba (22816 \$ppa contre 6876 \$ppa pour Cuba en 2008). Pourtant, l'IDH de Cuba est supérieur à celui d'Oman (0,863 contre 0,846 en 2008).

La différence s'explique par le fait que Cuba compense son PIB plus faible par des efforts de scolarisation, dans le système de santé et par une répartition moins inégalitaire des revenus.

On constate cependant que la relation entre l'IDH et le PIBppa par tête est très nette, surtout en dessous d'un certain niveau de revenu.



Pour améliorer la contribution des ressources naturelles à la production on a proposé l'idée d'un *PIB vert* ou *écologique* qui les intégrerait.

Les Etats-Unis ont ainsi introduit des *comptes écologiques*, ou *comptes verts*, dans leur comptabilité nationale en 1994. Pour l'instant ces comptes ne mesurent que la contribution du sous-sol et pas encore des ressources renouvelables. Ils ont cependant déjà permis de montrer que les découvertes de gisements avaient compensé les extractions des années 50 à 90.

III- L'inflation

La lutte contre l'inflation est un objectif de la politique économique. C'est particulièrement le cas en Europe où la stabilité des prix était un des critères de participation à l'union monétaire fixés en 1992 par le traité de Maastricht. Le même traité impose au

Système Européen de Banques Centrales (la Banque Centrale Européenne par abus de langage) de poursuivre cet objectif en priorité sur tous les autres.

La BCE a chiffré cet objectif en indiquant juste avant le lancement de l'euro qu'elle chercherait à atteindre un taux d'inflation inférieur à 2% par an. Mais comment mesure-t-on cet objectif ? C'est ce que nous allons voir dans le premier paragraphe de cette sous-section en décrivant la construction de l'*indice des prix à la consommation* ou *IPC* (A).

Nous verrons alors qu'on pourrait également utiliser un agrégat que nous avons déjà défini, le déflateur du PIB, et nous comparerons les résultats obtenus avec les deux mesures (B).

A- L'indice des prix à la consommation

Il est très facile de mesurer l'évolution d'un seul prix. Cependant, lorsqu'on souhaite décrire l'évolution du niveau des prix dans une économie, on retrouve les mêmes problèmes d'agrégation que pour mesurer les quantités.¹⁵ Comment résumer l'évolution du prix de biens différents, consommés en quantités différentes, tout en essayant de coller le plus possible à la réalité de l'évolution du pouvoir d'achat des ménages ?

La solution réside dans la construction d'un indice de prix. Celle-ci se fait en trois étapes :

- on définit le panier du consommateur moyen, c'est-à-dire la répartition de son budget entre les différents biens et services qu'il consomme.

- on calcule le prix de ce panier de biens, à intervalle régulier (mois, année...).

- on calcule l'indice des prix en rapportant le prix courant du panier à celui d'une année de base.

La méthode semble très simple mais il faut garder à l'esprit que l'économie est composée de milliers de biens. L'exercice est donc lourd.

Pour bien comprendre la méthode, nous allons calculer le prix du panier dans un exemple très simple. On suppose que le consommateur moyen consomme 4 baguettes de pain et 2 litres d'essence chaque année. L'évolution du prix de ces deux biens est donnée dans le tableau suivant :

¹⁵ Il y a une difficulté en moins pour les prix, parce qu'ils sont tous exprimés dans la même unité.

	Prix de la baguette	Prix de l'essence	Prix du panier de biens	IPC	Taux d'inflation
2008	1 €	2 €	8 €	100	-
2009	2 €	3 €	14 €	175	75 %
2010	3 €	4 €	20 €	250	43 %

Pour calculer l'IPC, on commence par calculer le prix du panier de biens au cours des trois années.

$$P_{\text{panier}2008} = 4 \times 1 + 2 \times 2 = 8 \text{ €}$$

$$P_{\text{panier}2009} = 4 \times 2 + 2 \times 3 = 14 \text{ €}$$

$$P_{\text{panier}2010} = 4 \times 3 + 2 \times 4 = 20 \text{ €}$$

On choisit alors une année de base et on rapporte le prix du panier de biens de chaque année à celui de l'année de base. C'est à ce moment là qu'on obtient l'indice des prix proprement dit. Ici on a choisi 2008 comme année de base :

$$IPC_{2008} = 100 \times P_{\text{panier}2008} / P_{\text{panier}2008} = 100$$

$$IPC_{2009} = 100 \times P_{\text{panier}2009} / P_{\text{panier}2008} = 175$$

$$IPC_{2010} = 100 \times P_{\text{panier}2010} / P_{\text{panier}2008} = 250$$

Comme l'indice des prix est un rapport entre deux variables mesurées dans la même unité, il n'a lui-même pas d'unité.

Il faut éviter les abus de langage. L'indice des prix ne mesure pas directement l'inflation. Il faut pour cela franchir la dernière étape et calculer le taux de variation annuel de l'IPC. Comme nous ne disposons que de données pour trois années, nous ne pouvons calculer que deux taux d'inflation :

$$\text{inflation}_{2009} = 100 \times (IPC_{2009} - IPC_{2008}) / IPC_{2008} = 75 \%$$

$$\text{inflation}_{2010} = 100 \times (IPC_{2010} - IPC_{2009}) / IPC_{2009} \cong 43 \%$$

Pour résumer, on peut définir le taux d'inflation de la façon suivante :

Taux d'inflation : taux de variation de l'IPC sur une période donnée.

Cet exemple est évidemment simplifié. L'indice des prix à la consommation belge est en fait calculé depuis 2004 à partir d'un panier de biens, regroupés en douze rubriques dont les pondérations sont indiquées dans le tableau ci-dessous.

Composition du panier de biens de l'indice des prix à la consommation belge, base 2004

Rubrique	Pondération	Exemples de produits de base
1. Produits alimentaires et boissons	19,23	Pommes Jonagold, Pommes golden, Pommes granny
2. Tabac	1,05	Cigarettes courtes, Cigarettes longues
3. Articles d'habillement et articles chaussants	6,24	Pull-over hiver, Pull-over été (homme puis femme)
4. Logement, eau, électricité, gaz et autres combustibles	15,70	Loyers sociaux, Loyers non sociaux, Consommat° d'eau
5. Ameublement, équipement ménager et entretien courant de la maison	7,28	Table de cuisine, Sommier à lattes, Miroir
6. Dépenses de santé	4,25	Consultation dentaire, Obturations de cavités
7. Transports	15,62	Essence sans plomb 98 RON ("superplus"), Essence sans plomb 95 RON ("eurosuper")
8. Communications	3,66	Communications téléphoniques (poste fixe), Communications GSM
9. Loisirs et culture	12,35	Match de football : 1re provinciale, Match de football : 1re division (nationale)
10. Enseignement	0,58	Université : minerval et droit d'inscription aux examens, Haute école : minerval et droit d'inscription aux examens
11. Hôtels, cafés et restaurants	7,03	Steak au poivre, Salade niçoise, Sole ou truite meunière
12. Autres biens et services	7,02	Dentifrice, Assurance incendie, Pompes funèbres, Crémation

Source : Institut National de la Statistique.

Comme les produits consommés peuvent évoluer rapidement, la Commission de l'Indice a décidé en 2004 qu'on pourrait dorénavant ajuster tous les deux ans la liste des produits contenus dans chaque catégorie. On peut donc ajouter ou supprimer des produits, ou encore modifier leur définition.

L'indice base 2004 est entré en vigueur en 2006. Il a donc déjà subi deux ajustements, l'un en 2008 et l'autre en 2010. Le nombre de produits recensés est donc passé de 507 en 2006, à 518 en 2008, puis 520 en 2010.

Lors de l'ajustement de 2010, on a par exemple ajouté les lecteurs MP3/MP4, mais supprimé les lecteurs de CD portables. On a aussi ajouté les ordinateurs portables à la définition des ordinateurs.¹⁶

Dans les statistiques des pays européens, vous rencontrerez probablement le sigle « *IPCH* » qui correspond à *Indice des Prix à la Consommation Harmonisé*. Cela signifie

¹⁶ On trouvera une description complète des deux ajustements sur le site du ministère des finances, à l'adresse : http://statbel.fgov.be/fr/binaries/actu2010_fr_tcm326-86545.pdf.

simplement que les pays européens ont harmonisé leurs méthodes de calcul de l'IPC pour rendre leurs taux d'inflation directement comparables. Le principe de construction de l'IPCH est donc rigoureusement celui que nous venons de voir.

Bien que sa construction soit en principe simple, l'IPC admet trois lacunes principales :

- problème de *substitution* : l'IPC repose sur l'hypothèse que le panier de biens représentatif est invariant. Or, si les prix relatifs évoluent, les consommateurs vont modifier la répartition de leur budget en substituant aux biens dont le prix a augmenté des biens moins coûteux. La composition du panier de biens va donc s'éloigner de celle de la consommation effective parce que les consommateurs se tourneront de préférence vers les biens les moins chers.

- problème des *biens nouveaux* : la composition du panier étant en principe fixée une fois pour toutes, il ne peut prendre en compte les biens nouveaux, alors que l'évolution de leur prix peut être sensiblement différente de celle des autres biens.

- problème de *l'évolution de la qualité* des biens : grâce au progrès technique, la qualité des biens tend en général à s'améliorer (ex : ordinateurs). On remplace sans s'en rendre compte des biens anciens par des biens de meilleure qualité. La composition du panier de biens représentatif n'est donc pas constante dans le temps.

Ces trois problèmes créent un biais dans le même sens : l'IPC tend à surestimer l'inflation. C'est ce qu'on appelle l'*effet Boskin*.

Michael Boskin est l'économiste qui a dirigé la commission sur l'indice des prix à la consommation chargée par le sénat américain d'estimer le biais dans la mesure de l'inflation dans les années nonante. Le comité a conclu que la surestimation de l'inflation s'élevait environ à un point de pourcentage du taux d'inflation. Un taux de variation de l'IPC de 1% correspondrait ainsi à une inflation nulle.

Pour pallier ces inconvénients, les instituts chargés de mesurer l'inflation modifient la composition des paniers de biens environ tous les dix ans.

Par ailleurs, on essaye de tenir compte de l'évolution de la qualité des biens grâce à la méthode des *prix hédoniques*.¹⁷ Cette méthode consiste à corriger l'augmentation de la qualité des biens, en tenant compte de ce que les consommateurs sont prêts à payer pour l'écart de qualité.

¹⁷ « *Hedonè* » signifie plaisir en grec. On cherche implicitement à tenir compte de l'utilité procurée par les biens.

ex : supposons que le prix des ordinateurs soit resté le même entre 2007 et 2006 mais que tous les ordinateurs de 2007 soient équipés d'un graveur de DVD intégré.

Supposons qu'on ait mesuré que les consommateurs seraient prêts à payer 10% de plus pour un ordinateur équipé d'un graveur de DVD que pour le même ordinateur sans graveur de DVD.

Cela signifie que le prix des ordinateurs de 2007 a en fait baissé de 10% par rapport à 2006.

Ces méthodes ne sont pas complètement satisfaisantes car si la composition du panier de biens change trop souvent, l'indice perd son sens. De plus la méthode des prix hédoniques ne peut s'appliquer à tous les biens parce qu'on ne dispose pas toujours d'une mesure objective de l'amélioration de la qualité (ex : soins médicaux).

B- La mesure de l'inflation grâce au déflateur du PIB

Compte tenu des lacunes de l'IPC, il est utile de le compléter par une autre mesure de l'inflation. C'est une des utilisations du déflateur du PIB. Pour mémoire, le déflateur du PIB, défini dans le II de cette section, est le rapport entre le PIB nominal et le PIB réel.

Comme l'IPC, le déflateur est un indice et n'a donc pas de sens en soi. En revanche, sa variation donne aussi une indication de l'évolution des prix. On doit donc s'attendre à ce que les évolutions du déflateur et de l'IPC soient très proches l'une de l'autre. Il existe cependant deux différences importantes entre les deux indices :

- le déflateur concerne l'évolution de tous les prix des biens et services produits sur le territoire alors que l'IPC n'intègre que le prix des biens consommés par les consommateurs présents sur ce territoire.

L'IPC n'intègre donc pas le prix des commandes publiques (matériel militaire, infrastructures etc.) alors que le déflateur le fait.

A l'inverse, l'IPC intègre le prix des biens importés (essence) alors que le déflateur ne le fait pas.

- on a vu que la composition du panier de biens sur lequel est calculé l'IPC n'est que rarement révisée.

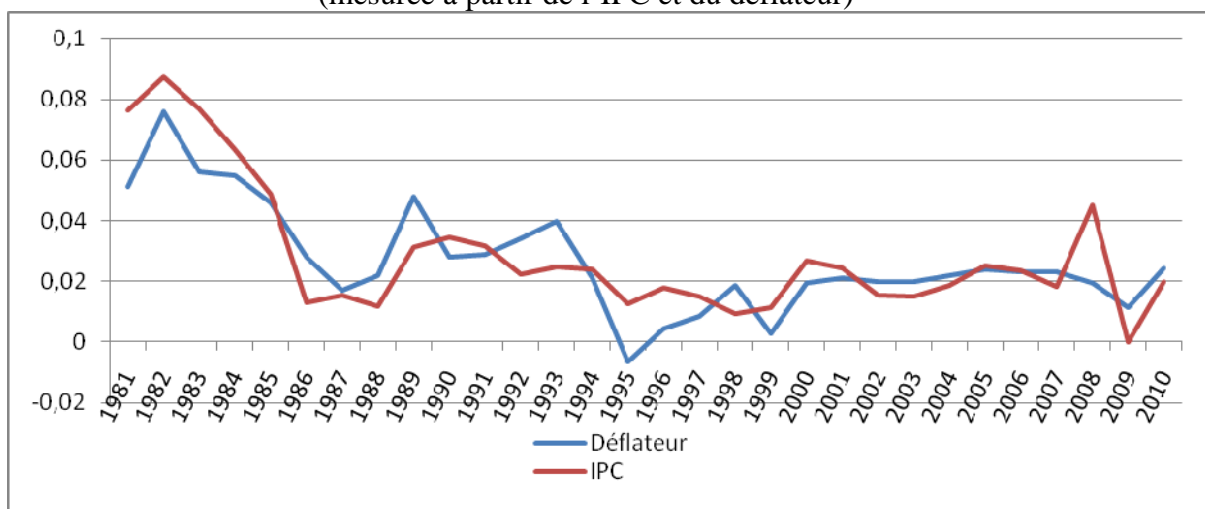
En revanche, la composition du PIB s'ajuste en permanence aux quantités produites. Le déflateur est donc calculé à partir de quantités qui s'adaptent en permanence.

Par conséquent, les taux d'inflation mesurés à partir de l'IPC et du déflateur se ressemblent mais ne coïncident pas.

Le graphique suivant compare l'évolution de l'inflation mesurée par l'IPC et le déflateur en Belgique, entre 1981 et 2006.

On voit que l'inflation mesurée par l'IPC ressemble beaucoup à celle obtenue à partir du déflateur. On repère cependant une exception majeure au début des années quatre-vingts. L'inflation mesurée par l'IPC y dépasse très nettement celle que donne le déflateur. Cette période est de plus marquée par une très forte inflation (presque 9% avec l'IPC en 1982).

L'évolution des prix en Belgique de 1981 à 2010
(mesurée à partir de l'IPC et du déflateur)



Source : FMI.

Ce phénomène est typique d'un choc pétrolier dans un pays importateur de pétrole. On l'aurait d'ailleurs retrouvé dans les autres pays européens au même moment et aussi dans les années septante si le graphique était allé suffisamment loin. En effet, un choc pétrolier augmente le prix des importations de pétrole, donc de tous les produits dérivés du pétrole. Comme ces produits, comme l'essence, constituent un poste important du budget des ménages, le choc pétrolier affecte très fortement l'IPC.

En revanche, comme le pétrole n'est pas produit sur le territoire national, il n'affecte qu'indirectement le PIB donc le déflateur. C'est pourquoi le déflateur reflète moins le choc pétrolier que l'IPC.

IV- Le chômage

Alors qu'on croyait s'en être débarrassé depuis la fin de la crise des années trente, le taux de chômage en Europe occidentale est redevenu dans les années septante l'un des soucis majeurs des gouvernements. Même si son évolution fait l'objet de commentaires réguliers dans les médias, sa mesure reste floue pour beaucoup de commentateurs.

Pour bien mesurer le chômage, il est nécessaire de préciser quelques concepts tels que ceux de population active, inactive et employée. (A). On verra alors que, même après ces précisions, le chômage reste difficile à mesurer (B).

A- La distinction entre chômage, inactivité et emploi

A tout instant, on peut classer les adultes selon trois catégories :

- ceux qui ont un emploi, c'est-à-dire une activité rémunérée.
- les chômeurs, qui n'ont pas d'activité rémunérée.
- les inactifs, c'est-à-dire ceux qui ont une activité non rémunérée (étudiants, femmes ou hommes au foyer...).

La *population active* regroupe les personnes qui occupent un emploi et les chômeurs. C'est par rapport à la population active qu'on mesure le taux de chômage.

$Taux\ de\ chômage = 100 \times \text{nombre de chômeurs} / \text{population active}.$

Il ne faut pas confondre le taux de chômage avec le taux d'activité de la population qui mesure la part des actifs dans la population adulte, c'est-à-dire en âge de travailler.

$Taux\ d'activité = 100 \times \text{population active} / \text{population adulte}.$

Malgré la simplicité des définitions, il reste délicat de mesurer concrètement le nombre de chômeurs.

B- La difficulté de mesurer le chômage

Pour mesurer le chômage, il est nécessaire de tracer une limite nette entre activité, inactivité et chômage, afin de pouvoir classer les individus dans l'une de ces catégories. Certaines situations sont cependant très floues :

- une personne qui travaille à temps partiel involontaire occupe un emploi mais est touchée partiellement par le chômage.

- un chômeur en formation peut être considéré comme un actif, puisqu'il est chômeur, ou comme inactif, puisqu'il suit une formation.

Il est donc nécessaire de donner une définition opérationnelle du chômage.

Une première définition consiste à considérer comme chômeurs tous les individus qui sont inscrits à l'ONEM (Office National pour l'EMPloi), mais cette définition pose quelques problèmes :

- tous les chômeurs ne sont pas inscrits.
- certains travailleurs au noir sont inscrits.
- les systèmes de prise en charge des chômeurs diffèrent d'un pays à l'autre, ce qui rend délicates les comparaisons internationales.

La définition que nous retiendrons est celle qu'a adoptée le Bureau International du Travail (B.I.T.) en 1982.

Chômeur au sens du B.I.T. : toute personne en âge de travailler, sans emploi, immédiatement disponible, et à la recherche d'un emploi ou en ayant trouvé un qui commence ultérieurement.

Plus précisément, le B.I.T. définit une personne en âge de travailler comme ayant 15 ans ou plus, le fait d'être sans emploi comme celui de ne pas avoir travaillé, ne serait-ce qu'une heure, durant une semaine de référence, et le fait d'être immédiatement disponible comme celui de pouvoir prendre un emploi dans les 15 jours.

Cette définition est précise et permet de clairement discriminer les individus actifs, inactifs et au chômage. Elle permet surtout les comparaisons internationales.

Elle a cependant un inconvénient majeur puisqu'elle nécessite l'organisation d'enquêtes régulières.

En Belgique, c'est l'I.N.S. qui réalise ces enquêtes. Il en publie chaque année les résultats dans l'Enquête sur les Forces de Travail.

Selon cette enquête, la population belge de 15 ans et plus se répartissait de la façon suivante en 2009 :

Population de 15 à 64 ans	7 125 504
Population active	4 768 748
Pop active occupée	4 389 369
Chômeurs BIT	379 378

Cela donne donc un taux d'activité de 66.93% et un taux de chômage de 7,96% en Belgique.

A titre de comparaison, en 2009, le taux de chômage était de 11,90% en Irlande et atteignait 8.2% en Pologne. Le taux de chômage de l'UE27 était de 8.9% en 2008. Il s'élevait à 9.3% aux Etats-Unis (données Eurostat).

La définition du BIT est relativement restrictive. En 2009, l'ONEM recensait 645 700 chômeurs indemnisés (soit presque le double de l'OIT).

Conclusion

Parce qu'elle s'intéresse à l'économie d'un point de vue global, la macroéconomie suppose d'utiliser des données spécifiques, parfois abstraites, les agrégats. Ces données permettent d'évaluer la pertinence des différentes théories macroéconomiques.

L'évolution des faits observés, mais aussi l'évolution des idées, ont été à l'origine d'une évolution importante des théories économiques. Après une période de synthèse dans les années cinquante et soixante, on a vu réapparaître des débats théoriques au sein des macroéconomistes.

Ce chapitre n'a fait que les évoquer. A partir du prochain chapitre, nous allons commencer à étudier plus en détail les théories débattues, les hypothèses sur lesquelles elles reposent et les résultats qui en découlent. Nous commencerons par le long terme, c'est-à-dire les théories de la croissance.

Références

Sur la méthode et l'histoire de la macroéconomie

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 1 et 26.

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 1.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 1.

Samuelson P.A. et W.D. Nordhaus : *Economie*, Economica, Paris, 2000 : chapitre 20.

Sur les agrégats

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 2.

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 2.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 2.

Samuelson P.A. et W.D. Nordhaus : *Economie*, Economica, Paris, 2000 : chapitre 21.

Pour les passionnés

Sur l'évolution de la macroéconomie :

Boncoeur J. et H. Thouément : *Histoire des idées économiques de Walras aux contemporains*, Nathan, Paris, 2000.

Le site « History of Economic Thought » : <http://cepa.newschool.edu/het/home.htm>.

Sur la comptabilité nationale :

Stiglitz, J.E., A. Sen et J.-P. Fitoussi, *Rapport de la Commission sur la mesure des performances économiques et du progrès social*, 2009.

<http://www.stiglitz-sen-fitoussi.fr/fr/index.htm>.

Piriou J.-P. : *La comptabilité nationale*, Collection Repères n°57, La Découverte, Paris, 2004.

Gadrey J. et F. Jany-Catrice : *Les nouveaux indicateurs de richesse*, Collection Repères n°404, La Découverte, Paris, 2007.

Première partie : La croissance

Lorsqu'on utilise le terme de croissance sans plus de précision, on a en tête la croissance de la production, qu'elle soit mesurée par la production totale ou par la production par habitant. On peut mesurer cette croissance sur des intervalles plus ou moins longs. Ainsi, les conjoncturistes commentent-ils les évolutions trimestrielles de la production.

Lorsqu'un macroéconomiste parle de la croissance, et a fortiori de la théorie de la croissance, il utilise le terme dans un sens plus restrictif, dont la définition pourrait être la suivante :

Croissance : augmentation pendant une période longue du produit d'un pays.

La croissance est donc considérée comme un phénomène de long terme. Elle concerne donc les tendances longues. On parlera parfois de « *trends* ». L'ordre de grandeur est au moins celui d'une décennie.

La croissance s'oppose donc aux fluctuations qui vont affecter la production à plus court terme, à l'échelle d'une année voire d'un trimestre.

Certains ajoutent qu'il s'agit d'une augmentation du *produit potentiel*, c'est-à-dire des capacités de production d'un pays. Ils soulignent par là que la croissance concerne l'évolution de ce que l'économie peut produire au maximum de ses capacités, c'est-à-dire au plein emploi.

A l'inverse, les fluctuations sont alors interprétées comme des phénomènes liés au sous emploi des ressources disponibles.

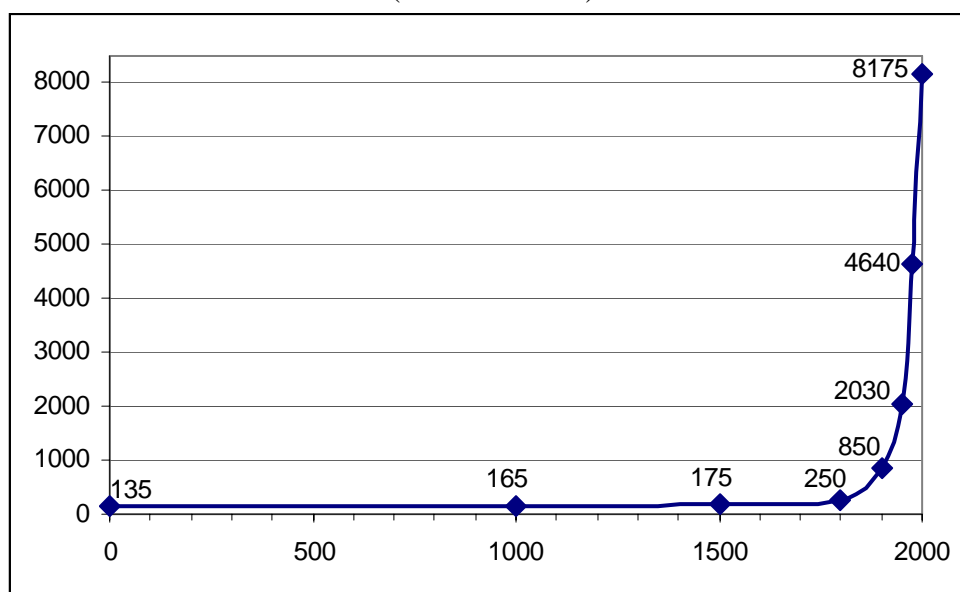
Les théories de la croissance ont par conséquent pour objet de comprendre par exemple pourquoi le niveau de vie d'un européen du XXI^{ème} siècle est en moyenne plus élevé que celui de ses grands-parents, qui eux-mêmes avaient connu un niveau de vie supérieur à celui de leurs grands-parents. (ex : entre 1960 et 2007, le PIB réel par habitant belge a été multiplié par 3,5).

C'est aussi grâce à l'analyse de la croissance sur longue période qu'on peut espérer comprendre pourquoi certains pays sont plus riches que d'autres, ou pourquoi certains pays croissent plus rapidement que d'autres. (ex : de 1960 à 2007, le PIB par tête de Madagascar a diminué de 13% alors que celui du Portugal de 394% et celui de la Corée du Sud de 1251%, selon les estimations des Penn World Tables).

Avant de se lancer dans l'analyse de la croissance, il est utile de se donner quelques idées générales du phénomène, en se donnant quelques faits stylisés.

Le premier fait stylisé est que la croissance est **un phénomène récent**. Ce constat semble paradoxal pour un occidental du XXI^{ème}. On considère en effet que le taux de croissance de 1,3% en Belgique en 2003 était décevant.

Graphique 1
Evolution du PIB par habitant de l'humanité depuis le début de notre ère
(en dollar 2000)



Source : Bradford DeLong d'après Angus Maddison.

Comme le montre le graphique 1, l'humanité est en fait passée par une période de longue stagnation avant de connaître un décollage de son revenu par habitant.

C'est ce que montrent les travaux d'Angus Maddison¹⁸, qui donnent des ordres de grandeur de l'évolution de la croissance en Europe depuis la fin de l'Empire romain et le début du Moyen Age.

Ainsi, le taux de croissance du revenu par habitant a été pratiquement nul pendant la période agraire, courant de l'an 500 à l'an 1500. De 1500 à 1700, pendant la période agraire progressive, le revenu par tête a connu une croissance moyenne de 0,1% par an. Au cours de la période dite du capitalisme commercial, de 1700 à 1820, le taux de croissance du revenu est

¹⁸ A. Maddison, *Les phases du développement capitaliste*, Economica, Paris, 1981.

passé à 0,2%. Ce n'est que lors de la dernière période dite capitaliste (1820-1980) que le taux de croissance moyen s'est approché de ce que nous connaissons.

Ces différences peuvent sembler négligeables, mais elles sont significatives. De petites différences dans les chiffres du taux de croissance annuel se traduisent par des écarts de revenu très importants au bout de quelques années. C'est ce que certains appellent la « force des intérêts composés ».

Pour s'en convaincre, il suffit de réaliser qu'avec un taux de croissance moyen de 1,6% (capitalisme de 1820 à 1980), le revenu par habitant double en 44 ans.

Avec un taux de 0,2 (capitalisme commercial 1700-1820), le revenu par tête double en 347 ans. Au bout de 44 ans, le PIB par tête n'augmente que de 9%.

Le deuxième fait stylisé est que **la croissance n'est pas universelle**.

De 1960 à 2007, le PIB par tête de Madagascar a diminué de 13% alors que celui du Portugal de 394% et celui de la Corée du Sud de 1251%, selon les estimations des Penn World Tables

A part le Venezuela, les pays qui ont vu leur PIB par habitant diminuer sur la période sont surtout des pays africains : Madagascar, Sierra Leone, Angola, Zambie. Dans certains cas, on peut accuser la guerre (Angola) d'être à l'origine du recul du PIB par habitant, mais d'autres pays comme Madagascar n'ont pas connu de conflit majeur.

Les disparités observées dans les taux de croissance se retrouvent dans les niveaux de PIB par habitant.

ex : En 2008, le PIB PPA par tête américain était 51 fois plus élevé que celui de l'Ethiopie. (cf. chapitre 1).

Le troisième fait stylisé est que **ni la croissance ni la richesse ne sont éternelles**. Pendant tout le 1^{er} millénaire de notre ère, la Chine était sans doute le pays le plus riche du monde. A partir du XVI^{ème} siècle, c'est l'Italie du Nord qui a pris sa place, puis les Pays-Bas, le Royaume Uni, et enfin les Etats-Unis, à partir du tournant du XIX^{ème} au XX^{ème} siècle.

En sens inverse, au début du XX^{ème} siècle, sur les 18 pays les plus développés et les plus riches, on comptait 3 pays d'Amérique Latine (Argentine, Uruguay, Chili).

Dans cette partie du cours, nous allons chercher à comprendre les déterminants de la croissance. Pour cela, il faut construire des théories de la croissance. Cela nous permettra ensuite non seulement d'expliquer ce qui s'est passé, mais aussi de tenter des prévisions et de réfléchir aux politiques susceptibles d'influencer la croissance.

Expliquer la croissance, revient à chercher à expliquer pourquoi, en moyenne, vous produirez plus que vos grands-parents lorsque vous aurez commencé à travailler, alors même que votre temps de travail sera le même que le leur, voire inférieur. Intuitivement, on peut suggérer deux explications :

- vous aurez à votre disposition une plus grande quantité de moyens de production. Plus précisément, vous pourrez utiliser plus de capital, parce que la génération de vos grands-parents et celle de vos parents en auront accumulé.

- vous bénéficierez du progrès technique qui est survenu depuis la génération de vos grands-parents. Vous pourrez utiliser des techniques qui n'étaient pas disponibles aux générations précédentes (ordinateurs, engrais plus performants, procédés industriels nouveaux, formes plus efficaces d'organisation du travail etc.)

Les théories de la croissance reposent sur ces deux intuitions et nous allons voir comment. Le prochain chapitre (chapitre 2) montrera jusqu'où on peut aller en s'intéressant à l'accumulation de capital. Dans le chapitre suivant (chapitre 3), nous verrons comment la prise en compte de la technologie est nécessaire pour comprendre la croissance.

Encadré 1 : Trois règles de calcul

Pour analyser les phénomènes de croissance, il est nécessaire de manipuler des taux de croissance. Il suffit pour cela de maîtriser trois règles de calcul simples qui permettent de déterminer facilement le taux de croissance d'une fonction de plusieurs variables dont on connaît le taux de croissance.

- Croissance du *produit de plusieurs variables* : Le taux de croissance d'un produit de plusieurs variables est approximativement égal à la somme des taux de croissance de ses composantes.

ex : Si la production par travailleur croît au taux de 2% par an et que le nombre de travailleurs croît au taux de 3% par an, la production totale croîtra au taux de $2 + 3 = 5\%$ par an.

- Croissance d'un *quotient* : Le taux de croissance d'un quotient est approximativement égal à la différence entre le taux de croissance du numérateur et le taux de croissance du dénominateur.

ex : Si la production totale croît au taux de 5% par an et que le nombre de travailleurs croît au taux de 3% par an, la production par travailleur croîtra au taux de $5 - 3 = 2\%$ par an.

- Croissance d'une variable élevée à la *puissance* α : Le taux de croissance d'une variable élevée à la puissance α est approximativement égal à α fois le taux de croissance de la variable.

ex : Si on estime que la production totale est une fonction du stock de capital du type $Y = K^{0,3}$, et que le stock de capital croît au taux de 2% par an, la production totale croîtra au taux de $0,3 \times 2 = 0,6\%$.

Ces trois règles vont nous permettre de manipuler les modèles de croissance que nous allons passer en revue. Il faut cependant garder à l'esprit qu'elles sont en fait des approximations, d'où l'adverbe « approximativement » qui s'y trouve. Si vous souhaitez les utiliser, sachez que l'approximation est d'autant plus fiable que les taux de croissance sont faibles. Cela signifie que l'erreur de calcul que suppose l'approximation augmente avec les taux de croissance.

Chapitre 2 : La croissance par accumulation de ressources

Introduction : Les origines : David Ricardo

C'est à David Ricardo qu'on doit la première théorie de la croissance fondée sur l'accumulation de capital.¹⁹ Dans sa théorie, la croissance est possible grâce à l'accumulation de capital, motivée par la perspective de profits.

La croissance permet alors l'augmentation de la population, qui nécessite l'augmentation de la production agricole. Mais les rendements sont décroissants dans l'agriculture, et le coût des grains augmente en même temps que le stock de capital. Les salaires vont donc augmenter, ce qui va réduire les profits et ralentir l'accumulation de capital.

Selon le même mécanisme, les profits vont finir par s'annuler. L'accumulation de capital va cesser. La production va alors stagner, de même que la population. L'économie aura atteint un *état stationnaire*.

Dans les années 1950, Robert Solow (Prix Nobel 1987) propose une théorie formalisée de la croissance dont les conclusions peuvent rappeler celles de Ricardo.²⁰ Le modèle de Solow, ou modèle de croissance néoclassique, repose lui aussi sur l'accumulation de capital et aboutit lui aussi à un état stationnaire à cause de l'existence de rendements décroissants. Il ne faut cependant pas confondre les deux théories. Celle de Solow est beaucoup moins large que celle de Ricardo. Il s'agit d'un modèle de croissance économique alors que Ricardo décrivait l'évolution de toute la société.

Quoi qu'il en soit, le modèle de Solow est la base de la réflexion contemporaine sur la croissance et permet de percevoir des intuitions importantes. Nous allons d'abord étudier ses hypothèses (section 1), avant d'en tirer les conséquences pour la croissance (section 2) et pour la relation entre croissance et épargne (section 3).

¹⁹ D. Ricardo : *Principes d'économie politique*, 1817.

²⁰ R. Solow "A contribution to the theory of economic growth", *Quarterly Journal of Economics*, vol 70 n°1, p.65-94, 1956.

Section 1 : Les hypothèses du modèle de Solow

Comme tout modèle, le modèle de Solow repose sur des hypothèses. Il est nécessaire de les avoir à l'esprit pour bien comprendre la logique des résultats du modèle. Les deux séries d'hypothèses les plus importantes concernent l'existence d'une fonction de production agrégée (I) et les décisions d'accumuler du capital (II).

I- La fonction de production agrégée

Le modèle de Solow repose sur l'hypothèse importante qu'on peut résumer le fonctionnement de l'économie par une fonction de production. Nous allons d'abord voir en quoi consiste cette fonction de production agrégée (A) avant de préciser sa forme (B).

A- Une représentation simplifiée de la production

Au cœur de la théorie de la croissance, on trouve la *fonction de production agrégée* ou *fonction de production globale*. Comme toute fonction de production, cette fonction exprime la relation entre des facteurs de production (inputs ou intrants) et une quantité produite (output).

Ici, on considère qu'il y a deux facteurs de production : le capital (K) et le travail (L). Le produit est le produit brut de l'économie que nous notons Y . Comme nous supposons que l'économie est fermée sur elle-même, elle n'a pas de relation avec le reste du monde. Il n'y a donc pas lieu de distinguer PIB et PNB. On peut alors écrire :

$$Y = F(K, L) \quad (1)$$

Analytiquement, il ne s'agit que d'une transposition à l'échelle macroéconomique du concept de fonction de production d'une entreprise de la théorie microéconomique. Cependant, il faut garder à l'esprit qu'on raisonne sur des agrégats :

- la production (Y) agrège la production d'un grand nombre de biens et services différents : produits agricoles, biens intermédiaires, biens de consommation, services etc.
- la population active agrège des travailleurs dont les compétences diffèrent : travailleurs qualifiés et non qualifiés, ingénieurs en informatique et en agronomie, médecins, agriculteurs, menuisiers etc.
- le capital agrège des facteurs de production hétérogènes : machines, bâtiments, terre, véhicules, ordinateurs, téléphones etc.

Cependant, la façon dont nous avons écrit la fonction de production nous oblige à raisonner comme s'il n'existait qu'un produit unique et deux facteurs de production homogènes. Il s'agit là d'une simplification, mais elle va nous permettre d'obtenir des résultats importants, comme le permet une parabole.

Une question importante est de se demander ce qui détermine la fonction de production. Pour faire court, on dit qu'elle est déterminée par *l'état de la technologie*. La technologie doit être comprise au sens large. Elle inclut les techniques de production, mais aussi les modes d'organisation du travail et de la société.

En bref, on appelle technologie tout ce qui permet de produire plus avec la même quantité de facteurs de production (ou autant avec moins de facteurs).

Dans ce chapitre, nous allons supposer que la technologie est donnée. Il n'y a donc pas de progrès et la fonction de production reste la même au fil du temps. Cette hypothèse va nous permettre de nous concentrer sur le rôle de l'accumulation de capital dans le processus de croissance. Nous nous empresserons de la lever dès le prochain chapitre.

B- Forme de la fonction de production

Pour pouvoir utiliser la fonction de production, il faut faire des hypothèses quant à sa forme.

Nous allons d'abord supposer que la productivité marginale des deux facteurs de production est positive. Ainsi, toute augmentation de la population active ou du stock de capital se traduira par une augmentation de la production.

Nous allons être plus précis et supposer que la fonction de production admet des *rendements d'échelle constants*. Cela signifie que si on double ou triple la quantité utilisée de tous les facteurs, la production doublera ou triplera également. Cela est vrai quelle que soit la proportion dans laquelle on augmente les facteurs de production. On pourra donc écrire :

$$F(aK, aL) = aY \quad (2)$$

L'égalité sera vérifiée quelle que soit la valeur de a . On dit que la fonction de production est *homogène de degré un*.

La justification de cette hypothèse est qu'il est raisonnable de penser qu'en dupliquant les facteurs de production, on dupliquerait aussi la production. En d'autres termes, si on doublait la population et le stock de capital de la Belgique, la production de cette « double Belgique » serait exactement égale à deux fois celle de la Belgique actuelle.

Cette hypothèse est importante parce qu'elle signifie qu'un petit pays a la possibilité de produire dans les mêmes conditions qu'un grand pays. Ainsi, il n'y a pas de différence entre les Etats-Unis et le Luxembourg à part leur taille.

La deuxième hypothèse relative à la forme de la fonction de production est celle des *rendements factoriels décroissants*. La production augmente lorsque la quantité d'un seul facteur de production augmente, mais cette augmentation est d'autant moins importante que la quantité initiale du facteur est élevée. Ceci est vrai pour le capital et le travail. On suppose donc à la fois des rendements décroissants du travail et des rendements décroissants du capital.

Les deux hypothèses de rendements d'échelle constants et de rendements factoriels décroissants vont nous permettre de décrire le fonctionnement du modèle sans avoir besoin de choisir une forme spécifique de fonction. La première étape pour cela est de représenter la relation entre la production par travailleur et le stock de capital par travailleur.

La production par travailleur n'est rien d'autre que la production totale divisée par le nombre de travailleurs. On utilise des lettres minuscules pour représenter les variables mesurées en unités par travailleur. On peut alors écrire :

$$y \equiv \frac{Y}{L} = \frac{1}{L} F(K, L) = F\left(\frac{K}{L}, \frac{L}{L}\right) = F(k, 1) \quad (3)$$

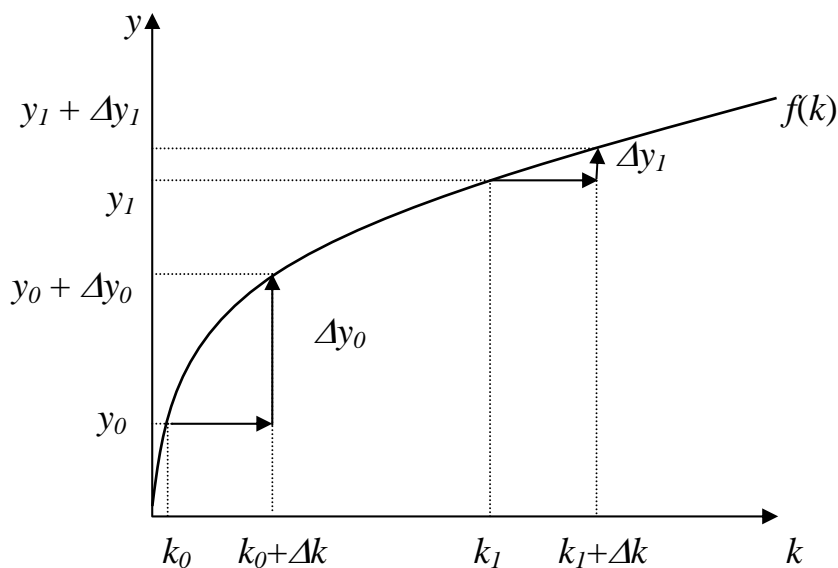
On a utilisé l'homogénéité de la fonction de production pour obtenir ce résultat. Pour vous en convaincre, il vous suffit de réécrire l'expression (2) en remplaçant le paramètre a par $1/L$.

Pour ne pas traîner un paramètre qui serait inutile et encombrant, on va appeler f la fonction F lorsque son deuxième argument est égal à un.

$$F(k, 1) \equiv f(k) \quad (4)$$

Cette notation souligne le fait que la production par travailleur ne dépend que d'une chose : la quantité de capital par travailleur. On peut alors représenter cette relation dans un simple graphique :

Graphique 1 : la fonction de production



La courbe passe par l'origine. Cela semble raisonnable dans la mesure où les travailleurs ne pourraient rien produire sans capital du tout. On constate aussi que la courbe est croissante, ce qui vient de l'hypothèse que la productivité marginale du capital est positive.

Surtout, on constate que la pente de la courbe diminue lorsque le stock de capital par travailleur augmente. Elle est *concave*. Cette propriété lui vient de l'hypothèse de rendements factoriels décroissants. Une même augmentation du stock de capital (Δk) par travailleur fera plus augmenter la production si le stock de capital par travailleur initial est faible que s'il est élevé ($\Delta y_0 > \Delta y_1$).

On voit à présent que si on suppose que la technologie est donnée, la fonction de production sera donnée, et la seule chose qui puisse faire augmenter le revenu par tête est une augmentation du stock de capital. Il faut à présent se demander comment ce stock peut évoluer.

II- L'accumulation de capital

C'est l'investissement qui permet d'accumuler du capital. Par conséquent, si on souhaite décrire la façon dont le capital s'accumule (B), il faut commencer par décrire ce qui détermine la quantité d'investissement (A).

A- La répartition de la production entre investissement et consommation

On sait depuis le premier chapitre que dans une économie fermée et sans Etat, la production peut être utilisée soit pour la consommation soit pour l'investissement. On sait aussi que dans ces conditions l'identité comptable fondamentale implique que l'investissement soit égal à l'épargne. Il faut donc déterminer comment se prennent les décisions d'épargne pour déterminer l'investissement.

Nous allons utiliser pour ce faire une hypothèse très simple : l'épargne est simplement proportionnelle au revenu. On écrira une *fonction d'épargne* de la forme :

$$S = s.Y \quad (5)$$

Ici, S représente l'épargne totale du pays. Le paramètre s mesure le taux d'épargne ou encore la propension à épargner. Comme s est un paramètre, il est indépendant de toutes les autres variables du modèle.

L'hypothèse que le taux d'épargne est constant n'est pas sans importance. Elle implique que les pays riches et les pauvres épargneront la même proportion de leur revenu, ce qui peut sembler irréaliste. Pour l'instant nous nous contenterons de cette hypothèse parce qu'elle est pratique. Nous réfléchirons plus en détail au rôle que peuvent jouer les variations du taux d'épargne, dans la troisième section de ce chapitre.

Comme l'investissement est égal à l'épargne, on peut écrire que l'investissement est lui aussi proportionnel au revenu :

$$\begin{aligned} I &= S = s.Y \\ \Rightarrow I &= s.Y \end{aligned} \quad (6)$$

On a alors tous les éléments en main pour décrire l'évolution du stock de capital.²¹

B- La variation du stock de capital

Le stock de capital est un stock, comme son nom l'indique. L'investissement est un flux. Le stock de capital est donc constitué des investissements passés. L'investissement d'une année vient donc s'ajouter au stock de capital initial.

²¹ On peut remarquer que l'investissement consiste à consacrer une partie de la production à augmenter le stock de capital. Comme la production est homogène, l'investissement est constitué du même type de biens que la consommation. On résume parfois cette hypothèse en disant que le modèle de Solow décrit une économie dans laquelle on produit du blé, avec du travail et du blé.

Est-ce là le seul mécanisme qui fasse varier le stock de capital ? Hélas non. En effet, il faut aussi tenir compte du fait que, chaque année, une partie du stock de capital doit être remplacée. C'est ce qu'on appelle la *dépréciation du capital*. Ce phénomène décrit le fait que les machines utilisées s'usent et vieillissent, et qu'elles finissent par partir à la casse. La quantité de capital qui quitte le service chaque année est mesurée par ce qu'on appelle les *amortissements*.

Pour tenir compte de ce phénomène de dépréciation du capital, nous allons supposer qu'il se fait à un taux constant δ . En d'autres termes, nous allons donc supposer que chaque année les amortissements représentent une part constante δ du capital :

$$Am = \delta.K \quad (7)$$

Nous pouvons à présent dire que l'évolution du stock de capital dépend de deux flux : les investissements et les amortissements. Les investissements font augmenter le stock alors que les amortissements le font diminuer.²²

Le stock de capital au début d'une année sera donc égal au stock de capital du début de l'année précédente auquel s'ajoutera l'investissement de l'année précédente mais se retranchera l'amortissement de l'année précédente. Pour transcrire cela en termes algébriques, on va introduire un indice pour le temps, t , qui numérote les années. On pourra alors écrire :

$$K_{t+1} = K_t + I_t - Am_t \quad (8)$$

Si on remplace l'investissement et les amortissements par leur valeur, on peut réécrire l'expression (8) :

$$\begin{aligned} K_{t+1} &= K_t + sY_t - \delta K_t \\ K_{t+1} - K_t &= sY_t - \delta K_t \\ \Rightarrow \Delta K_t &= sY_t - \delta K_t \end{aligned} \quad (9)$$

Où ΔK_t représente la variation du stock de capital au cours de l'année t , qui est aussi un flux. L'intérêt de l'expression (9) est qu'elle ne fait plus intervenir que des flux.

²² L'évolution du stock de capital est comparable à celle d'une baignoire dans un problème de robinet. L'investissement est comparable au flux d'eau qui coule dans la baignoire alors que l'amortissement est l'équivalent des fuites de la baignoire. Comme dans un problème de robinet, il s'agit de déterminer à quelle vitesse et jusqu'où la baignoire, c'est-à-dire le stock de capital, se remplit ou se vide.

Nous avons pris l'habitude de raisonner en termes de production ou de stock de capital par travailleur. Ce serait dommage d'en changer. Pour traduire l'expression (9) en termes de flux par travailleur, il suffit d'en diviser les deux côtés par le nombre de travailleurs L .

Nous allons supposer que la population est constante. C'est pourquoi L n'est pas suivi d'un indice de temps. L'expression (9) implique alors :

$$\frac{\Delta K_t}{L} = s \frac{Y_t}{L} - \delta \frac{K_t}{L}$$

$$\Rightarrow \Delta k_t = sy_t - \delta k_t \quad (10)^{23}$$

Cette expression a une interprétation très intuitive. En effet, elle signifie que la variation du stock de capital par tête correspond exactement à la différence entre l'investissement total par tête (sy_t) et l'investissement nécessaire à remplacer la fraction du stock de capital par tête qui s'est dépréciée (δk_t). En effet, si on souhaite simplement garder constant le stock de capital par tête, on doit investir tout juste suffisamment pour compenser l'amortissement, qui est égal à δk_t .²⁴

Si on se rappelle que la production par tête est une fonction du stock de capital par tête, l'expression (10) peut encore se simplifier :

$$\boxed{\Delta k_t = sf(k_t) - \delta k_t} \quad (11)$$

Cette équation s'interprète comme la précédente. Elle signifie que la variation du capital est égale à la différence entre l'investissement total et l'investissement de remplacement.

Il s'agit de l'équation centrale du modèle de Solow. Elle a le bon goût d'être très simple, puisqu'elle ne fait intervenir qu'une seule variable et deux paramètres. Elle est fondamentale, puisqu'elle implique que la variation du stock de capital par tête ne dépend *que* du stock de capital par tête. Toute l'évolution de l'économie va donc pouvoir être analysée à partir de celle du stock de capital par travailleur. C'est ce que nous allons faire dans la section suivante.

²³ Pour mémoire : $\Delta K_t = K_{t+1} - K_t$. Par conséquent : $\Delta K_t/L = K_{t+1}/L - K_t/L = k_{t+1} - k_t \equiv \Delta k_t$.

²⁴ En comptabilité nationale, on dit que la formation nette de capital fixe est égale à la formation brute de capital fixe moins les amortissements.

Section 2 : La croissance dans le modèle de Solow

Dans cette section, nous allons décrire comment l'accumulation de capital se traduit en termes de production dans le modèle de Solow. A cette fin, nous allons commencer par nous intéresser à la situation vers laquelle l'économie converge (I) avant de regarder ce qui lui arrive en amont de cette situation (II).

I- L'état stationnaire

L'état stationnaire se définit comme l'état dans lequel plus aucune variable endogène ne varie. En particulier, le stock de capital par travailleur y reste constant. Nous allons baptiser \bar{k} la valeur stationnaire du stock de capital. L'équation d'accumulation (11) nous permet de préciser cette valeur :

$$\Delta \bar{k} = sf(\bar{k}) - \delta \bar{k} = 0$$

$$\Rightarrow \boxed{sf(\bar{k}) = \delta \bar{k}} \quad (12)$$

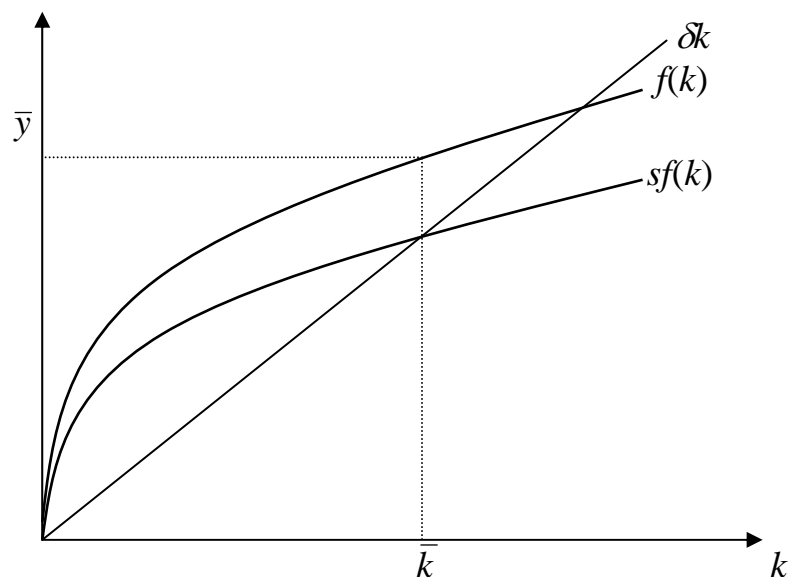
Cette expression montre que le stock de capital stationnaire est déterminé par le fait qu'il permet une production telle que l'épargne va tout juste permettre de compenser les amortissements. Comme l'investissement compense exactement les amortissements, le stock de capital reste constant.

Cette condition peut se traduire graphiquement de façon très simple. L'épargne étant proportionnelle au revenu, elle va se déduire de la fonction de production.

L'amortissement quant à lui va être représenté par une droite passant par l'origine et dont la pente est égale à δ .

L'état stationnaire correspond alors à l'intersection entre la courbe qui représente l'épargne et la droite qui représente l'amortissement.

Graphique 2a : L'état stationnaire



Le graphique 2a permet de constater qu'il y a un état stationnaire et un seul, et qu'il correspond à une seule valeur du stock de capital par travailleur. Il y a donc aussi une seule valeur du revenu par tête.

On peut aussi remarquer que, puisque le stock de capital par tête est par définition constant à l'état stationnaire, la production par tête sera elle aussi constante. L'état stationnaire correspond donc à une situation de croissance nulle.

Ce qui nous assure que l'état stationnaire est unique est la forme de la fonction de production. Comme le rendement du capital est décroissant, la courbe qui représente l'épargne devient de moins en moins raide au fur et à mesure que le stock de capital augmente et finit donc forcément par couper la droite qui représente l'amortissement.

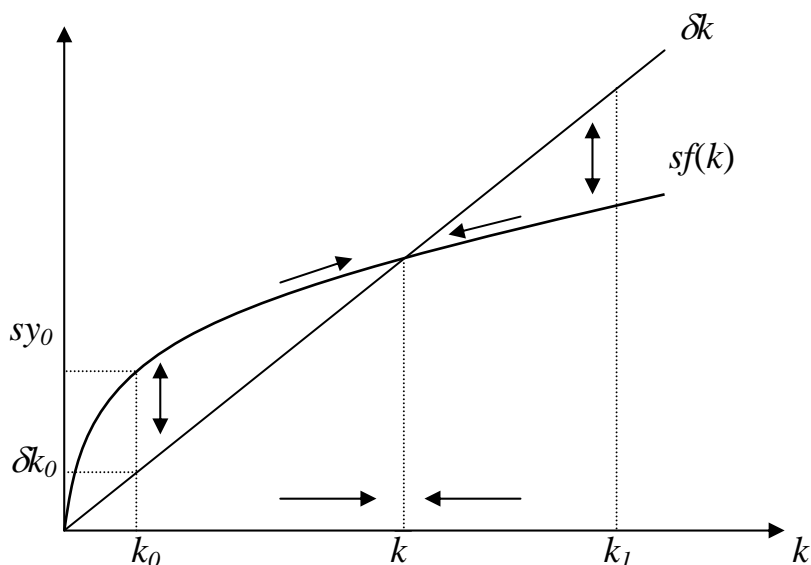
Intuitivement, chaque unité de capital supplémentaire augmente à la fois le revenu, donc l'épargne et l'investissement, et l'amortissement. Cependant, l'amortissement augmente proportionnellement au stock de capital alors que le revenu augmente moins que proportionnellement au stock de capital, à cause des rendements décroissants. L'amortissement finit fatalement par rattraper l'investissement.

La question est à présent de savoir comment on arrive à cette situation. Ceci est l'objet du prochain paragraphe.

II- La convergence vers l'état stationnaire

Supposons qu'une économie démarre avec un stock de capital inférieur à celui de l'état stationnaire (k_0). Quelle va être son évolution ? Le graphique 2 permet de répondre à cette question.

Graphique 2b : la convergence vers l'état stationnaire



On constate qu'en k_0 , l'épargne est supérieure à l'amortissement ($sy_0 > \delta k_0$). Cela signifie donc qu'au cours de l'année le pays va investir plus que ce qui est nécessaire pour compenser l'amortissement. Le stock de capital va donc augmenter et se rapprocher de sa valeur stationnaire \bar{k} . Le même raisonnement s'applique tant que le stock de capital n'a pas atteint sa valeur stationnaire, c'est-à-dire à gauche du point d'intersection entre les deux courbes.

Supposons à présent que le pays a dépassé le stock de capital stationnaire. On voit qu'à présent, l'amortissement est supérieur à l'investissement. Le stock de capital va donc diminuer et se rapprocher de son niveau stationnaire.

On voit donc que toutes les économies, quel que soit leur stock de capital de départ, vont se diriger vers le même état stationnaire. Elles aboutiront donc au même stock de capital par travailleur donc au même revenu par tête. Ce résultat est un des résultats principaux du modèle de Solow. C'est ce qu'on appelle la *convergence* des économies.

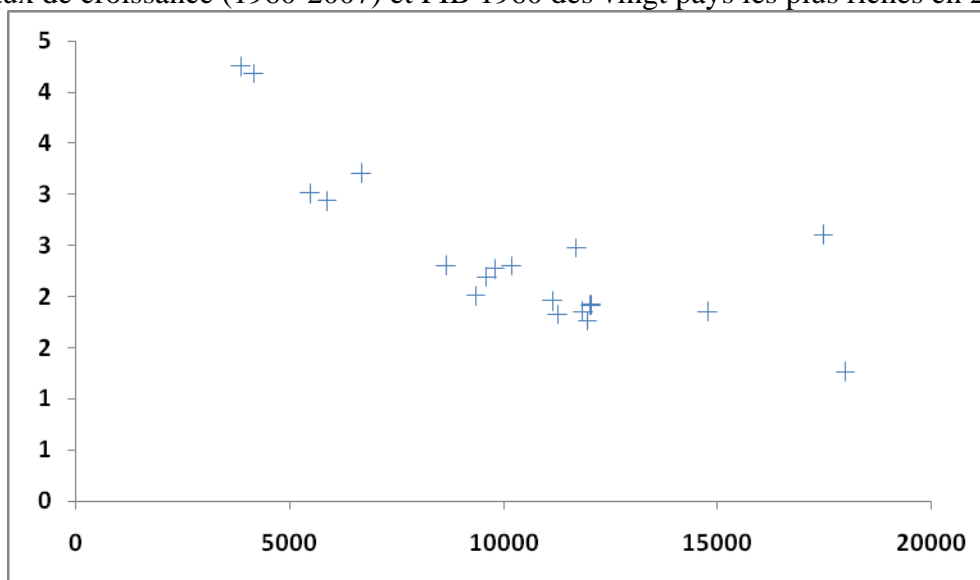
On peut remarquer de plus, que plus le stock de capital est proche de sa valeur stationnaire, moins l'écart entre l'épargne et l'amortissement est élevé. On peut donc conclure que la croissance d'un pays sera d'autant plus lente que son stock de capital sera proche de sa valeur stationnaire.

Ce résultat est un corollaire de la convergence. C'est bien parce que les économies les plus pauvres croissent plus vite que les autres qu'elles peuvent espérer les rattraper un jour.

Une question majeure de la recherche empirique sur la croissance est d'évaluer la pertinence de cette hypothèse de convergence. On peut s'en faire une première idée en jetant un œil au graphique suivant :

Graphique 3

Taux de croissance (1960-2007) et PIB 1960 des vingt pays les plus riches en 2007



Source : Penn world Tables 6.3.

Ce graphique montre que, dans l'échantillon des vingt pays les plus riches en 1960, ceux qui partaient avec un niveau de revenu initial plus faible en 1960 ont connu en moyenne un taux de croissance plus élevé que les autres. On voit nettement une relation négative entre le taux de croissance moyen et le revenu initial. Ce résultat correspond bien à l'hypothèse de convergence du modèle de Solow. Les revenus par habitant des vingt pays qui étaient les plus riches en 1960 ont effectivement convergé.

Nous avons à présent décrit l'état vers lequel tend l'économie dans le modèle de Solow. Il serait intéressant de nous demander quelles sont les variables qui sont susceptibles

de l'affecter. Comme l'épargne est la principale d'entre elles, la prochaine section lui est consacrée.

Section 3 : La relation entre croissance et épargne

Lorsqu'on utilise le modèle de Solow pour réfléchir à la relation entre une variable et une autre, on doit se rappeler que le modèle peut connaître deux états. Le premier est l'état stationnaire. Le second est la convergence vers l'état stationnaire. Or l'économie se comporte différemment dans ces deux états. Il est donc nécessaire de les étudier séparément.

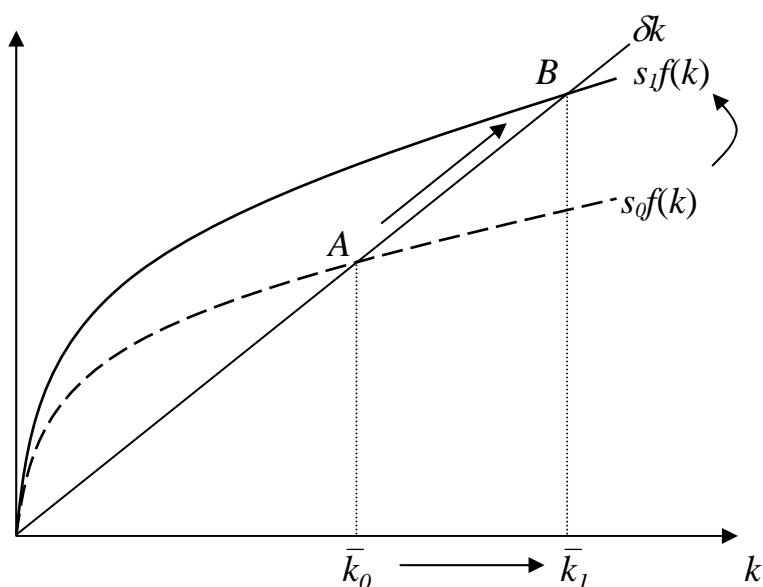
C'est pourquoi nous allons d'abord étudier l'état stationnaire (A) avant de nous placer dans la période de convergence (B).

I- L'absence de relation entre épargne et croissance à l'état stationnaire

Si on se rappelle ne serait-ce que la définition de l'état stationnaire, on obtient directement la réponse à notre question. En effet, en absence de progrès technique, la croissance est forcément nulle à l'état stationnaire. L'épargne ne peut donc pas avoir d'effet sur la croissance à l'état stationnaire.

Graphique 4a

Impact d'une modification de l'épargne sur l'état stationnaire



Les notes de cours ne sont transmises qu'à titre indicatif.

Cependant, une modification de l'épargne doit bien avoir des répercussions sur l'économie. Pour les analyser, on peut utiliser le graphique 4a, en supposant que le taux d'épargne augmente pour passer d'un niveau s_0 à un niveau s_1 . Cette augmentation du taux d'épargne se traduit par un déplacement vers le haut de la courbe d'épargne. Elle coupe désormais la droite qui représente les amortissements en un nouveau point, situé au-dessus et à droite du précédent.

L'augmentation du taux d'épargne se traduit donc par le passage à un nouvel état stationnaire qu'on peut comparer au précédent. Par définition, le taux de croissance y est le même et il est nul.

En revanche, le nouveau stock de capital stationnaire est clairement supérieur au précédent. Cela implique que la production a elle aussi augmenté. On peut donc conclure que l'augmentation du taux d'épargne dans le modèle de Solow se traduit à terme par une augmentation du stock de capital et de la production par habitant.

Ce résultat a une implication pour l'hypothèse de convergence. On peut en effet faire remarquer que deux économies ne convergeront vers le même revenu que si elles ont le même taux d'épargne. Si ce n'est pas le cas, elles convergeront vers des états stationnaires différents. On pourrait tenir le même raisonnement pour deux économies dont les taux de dépréciation sont différents. C'est pourquoi on parle de *convergence conditionnelle*.

Il faut souligner qu'il s'agit ici de la convergence des revenus. A l'état stationnaire, toutes les économies aboutiront à un taux de croissance nul à l'état stationnaire.

On peut se demander comment la consommation est affectée par les variations du taux d'épargne. La réponse n'est pas aussi tranchée que pour la production. En effet, la consommation est soumise à deux effets contradictoires :

- l'augmentation de la production augmente le revenu ;
- mais une part plus importante du revenu est épargnée et investie.

Par conséquent, l'effet total d'une augmentation du taux d'épargne sur la consommation est ambigu. On peut cependant prédire qu'il sera positif si le taux d'épargne initial est faible. En effet, l'état stationnaire initial sera caractérisé par un stock de capital faible. La productivité marginale du capital sera donc élevée. L'augmentation du revenu permise par l'augmentation du stock de capital sera donc importante et compensera l'augmentation du taux d'épargne. L'augmentation du taux d'épargne se traduira alors par une augmentation de la consommation.

En revanche, si le taux d'épargne initial est déjà élevé, le stock de capital initial sera élevé et la productivité marginale du capital très faible. L'augmentation du stock de capital se traduira donc par une très faible augmentation du revenu qui ne suffira pas à compenser l'augmentation du taux d'épargne. L'augmentation du taux d'épargne se traduira alors par une baisse de la consommation.

On peut déterminer le taux d'épargne qui maximise la consommation. On obtient alors ce qu'on a appelé *la règle d'or d'accumulation du capital*. La dérivation et l'interprétation de cette règle d'or nous emmèneraient cependant trop loin, et il vaut mieux les laisser pour la suite de vos études.

Pour l'instant, on peut retenir l'idée qu'il n'est pas optimal d'épargner ni trop ni trop peu. Si on épargne trop peu, l'économie manquera de capital et la production sera faible, ce qui limitera la consommation. Si on épargne trop, on produira beaucoup mais cette production ne servira qu'à financer des investissements et pas à être consommée.

Ces conclusions ne sont cependant valables que si on se place à l'état stationnaire, c'est-à-dire si on laisse le temps à l'économie de s'adapter à son nouveau taux d'épargne. Hors de l'état stationnaire, les choses peuvent être différentes.

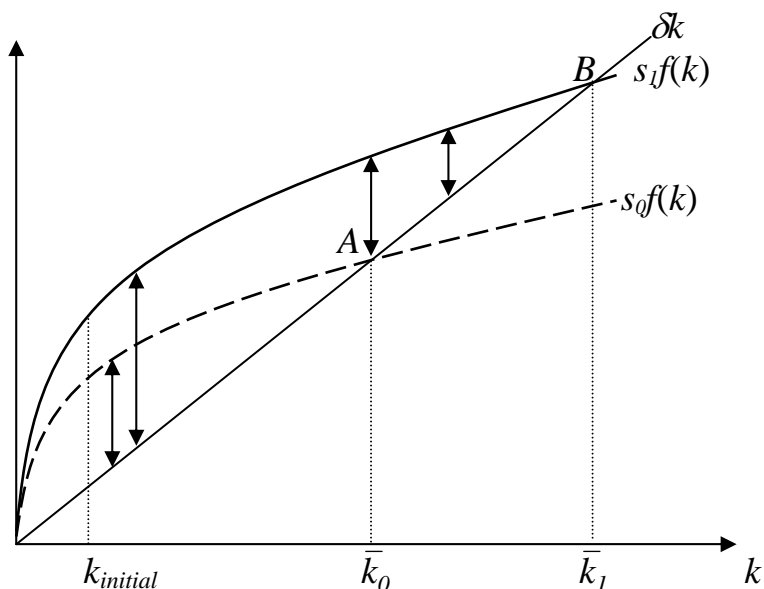
II- Une relation croissante entre épargne et croissance hors de l'état stationnaire

Supposons qu'il existe deux économies absolument identiques et qui disposent au départ du même stock de capital $k_{initial}$, inférieur au stock stationnaire. La seule différence entre elles est leur taux d'épargne. La première connaît un taux d'épargne faible (s_0) alors que la seconde a un taux d'épargne élevé (s_1).

On sait que ces deux économies finiront à des états stationnaires différents où la première aura accumulé moins de capital que la seconde. Mais que peut-on dire de leur taux de croissance en amont de leur état stationnaire ?

Graphique 4b

Impact d'une différence de taux d'épargne sur l'état stationnaire et la croissance hors de l'état stationnaire



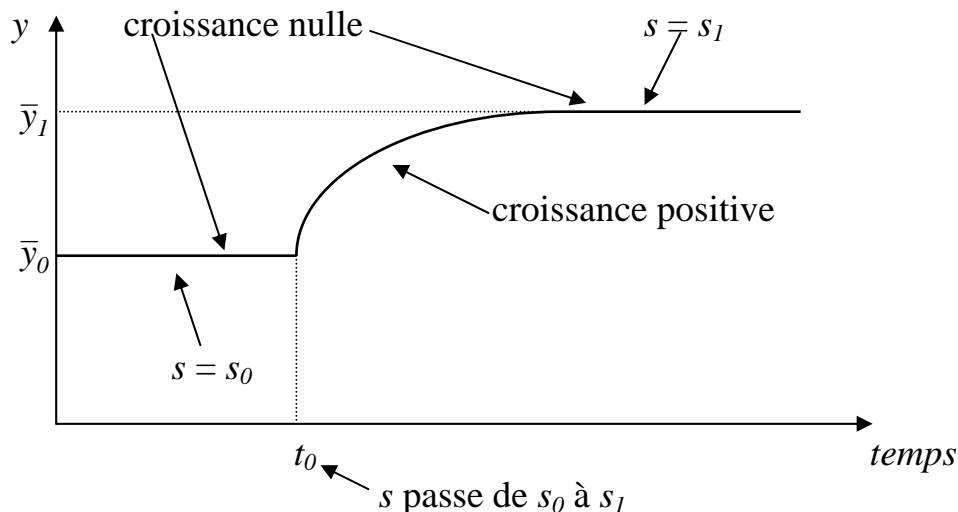
Le graphique permet de voir que l'économie qui épargne le moins sera aussi celle qui accumulera du capital le moins vite. En effet, les deux économies partent avec le même stock de capital, donc la même production et les mêmes amortissements. Celle qui épargne le plus va investir plus que l'autre, ce qui se traduira par une formation nette de capital plus élevée. Comme c'est l'accumulation de capital qui détermine la croissance, on peut dire sans ambiguïté que l'économie qui épargnera le plus croîtra plus vite que l'autre.

On peut donc conclure que, hors de l'état stationnaire, il existe une relation positive entre le taux d'épargne et le taux de croissance.

On peut utiliser le même graphique pour analyser la transition entre un état stationnaire à épargne faible et un nouvel état stationnaire lorsque le taux d'épargne a augmenté (passage du point A au point B).

On voit que lorsque le taux d'épargne augmente, l'investissement devient instantanément plus élevé que l'amortissement. Le stock de capital se met donc à augmenter. Cependant, au fur et à mesure que le stock de capital augmente, la productivité marginale du capital diminue, et l'accumulation de capital se ralentit. Après un saut initial, le taux de croissance diminue pour tendre vers zéro lorsque l'économie se rapproche de son état stationnaire. Cela peut être résumé par un graphique :

Graphique 5
Evolution du revenu lorsque le taux d'épargne augmente



La croissance est nulle au départ puisque l'économie est à l'état stationnaire. Soudain, à la date t_0 , le taux d'épargne augmente. L'accumulation de capital redémarre et la production se met à augmenter. La croissance est positive. Peu à peu cependant, l'accumulation de capital se ralentit et l'économie atteint son nouvel état stationnaire. La croissance redevient nulle.

On peut donc conclure qu'une augmentation définitive du taux d'épargne ne provoque qu'une croissance temporaire.

Conclusion : Un retour sur la convergence

Malgré sa simplicité, le modèle de Solow nous permet d'obtenir plusieurs intuitions importantes sur le processus de croissance :

- l'accumulation de capital est un ressort important de la croissance, mais il ne peut provoquer qu'une croissance temporaire. Les rendements décroissants du capital font converger les économies vers un état stationnaire.

- il existe une relation positive entre le taux d'épargne et le taux de croissance, mais cette relation ne peut être que temporaire. Elle disparaît à l'état stationnaire.

- il existe une relation positive entre le taux d'épargne et le revenu. Cette relation n'implique pas de relation claire entre le taux d'épargne et la consommation.

- toutes les économies devraient converger vers le même taux de croissance (nul). Elles convergeront de plus vers le même niveau de revenu si elles ont le même taux d'épargne et le même taux d'amortissement.

La convergence est l'un des résultats les plus importants du modèle. C'est aussi un résultat plutôt optimiste. En effet, il implique que toutes les économies devraient spontanément aboutir au niveau de revenu des plus riches en croissant d'autant plus vite qu'elles sont initialement pauvres. C'est pourquoi il est important de tester cette conclusion.

Dans la section 2, nous avons vu qu'il existait une relation clairement négative entre le taux de croissance sur la période 1960-99 et le niveau de revenu par habitant, dans l'échantillon des vingt pays les plus riches au début de la période.

Ce résultat est-il généralisable ? Il pourrait être dû à l'échantillon retenu, qui est précisément composé des pays qui partaient avec un revenu relativement élevé puisqu'il s'agissait des vingt pays les plus riches en 1960. Notre observation serait due à un biais de sélection.

On obtient effectivement une impression beaucoup plus mitigée si on utilise un échantillon plus large. Le graphique 6 est identique au graphique 3 présenté dans la deuxième section (2.II), mais représente cette fois tous les pays pour lesquels il existe des données (l'échantillon contient 100 pays).

Graphique 6 : Taux de croissance (1960-2007) et PIB 1960



Source : Penn World Tables 6.3.

Les notes du cours ne sont transmises qu'à titre indicatif.

On constate que la relation négative entre revenu initial et taux de croissance moyen devient beaucoup plus floue. Ce n'est qu'en faisant preuve de bonne volonté qu'on repère encore une telle relation. Certains pays qui partaient avec un revenu très faible ont connu des taux de croissance très faibles, voire négatifs.

Si convergence il y a, elle semble donc se faire au sein de groupes de pays. On parle de *clubs de convergence* pour décrire ce phénomène. Ce qui explique l'existence des clubs de convergence n'est pas encore très clair. La prise en compte de l'évolution de la productivité est une des pistes qui permettent d'en rendre compte. Ce sera l'objet du prochain chapitre.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 3 et chapitre 4 (section 4.1 uniquement).

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 3 (sections 3.1 à 3.3).

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 7 (section 7.1. uniquement. Eventuellement section 7.2. sur la règle d'or).

Samuelson P.A. et W.D. Nordhaus : *Economie*, Economica, Paris, 2000 : chapitre 27 (très basique).

Pour les passionnés

Guellec D. et P. Ralle, *Les nouvelles théories de la croissance*, Repères, La Découverte, Paris, 2003 : chapitre 1 et chapitre 2.

Weil D.N., *Economic Growth*, Pearson Education, Boston, 2005.

Chapitre 3 : Le rôle du progrès technique dans la croissance

Introduction : Les origines : Adam Smith

On peut résumer le progrès en le définissant comme tout ce qui permet d'améliorer le bien-être des consommateurs. Le progrès a donc à la fois une dimension quantitative et une dimension qualitative.

Dans sa dimension quantitative, il permet de produire des quantités de biens plus importantes avec la même quantité de travail. A l'inverse, il permet aussi de produire la même quantité de biens avec moins de travail. La tendance de long terme à la réduction du temps de travail en est une conséquence.

Dans sa dimension qualitative, le progrès permet de produire des biens de meilleure qualité, mais aussi de nouveaux biens qui n'existaient pas auparavant. Il augmente donc la variété des biens disponibles, ce qui augmente le choix offert aux consommateurs.

L'intuition du rôle du progrès technique remonte au moins à Adam Smith²⁵ qui utilise l'exemple d'une manufacture d'épingles pour appuyer son raisonnement. Smith observe que la production par travailleur dans cette manufacture est sans commune mesure avec celle d'un travailleur isolé, grâce à la division des tâches.

Selon Smith, le gain de productivité est obtenu grâce à un gain en « dextérité » de chaque travailleur, à l'économie du temps passé à changer de tâche et à la possibilité d'utiliser des « machines qui facilitent et abrègent le travail ».

On voit bien apparaître l'importance des gains de productivité et du progrès technique.

Dans ce chapitre, nous allons étudier le rôle du progrès technique dans la croissance. Pour ce faire, nous allons commencer par le mesurer (section 1), puis nous l'intégrerons dans le modèle de Solow vu dans le chapitre précédent (section 2), avant d'essayer de comprendre ce qui le détermine (section 3).

²⁵ A. Smith *Recherche sur la nature et les causes de la Richesse des Nations*, 1776.

Section 1 : Une mesure de la part du progrès dans la croissance

C'est encore à Robert Solow qu'on doit la méthode qui permet de quantifier la part du progrès dans la croissance.²⁶ Tout le problème est d'arriver à quantifier le progrès alors qu'on ne peut pas l'observer directement. En pratique, on n'observe en effet jamais que l'évolution de la production et des quantités de facteurs de production utilisées. Il faut alors trouver une astuce pour déduire la part du progrès technique à partir des variables observées.

Nous allons d'abord voir en quoi consiste cette astuce (I) avant de l'appliquer pour apprécier le rôle du progrès technique (II).

I- Construction du résidu de Solow

Lorsqu'on connaît son modèle, on n'est pas étonné d'apprendre que l'astuce proposée par Solow repose sur une fonction de production agrégée. Cependant, on prend à présent explicitement compte du progrès technique en écrivant :

$$Y = A F(K, L) \quad (1)$$

La variable A représente le progrès technique. L'idée est ici que le progrès vient augmenter la productivité des autres facteurs de production. On appelle ce paramètre la *productivité globale des facteurs* ou *productivité totale des facteurs* (PTF). C'est cette PTF qu'on cherche à mesurer.

La méthode proposée par Solow pour mesurer la croissance de la PTF repose sur la fonction de production (1) et l'hypothèse de rendements d'échelle constants. Le principe en est très simple. Supposons que sur une période donnée le stock de capital et la population active aient augmenté dans la même proportion, par exemple 1%. En raison des rendements d'échelle constants, la production devrait avoir également augmenté de 1%.

Supposons que sur la même période le taux de croissance observé ait en fait été de 2%. Cela signifie que les variations du stock de capital et de la population active ne permettent pas d'expliquer la totalité de la croissance, puisqu'ils ne peuvent être responsables que de 1% de croissance. Il reste un point de croissance résiduel qui n'est pas expliqué.

Si on reprend la fonction de production, ce point de croissance supplémentaire ne peut être dû qu'à une seule chose : une augmentation de la productivité. C'est le principe de la

²⁶ R. Solow "Technical change and the aggregate production function", *Review of Economics and Statistics*, vol. 39 n°3, p. 312-320, 1957.

méthode proposée par Solow pour mesurer la croissance de la productivité. On mesure donc la part de la productivité dans la croissance en lui attribuant toute la croissance que l'évolution des autres facteurs de production ne peut expliquer. Voilà pourquoi on parle autant du *résidu de Solow* que de la *croissance de la productivité totale des facteurs*.

Le problème est que les quantités de facteurs de production ne croissent en général pas dans les mêmes proportions. Les choses se compliquent donc puisque les facteurs ne contribuent pas dans les mêmes proportions à la croissance. Il faut donc attribuer à chacun la part de la croissance qui lui revient. On a donc besoin de connaître sa productivité marginale.

C'est là qu'intervient la deuxième hypothèse importante de la méthode de Solow. On va supposer que l'économie est parfaitement concurrentielle. Chaque facteur doit donc être rémunéré en fonction de sa productivité marginale, comme vous l'avez vu dans le cours de microéconomie. A partir de cette hypothèse, on peut montrer que le taux de croissance peut être décomposé pour donner *l'équation fondamentale de décomposition de la croissance* :²⁷

$$g_y = \alpha_K g_K + (1 - \alpha_K) g_L + g_{PTF} \quad (2)$$

g_y , g_L , g_K et g_{PTF} représentent respectivement les taux de croissance du revenu, de la population active, du stock de capital et de la productivité totale des facteurs. Ils sont tous exprimés en pourcents.

Le coefficient α_K est compris entre zéro et un. Il mesure la part du capital dans le revenu, c'est-à-dire rK/Y , où r est la rémunération du capital. $(1 - \alpha_K)$ mesure donc la part du travail dans le revenu. On estime en général que α_K est environ égal à 0,3 et $(1 - \alpha_K)$ à 0,7.²⁸

Or ces parts sont observables, de même que g_y , g_L , et g_K . Cela signifie qu'on peut très simplement déduire la croissance de la productivité à partir de l'équation (2) :

$$g_{PTF} = g_y - \alpha_K g_K - (1 - \alpha_K) g_L \equiv \text{Résidu} \quad (3)$$

C'est cette équation fondamentale qui permet de déduire la part du progrès technique dans la croissance ou résidu de Solow.

²⁷ Une démonstration intuitive de cette décomposition se trouve dans l'annexe du chapitre 4 de Blanchard (2004) et dans le chapitre 27 de Nordhaus et Samuelson (2000). Une démonstration plus rigoureuse en est donnée dans l'annexe du chapitre 4 de Mankiw (2004).

²⁸ Pour la Belgique, ces valeurs ont été de 0,29 et 0,71 en moyenne entre 1980 et 2001. Calculs de M.P. Timmer, G. Ypma et B. van Ark, "IT in the European Union: Driving Productivity Divergence?", Research Memorandum GD-67, Groningen Growth and Development Centre, octobre 2003.

L'expression de résidu est probablement préférable à celle de taux de croissance de la PTF ou progrès technique, parce qu'il s'agit bien ici de la part de la croissance qu'on ne parvient pas à expliquer. C'est par défaut qu'on parle de progrès technique, mais ce résidu peut avoir des origines très diverses : économies d'échelle, éducation... sans oublier les erreurs de mesure. Dans les années cinquante, l'économiste américain Moses Abramowitz a souligné que le résidu était « une mesure de notre ignorance » des causes de la croissance.²⁹

Il faut donc rester prudent dans l'interprétation de ce résidu. Cependant, il fournit la seule estimation disponible de l'évolution de la productivité d'une économie au niveau agrégé. Il permet en outre d'évaluer la pertinence des théories économiques de la croissance en se livrant à une *comptabilité de la croissance*.

Encadré 1 : Démonstration de l'expression (2)

Le point de départ de la démonstration est la fonction de production donnée par l'expression (1) : $Y = A F(K, L)$. Cette fonction nous permet de calculer l'augmentation de la production qui suivrait une augmentation du stock de capital d'un montant ΔK . Il suffit pour cela de multiplier ΔK par la productivité marginale du capital, que nous notons Pm_K . On peut alors écrire :

$$\Delta Y = Pm_K \times \Delta K \quad (\text{E1})$$

Où ΔY mesure tout simplement la variation absolue de la production consécutive à l'augmentation du stock de capital.

L'expression (E1) signifie simplement que si la productivité marginale du capital est de 3 unités de production et que le stock de capital augmente de 2 unités de capital, l'augmentation de la production sera de $2 \times 3 = 6$ unités.

Supposons à présent que ce n'est pas le stock de capital mais la population active qui augmente. On peut alors écrire que la variation de la production consécutive à cette augmentation de la population active est donnée par le produit de la productivité marginale du travail (Pm_L) et de la variation de la population active :

$$\Delta Y = Pm_L \times \Delta L \quad (\text{E2})$$

Enfin, on peut encore déterminer l'effet sur la production d'une augmentation de la productivité mesurée par A , en supposant que la population active et le stock de capital restent constants. Il faut pour cela multiplier la variation de A par la dérivée de la fonction de production par rapport à A . Ainsi :

²⁹ M. Abramowitz "Resource and output trends in the US since 1870", *American Economic Review*, vol 46 n°2, p.5-23, 1956.

$$\begin{aligned}\Delta Y &= \frac{\partial F}{\partial A} \Delta A = \frac{\partial [AF(K, L)]}{\partial A} \Delta A \\ \Delta Y &= F(K, L) \times \Delta A\end{aligned}\quad (E3)$$

Si à présent on considère que les trois paramètres peuvent varier en même temps, on peut dire que la variation totale de la production sera la somme de la variation provoquée par la variation du stock de capital, de celle provoquée par la variation de la population active et de celle provoquée par l'augmentation de la productivité. On obtient cette expression en sommant les membres de droite des expressions (E1), (E2) et (E3) :

$$\Delta Y = Pm_K \times \Delta K + Pm_L \times \Delta L + F(K, L) \times \Delta A \quad (E4)$$

Pour exprimer la variation de la production en pourcentage, on peut diviser les deux côtés de l'expression ci-dessus par la valeur initiale de la production.

$$\frac{\Delta Y}{Y} = \frac{Pm_K}{Y} \times \Delta K + \frac{Pm_L}{Y} \times \Delta L + \frac{F(K, L)}{Y} \times \Delta A \quad (E5)$$

Enfin, il serait préférable d'exprimer la variation en pourcentage de la production en fonction de la variation en pourcentage des trois variables de droite. Pour ce faire, on va utiliser une astuce qui consiste à multiplier le premier terme du membre de droite de l'expression ci-dessus par K/K et le deuxième par L/L . On va également remplacer Y par sa valeur au dénominateur du troisième terme de l'expression.

$$\begin{aligned}\frac{\Delta Y}{Y} &= \frac{Pm_K}{Y} \frac{K}{K} \times \Delta K + \frac{Pm_L}{Y} \frac{L}{L} \times \Delta L + \frac{F(K, L)}{AF(K, L)} \times \Delta A \\ \frac{\Delta Y}{Y} &= \frac{Pm_K K}{Y} \frac{\Delta K}{K} + \frac{Pm_L L}{Y} \frac{\Delta L}{L} + \frac{\Delta A}{A}\end{aligned}\quad (E6)$$

Par définition :

$$\begin{aligned}g_y &\equiv \frac{\Delta Y}{Y} ; \quad g_K \equiv \frac{\Delta K}{K} ; \quad g_L \equiv \frac{\Delta L}{L} ; \quad g_{PTF} \equiv \frac{\Delta A}{A} ; \\ \alpha_K &\equiv \frac{Pm_K K}{Y} \quad \text{et} \quad \alpha_L \equiv \frac{Pm_L L}{Y} .\end{aligned}$$

$$\text{D'où : } g_y = \alpha_K g_K + \alpha_L g_L + g_{PTF}$$

Ou encore :

$$g_y = \alpha_K g_K + (1 - \alpha_K) g_L + g_{PTF} \quad (2)$$

Comme la rémunération d'un facteur est égale à la productivité marginale de ce facteur, le produit de cette productivité marginale par le stock de ce facteur mesure le revenu total de ce facteur. Si on divise le revenu de ce facteur par la production totale, on obtient bien la part de ce facteur dans le revenu national. α_K et α_L mesurent donc respectivement la part du capital et du travail dans le revenu national. Enfin, comme il n'y a que deux facteurs de production dans l'économie, ils se partagent forcément tout le revenu. On peut donc écrire :

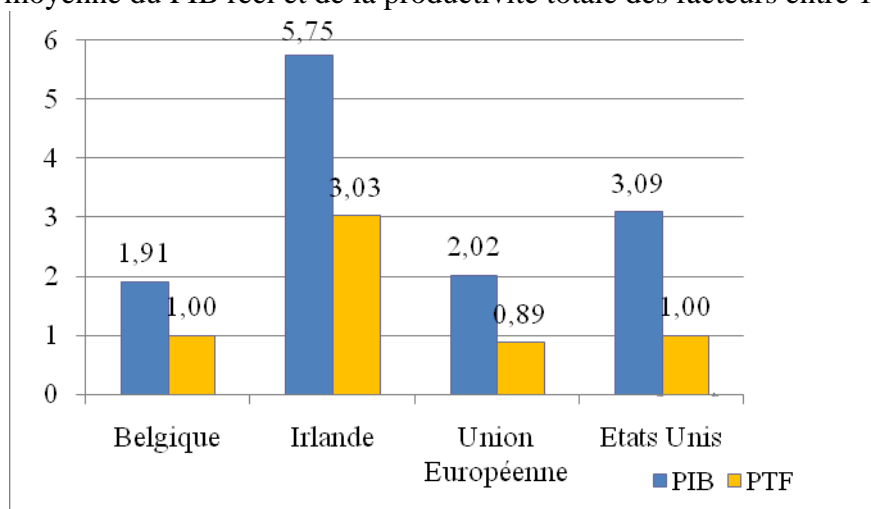
$$\alpha_L = 1 - \alpha_K.$$

II- La part du progrès technique dans la croissance

On peut utiliser le résidu pour mesurer et comparer la part du progrès technique dans la croissance de plusieurs pays. Le graphique ci-dessous compare les taux de croissance du PIB et de la productivité totale des facteurs en Belgique, Irlande, en Europe et aux Etats-Unis, de 1980 à 2001.³⁰

Graphique 1

Croissance moyenne du PIB réel et de la productivité totale des facteurs entre 1980 et 2001.



Source: D'après Timmer, Ypma et van Ark (2003).

Il ressort de ces données deux faits stylisés qui soulignent les limites des explications de la croissance qui reposent uniquement sur l'accumulation de capital :

- d'abord, le progrès technique est responsable d'une part très importante de la croissance. Le résidu est ainsi responsable d'un quart à une moitié (ou plus même) du taux de croissance des quatre régions représentées.

³⁰ Des données plus complètes sont données en annexe, dans le tableau A1.

- deuxièmement, la convergence des productions par travailleur est largement due au fait que les économies initialement plus pauvres connaissent un taux de croissance de leur productivité plus rapide. Ici encore, l'accumulation de capital ne joue qu'un rôle secondaire.

Dans l'échantillon, on voit que c'est l'Irlande qui a effectué son rattrapage. Elle a connu une croissance moyenne de son PIB de 5.75% sur la période. La croissance de la productivité s'est élevée, à elle seule, à 3.03%. Dans les autres pays, la croissance de la productivité est restée comprise entre 0.89% et 1.00%.

On peut généraliser ces observations. On obtient le même genre de résultats lorsqu'on explique les différences de revenu entre pays industrialisés et en développement. Ce qui les distingue ce sont plus des différences de productivité que des différences de stock de capital.

On arrive aux mêmes conclusions lorsque l'on observe des périodes plus longues.³¹ C'est notamment le cas pour les trente glorieuses, ce qui peut être contre intuitif. En effet, on aurait pu croire que cette période de croissance forte était due aux reconstructions d'après-guerre, donc à l'accumulation de capital. Ces reconstructions ont sans doute joué un rôle mais c'est un rôle secondaire. Même à cette époque, c'est le progrès technique qui était le facteur principal de croissance.

De plus, il apparaît que le ralentissement de la croissance observé depuis les années septante dans les pays industrialisés est dû principalement à un ralentissement du progrès technique et non à une moindre accumulation du capital. Cela remet en cause certaines explications de ce ralentissement qui reposent sur la baisse du taux d'épargne.³²

³¹ Une illustration est proposée en annexe, dans le tableau A2.

³² On peut se demander quel sera l'impact des nouvelles technologies de l'information et de la communication (NTIC) sur cette tendance. Comme l'avait fait remarquer Robert Solow dans un article grand public de 1987, on a mis du temps à repérer l'impact de ces NTIC sur la croissance de la productivité aux Etats-Unis. C'est ce qu'on a appelé le « paradoxe de Solow ». Ce paradoxe semble avoir été résolu depuis dans le cas des Etats-Unis, mais pas encore tout à fait en Europe. On a par exemple argué que le paradoxe de Solow pouvait s'expliquer par le fait que les NTIC n'augmentaient la productivité qu'à condition d'être associées à une nouvelle organisation des tâches dans les entreprises. En effet, les entreprises qui se contentent d'acheter des ordinateurs sans modifier leurs méthodes de travail font donc baisser leur productivité puisqu'elles produisent autant qu'auparavant mais avec plus de capital. A l'échelle d'un pays, il est possible que la diminution de la productivité des entreprises qui n'ont pas modifié leur organisation du travail compense les gains de productivité des entreprises qui se sont adaptées à l'informatique. Une référence utile sur le sujet est l'article de W.D. Nordhaus « Productivity Growth and the New Economy » », *Brookings Papers on Economic Activity*, n°2, p. 211-244, 2002.

Ces faits stylisés nous amènent à revoir à la baisse le pouvoir explicatif du modèle que nous avons étudié jusqu'à présent. Pour pouvoir l'utiliser, il faut y intégrer le progrès technique.

Section 2 : Le progrès technique dans le modèle de Solow

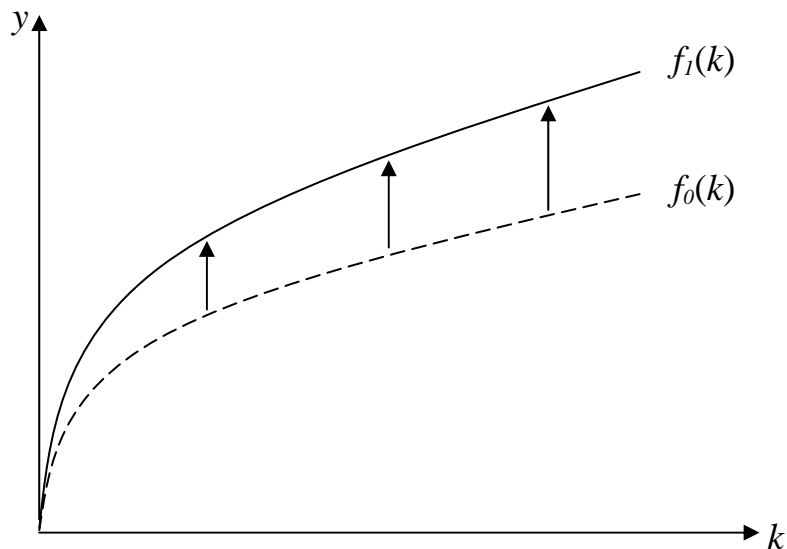
Dans cette section, nous allons étudier les effets du progrès technique, c'est-à-dire de la croissance de la productivité, dans le modèle de Solow. Pour bien comprendre ce qui s'y passe, nous allons d'abord déterminer les effets d'une augmentation ponctuelle de la productivité (I) avant de décrire le comportement du modèle lorsqu'on suppose que la productivité augmente de façon continue (II).

I- Effet d'une croissance ponctuelle de la productivité

Supposons qu'une innovation quelconque vienne augmenter la productivité des facteurs de production. Il peut s'agir de l'apparition d'une nouvelle technique de production (engrais, machine à vapeur, moteur à explosion, téléphone, ordinateur), d'une nouvelle organisation du travail (taylorisme), ou d'une combinaison des deux.

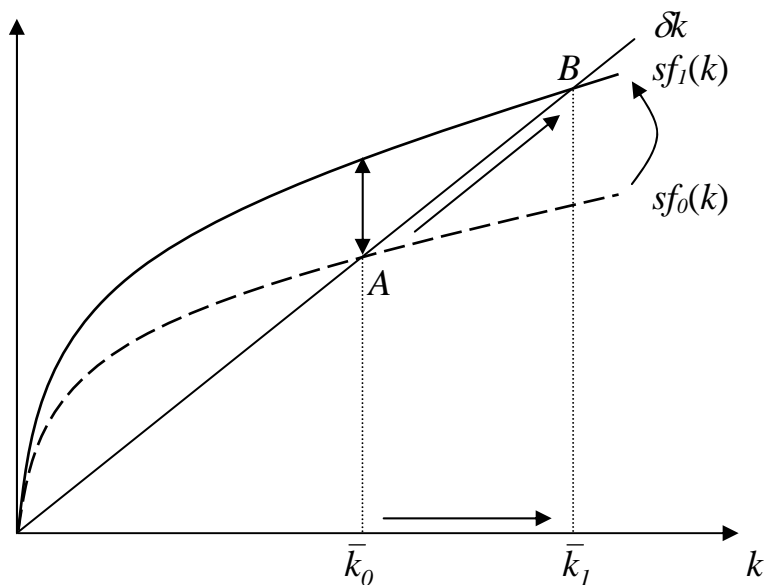
Quelle qu'en soit la raison, cette innovation signifie qu'on va pouvoir produire davantage avec la même quantité de facteurs de production. Dans le cadre du modèle de Solow, on peut représenter cet événement en supposant qu'une nouvelle fonction de production remplace l'ancienne et augmente la production pour chaque valeur du stock de capital par travailleur. Cette modification est représentée par le graphique 2, qui reprend le graphique 1 du chapitre précédent.

Graphique 2
Représentation d'une augmentation de la productivité dans le modèle de Solow



Une fois qu'on a compris comment le progrès technique peut être représenté dans le modèle de Solow, il devient très facile d'en déterminer les conséquences. Ainsi, la déformation de la fonction de production va affecter la fonction d'épargne. On peut alors déterminer l'effet des gains de productivité sur l'état stationnaire, en reprenant la méthode vue dans le chapitre précédent. On adapte alors simplement le graphique 2a du chapitre 2.

Graphique 3
Impact d'une augmentation de la productivité sur le stock de capital à l'état stationnaire



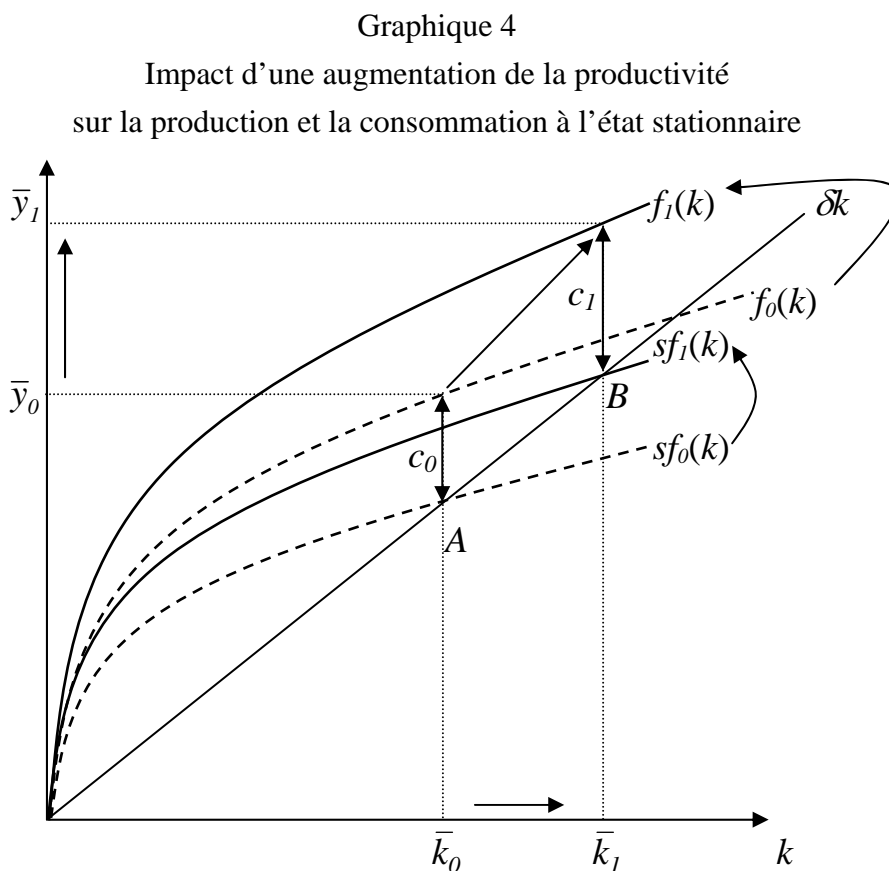
Les notes du cours ne sont transmises qu'à titre indicatif.

L'augmentation de la productivité déforme la fonction d'épargne. En d'autres termes, l'épargne par travailleur est à présent systématiquement supérieure à ce qu'elle était auparavant pour un même stock de capital par travailleur. Cela implique que l'état stationnaire se modifie (on passe du point *A* au point *B*). On constate que le stock de capital par travailleur à l'état stationnaire a augmenté grâce à l'augmentation de la productivité.

On peut de plus décrire étape par étape ce qui s'est passé entre les points *A* et *B*. Le point *A* correspond à l'état stationnaire avant l'augmentation de la productivité. Cela signifie que l'épargne finance un investissement par travailleur tout juste suffisant pour compenser les amortissements.

Lorsque la productivité augmente, la production augmente. L'épargne et l'investissement deviennent alors supérieurs à l'investissement de remplacement. Le stock de capital par travailleur commence alors à augmenter et l'économie se déplace vers le point *B*.

Cependant, au fur et à mesure que le stock de capital augmente, la productivité marginale du capital diminue. Il devient donc de plus en plus difficile d'accumuler du capital. A terme, on atteint donc un nouvel état stationnaire, représenté par le point *B*.



Les notes du cours ne sont transmises qu'à titre indicatif.

L'augmentation du stock de capital n'est pas sans effet sur la production. En effet, puisque la productivité et le stock de capital ont augmenté, on peut déduire que la production par travailleur a elle aussi augmenté. C'est ce que montre le graphique 4, qui complète le graphique 3 en y ajoutant les deux fonctions de production, ce qui permet de représenter l'évolution de la production. La nouvelle valeur stationnaire du revenu est bien supérieure à ce qu'elle était avant l'augmentation de la productivité. De plus, la consommation par travailleur augmente sans ambiguïté, puisque le revenu par travailleur augmente alors que la propension à épargner reste constante.

On peut donc conclure que les gains de productivité permettent de faire augmenter le stock de capital, la production et la consommation par travailleur, même lorsque l'état stationnaire est atteint. Ce résultat ne concerne cependant qu'une augmentation ponctuelle de la productivité. La section prochaine montre comment on peut intégrer une croissance continue de la productivité dans le modèle de Solow, afin de le rendre compatible avec les caractéristiques observées de la croissance.

II- La prise en compte d'une augmentation tendancielle de la productivité

Pour intégrer le progrès technique de façon systématique, il faut modifier la fonction de production agrégée pour y faire apparaître un paramètre de productivité. Il y aurait plusieurs façons de le faire, mais Solow choisit la suivante :

$$Y = F(K, AL) \quad (4)$$

Cette fonction est légèrement différente de celle que nous avons utilisée pour décomposer la croissance dans la première section du chapitre. En effet, le paramètre qui mesure la productivité n'affecte plus à présent que le travail.

Cette spécification a le gros avantage d'être plus facile à manipuler. Elle a d'autres propriétés que nous n'utiliserons pas ensemble mais qui permettent par exemple d'obtenir une évolution réaliste de la rémunération des facteurs de production. Ces questions seront traitées plus tard dans votre cursus.

L'hypothèse que le progrès affecte la productivité du travail peut s'interpréter de deux façons équivalentes :

- d'abord, elle signifie que le progrès technique réduit le nombre de travailleurs nécessaires pour produire une quantité donnée. Si A double, il faudra deux fois moins de travailleurs pour produire la même chose.

- ensuite, elle signifie que le progrès augmente la production des travailleurs. Le doublement de A a alors le même effet sur la production totale qu'un doublement de la population active. Pour tenir compte de ce phénomène, on appelle AL la *quantité de travail effectif*.

A part le paramètre A , la fonction de production est en tout point identique à celle que nous avons utilisée dans le chapitre précédent. Les rendements d'échelle sont donc constants et les rendements factoriels décroissants. Cependant, il faut garder à l'esprit que ces propriétés se conçoivent pour un état donné de la technologie.

Pour tenir compte de l'augmentation continue de la productivité, Solow suppose que le paramètre A augmente à un taux constant g_A . Ce taux est exogène. En d'autres termes, il n'est pas expliqué et fait partie des hypothèses du modèle.

La conséquence de cette hypothèse est que la fonction de production se déforme en permanence comme dans le graphique 2. Si on applique les résultats que nous venons d'obtenir, cela implique que le stock de capital par travailleur et la production par travailleur augmentent en permanence.

On peut même calculer leur taux de croissance à l'état stationnaire. Nous ne le ferons pas ici pour éviter de rentrer dans des calculs fastidieux, mais on peut montrer que le taux de croissance du stock de capital est le même que celui de la production par tête, et qu'ils correspondent tous les deux au taux de croissance de la productivité : g_A .

Ainsi, le modèle de Solow est compatible avec une croissance soutenue de la production par travailleur, même à l'état stationnaire. On doit en revanche souligner que la croissance à l'état stationnaire n'est pas expliquée par le modèle.

En effet, c'est uniquement parce qu'on suppose une augmentation exogène de la productivité que la croissance se poursuit à l'état stationnaire. Si on pose $g_A = 0$, la croissance à l'état stationnaire est nulle. On retrouve ainsi le modèle de Solow sans progrès technique, qui n'est qu'un cas particulier du modèle avec progrès technique.

En amont de l'état stationnaire, le comportement du modèle est tout à fait comparable à ce qu'il était sans progrès. Le taux de croissance s'explique alors à la fois par l'accumulation de capital et le progrès technique.

En amont de l'état stationnaire, on retrouve alors les résultats de la première section :

- le taux de croissance est une fonction croissante du taux d'épargne ;
- les pays initialement plus pauvres croissent plus vite que les autres.

On voit donc qu'en dehors des périodes transitoires où le taux d'épargne joue un rôle, les différences de taux de croissance du revenu par habitant dépendent uniquement du taux de croissance de la productivité.

Cette conclusion est également applicable aux différences entre les pays. Si un pays croît durablement plus vite qu'un autre, c'est que sa productivité croît plus vite.

Paradoxalement, alors que le modèle de Solow représente explicitement la croissance comme un processus d'accumulation du capital, ses conclusions soulignent que la cause fondamentale de la croissance est à chercher ailleurs. Il est donc nécessaire de compléter le modèle de Solow par une réflexion sur les déterminants du progrès technique.

Section 3 : Les déterminants du progrès technique

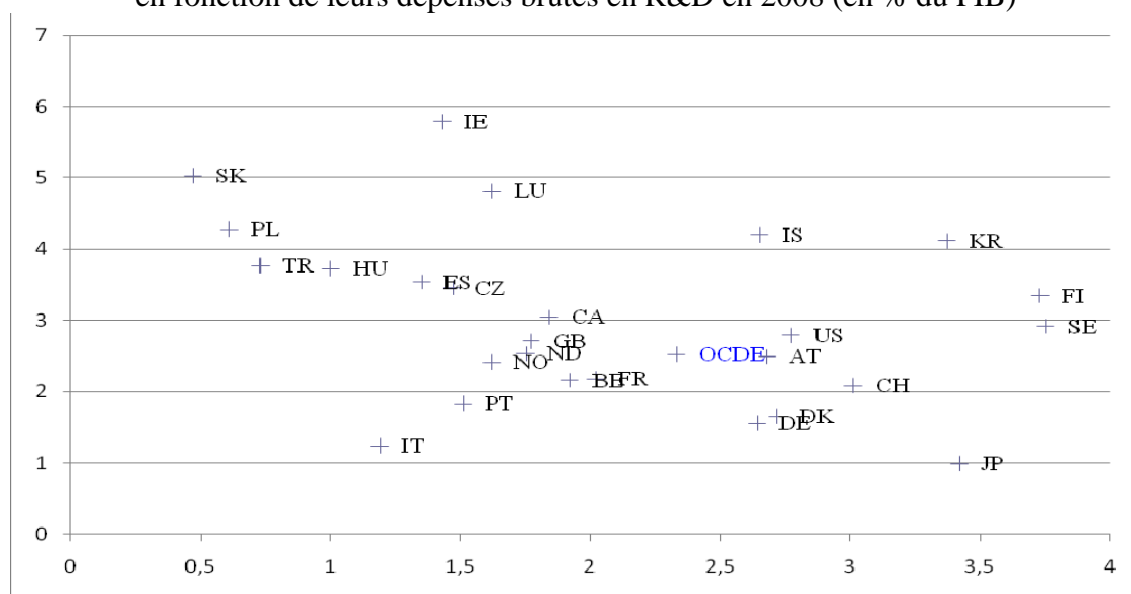
Nous avons jusqu'à présent mesuré le progrès technique et envisagé ses conséquences sur l'évolution de la production. En revanche, nous n'avons pas cherché à l'expliquer. Dans le modèle de Solow, la croissance de la productivité est exogène et prend la forme d'une manne tombée du ciel, sans effort. C'est un peu frustrant lorsqu'on connaît son importance quantitative.

Pourtant, le progrès technique est le résultat de l'activité humaine et devrait donc pouvoir être compris par les économistes.

On peut envisager une foule de déterminants de la croissance de la productivité. L'attention s'est par exemple portée sur le rôle de l'éducation et sur la qualité des infrastructures publiques. Il n'est pas question ici de sous-estimer l'importance de ces phénomènes mais nous ne les traiterons pas explicitement dans ce chapitre, par manque de temps.

Nous allons en revanche nous concentrer sur l'activité qui est la plus intuitivement liée au progrès technique, en particulier dans les pays les plus riches : la recherche et l'innovation.

Graphique 5 : Taux de croissance annuel moyen du PIB réel (1998-2008) des pays de l'OCDE en fonction de leurs dépenses brutes en R&D en 2008 (en % du PIB)



Source : OCDE, Main Science and Technology Indicators 2010.

C'est à l'économiste autrichien Joseph Aloïs Schumpeter³³ que l'on doit les intuitions fondamentales sur lesquelles repose encore aujourd'hui l'analyse économique de la recherche et de l'innovation. Il est le premier économiste à avoir systématiquement souligné et étudié le rôle de l'innovation dans la croissance économique à long terme.

Schumpeter distingue deux formes d'innovations qui vont retenir notre attention dans ce chapitre : l'invention de produits nouveaux et la mise au point de procédés nouveaux.³⁴ Son idée principale est qu'on peut analyser ces activités comme le résultat d'investissements. Nous allons d'abord voir comment (I). Il sera alors possible d'étudier ce qui détermine leur efficacité donc leur impact sur la croissance (II).

I- L'investissement dans la recherche

Comme tout investissement, les investissements de recherche et de développement (R&D) sont motivés par la perspective d'un rendement (A). Dans le cas de la R&D, la question de l'appropriation de ce rendement se pose avec une acuité particulière (B).

³³ J.A. Schumpeter *The Theory of Economic Development: An inquiry into profits, capital, credit, interest and the business cycle*, 1911.

³⁴ Pour être juste, Schumpeter y ajoute la découverte de nouveaux débouchés, la découverte de nouvelles sources de matières premières et les changements dans l'organisation des entreprises.

A- Motivation des investissements en R&D

Pour Schumpeter, l'innovation est une activité économique qui a pour objet la recherche du profit. C'est parce que les *entrepreneurs* espèrent augmenter leur profit qu'ils investissent des ressources dans la production de connaissances nouvelles. Ces connaissances nouvelles vont déboucher sur des produits nouveaux et des procédés nouveaux.

Ces produits et procédés nouveaux vont conférer à leur inventeur une position de monopole. Grâce à cette position de monopole, les entrepreneurs vont pouvoir obtenir des profits de monopole au lieu d'un profit nul sur un marché concurrentiel. C'est pour obtenir une telle rente de monopole, que les entrepreneurs consacrent des ressources à la R&D.

Cependant, la rente de monopole n'est pas éternelle. Elle finit par s'éroder à cause de la réaction des concurrents de l'entreprise qui a innové.

Ses concurrents consacrent eux aussi des ressources à la recherche et finissent eux aussi par découvrir des produits nouveaux qui peuvent supplanter l'innovation initiale.

ex : les vieux 78 tours sont remplacés par le microsillon, qui est remplacé à son tour par le CD, qui cède la place au minidisc et au MP3 etc.

La conséquence de cette érosion des rentes dues à l'innovation passée est que les entreprises ne peuvent se contenter d'avoir innové une fois pour toutes mais doivent constamment continuer d'investir dans la R&D. La concurrence pour les rentes de monopole engendre donc une tendance continue à l'innovation qui se traduit par une augmentation de la productivité.

B- L'appropriabilité des résultats de la R&D

Le raisonnement de Schumpeter suppose que l'entrepreneur qui a innové va pouvoir profiter d'une rente. En d'autres termes, il est important qu'il puisse s'approprier les bénéfices de son investissement. Aucune entreprise n'investirait dans quoi que ce soit si elle ne pouvait pas en retirer des bénéfices.

Or l'investissement en R&D a une spécificité qui menace la capacité des entreprises à accaparer ses résultats. Bien souvent, l'innovation sur laquelle il débouche consiste dans une *connaissance* nouvelle (procédé nouveau, formule ou idée nouvelle). Or la connaissance possède les caractéristiques d'un bien public :

- il n'y a pas de rivalité dans les consommations. Ce n'est pas parce qu'une entreprise utilise une formule qu'elle en réduit l'intérêt pour les autres entreprises.

ex : ce n'est pas parce qu'un laboratoire utilise la formule de l'aspirine qu'un autre laboratoire ne peut pas produire en même temps de l'aspirine d'aussi bonne qualité.

La formule d'un produit pharmaceutique ne s'use pas lorsqu'elle est utilisée, d'où la possibilité de produire des médicaments *génériques*.

- il est difficile d'exclure une entreprise de l'utilisation d'une connaissance.

ex : si une entreprise a l'idée de livrer des pizzas à domicile, elle aura du mal à empêcher d'autres entreprises de l'imiter.

Un laboratoire pharmaceutique peut analyser la composition du dernier produit commercialisé par son concurrent pour en faire une copie. On appelle cela de l'*ingénierie à rebours*.

On voit donc, que l'innovation va finir par se diffuser. Elle va non seulement être concurrencée par des innovations nouvelles, mais aussi par des imitations. C'est le processus de *diffusion*.

Au cours du processus de diffusion, la rente de monopole s'érode peu à peu. La question est de savoir à quelle vitesse.

Si la diffusion est immédiate, aucune entreprise n'aura intérêt à investir dans la R&D parce qu'elle n'aura pas la possibilité de rentabiliser son investissement. Aucune entreprise n'aura la possibilité de s'approprier les fruits de son investissement.

La capacité des entrepreneurs à s'approprier les fruits de leur innovation détermine donc leur incitation à investir dans la R&D. In fine, c'est le progrès technique et la croissance qui vont dépendre de l'appropriabilité du rendement de l'innovation.

Si certaines innovations sont difficiles à copier parce qu'elles peuvent être cachées ou qu'elles reposent sur un savoir-faire spécifique, d'autres peuvent être copiées facilement (logiciels, musique, accessoires de mode etc.).

Cela souligne l'importance de la protection de la propriété intellectuelle (brevets, droits d'auteurs, marques déposées etc.).

La protection de la propriété intellectuelle a un effet direct sur l'innovation, donc sur la croissance.

La protection de la propriété intellectuelle est cependant un dilemme fondamental.

En l'absence de protection, on sait qu'il n'y aura pas d'innovation.³⁵

³⁵ Cet argument doit être nuancé. Dans certains cas, le dépôt d'un brevet peut faciliter la diffusion d'une technologie. En effet, lorsqu'une entreprise dépose un brevet, elle doit fournir une description précise du produit breveté. Cette description permet aux entreprises concurrentes de copier le produit sans gros investissement lorsque le brevet tombe dans le domaine public, elle permet également aux entreprises concurrentes de proposer

En revanche, il n'est pas optimal d'accorder un monopole définitif aux innovateurs. Vous avez en effet vu en cours de microéconomie qu'il est plus efficace que le marché soit concurrentiel. De plus, si une entreprise innovante obtient un monopole définitif sur un produit, elle ne sera plus incitée à rechercher de nouvelles innovations, ce qui n'est pas bon pour la croissance. Enfin, la diffusion de la technologie permet d'améliorer la productivité des autres entreprises.³⁶

=> la protection de la propriété intellectuelle doit être temporaire.

Cependant, la relation entre investissement en R&D et innovation n'est pas mécanique. Nous venons de voir ce qui pouvait motiver l'investissement. Il faut également se demander ce qui détermine son efficacité.

II- L'efficacité des efforts de recherche

Les économistes sont mal armés pour comprendre en détail l'efficacité des efforts de recherche. Ils peuvent cependant souligner le rôle de certains facteurs institutionnels (A) et insister sur le caractère cumulatif des connaissances, qui peut avoir des conséquences cruciales pour les théories de la croissance (B).

A- Facteurs institutionnels

La fécondité du processus de recherche se définit comme la façon dont les investissements en R&D vont se traduire par des innovations. Deux pays qui investissent des sommes comparables dans la recherche pourront obtenir des résultats en termes de croissance très différents en fonction de la fécondité de leur système de recherche.

La fécondité de la recherche mesure en quelque sorte le rendement des investissements en R&D. Il faut cependant souligner une différence importante entre le rendement d'un investissement en R&D et celui des investissements en capital physique. Le rendement des investissements physiques est en général connu avec certitude ou avec peu d'aléas. On sait que si on acquiert telle machine, la production augmentera de tant d'unités.

plus facilement des substituts. C'est pourquoi certaines entreprises préfèrent protéger leurs produits par le secret. La formule du Coca-Cola n'a par exemple jamais été brevetée et reste pourtant secrète depuis plus d'un siècle.

³⁶ Le même phénomène se produit à l'échelle internationale. La diffusion des technologies est un ressort de la croissance dans la mesure où elle permet à chaque pays de bénéficier des innovations réalisées à l'étranger. Une question importante, que nous n'aborderons pas dans le cours, consiste à se demander ce qui détermine la capacité d'un pays à absorber les techniques venues de l'étranger et la vitesse à laquelle cette absorption peut se faire.

La question ne se pose pas de cette façon dans le cas des investissements en R&D. En effet, on ne sait en général ni ce qu'on va découvrir, ni combien de temps cela va prendre. La décision d'investir dans la recherche consiste donc à augmenter ses chances de découvrir quelque chose. On sait simplement que plus on investira dans la recherche, plus on aura de chances de découvrir quelque chose.³⁷ Le rendement de la recherche est donc délicat à mesurer.

A l'échelle d'un pays, c'est principalement l'organisation de la recherche qui va déterminer sa fécondité. En particulier, la relation entre la recherche fondamentale et la recherche appliquée est cruciale. La première développe des théories plus générales et parfois abstraites. La seconde utilise ces résultats pour développer de nouveaux usages. Ces deux branches de la recherche sont indissociables.

ex : Les travaux d'Albert Einstein relevaient de la recherche fondamentale. Ils ont cependant permis de développer les centrales nucléaires grâce à la recherche appliquée.

La relation entre recherche appliquée et recherche fondamentale est difficile à cerner et le dosage optimal des deux est donc une question en débat. Il va dépendre de la façon dont sont formés les scientifiques et les ingénieurs, des relations entre les universités et les entreprises, des incitations des uns et des autres à rechercher une application de leurs découvertes.

On dit souvent que le système français valorise les études et la recherche théoriques, aux dépens de l'enseignement technique et de la recherche appliquée. En termes de croissance, cela n'est probablement pas optimal car cela limite les innovations.

On a aussi invoqué l'impact de la culture. Le rôle des entrepreneurs serait valorisé aux Etats-Unis et moins en Europe. Ce type d'explication peut sembler séduisant mais il faut l'utiliser avec prudence. Si la notion de culture n'est pas définie très précisément, elle ne sera pas très utile. Elle pourra même ne fournir qu'une explication a posteriori des différences observées dans les performances économiques.

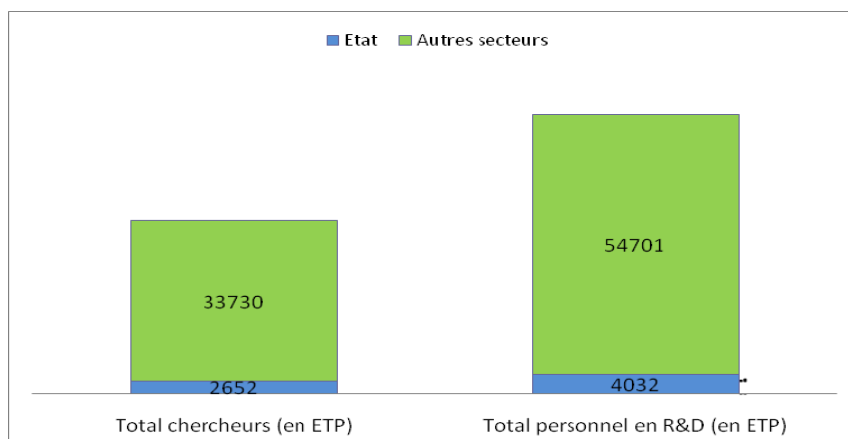
ex : Lorsque l'on dit que la culture américaine est favorable à la croissance, cela revient à constater que les Etats-Unis sont aujourd'hui l'économie la plus efficace du monde. On ne dit pas pourquoi, ni comment cela pourrait changer, et encore moins ce que pourraient faire les autres pays.³⁸

³⁷ On peut utiliser l'image d'un forage pétrolier : plus on fore plus on a de chances de trouver du pétrole, mais on ne sait pas au départ où, ni combien de forages on va devoir réaliser avant de trouver un puits de pétrole rentable.

³⁸ Ce phénomène relève de ce qu'on appelle le « biais de l'observateur. Pour illustrer ce que peut produire ce biais, l'historien économique David Landes rapporte les propos d'un observateur occidental qui décrivait le

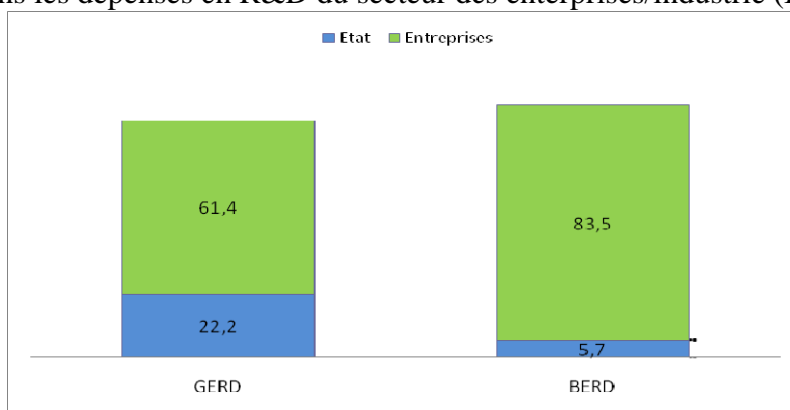
Enfin, un facteur institutionnel fondamental est la participation de l'Etat dans la recherche. L'Etat joue en effet un rôle important non seulement en accordant des brevets mais aussi en participant directement à l'effort de recherche.

Graphique 6 : Part de l'Etat dans le total des chercheurs et du personnel R&D en Belgique en 2008



Source : OECD, Main science and technology indicators database 2010.

Graphique 7: Part de l'Etat et des entreprises/industrie dans les dépenses brutes en R&D (GERD) et dans les dépenses en R&D du secteur des entreprises/industrie (BERD) en 2008



Source : OECD, Main science and technology indicators database 2010.

La question qui se pose alors est double. Il s'agit d'une part de déterminer le volume de l'intervention de l'Etat dans l'effort de recherche. Il s'agit d'autre part de déterminer quel

Japon en 1881 : « Les Japonais sont une race heureuse, et se contentant de peu, ont peu de chance d'arriver à grand' chose. ». En 1992, une enquête du magazine Times révélait que 94% des Américains considéraient les Japonais comme « travailleurs ». Sur l'impact de la culture sur la croissance, vous pouvez consulter le chapitre 14 du manuel de D.N. Weill.

type de recherche il doit financer. L'idée est alors de financer les activités qui ne seraient pas entreprises par le secteur privé en l'absence de financement public. Cet argument suggère que l'Etat prenne en charge au moins la recherche fondamentale, dont les applications à court terme sont rares.

B- Le caractère cumulatif des connaissances

On peut se demander si l'un des facteurs décisifs de l'accumulation de connaissances n'est pas l'existence d'un stock de connaissances initial. C'est ce qu'implique le *caractère cumulatif des connaissances*. Cela signifie que toute innovation s'appuie sur les connaissances disponibles, donc les innovations passées.

Parfois, une innovation ne consiste que dans l'idée de rapprocher deux techniques ou produits existants.

ex : la carte à puce consiste fondamentalement à jumeler une puce et une carte de crédit.

La livraison de pizza à domicile consiste à utiliser à la fois les connaissances des pizzerias et des coursiers.

Globalement, pratiquement tous les produits nouveaux mobilisent des connaissances et des technologies très diverses. La construction d'une automobile moderne suppose de maîtriser les techniques de carrosserie, de motorisation, d'électronique etc., qui ont souvent été développées indépendamment les unes des autres.

Le caractère cumulatif des connaissances a des implications de première importance pour la dynamique des économies. Il implique en effet qu'une économie pourra innover d'autant plus qu'elle disposera déjà d'un stock de connaissances important. Si on suppose que le stock de connaissances détermine la productivité, cela signifie qu'une économie initialement plus productive verra sa productivité croître plus rapidement que les autres.

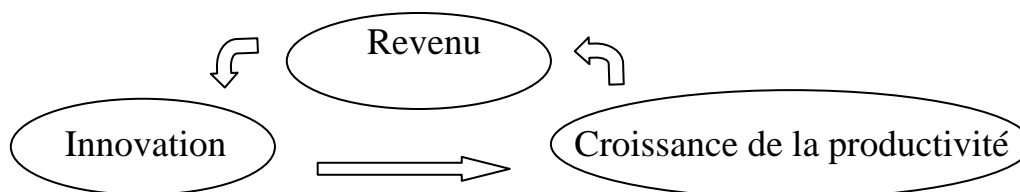
En d'autres termes, une économie plus riche au départ croîtra plus rapidement que les autres.

=> on n'observera pas un phénomène de convergence mais un phénomène de divergence.

Cette intuition se trouve à la base des théories de la croissance développées à la fin des années quatre-vingts et au début des années nonante et qu'on a baptisées *théories de la croissance endogène*.³⁹

Ces théories reposent sur des mécanismes par lesquels la croissance de la productivité est d'autant plus rapide que le revenu est élevé au départ. Ces mécanismes reposent notamment sur l'incitation à innover ou l'apprentissage par la pratique, c'est-à-dire le fait qu'un travailleur accumulera d'autant plus d'expérience qu'il aura produit beaucoup.

Ces modèles ont constitué une percée significative pour les théories de la croissance. En effet, ils proposent une explication cohérente au fait que les pays pauvres ne convergent pas vers les niveaux de revenu des pays riches. Cette explication peut se résumer par le graphique suivant :



Les pays riches bénéficient d'un cercle vertueux. Comme leur revenu est élevé au départ, ils innoveront beaucoup. Comme ils innoveront beaucoup, leur productivité augmente rapidement, ce qui augmente leur revenu. Ils innoveront encore plus vite par la suite etc.

Les pays pauvres sont au contraire enfermés dans un cercle vicieux. Comme leur revenu est faible au départ, ils innoveront peu. Comme ils innoveront peu, leur productivité augmente lentement ou pas du tout. Leur revenu stagne donc ou augmente lentement. Ils innoveront toujours aussi lentement par la suite etc.

Ce cercle vicieux correspond à ce qu'on appelle la *trappe à pauvreté*. Il implique que les pays pauvres sont pauvres... parce qu'ils sont pauvres.

³⁹ On considère que le premier modèle de croissance endogène a été proposé par Paul Romer dans l'article intitulé "Increasing returns and long run growth", *Journal of Political Economy*, vol 94 n°5, pp.1002-1037, 1986. Ce premier modèle a été complété par d'autres qui ont amélioré la description du progrès technique et proposé d'autres mécanismes qui reposent notamment sur l'éducation et la qualité des infrastructures publiques.

Conclusion

Lorsque l'on mesure la part du progrès technique dans la croissance, on réalise qu'elle est très élevée. De plus, seul le progrès technique permet d'expliquer l'augmentation des niveaux de vie observée sur le long terme. Le modèle de Solow montre en effet que l'accumulation de capital ne peut être à elle seule qu'un facteur temporaire de croissance.

Pour comprendre les déterminants de la croissance à plus long terme, il est donc nécessaire de comprendre les déterminants du progrès technique. La protection de la propriété intellectuelle et l'organisation de la recherche jouent là un rôle crucial.

De plus, comme la connaissance est un bien cumulatif, le progrès sera plus rapide dans les économies disposant d'un stock initial de connaissances plus élevé. Les théories de la croissance endogène reposent sur cette intuition. Elles mettent en cause l'hypothèse optimiste de convergence spontanée des économies.

Ces théories ne sont cependant pas tout à fait satisfaisantes, mais elles ouvrent des pistes de réflexion :

- pourquoi les pays pauvres sont-ils pauvres au départ ?
- quelles sont les politiques à mener pour sortir de la trappe à pauvreté ?

Ces questions préoccupent encore les économistes spécialistes de la croissance qui affinent les théories et accumulent les études empiriques.

Pour notre part, nous allons abandonner ici les théories de la croissance. Vous aurez l'occasion de les étudier plus en détail par la suite de votre scolarité. Pour l'heure nous allons nous rapprocher du court terme et entamer la deuxième partie du cours.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2004 : chapitre 4 (section 4.2 et suivantes).

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 4.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 8.

Samuelson P.A. et W.D. Nordhaus : *Economie*, Economica, Paris, 2000 : chapitre 27 (très basique).

Pour les passionnés

Guellec D. et P. Ralle, *Les nouvelles théories de la croissance*, Repères, La Découverte, Paris, 2003 : chapitres 3 et 4.

Annexes

Tableau A1

Croissance et productivité entre 1980 et 2001

	Belgique		Irlande		Union Européenne		Etats-Unis	
	PIB	PTF	PIB	PTF	PIB	PTF	PIB	PTF
1980-1990	1,9	0,7	3,6	2,9	2,4	1,1	3,2	0,7
1990-1995	1,6	1,0	4,7	2,9	1,6	1,1	2,4	0,6
1995-2001	2,5	1,1	8,9	3,6	2,4	0,5	3,5	0,8
1980-2001	2,01	0,90	5,33	3,11	2,20	0,94	3,10	0,73
1990-2001	2,06	1,06	6,94	3,30	2,04	0,77	3,02	0,72

Source: Timmer, Ypma et van Ark (2003).

On doit à Angus Maddison d'avoir réalisé ce travail sur la période de l'après-guerre pour plusieurs économies développées. Le tableau A1 reprend certains de ses résultats et compare les périodes 1950-1973 et 1973-1987 :

Tableau A2

Croissance et productivité dans l'après-guerre

	Croissance de la production par habitant		Taux de progrès technique	
	1950-73	1973-87	1950-73	1973-87
	France	4,0	1,8	4,9
Allemagne	4,9	2,1	5,6	1,9
Japon	8,0	3,1	6,4	1,7
Royaume Uni	2,5	1,8	2,3	1,7
Etats-Unis	2,2	1,6	2,6	0,6

Source : A. Maddison, *Les phases du développement capitaliste*, Paris, Economica, 1981.

Les notes du cours ne sont transmises qu'à titre indicatif.

Deuxième partie : L'économie dans le court terme : les fluctuations

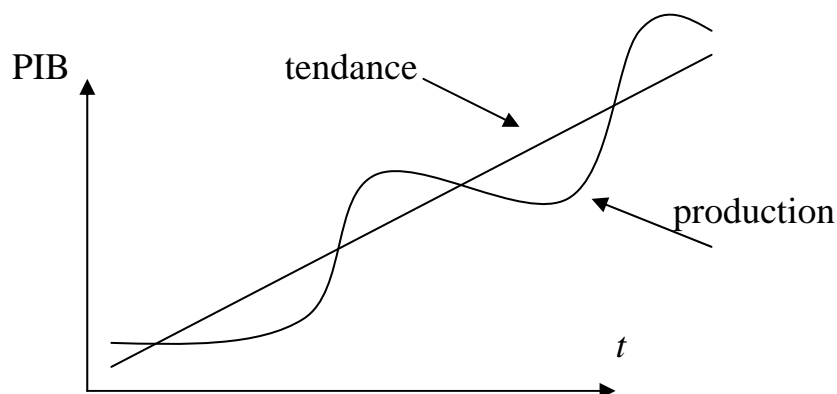
Dans la partie précédente, nous avons étudié l'économie en supposant qu'elle connaissait un taux de croissance constant sur de longues périodes. Nous avons proposé des explications au fait que certaines économies connaissent une croissance moyenne forte alors que d'autres stagnaient.

Pourtant, même les économies les plus performantes connaissent des périodes de croissance plus ou moins rapide. On observe en effet des écarts réguliers par rapport à la tendance. On appelle *fluctuations* ces variations de la croissance autour de sa tendance de long terme.

On a longtemps utilisé, et on utilise encore, le terme de *cycle* mais celui-ci suggère une parfaite régularité des variations de la croissance à court et moyen terme. Le terme de *crise* est aussi employé mais il désigne une phase du cycle et a une connotation pathologique. Ainsi préfère-t-on aujourd'hui le terme de fluctuations qui est plus neutre.

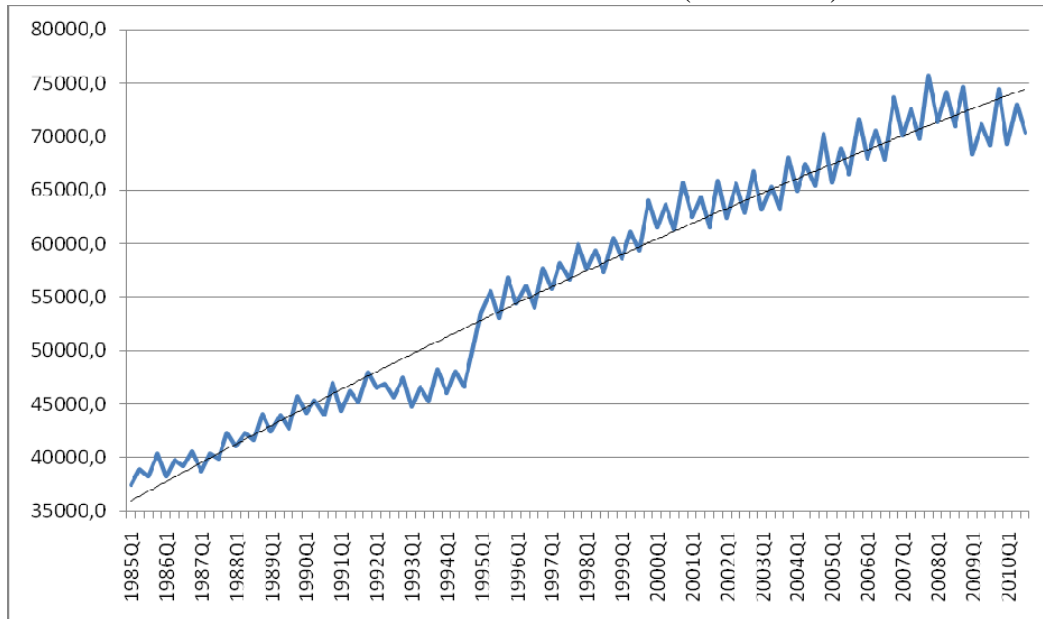
En bref, la distinction entre croissance et fluctuations peut se résumer par un graphique qui représente l'évolution de la production au cours du temps :

Graphique 1 : tendance et fluctuations



On constate que sur un horizon de quelques années le taux de croissance est beaucoup plus instable que ce que laisserait supposer les modèles de croissance. L'exemple de la Belgique n'échappe pas à la règle.

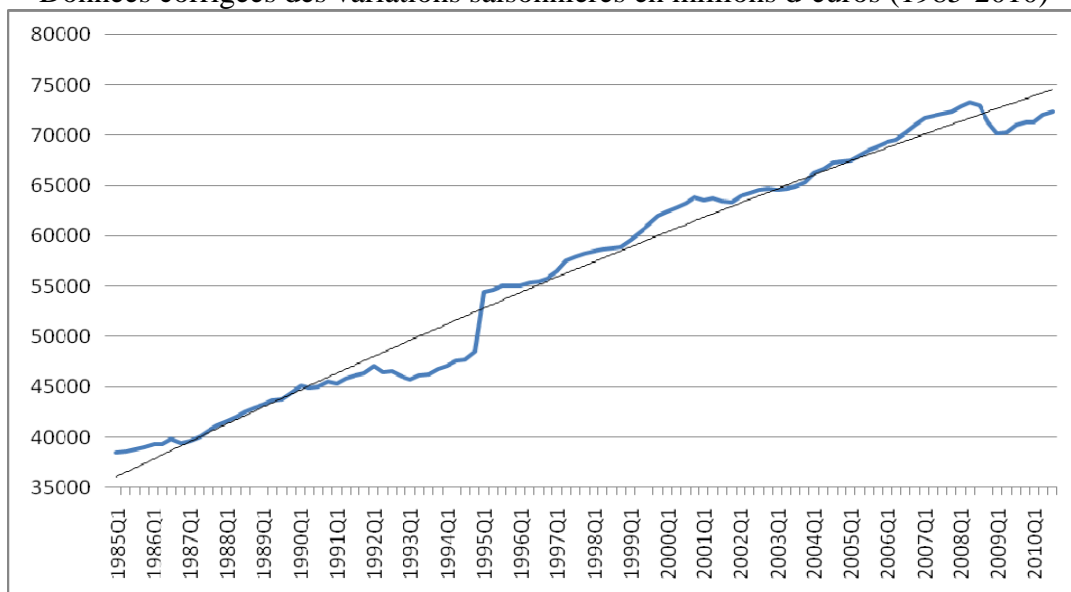
Graphique 2a : Evolution du PIB de la Belgique aux prix de 2000
Données brutes en millions d'euros (1985-2010)



Source : Eurostat.

Une partie des fluctuations se répète régulièrement d’une année à l’autre. C’est ce qu’on appelle les variations saisonnières. Ces variations peuvent se corriger assez facilement. Mais même quand les données sont corrigées des variations saisonnières, on observe une forte variabilité de la croissance. Le taux de croissance peut être soit supérieur soit inférieur à la tendance pendant des périodes assez longues.

Graphique 2b : Evolution du PIB de la Belgique aux prix de 2000
Données corrigées des variations saisonnières en millions d'euros (1985-2010)



Source : Eurostat.

Les notes du cours ne sont transmises qu’à titre indicatif.

L'analyse des fluctuations consiste donc à chercher à expliquer la succession de phases d'expansion et de récession. On peut se demander s'il faut pour cela construire des théories spécifiques ou essayer de raffiner les théories que nous avons déjà vues. Certains économistes considèrent qu'il est possible d'expliquer les fluctuations en restant dans le cadre concurrentiel que nous avons déjà utilisé. Selon eux, il faudrait mieux prendre en compte les décisions intertemporelles et les délais d'ajustement qui apparaissent dans la gestion de la production et de l'investissement.⁴⁰

Pourtant, il existe au moins deux différences importantes entre le court et le long terme :

- à court terme les prix sont rigides. En effet, on constate que beaucoup de prix ne s'ajustent pas instantanément aux fluctuations de l'offre et de la demande. Les raisons de cette rigidité sont encore l'objet de débats mais son existence ne fait pas de doute.⁴¹

L'exemple le plus frappant est celui des journaux quotidiens. Leur prix n'est révisé que très rarement, après des mois de stabilité, alors que leur demande fluctue énormément en fonction de l'actualité. Ainsi, on peut remarquer qu'aucun quotidien n'a augmenté son prix au lendemain du 11 septembre ou pendant les jeux olympiques, qui ont pourtant fait augmenter les tirages. Ce constat est cependant plus général. En Belgique, seulement dix-sept pourcents des prix à la consommation changent chaque mois. La durée médiane d'un prix en Belgique est de plus estimée à treize mois.⁴²

Il faut de plus garder à l'esprit que cette rigidité ne concerne pas seulement les biens mais aussi, par exemple, les salaires.

- à court terme on constate également que la production peut être inférieure à son niveau potentiel. Les récessions ne sont pas dues à une réduction des capacités de production de l'économie mais à une utilisation moins intensive de ses capacités de production. On a du mal à croire que le fait que la production industrielle des Etats-Unis ait été divisée par deux

⁴⁰ Cette voie de recherche est à la base de la théorie des cycles réels, évoquée dans le chapitre 1.

⁴¹ Il existe beaucoup de raisons pour lesquelles les prix et les salaires sont rigides. D'abord, il est coûteux de modifier les prix car il faut y consacrer du temps et d'autres ressources. Entre autres, il faut par exemple modifier les étiquettes et changer le menu des restaurants. C'est pourquoi on parle de *coûts de menu*, même si la restauration n'est pas forcément un secteur primordial dans l'économie. Deuxièmement, les agents peuvent souffrir d'*illusion monétaire*, c'est-à-dire attacher de l'importance aux variables nominales en tant que telles au lieu de ne réagir qu'aux fluctuations des variables réelles. On a récemment proposé des explications sociologiques à l'illusion monétaire fondées sur le fait que les prix et les salaires ont une dimension symbolique et sont liés à des considérations d'équité qui en limitent la flexibilité. Troisièmement, les agents peuvent disposer d'une *information imparfaite* sur l'évolution de l'économie et ne pas toujours ajuster les prix de façon adéquate. Nous reviendrons sur certaines de ces explications dans la suite du cours.

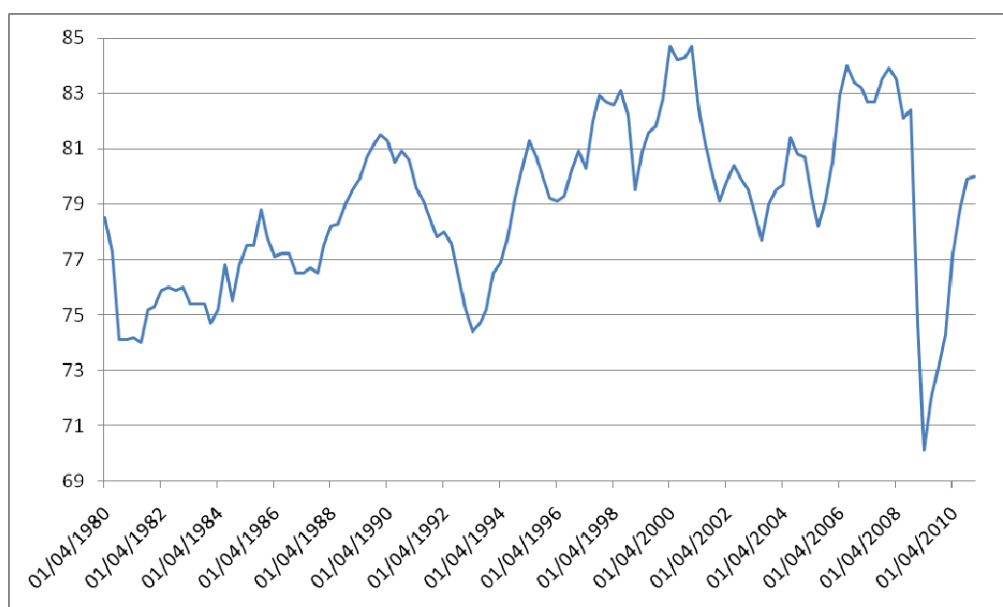
⁴² L. Aucremanne et E. Dhyne « How frequently do prices change? Evidence based on the micro data underlying the Belgian CPI », Banque centrale européenne, Working Paper Series, n°331, avril 2004.

pendant les trois premières années de la crise de 1929 soit dû à une réduction de même ampleur des capacités de production du pays.

Plus généralement, on a développé une mesure de l'utilisation du potentiel productif d'une économie appelée *taux d'utilisation des capacités (TUC)*. On constate que le TUC peut fluctuer de façon significative.

Ainsi, en Belgique sur les trente dernières années, le TUC a connu un maximum de 85% au troisième trimestre 2000 et un minimum de 69,7% au deuxième trimestre 2009. Ces données permettent de mesurer l'ampleur de la récession actuelle. Le TUC n'était jamais passé sous la barre des 70% depuis trente ans. Avant 2009, la valeur la plus faible que le TUC avait atteinte était celle de 73%, au premier trimestre 1981.

Graphique 3
Evolution du taux d'utilisation des capacités industrielles en Belgique (1980-2010)



Source : BNB : Industrie manufacturière: résultats trimestriels bruts nationaux par secteur: degré d'utilisation de la capacité de production en %.

Dans cette partie, nous allons développer des théories qui permettent de comprendre ces phénomènes. Nous verrons que le second est dans une large mesure une conséquence du premier. Dans tous les cas, si on veut étudier la possibilité d'une situation de sous-emploi des capacités, il faut intégrer la demande dans notre raisonnement.

Le premier chapitre de cette partie (chapitre 4 du cours) sera donc consacré à la demande de biens et à l'équilibre sur le marché des biens. Cependant, une théorie macroéconomique ne peut être satisfaisante si elle n'intègre pas les interactions entre

marchés. C'est pourquoi le deuxième chapitre (chapitre 5) sera consacré à la monnaie et aux marchés financiers.

Le troisième chapitre (chapitre 6) sera peut-être le plus important du cours. Il montrera comment on peut analyser conjointement l'équilibre du marché des biens et du marché de la monnaie et en tirer des leçons pour la politique économique.

Nous concluons alors la partie en réintégrant l'offre, c'est-à-dire les contraintes de production, dans l'analyse, ce qui supposera de tenir compte du marché du travail (chapitre 7).

Chapitre 4 : Le marché des biens

Introduction : L'importance de la demande de biens

Dans la première partie du cours, consacrée à la croissance, nous avons décrit l'évolution de l'économie en nous concentrant sur l'offre de biens et en négligeant les effets de la demande. Comme nous raisonnions à long terme, c'était l'évolution des capacités de production qui déterminait l'évolution de la production. On pouvait supposer que la demande absorbait forcément ce qui était produit.

Si on veut s'intéresser au court terme, et donc à la possibilité de surproduction, il faut intégrer la demande de biens dans son raisonnement. Il est en effet nécessaire de comprendre pourquoi la demande de biens peut être insuffisante pour absorber la production. C'est ce que fait ce chapitre.

La première étape dans la description de la demande est de décrire les composantes de la demande de biens. Nous l'avons déjà fait en comptabilité nationale. Nous avons en effet défini le PIB comme la somme des demandes de biens finaux, c'est-à-dire la consommation, l'investissement et les dépenses de l'Etat, en économie fermée. Cependant, nous n'avons pas cherché à décrire les déterminants de ces trois agrégats. C'est précisément à cela qu'est consacrée la première section de ce chapitre. Il s'agit ainsi d'y spécifier les *fonctions de comportement* des composantes de la demande.

La deuxième section prend ces fonctions de comportement comme données et les utilise pour construire le *diagramme à 45 degrés*. Il s'agit de la représentation graphique du premier modèle que nous pourrons utiliser pour comprendre l'évolution de l'économie dans le court terme et la politique économique.

Section 1 : Les composantes de la demande globale

Notre raisonnement se situe pour l'instant dans le cadre d'une économie fermée. La demande globale a donc trois composantes : la consommation (I), l'investissement (II) et les dépenses publiques (III).

I- La consommation

En 2009, la consommation des ménages et des ASBL représentait 52,4% du PIB en Belgique. C'était donc la composante principale du PIB. C'est pourquoi l'étude de la consommation est le point de départ de toute analyse de la demande.

On entend souvent dire que la consommation est le « moteur de la croissance ». On le verra dans la section 2, cette expression repose implicitement, et peut-être inconsciemment, sur une vision keynésienne du fonctionnement de l'économie. Pour pouvoir formaliser cette vision, il est nécessaire de déterminer les variables qui peuvent affecter la consommation.

On admettra facilement que le revenu est le déterminant principal de la consommation. Le débat porte sur le revenu à prendre en compte. Si on s'est au départ concentré sur le revenu courant (A), on a par la suite tenu compte des revenus futurs (B).

A- La consommation : une fonction du revenu courant

Le pilier de notre analyse sera la fonction de consommation keynésienne. Dans la Théorie Générale, Keynes ne donne pas de définition algébrique de cette fonction mais fonde son raisonnement sur une *loi psychologique fondamentale* qui a été interprétée et formalisée par la suite.

C'est dans le chapitre 8, section 3 de la théorie générale que Keynes définit sa loi psychologique fondamentale :

« La loi psychologique fondamentale sur laquelle nous pouvons nous appuyer en toute sécurité, à la fois a priori en raison de notre connaissance de la nature humaine et a posteriori en raison des enseignements de l'expérience, c'est qu'en moyenne et la plupart du temps, les hommes tendent à accroître leur consommation à mesure que leur revenu croît, mais non d'une quantité aussi grande que l'accroissement du revenu ».

On peut remarquer que la justification de l'évolution de la consommation en fonction du revenu est un peu sommaire. Si on veut spécifier la fonction de consommation, il faut interpréter la citation de Keynes.

Par commodité, nous allons retenir une forme très simple de la fonction de consommation, et vérifier si elle peut correspondre à l'idée de Keynes. La fonction de consommation la plus simple est linéaire :

$$\boxed{C = c.Y + C_0} \quad (1)$$

c est appelé la propension marginale à consommer (PmC). Ce paramètre mesure l'augmentation de la consommation provoquée par une augmentation du revenu d'une unité. La propension marginale à consommer est définie de façon générale comme la dérivée de la fonction de consommation par rapport au revenu. On vérifie que c'est bien le cas ici :

$$PmC = \frac{\partial C}{\partial Y} = c \quad (2)$$

C_0 mesure la consommation qui serait observée si le revenu était nul. On l'appelle donc la *consommation incompressible*. En réalité, le revenu n'est jamais nul. La consommation incompressible est donc surtout un concept théorique qui apparaît à cause de la forme retenue pour la fonction de consommation.

D'après la loi psychologique fondamentale, la consommation doit augmenter avec le revenu. Cela signifie que la propension marginale à consommer doit être positive : $c > 0$.

De plus, cette augmentation n'est pas « aussi grande que l'accroissement du revenu ». Cela implique que la propension marginale à consommer doit aussi être inférieure à un : $0 < c < 1$.

On considère en général qu'une valeur réaliste de c tourne autour de 0,8.

Dans d'autres passages de la théorie générale, Keynes suppose que la propension marginale à consommer est décroissante. Ce n'est pas le cas lorsque la consommation est une fonction linéaire puisque c est alors constant. Cette simplification ne va pas jouer de rôle important dans notre raisonnement, mais elle pourrait en jouer un si nous nous intéressions aux effets de la redistribution des revenus.

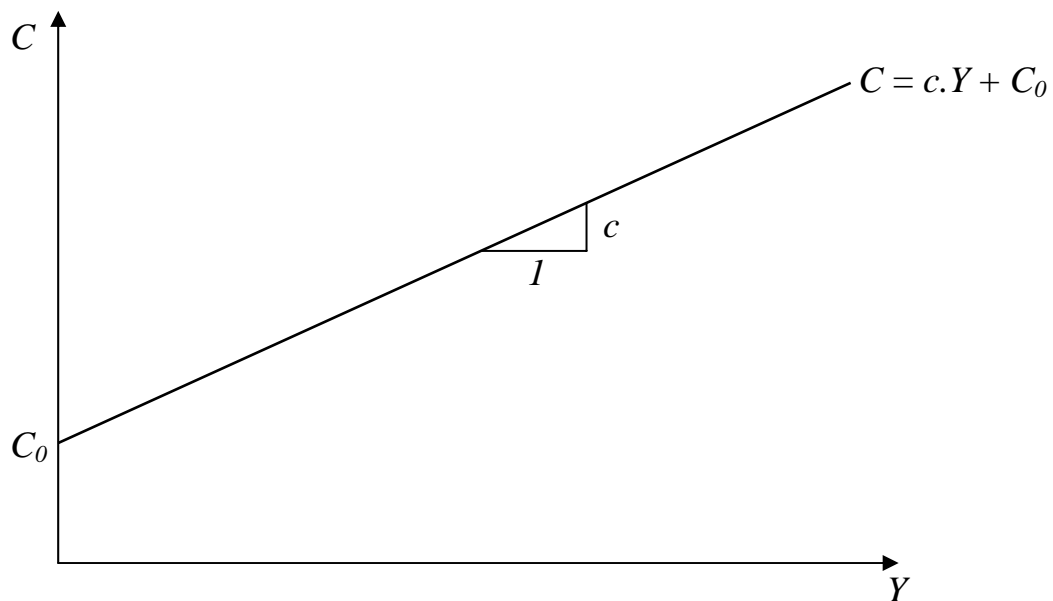
On définit également la *propension moyenne à consommer* (PMC). Il s'agit de la part du revenu qui est consacrée à la consommation :

$$PMC \equiv \frac{C}{Y} = c + \frac{C_0}{Y} \quad (3)$$

On constate alors que la propension moyenne à consommer est une fonction décroissante du revenu. Cela ne découle pas de la loi psychologique de Keynes mais ne la contredit pas non plus. En revanche, cette propriété est partagée avec des fonctions de consommation plus complexes qui impliqueraient une propension marginale à consommer décroissante.

La représentation graphique de la fonction de consommation que nous avons retenue est très simple. Il s'agit d'une droite de pente c et d'ordonnée à l'origine C_0 :

Graphique 1 : la fonction de consommation keynésienne



La décision de consommer est la contrepartie de la décision d'épargner. On peut donc déduire la fonction d'épargne de la fonction de consommation :

$$\begin{aligned} S &\equiv Y - C = Y - (c.Y + C_0) \\ \Rightarrow S &= (1 - c)Y - C_0 \end{aligned} \quad (4)$$

On constate que l'épargne ne dépend que du revenu courant. Tout se passe comme si l'épargne n'était que ce qui restait après la consommation. Les consommateurs ne tiennent ainsi pas compte de l'évolution à venir de leur revenu. Cela a été critiqué très tôt et a amené à nuancer les hypothèses keynésiennes.

NB : En toute rigueur, la consommation ne dépend pas du revenu mais du revenu disponible, Y^d , c'est-à-dire ce qui reste aux consommateurs une fois qu'ils ont payé leurs impôts. Comme nous n'avons pas introduit les impôts pour l'instant, nous avons pu nous passer de cette complication. Dans le cas contraire, il faudrait remplacer Y par $Y^d = Y - T$ dans la fonction de consommation. Nous aurons l'occasion d'utiliser cette formule à la fin du chapitre.

B- La consommation : une fonction des revenus futurs

On reproche à Keynes d'avoir utilisé une fonction de consommation ad hoc, c'est-à-dire créée spécialement pour servir son raisonnement, et de ne pas s'être embarrassé de fondements microéconomiques.

Or l'analyse microéconomique montre que la décision de consommer repose sur un arbitrage entre consommation présente et future. Lorsqu'ils font cet arbitrage, les consommateurs tiennent compte de leur contrainte budgétaire intertemporelle.⁴³ Par conséquent, la consommation devrait dépendre non seulement du revenu courant, comme le suppose Keynes, mais aussi de tous les revenus futurs et du taux d'intérêt.

L'effet du taux d'intérêt sur la consommation est ambigu d'un point de vue théorique et difficile à mettre en évidence empiriquement. Cela vient du fait que l'effet de revenu et l'effet de substitution d'une variation du taux d'intérêt ont tendance à se compenser.

Pour illustrer cette idée, considérons l'exemple d'un agent qui est un prêteur net et qui constate une diminution du taux d'intérêt.⁴⁴ Comme le taux d'intérêt a diminué, le rendement de l'épargne s'est également réduit, ce qui incite notre agent à épargner moins qu'auparavant, donc à consommer davantage. C'est l'effet de substitution.

Cependant, la diminution de la rémunération de l'épargne réduit le revenu des prêteurs, et notre agent se retrouve donc moins riche qu'auparavant. Il sera donc incité à réduire sa consommation. C'est l'effet de revenu.

On voit bien que l'effet de substitution et l'effet de revenu s'opposent. L'effet total d'une variation du taux d'intérêt est donc indéterminé et probablement d'une ampleur limitée. Par conséquent, nous négligerons cet effet, et supposerons que la consommation est indépendante du taux d'intérêt.

En revanche, le rôle des revenus futurs a été reconnu, dès les années cinquante. Les deux économistes à l'origine de cette prise de conscience sont Milton Friedman (prix Nobel 1976) et Franco Modigliani (prix Nobel 1985). Les raisonnements proposés par ces deux auteurs sont légèrement différents mais reposent tous les deux sur l'hypothèse que les

⁴³ C'est à l'économiste américain Irving Fisher que l'on doit la représentation de la décision d'épargner comme un arbitrage entre consommation présente et à venir telle qu'elle est de nos jours présentée dans la plupart des cours de microéconomie. Elle apparaît dans *The Nature of Capital and Income* (New York, Macmillan 1906), puis *The rate of interest* (New York, Macmillan 1907), et est résumée dans *The theory of interest* (New York, Macmillan, 1930). Le prestige d'Irving Fisher se ternit parce qu'il continua de soutenir que l'économie américaine était en bonne santé jusqu'au lendemain du crash boursier de 1929. Une interview où il défend ce point de vue pour les nouvelles de 1929 est disponible à l'adresse suivante : http://fr.youtube.com/watch?v=MTCKxye9_so.

⁴⁴ Vous pourrez considérer le cas d'un emprunteur net et d'une augmentation du taux d'intérêt à titre d'exercice.

consommateurs sont prévoyants. Par conséquent, ils ne peuvent déterminer leur consommation sans tenir compte de l'évolution de leurs revenus.

Friedman⁴⁵ argue que si le consommateur est prévoyant, il ne fondera pas ses décisions de consommer sur son revenu courant, qui peut être très variable, mais sur l'évolution prévisible de son revenu. Plus précisément, il convient de distinguer le *revenu permanent* (Y^P) et le *revenu transitoire* (Y^T). Le revenu permanent est en quelque sorte le revenu que le consommateur peut anticiper en moyenne. Le revenu transitoire est un revenu accidentel, par exemple un gain à la loterie, ou des heures supplémentaires dues à une relance temporaire de l'activité. Le revenu courant est la somme du revenu permanent et du revenu transitoire ($Y = Y^P + Y^T$). Le revenu courant peut dans certains cas être inférieur au revenu permanent, par exemple pendant une récession transitoire.

Dans tous les cas, si le consommateur est prévoyant, il fondera sa consommation sur son revenu permanent. S'il bénéficie d'un revenu courant supérieur à son revenu permanent (le revenu transitoire est positif), il choisira d'épargner davantage. Si le revenu courant est plus faible que le revenu permanent (le revenu transitoire est négatif), le consommateur empruntera ou puisera dans son épargne. Par conséquent, la seule relation stable qui existe est celle qui relie la consommation au revenu permanent : $C = k.Y^P$. La relation entre le revenu transitoire et la consommation doit être nulle ou très faible.

L'hypothèse du revenu permanent suggère donc que la fonction de consommation keynésienne est mal spécifiée. On peut toujours l'estimer par des méthodes statistiques mais elle ne sera pas utilisable.

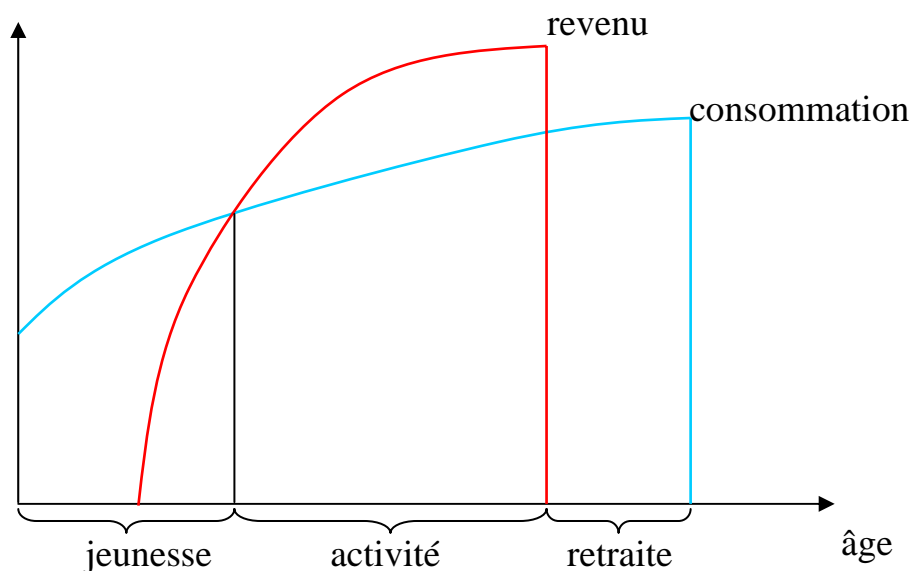
Franco Modigliani⁴⁶ va plus loin que la théorie du revenu permanent, en faisant remarquer que l'évolution du revenu au cours de la vie est largement prévisible. En effet, un consommateur va commencer son existence avec des revenus nuls, puisqu'il ne travaille pas, puis verra ses revenus augmenter avec l'âge et l'expérience. A sa retraite, ses revenus deviendront nuls et il devra puiser dans son épargne pour continuer à consommer. Si le consommateur est prévoyant et qu'il souhaite maintenir une consommation à peu près constante, il tiendra compte de l'évolution de son épargne au cours de sa vie. C'est la base de *l'hypothèse du cycle de vie*.

⁴⁵ M. Friedman *A theory of the consumption function*, Princeton University Press, Princeton, 1957.

⁴⁶ A.K. Ando et F. Modigliani « The life-cycle hypothesis of saving : aggregate implications and tests », *American Economic Review*, vol 53 n°1, p.55-84, 1963. F. Modigliani et R. Brunberg « Utility analysis and the consumption function : an interpretation of cross-section data », in K. Kurihara (dir.), *Post-keynesian economics*, Rutgers University Press, New Brunswick, 1954.

On peut alors décrire l'évolution de son épargne et de son patrimoine :

Graphique 2 : le cycle de vie



Au cours de sa jeunesse, le consommateur a un revenu très faible et va donc emprunter ou puiser dans son patrimoine. Lorsqu'il se met à travailler, son revenu augmente, il finit donc par pouvoir couvrir sa consommation courante. Le reste lui sert à rembourser ses dettes et à se constituer une épargne. Lorsqu'il part à la retraite, le consommateur va puiser dans son épargne pour maintenir sa consommation jusqu'à son décès.

Cette théorie a deux implications essentielles au niveau agrégé :

- elle suggère qu'il existe une relation entre la structure démographique d'un pays et son taux d'épargne. Les pays très jeunes ou très vieux auront tendance à moins épargner.

- surtout, elle suggère que les fluctuations du revenu de court terme n'ont qu'une importance secondaire dans la détermination de la consommation. Elle complète donc l'hypothèse du revenu permanent de Friedman.

Il existe donc des arguments très forts pour remettre en cause la relation entre le revenu courant et la consommation courante. Cette relation reste cependant à la base de beaucoup de théories macroéconomiques et de modèles agrégés. Il y a plusieurs raisons à cela :

- lorsqu'on considère qu'un consommateur n'est contraint que par une contrainte budgétaire intertemporelle qui n'est que la somme actualisée de ses revenus futurs, on fait

implicitement l'hypothèse qu'il a un accès illimité au crédit. Il peut en effet emprunter pour consommer aujourd'hui et rembourser lorsque ses revenus seront plus élevés.

Or le marché du crédit ne fonctionne pas de façon parfaite, ce qui se traduit par le fait que certains consommateurs se voient refuser des prêts alors qu'ils seraient tout à fait capables de les rembourser. On dit qu'ils sont *rationnés*. Ils subissent alors une *contrainte de liquidité*.

Lorsqu'un consommateur est rationné, sa consommation dépend principalement de son revenu courant. On retrouve alors des comportements compatibles avec la fonction de consommation keynésienne.

- tant la théorie du revenu permanent que la théorie du cycle de vie amènent à distinguer les variations temporaires et permanentes du revenu.

Seules les variations permanentes du revenu vont se répercuter de façon proportionnelle sur la consommation car elles correspondent bien à une augmentation du niveau de vie. Les variations temporaires du revenu vont être en partie absorbées grâce à l'épargne. La consommation va donc répondre de façon moins que proportionnelle aux variations du revenu courant.

=> la fonction de consommation keynésienne qui implique que la propension moyenne à consommer diminue (elle est donc moins que proportionnelle) avec le revenu est une approximation convenable pour étudier les fluctuations temporaires du revenu.

Comme nous nous intéressons aux fluctuations, donc au court terme, cette approximation n'est pas abusive. Elle l'aurait été si nous l'avions appliquée à l'étude de la croissance. C'est pourquoi nous avons supposé que la consommation était proportionnelle au revenu dans les deux chapitres précédents.

II- L'investissement

L'investissement est l'un des agrégats les plus variables. Compte tenu de sa part dans le PIB, on doit s'attendre à ce qu'il soit à l'origine d'une partie des fluctuations conjoncturelles. En effet, l'investissement (Formation Brute de Capital Fixe) représentait 24,7% du PIB en Belgique en 2008. Par conséquent, une explication convaincante des fluctuations passe nécessairement par une description de l'investissement.

La variabilité de l'investissement pourrait laisser croire qu'il dépend de décisions arbitraires des entreprises et d'effets de mode. Même si le « moral des investisseurs » joue peut-être un rôle, il ne supprime pas les contraintes auxquelles les entreprises sont confrontées

ni l'objectif de profit des actionnaires. On peut donc essayer de comprendre les déterminants de l'investissement grâce aux outils de l'analyse économique.

Comme la consommation, l'investissement est une décision intertemporelle puisqu'elle affecte les capacités de production des entreprises de façon durable. Il n'est donc pas étonnant que le niveau du taux d'intérêt (A) et les anticipations (B) soient les déterminants principaux de l'investissement.

A- L'investissement : une fonction du taux d'intérêt

La microéconomie nous apprend que le *coût d'opportunité* de l'investissement est constitué par le taux d'intérêt des titres.

En effet, un entrepreneur qui cherche à investir un euro va se demander s'il doit l'utiliser pour acheter des titres, par exemple des obligations, ou bien pour augmenter le stock de capital de son entreprise. Il va donc comparer le rendement de l'obligation, le taux d'intérêt, avec le rendement du capital, c'est-à-dire la productivité marginale du capital.

Lorsqu'on détermine l'équilibre du producteur, on montre que le stock de capital de l'entreprise va être déterminé de façon à ce que la productivité marginale du capital soit égale au taux d'intérêt, c'est-à-dire au rendement de l'obligation. Comme la productivité marginale du capital est décroissante, la quantité de capital utilisée va être une fonction décroissante du taux d'intérêt. L'investissement sera donc lui aussi une fonction décroissante du taux d'intérêt.

Il existe une autre raison pour laquelle l'investissement dépend du taux d'intérêt. C'est le *canal du crédit*. Le canal du crédit apparaît en situation de *rationnement du crédit*, c'est-à-dire dans une situation où les banques ne satisfont pas l'ensemble des demandes de prêts qui leur sont adressées. Cela signifie donc que les entreprises ne pourront pas financer tous les investissements qu'elles souhaitent réaliser.

Or le taux d'intérêt représente pour les banques le coût marginal des prêts. Si le taux d'intérêt augmente, les banques réduiront encore leur offre de prêts. Elles empêcheront donc davantage d'entreprises de financer leurs investissements. Le canal du crédit renforce donc l'impact négatif du taux d'intérêt sur l'investissement.

B- L'investissement : une fonction de la demande et des anticipations

Une entreprise investit pour augmenter ses capacités de production.⁴⁷ Quel que soit le taux d'intérêt, il ne lui sert à rien d'augmenter son stock de capital si elle ne parvient pas à écouler sa production.

En revanche, si la demande augmente et que l'entreprise n'est pas capable de la satisfaire, elle laissera passer des occasions d'augmenter ses profits. On voit donc que l'investissement va être une fonction croissante de la demande de biens. A l'échelle macroéconomique, la demande de biens dépend du revenu donc de la production. On pourra donc supposer que l'investissement est une fonction de la production.

Enfin, lorsqu'une entreprise investit, elle acquiert du capital qu'elle pourra utiliser pendant plusieurs années. Elle doit donc tenir compte de l'effet de ce capital non seulement sur ses profits courants, mais sur ses profits à venir. L'investissement va donc être une fonction des profits anticipés par l'entreprise. Ces profits anticipés dépendent à leur tour de l'évolution anticipée de la demande, des taux d'intérêts, du prix des inputs etc. mais aussi dans une large mesure de l'optimisme des investisseurs.⁴⁸

En résumé, l'investissement est donc une fonction décroissante du taux d'intérêt, et croissante du revenu et des profits anticipés :

$$I = I\left(r^-, Y^+, profits^e\right) \quad (5)$$

III- Les dépenses de l'Etat

Dans une économie fermée, les dépenses publiques constituent la dernière composante de la demande. Cette composante est quantitativement importante. Elle représentait 21,3% du PIB en 2009 en Belgique.

Par ailleurs, c'est cette composante qui peut être utilisée par le gouvernement à des fins de politique économique. Le niveau des dépenses et des prélèvements publics est en effet un choix politique, même s'il doit respecter certaines contraintes. En effet, il existe des règles institutionnelles qui contraignent l'évolution des dépenses publiques. Dans l'Union européenne par exemple, les gouvernements doivent respecter le Pacte de Stabilité et de

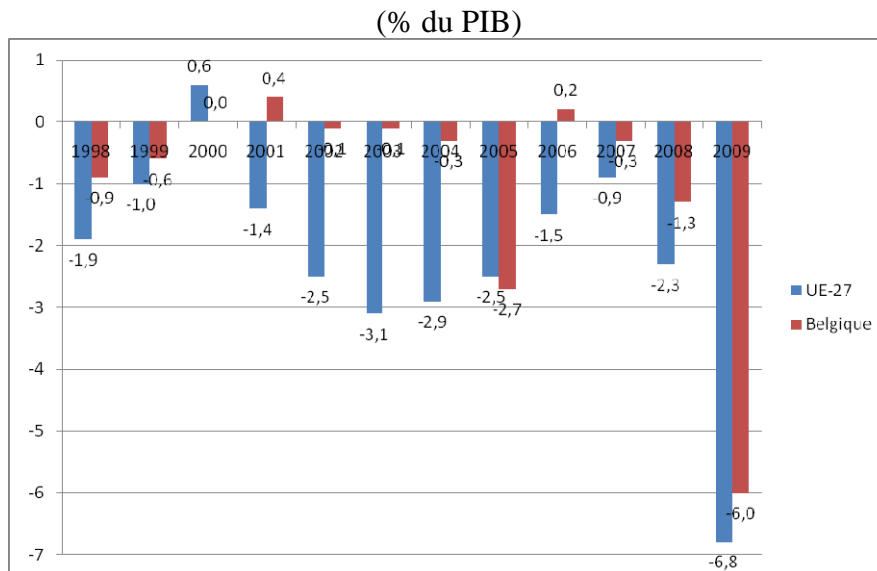
⁴⁷ L'investissement dont il est question ici n'inclut pas la variation des stocks. Il correspond à ce que la comptabilité nationale appelle la *formation brute de capital fixe* (FBCF).

⁴⁸ C'est ce que Keynes appelait les « esprits animaux » des investisseurs.

Croissance qui leur donne des objectifs en termes de déficit (3% du PIB) et d'endettement (60% du PIB).

Le graphique ci-dessous compare le déficit public de la Belgique avec celui de l'Union Européenne dans son ensemble.

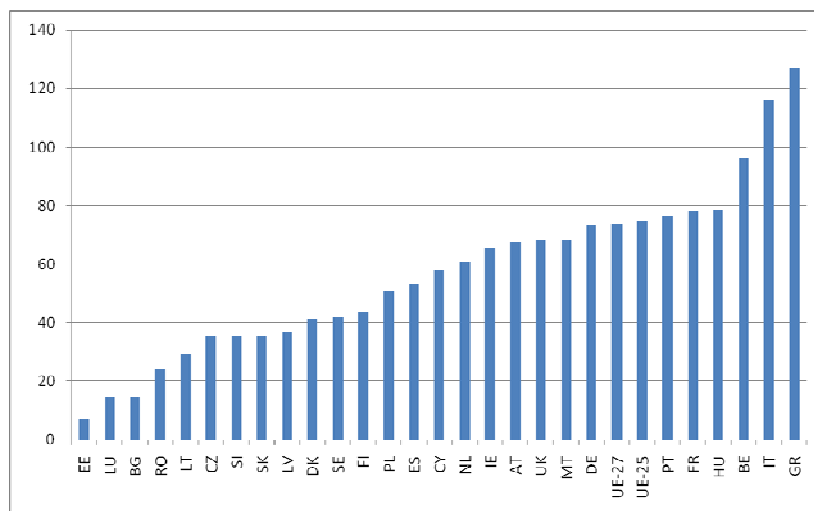
Déficit public de la Belgique et déficit public consolidé des membres de l'Union Européenne



Source : Eurostat

On constate que le budget des deux entités est en général en déficit, ce qui se traduit par une augmentation de la dette. Tous les états ne partent cependant pas avec le même stock de dette, ce qui peut contraindre plus ou moins leur politique budgétaire, comme le montre le graphique ci-dessous :

Stock de dette des pays membre de l'Union Européenne en 2009 (pourcentage du PIB)



Source : Eurostat

Les notes du cours ne sont transmises qu'à titre indicatif.

Dans tous les cas, tous les gouvernements doivent en principe respecter leur contrainte budgétaire intertemporelle s'ils ne veulent pas se retrouver en situation de devoir faire défaut sur leur dette.

Nous ne tiendrons pas explicitement compte de ces contraintes dans nos raisonnements. Il faudra cependant les garder à l'esprit pour comprendre pourquoi on ne peut pas augmenter indéfiniment le déficit ou la dette. Vous aurez l'occasion d'étudier ces questions plus en détail dans la suite de vos études.

Pour l'instant, nous nous contenterons de supposer que les dépenses publiques et les impôts sont choisis librement par le gouvernement et constituent donc des paramètres du modèle. Dans la prochaine section, nous verrons comment une modification de ces paramètres peut modifier l'équilibre sur le marché des biens.

Section 2 : L'équilibre sur le marché des biens

Inventé par Paul Samuelson en 1939, le diagramme à 45 degrés décrit l'équilibre sur le marché des biens. Il peut être considéré comme le premier modèle d'inspiration keynésienne. En un sens, on pourrait dire qu'il est même le premier modèle macroéconomique grand public, puisque Samuelson le présente dès la première édition, en 1948, de son manuel, qui reste encore aujourd'hui un best-seller.⁴⁹

Il est donc important de connaître son fonctionnement (A). On pourra alors faire apparaître l'un des résultats les plus novateurs de l'analyse keynésienne : le multiplicateur (B). Ce résultat est tellement important qu'on parle aussi bien du modèle du multiplicateur que du diagramme à 45 degrés pour désigner ce modèle.

I- Le diagramme à 45 degrés

Le diagramme à 45 degrés repose sur l'égalité entre l'offre et la demande sur le marché des biens. Cependant, cette égalité est obtenue de façon très différente de celle qui est supposée par la microéconomie.

En effet, le diagramme repose sur l'hypothèse que les prix sont fixes et que l'économie se trouve en situation de sous-emploi. Cela implique que l'ajustement de l'offre et de la demande ne passe pas par l'ajustement des prix mais par celui des quantités produites.

⁴⁹ P.A. Samuelson *Economics: an introductory analysis*, McGraw-Hill, 1948.

En situation de sous-emploi, il existe par définition des capacités excédentaires. Les producteurs pourraient donc augmenter facilement leur production s'ils pouvaient l'écouler. Ils sont cependant contraints par la demande qui est insuffisante. On dit que la demande constitue le *côté court* du marché.

Dans ces circonstances, c'est la demande qui va déterminer la production. C'est pourquoi on peut étudier l'équilibre de l'économie sans décrire les conditions de production ni définir de fonction d'offre.

La demande globale correspond à la somme de la consommation, de l'investissement et des dépenses publiques. On peut donc écrire :

$$D^G = C + I + G \quad (6a)$$

Parmi ces trois composantes, deux sont déterminées en dehors du marché des biens. Nous négligeons en effet pour l'instant le rôle des anticipations. Par ailleurs, comme l'économie est en situation de surcapacité, les entreprises pourront augmenter leur production sans avoir besoin d'augmenter leurs capacités de production. L'investissement dépend alors surtout du taux d'intérêt qui est déterminé sur le marché financier et est donc exogène au modèle.

Par conséquent, l'investissement peut être considéré comme une constante : $I = I_0$

De même, les dépenses publiques sont déterminées par le gouvernement. Elles sont donc elles aussi un paramètre du modèle : $G = G_0$

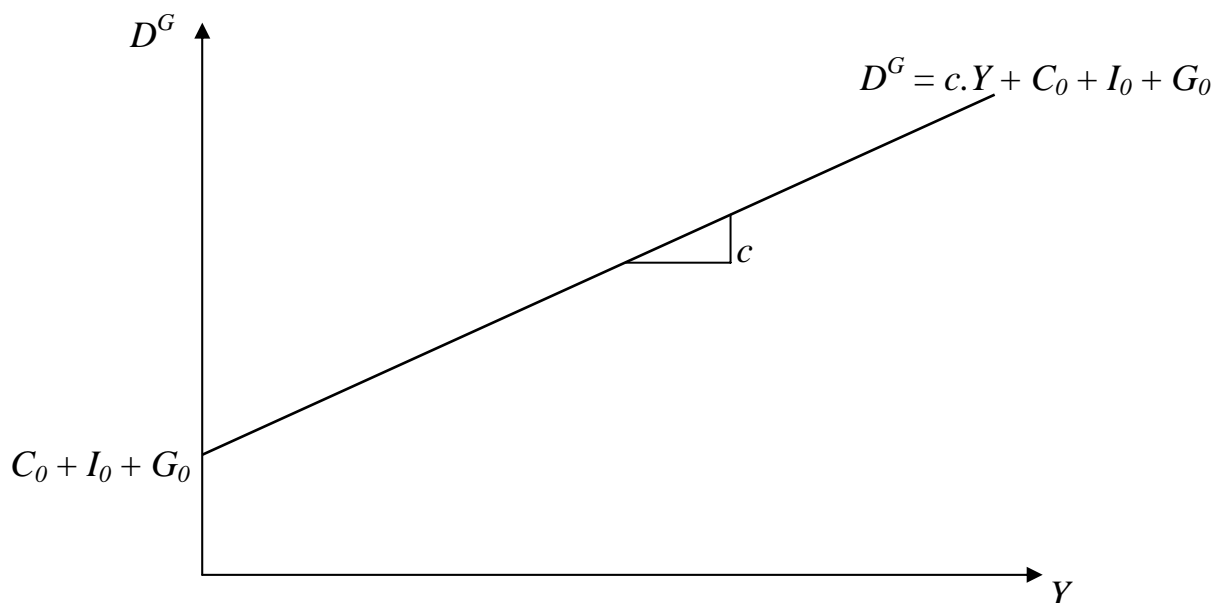
Par conséquent, la seule composante de la demande qui soit endogène est la consommation. Nous supposons qu'elle est décrite par une fonction de consommation keynésienne : $C = c.Y + C_0$

On peut alors réécrire la demande globale en remplaçant chacune de ses composantes par sa valeur :

$$D^G = c.Y + C_0 + I_0 + G_0 \quad (6b)$$

On peut alors représenter la demande globale en fonction du revenu de façon très simple :

Graphique 3a : la demande globale



La demande globale est donc représentée par une droite croissante dont l'ordonnée à l'origine est égale à la somme de l'investissement, des dépenses publiques et de la consommation incompressible, et la pente à la propension marginale à consommer. Cette pente est donc inférieure à un selon nos hypothèses.

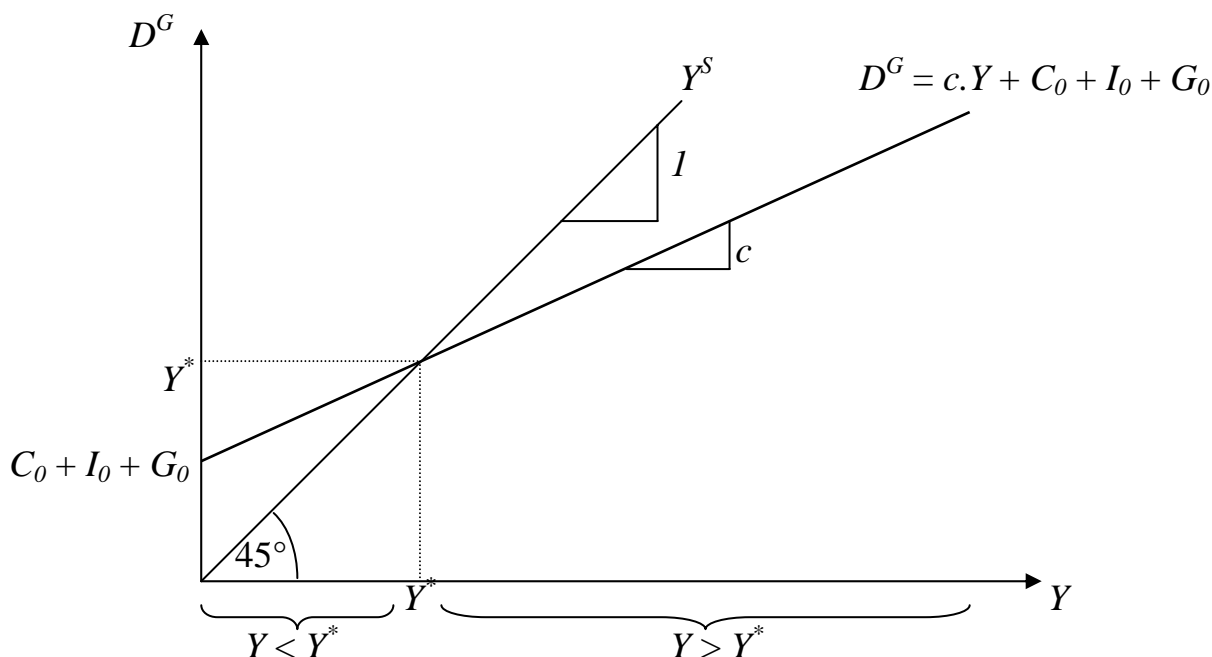
L'équilibre sur le marché des biens sera atteint lorsque l'offre sera tout juste égale à la demande. Les producteurs vont en effet adapter leur production de façon à satisfaire la demande et à éviter de se retrouver avec des excédents.

Au niveau agrégé, l'offre de biens est égale à la production : $Y^s = Y$.

Sur le graphique 3a, cette équation correspond à une droite passant par l'origine et de pente égale à un. Si on utilise la même échelle en abscisse et en ordonnée, cette droite correspondra à la première bissectrice du repère et elle formera un angle de 45 degrés avec les deux axes. C'est pourquoi on parle de diagramme à 45 degrés.

Une fois que l'offre est représentée sur le graphique, on détermine facilement l'équilibre sur le marché des biens. Il correspond à l'intersection entre la courbe de demande et la première bissectrice, qui correspond à la courbe d'offre.

Graphique 3b : l'équilibre sur le marché des biens



Comme la courbe de demande a une pente inférieure à un et que la première bissectrice a par définition une pente égale à un alors que son ordonnée à l'origine est inférieure à celle de la courbe de demande, les deux courbes vont forcément se couper. Il y a donc forcément un équilibre.

Le revenu d'équilibre, noté Y^* , peut se mesurer tant en abscisse qu'en ordonnée.

On peut se demander ce qui se passerait si les producteurs avaient surestimé la demande globale et produit une quantité de biens supérieure à Y^* . On voit que dans ce cas la demande de biens est inférieure à l'offre. Les producteurs ne peuvent alors pas écouler leur production. Ils voient donc augmenter leurs stocks. Par la suite, ils réduiront leur production, ce qui les rapprochera de l'équilibre.

Si les producteurs ont sous-estimé la demande, il apparaît une demande excédentaire. Les producteurs vont donc devoir puiser dans leurs stocks pour satisfaire la demande. Par la suite, ils augmenteront leur production, ce qui les rapprochera aussi de l'équilibre.

On peut alors faire plusieurs remarques :

- l'équilibre sur le marché des biens est stable. En effet, si on s'en écarte à cause d'une erreur de prévision de la part des producteurs, leur réaction spontanée rapprochera l'économie de l'équilibre.

Les notes du cours ne sont transmises qu'à titre indicatif.

- l'ajustement à très court terme est assuré par les variations des stocks. On retrouvera alors l'égalité entre investissement et épargne, puisque les variations des stocks font partie de l'investissement.

- surtout, il n'existe aucun mécanisme qui permette d'affirmer que le revenu d'équilibre correspond au revenu de plein emploi. Par conséquent, si pour une raison ou une autre la demande de biens ne permet pas le plein emploi, l'économie restera durablement en situation de sous-emploi.

On obtient là le premier résultat fondamental de l'analyse keynésienne : l'économie peut durablement se trouver en situation de sous emploi. Cette conclusion s'oppose à celle de la théorie économique telle qu'elle existait dans les années trente. Selon cette théorie, si tous les marchés fonctionnent de façon concurrentielle, ils seront tous en équilibre et le sous emploi ne peut être que transitoire.

Jusqu'à présent, nous avons eu recours à un raisonnement graphique. Cependant, on peut aussi écrire la condition d'équilibre sur le marché des biens et la résoudre pour déterminer le niveau d'équilibre du revenu de façon algébrique. Il suffit pour cela de réécrire la condition d'équilibre et de remplacer l'offre et la demande de biens par leur valeur :

$$Y^S = D^G$$

$$Y = c.Y + C_0 + I_0 + G_0$$

$$Y - c.Y = C_0 + I_0 + G_0$$

$$Y^* = \frac{C_0 + I_0 + G_0}{1 - c}$$

(7)

Le revenu d'équilibre est égal à la somme des dépenses exogènes aux marché des biens divisée par la propension marginale à épargner ($s = 1 - c$). Nous allons à présent utiliser cette expression pour définir le multiplicateur des dépenses.

II- Le multiplicateur

L'expression du revenu d'équilibre montre que le revenu est une fonction croissante des dépenses. Par conséquent, l'augmentation de l'une des composantes de la demande se traduirait par une augmentation du revenu d'équilibre. Mais quel serait le montant de cette augmentation ?

C'est précisément ce que permet de déterminer le multiplicateur. Il mesure plus précisément le rapport entre la variation du revenu et la variation des dépenses qui l'a provoquée. Intuitivement, on peut dire qu'il mesure l'augmentation du revenu provoquée par un euro d'augmentation des dépenses.

Pour le calculer, nous allons étudier plus précisément l'impact sur le revenu d'une augmentation des dépenses budgétaires, mais on pourrait en faire autant avec les autres composantes de la demande agrégée. Supposons par exemple que le gouvernement lance un programme de grands travaux, d'augmentation des allocations ou encore de revalorisation de la rémunération des fonctionnaires. Cette politique se traduit par une augmentation de G . On note ΔG la variation des dépenses budgétaires, et ΔY la variation du revenu. On peut alors écrire :

$$\Delta Y^* = \frac{\partial Y^*}{\partial G_0} \cdot \Delta G \quad (8)$$

Ce qui d'après l'expression (7) est équivalent à :

$$\Delta Y^* = \frac{1}{1-c} \cdot \Delta G \quad (9)$$

Le multiplicateur est donc égal à la dérivée de la valeur du revenu d'équilibre par rapport aux dépenses. On retiendra donc :

$$\boxed{\text{multiplicateur} = \frac{1}{1-c}} \quad (10)$$

Encadré 1 : Une autre méthode pour déterminer le multiplicateur

Puisque par définition $\Delta G \equiv G_1 - G_0$, on peut aussi écrire $G_1 = G_0 + \Delta G$, où G_0 est le niveau initial des dépenses et G_1 son niveau final. On peut alors exprimer le nouveau niveau du revenu d'équilibre en fonction de la variation des dépenses budgétaires :

$$Y_1^* = \frac{C_0 + I_0 + G_1}{1 - c} \quad (\text{E1})$$

$$Y_1^* = \frac{C_0 + I_0 + (G_0 + \Delta G)}{1 - c} \quad (\text{E2})$$

$$Y_1^* = \frac{C_0 + I_0 + G_0}{1 - c} + \frac{\Delta G}{1 - c} \quad (\text{E3})$$

Si on note Y_0^* le niveau initial du revenu, Y_1^* son niveau final et ΔY la variation du revenu, on peut alors réécrire l'expression ci-dessus :

$$Y_1^* = Y_0^* + \frac{\Delta G}{1 - c} \quad (\text{E4})$$

$$\Rightarrow \Delta Y = Y_1^* - Y_0^* = \frac{1}{1 - c} \Delta G \quad (\text{E5})$$

On constate alors bien que la variation du revenu est proportionnelle à la variation des dépenses budgétaires. Le coefficient multiplicateur s'écrit alors également :

$$\text{multiplicateur} = \frac{1}{1 - c}$$

Attention : La valeur du multiplicateur obtenue par cette méthode n'est égale à celle obtenue en dérivant la valeur d'équilibre du revenu que lorsque toutes les relations du modèle sont linéaires. Quand ce n'est pas le cas, seule la méthode de la dérivée est correcte.

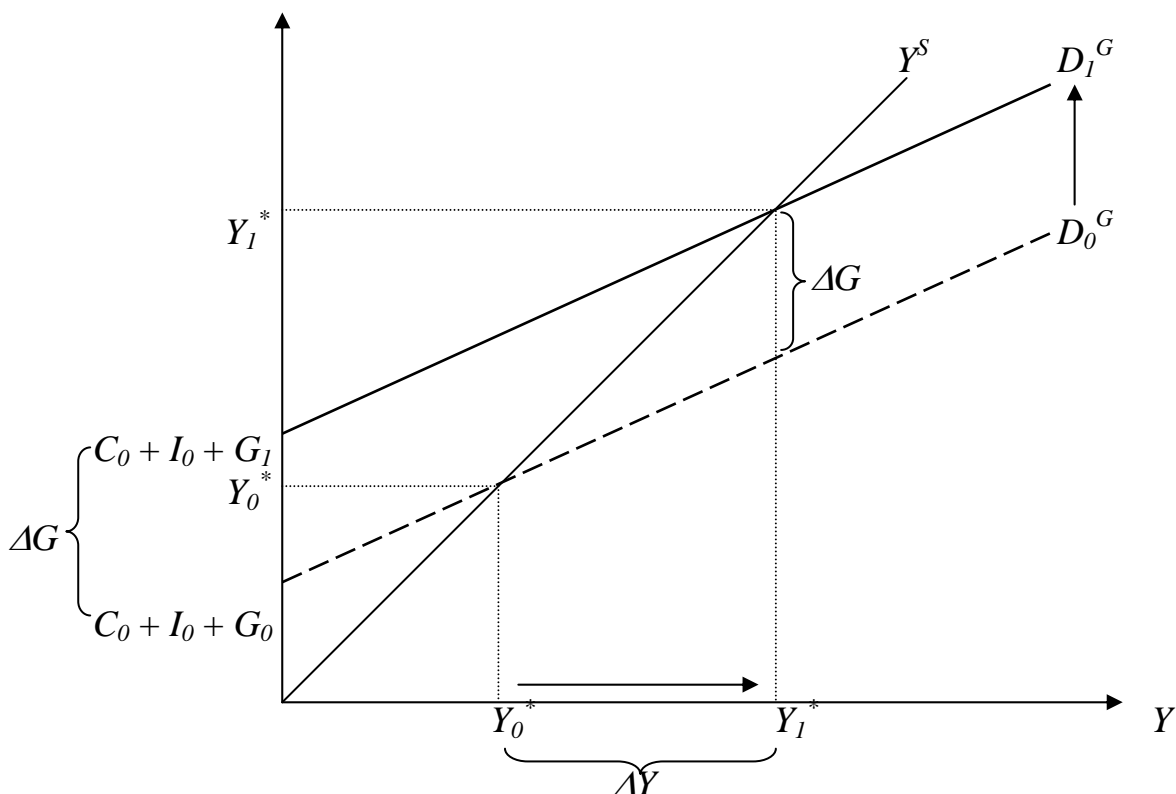
Quelle que soit la méthode utilisée, on constate que le multiplicateur budgétaire est égal à un divisé par un moins la propension marginale à consommer, c'est-à-dire par la propension marginale à épargner.

Comme la propension marginale à épargner est comprise entre zéro et un, *le multiplicateur est supérieur à un.*

Ce résultat est essentiel car il montre qu'un euro de dépenses publiques supplémentaires va se traduire par une augmentation du revenu supérieure à un euro. L'Etat peut donc relancer l'économie grâce à l'augmentation de ses dépenses.

On peut retrouver le même résultat de façon graphique en se rappelant qu'une augmentation des dépenses publiques se traduit par une translation de la courbe de demande vers le haut.

Graphique 4 : Une augmentation des dépenses publiques

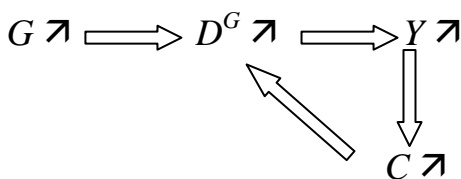


On observe ici aussi que l'augmentation du revenu est supérieure à celle des dépenses publiques. Lorsque le graphique est précis, on peut vérifier que le rapport entre ΔY et ΔC est exactement égal à l'inverse de la propension marginale à épargner, qui est égale au complément à un de la pente de la courbe de demande globale.

Ce résultat est tellement étonnant qu'il faut le décortiquer pour bien en comprendre l'origine. Comment une augmentation d'un euro des dépenses publiques peut-elle se traduire par une augmentation de plus d'un euro du revenu ?

La réponse à cette question est à trouver dans les composantes de la demande et en particulier dans la consommation. Nous avons en effet supposé que la consommation était une fonction croissante du revenu. Par conséquent, lorsque le gouvernement augmente la demande et le revenu d'un euro, il relance aussi la consommation qui augmente de $1 \times c$ euro. La demande augmente encore, ce qui augmente le revenu, ce qui augmente la consommation etc.

Le multiplicateur keynésien repose donc sur un cercle vertueux :



Les notes du cours ne sont transmises qu'à titre indicatif.

A chaque tour du cercle, la demande augmente d'un montant supplémentaire inférieur à l'augmentation précédente du revenu parce qu'une partie de ce revenu supplémentaire est épargnée. Elle sort donc du marché des biens. C'est pourquoi l'épargne est parfois considérée comme une fuite.

Cette fuite est d'autant plus importante que la propension marginale à épargner est élevée. C'est pourquoi le multiplicateur est une fonction décroissante de la propension marginale à épargner.

Un exemple numérique permet de mieux visualiser ce qui se passe. Supposons que la propension marginale à consommer soit égale à 0,8. La propension marginale à épargner est donc égale à 0,2 et le multiplicateur à 5.

Supposons que le gouvernement augmente ses dépenses de cent euros. Compte tenu de la valeur du multiplicateur, le revenu devrait donc augmenter en tout de $100 \times 5 = 500$ €. On peut aussi décomposer le phénomène. Le revenu augmente d'abord de 100 €, mais la consommation augmente de $0,8 \times 100 = 80$ €. Le revenu augmente alors de $0,8 \times 80 = 64$ € etc. Le tableau ci-dessous décrit l'évolution du revenu pour un nombre plus élevé de cycles :

Tableau 1 : Simulation d'une relance budgétaire

	ΔY	ΔC	Variation cumulée du revenu
1	100	80	100
2	80	64	180
3	64	51,2	244
4	51,20	40,96	295,20
5	40,96	32,77	336,16
6	32,77	26,21	368,93
7	26,21	20,97	395,14
8	20,97	16,78	416,11
9	16,78	13,42	432,89
10	13,42	10,74	446,31
...
20	1,44	1,15	494,24
...
50	0,001784	0,001427	499,99

La première colonne indique le nombre d'itérations. Les deux colonnes suivantes donnent l'augmentation du revenu et de la consommation au cours de la dernière itération. La dernière colonne indique la variation cumulée du revenu.

Les notes de cours ne sont transmises qu'à titre indicatif.

On voit que l'augmentation du revenu et de la consommation au cours de chaque cycle diminue, mais qu'au bout d'un nombre suffisant de cycles, la variation cumulée du revenu est devenue très proche de 500.

Le résultat du multiplicateur a été obtenu en se concentrant sur les dépenses publiques, mais il est applicable à n'importe quelle autre composante de la demande. On constate en effet que l'investissement et la demande incompressible occupent des places symétriques à celle des dépenses publiques dans l'expression du revenu d'équilibre (7). Par conséquent, leur effet sur le revenu sera le même que celui des dépenses publiques.

C'est pourquoi on parle autant du *multiplicateur de dépenses* que du multiplicateur budgétaire.

Le multiplicateur de dépenses et le multiplicateur budgétaire diffèrent cependant à partir du moment où on tient compte du financement des dépenses publiques.

Jusqu'à présent, nous avons supposé que le gouvernement augmentait ses dépenses sans se soucier de les financer. La relance se faisait donc uniquement grâce à un creusement du déficit public. L'Etat ne peut cependant pas aggraver indéfiniment le déficit s'il veut respecter sa contrainte budgétaire.

On peut se demander si le multiplicateur keynésien fonctionnerait dans le cas où le gouvernement maintiendrait son budget équilibré. Il faudrait pour cela qu'il augmente les impôts en même temps que ses dépenses. Il reprendrait alors d'une main ce qu'il donnerait de l'autre. Cela affecterait forcément les effets d'une relance budgétaire. Pour le vérifier, nous allons supposer que le budget est équilibré en permanence, donc que les impôts couvrent exactement les dépenses : $G = T$.

Pour que les impôts soient correctement intégrés dans notre raisonnement, il faut tenir compte de leur effet sur le revenu disponible ($Y^d = Y - T$). La consommation sera alors donnée par :

$$C = c.(Y - T) + C_0 \quad (10)$$

A cette différence près, on détermine le revenu d'équilibre comme précédemment, en partant de l'égalité entre le revenu et la demande globale de biens :

$$\begin{aligned} Y^S &= D^G \\ Y &= c.(Y - T) + C_0 + I_0 + G \\ Y - c.Y &= -cT + C_0 + I_0 + G \end{aligned}$$

$$Y^* = \frac{C_0 + I_0 + G - cT}{1 - c} \quad (11)$$

Or on sait à présent que les impôts couvrent les dépenses ($G = T$). On peut donc réécrire l'expression du revenu d'équilibre de la façon suivante :

$$Y^* = \frac{C_0 + I_0 + G - cG}{1 - c} = \frac{C_0 + I_0}{1 - c} + \frac{(1 - c)G}{1 - c}$$

$$\boxed{Y^* = \frac{C_0 + I_0}{1 - c} + G} \quad (12)$$

On constate alors que le multiplicateur budgétaire est égal à un. Ce résultat est dû à l'économiste norvégien Trygve Haavelmo (prix Nobel 1989).⁵⁰ Le théorème d'Haavelmo stipule que le multiplicateur budgétaire est égal à un lorsque le budget de l'Etat est équilibré.

Ce résultat est un résultat très optimiste. Il signifie qu'on peut relancer l'économie par la politique budgétaire même en conservant un budget équilibré. On peut donc atteindre le plein emploi uniquement par la politique budgétaire sans mettre en cause la soutenabilité de la dette publique.

Conclusion

L'étude de l'équilibre du marché des biens en situation de sous-emploi montre que le revenu global peut rester durablement à un niveau inférieur à son niveau de plein emploi.

Une augmentation des dépenses publiques peut alors augmenter la demande donc la production. L'effet de cette augmentation sera amplifié par le mécanisme du multiplicateur. Le théorème de Haavelmo montre de plus que la politique budgétaire reste efficace, même quand le budget est équilibré.

Pourtant toutes ces conclusions sont fragiles parce qu'elles ne reposent que sur un raisonnement d'équilibre partiel. D'autres marchés importants sont négligés alors qu'ils pourraient être affectés par la politique budgétaire et affecter en retour le marché des biens. Nous avons par exemple vu que l'investissement dépendait du taux d'intérêt. Or le taux d'intérêt est déterminé sur les marchés financiers.

⁵⁰ T. Haavelmo "Multiplier effects of a balanced budget", *Econometrica*, vol 12 n°4, p.311-318, 1945.

Le prochain chapitre décrit donc le fonctionnement des marchés financiers et du marché de la monnaie.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 6 (section 6.1 uniquement) et chapitre 13.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 10 (section 10.1 uniquement).

Pour les passionnés

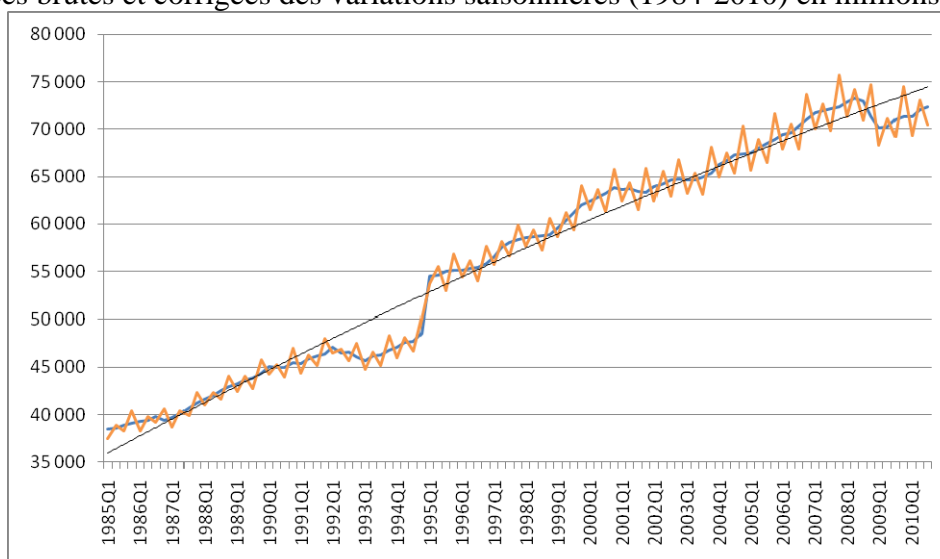
Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : Chapitre 8.

Villieu P. : *Macroéconomie : consommation et épargne*, Repères, La Découverte, Paris, 1997.

Villieu P. : *Macroéconomie : l'investissement*, Repères, La Découverte, Paris, 2000.

Annexes

Graphique A1 : Evolution du PIB de la Belgique aux prix de 2000
Données brutes et corrigées des variations saisonnières (1984-2010) en millions d'euros



Les notes du cours ne sont transmises qu'à titre indicatif.

Chapitre 5 : Le marché de la monnaie et les marchés financiers

Introduction : Ce qu'est la monnaie et ce qu'elle n'est pas

Alors que la plupart des variables que nous avons évoquées jusqu'à présent pouvaient sembler abstraites, la monnaie nous est familière. Non seulement nous l'utilisons tous les jours mais en plus elle fait partie de notre vocabulaire quotidien.

Cette familiarité est en fait trompeuse. En effet, elle recèle des pièges sémantiques qui peuvent venir brouiller notre raisonnement. C'est pourquoi il est préférable d'oublier ce que nous croyons savoir de la monnaie pour nous mettre d'accord sur une définition.

Monnaie : tout actif qui est généralement accepté en paiement de biens et services ou en remboursement de dettes.

Cette définition est très large. Elle inclut évidemment les pièces et les billets, c'est-à-dire la *monnaie fiduciaire*. Elle inclut aussi les *dépôts à vue*, puisqu'on peut les utiliser directement pour régler ses achats grâce aux cartes de paiement ou aux chèques.

Elle inclut également les monnaies traditionnelles : *monnaie métallique* (en métaux précieux), et *monnaie marchandise* (plaques de sel, coquillages, bétail, etc.).

En revanche, cette définition souligne que la monnaie n'est synonyme ni de revenu, ni de patrimoine ou de richesse. Le revenu est le flux qu'un individu peut consacrer à ses dépenses et son épargne. Le patrimoine et la richesse représentent l'ensemble des actifs possédés par un individu (immobilier, titres financiers, voiture, œuvres d'art etc.). Seule une partie du patrimoine est conservée sous forme de monnaie.

Toute la question est de comprendre comment un consommateur décide de conserver une part plus ou moins importante de son patrimoine sous forme de monnaie, plutôt que sous la forme d'un autre actif. L'une des alternatives principales est la détention d'actifs financiers. On voit donc que le marché de la monnaie est indissociable des marchés financiers.

C'est pourquoi il est si important. En effet, c'est sur ce marché que va être déterminé le taux d'intérêt, qui va à son tour affecter l'investissement et la consommation, et avoir par conséquent des conséquences réelles.⁵¹

L'objet de ce chapitre est donc de comprendre les déterminants du taux d'intérêt, en partant de l'équilibre sur le marché de la monnaie (section 3). Pour cela, nous aurons besoin de comprendre ce qui détermine l'offre de monnaie (section 1) et sa demande (section 2). Quand tout cela sera fait, nous pourrons nous intéresser à une application de nos connaissances à la relation entre l'offre de monnaie et les prix (section 3).

Section 1 : L'offre de monnaie

Pour comprendre l'offre de monnaie, il faut d'abord préciser les formes que prend la monnaie dans une économie moderne. Ces formes sont le résultat d'une évolution qui a duré des siècles. La monnaie a pourtant assuré les mêmes fonctions depuis ses origines (I).

Nous verrons alors que la création de monnaie est un phénomène complexe qui fait intervenir les banques centrales et les banques commerciales (II), même si ce sont les banques centrales qui ont les moyens de contrôler l'offre de monnaie (III).

I- Fonctions et formes de la monnaie

Même si les formes de la monnaie ont évolué (B), ce qui était utilisé comme monnaie à une époque donnée a toujours assuré les trois mêmes fonctions (A).

A- Les trois fonctions de la monnaie

La monnaie assure trois fonctions essentielles : intermédiaire des échanges (1), unité de compte (2) et réserve de valeur (3).

⁵¹ L'adjectif « réel » est ici utilisé dans son sens économique. Il s'oppose à « monétaire » ou « nominal ». Une conséquence réelle est une conséquence sur les variables réelles : quantités, prix relatifs... par opposition aux variables nominales ou monétaires : prix, inflation...

1) La fonction d'intermédiaire des échanges

Dans les économies contemporaines, on utilise de la monnaie à chacun de ses achats. Le troc ne joue qu'un rôle exceptionnel. Au lieu d'échanger un bien contre un autre, on échange de la monnaie contre un bien.

Cependant, cela n'est possible que parce que le vendeur sait qu'il pourra échanger par la suite la monnaie qu'il a acquise contre des biens qui l'intéressent. La monnaie ne sert donc que *d'intermédiaire des échanges* puisqu'elle n'est pas échangée pour elle-même.

Pour réaliser l'utilité de la monnaie comme intermédiaire des échanges, il faut imaginer ce que serait une économie où les échanges se feraient uniquement par troc.

Dans une telle économie le *problème de la double coïncidence des besoins* se poserait en permanence. Un producteur de blé souhaitant acheter un ordinateur devrait en effet trouver un vendeur d'ordinateur souhaitant acquérir du blé. On peut imaginer que le vendeur d'ordinateur accepte le blé dans l'espoir de l'échanger contre d'autres biens qu'il souhaite acquérir, mais il faudrait qu'il le stocke et qu'il consacre beaucoup d'énergie pour trouver un vendeur intéressé par du blé.

Dans une économie de troc, les coûts de transaction seraient donc très élevés. Certaines transactions ne se feraient pas ou trop tard.

Si on décide de choisir un bien qui sera accepté par tous en échange de n'importe quel autre bien, on supprime le problème de la double coïncidence des besoins. C'est cette fonction que remplit la monnaie lorsqu'elle sert d'intermédiaire des échanges.

La monnaie augmente donc l'efficacité d'une économie en supprimant le problème de la double coïncidence des besoins, ce qui réduit les coûts de transaction.

2) La fonction d'unité de compte

La deuxième fonction de la monnaie est celle *d'unité de compte*. Cette fonction implique que la monnaie soit utilisée pour mesurer la valeur de tous les biens et services de l'économie. Les unités monétaires (euro, dollar, franc...) servent d'unité pour mesurer la valeur au même titre que les kilogrammes servent d'unité pour mesurer les masses, et les litres d'unité de mesure des volumes.

Ici encore, pour comprendre l'intérêt de cette fonction de la monnaie, il faut imaginer une économie qui fonctionnerait sur le principe du troc. Dans une telle économie, il serait nécessaire de connaître le prix relatif de tous les biens entre eux (combien d'ordinateurs pour un quintal de blé, combien de quintaux de blé pour une coupe de cheveux, combien d'ordinateurs pour une coupe de cheveux etc.).

Or le nombre de prix relatifs augmente très rapidement avec le nombre de biens : Dans une économie à trois biens, il y a trois prix relatifs. Dans une économie à dix biens, il y en a 45. Dans une économie à 100 biens, il y en a 4950 !⁵² Et on sait qu'il y a beaucoup plus de 100 biens dans une économie moderne.

Comme personne ne peut retenir une telle quantité de prix relatifs, il est plus facile de choisir un bien comme unité de compte. Il suffit alors d'exprimer tous les prix en fonction de ce bien, qu'on appelle le *numéraire*. Non seulement on n'a plus à retenir qu'un seul prix par bien mais on peut de plus facilement en déduire tous les prix relatifs.

Comme la monnaie joue le rôle du numéraire dans les économies contemporaines, elle réduit de ce fait les coûts de transaction et rend plus efficace l'économie. Ce gain est d'autant plus appréciable que l'économie est complexe et qu'elle produit un nombre important de biens différents.

3) La fonction de réserve de valeur

La troisième fonction économique de la monnaie est celle de *réserve de valeur*. Cela signifie que la monnaie permet d'épargner du pouvoir d'achat entre le moment où un revenu est perçu et celui où il est dépensé.

La monnaie est certainement une meilleure réserve de valeur que beaucoup d'autres biens qui se dégradent rapidement (ex : poisson frais, bien qui se démode ou devient obsolète etc.). Elle ne remplit cependant ce rôle que de façon imparfaite. En effet, elle perd rapidement de la valeur en période d'inflation.

D'autres biens remplissent cette fonction de façon beaucoup plus satisfaisante : immobilier, métaux précieux, actions etc.

Plus généralement, on peut trouver des biens susceptibles d'assurer chacune des fonctions de la monnaie, tout aussi bien sinon mieux. On a d'ailleurs utilisé des biens très hétéroclites pour les assurer.

ex. : Dans les camps de prisonniers pendant la guerre, on utilisait des cigarettes.

Il suffit pour cela qu'il présente certaines qualités (il doit entre autres être facilement divisible, transportable, standardisé et assez résistant). La monnaie telle que nous la connaissons présente certes ces qualités, mais elle n'est pas la seule. Pour comprendre

⁵² La formule qui donne le nombre de prix relatifs en fonction du nombre de biens est $N_{prix} = N_{biens} \times (N_{biens} - 1) / 2$. Vous pouvez à titre d'exercice chercher à la redémontrer.

comment nous en sommes arrivés à ne plus utiliser que la monnaie que nous connaissons, il faut décrire son évolution.

B- L'évolution des formes de la monnaie

Au fil des siècles, les formes de monnaie ont évolué et ont suivi un processus de *dématérialisation*. Au départ, ce qui servait de monnaie était des marchandises qui avaient une valeur intrinsèque (plaques de sel, coquillages etc.). Ces monnaies ont été remplacées par les monnaies métalliques (or, argent, cuivre) qui avaient encore une valeur intrinsèque puisque les métaux ont d'autres utilisations. Elles étaient cependant plus pratiques à manipuler.

Les monnaies métalliques furent complétées par le *papier-monnaie*. Il s'agissait au départ de titres de propriété sur une certaine quantité de métaux précieux. Le transfert d'une certaine quantité de papier-monnaie était équivalent au transfert du métal précieux mais le transfert n'avait pas besoin de s'opérer physiquement. C'était plus commode. Toutefois, on pouvait toujours demander la conversion du papier-monnaie dans la quantité équivalente de métal précieux.

Une grande innovation fut l'invention de la *monnaie fiduciaire*. Dans la mesure où le papier-monnaie est très rarement converti, il suffit qu'il soit accepté en paiement pour avoir de la valeur. Il n'est alors pas nécessaire que le billet corresponde à une certaine quantité de métal précieux tant que ses utilisateurs ont confiance dans sa capacité à être accepté en règlement d'une dette.⁵³

L'invention de la monnaie fiduciaire marque l'apparition de la monnaie moderne. Sa caractéristique est de n'avoir pas de valeur intrinsèque mais de reposer sur le fait qu'elle est acceptée comme moyen de paiement.

Une bonne façon pour l'Etat d'assurer de la valeur à une monnaie est de l'accepter en paiement des impôts. Tout un chacun sait qu'il trouvera toujours au moins son percepteur pour accepter ses billets, ce qui leur confère mécaniquement de la valeur.

Parallèlement à la monnaie fiduciaire, est apparue la *monnaie scripturale*. Cette forme de monnaie repose sur de simples jeux d'écriture entre créditeurs et débiteurs. C'est le principe du chèque, qui descend de la lettre de change du Moyen Age.

Pour peu qu'il existe des relations suffisamment fréquentes et que la confiance règne entre ses utilisateurs, la monnaie scripturale est plus pratique et moins risquée que la monnaie

⁵³ « *Fiducia* » signifie « confiance » en latin.

fiduciaire. Elle ne suppose pas le transport physique des billets et ne peut être volée. Elle est donc plus sûre.

Elle s'est répandue avec le développement du système bancaire et avec le progrès technique. Grâce à l'informatique, on peut remplacer les chèques par des virements automatiques. La carte de paiement participe ainsi au développement de la monnaie scripturale. Les différents types de monnaie électronique, comme Proton, relèvent du même mécanisme.

A l'heure actuelle, la monnaie fiduciaire a remplacé la monnaie métallique dans la plupart des économies du monde. Toute référence à l'or a en effet été officiellement abandonnée, avec l'abandon du système de Bretton Woods dans les années septante. Si on s'échange encore des pièces d'or et d'argent dont la valeur est légèrement différente de celle du métal qui les compose, c'est principalement pour des raisons de collection.⁵⁴

Il s'agit donc à présent de comprendre comment fonctionne l'offre de monnaie fiduciaire.

C- La mesure de l'offre de monnaie

Mesurer concrètement l'offre de monnaie est une tâche difficile parce que beaucoup d'actifs correspondent plus ou moins bien à sa définition et à ses trois fonctions. Par exemple, un billet est sans aucun doute de la monnaie, de même qu'un solde créditeur sur un compte chèque. On peut les utiliser directement pour régler un achat.

Que faut-il penser d'un solde créditeur sur un compte épargne ? D'un côté, on ne peut utiliser ce solde pour payer un achat. En revanche, un simple virement suffit pour pouvoir le faire. Le solde du compte épargne peut de ce point de vue être considéré comme une forme de monnaie.

Ce qui le distingue d'un compte courant est sa *liquidité*, c'est-à-dire sa capacité à être converti rapidement en moyen d'échange.

C'est pourquoi on ne définit pas un mais plusieurs *agrégats monétaires* qui reposent sur une définition de plus en plus large de la monnaie. Ces agrégats fonctionnent donc comme des poupées russes qui englobent des actifs de moins en moins liquides :

⁵⁴ La valeur du Napoléon, une pièce d'or cotée à la bourse de Paris, a traditionnellement tendance à augmenter au mois de décembre. La raison en est qu'elle est utilisée pour confectionner des boutons de manchette qui servent de cadeaux lors de fêtes de fin d'année.

M1 : monnaie fiduciaire et dépôts à vue.

M2 : M1

+ dépôts à terme d'une durée inférieure ou égale à 2 ans

+ dépôts remboursables avec préavis inférieur ou égal à 3 mois.

M3 : M2

+ pensions⁵⁵

+ titres d'OPCVM monétaires⁵⁶

+ titres de créance émis pour une durée inférieure ou égale à 2 ans.

Comme des innovations financières et des évolutions imprévues se produisent régulièrement, la définition des agrégats monétaires est révisée pour en tenir compte.⁵⁷

Tableau 1 :
Agrégats monétaires dans la zone euro en décembre 2010

	Montant (milliards d'euros)	Par rapport à M3 (%)
M1	4701,7	49,36
M2	8385,9	88,04
M3	9525,1	100

Source : Banque Centrale Européenne : <http://sdw.ecb.europa.eu/reports.do?node=100000141>.

II- La création monétaire

Comme la création de monnaie est une opération complexe, nous allons commencer par l'envisager dans une économie simplifiée dans laquelle il n'existe qu'une seule banque, la banque centrale, et qu'une seule forme de monnaie, la monnaie fiduciaire (A). On pourra alors intégrer le rôle des banques et de la monnaie scripturale (B).

⁵⁵ Une prise en pension désigne le transfert provisoire de la propriété d'un titre de créance de son détenteur initial vers un tiers.

⁵⁶ Organismes de Placement Collectif en Valeurs Mobilières : SICAV et FCP.

⁵⁷ Par exemple, la définition de M3 a déjà été révisée depuis le lancement de l'euro (1^{er} janvier 1999). On avait observé une forte augmentation des détentions d'OPCVM monétaires par les non-résidents qui biaisait l'évolution de M3. Les titres d'OPCVM détenus par des non-résidents ont alors été exclus du calcul de M3 à partir de mai 2001.

A- La création monétaire en l'absence de banques commerciales

Supposons qu'il n'existe qu'une seule banque, la banque centrale. Cette banque détient le monopole d'émission de la monnaie fiduciaire, c'est-à-dire des billets et des pièces.

Pour mettre de la monnaie en circulation, elle ne peut se contenter de distribuer des billets au coin des rues. Elle a besoin d'une *contrepartie*. Cette contrepartie prend la forme de titres financiers (obligations, bons du trésor, actions...) que les agents vont lui vendre. Elle les paiera alors en leur donnant des billets qu'elle aura imprimés pour l'occasion.

Cette forme de mise sur le marché de monnaie s'appelle une opération d'*open market*, parce que la banque centrale intervient directement sur les marchés financiers pour acquérir des titres et distribuer de la monnaie.

Les opérations d'*open market* permettent aussi de réduire la quantité de monnaie en circulation. Il suffit pour cela que la banque centrale vende des titres au lieu d'en acheter. Elle récupère ainsi de la monnaie qu'elle a donc retirée de la circulation.

On peut retrouver la trace de ces opérations dans le bilan de la banque centrale. Ce bilan est le document comptable qui résume l'ensemble des actifs que possède la banque ainsi que l'ensemble de ses dettes dans deux colonnes : *actif* et *passif*. Une autre façon de décrire un bilan consiste à dire que l'actif montre comment une entreprise a utilisé les ressources dont elle dispose et qui sont présentées au passif.⁵⁸

Pour une banque, l'actif présente l'ensemble des créances qu'elle détient et le passif l'ensemble des dettes qu'elle a contractées.

Supposons à présent que la banque centrale ait mis en circulation un millier d'euros de billets. Cela suppose qu'elle ait acheté pour un millier d'euros de titres. Il apparaît donc un millier d'euros de titres à son actif. Pour acquérir ces titres, elle a dû créer un millier d'euros de billets qui vont donc apparaître à son passif. On peut aussi faire remarquer que les billets sont une forme de dette de la banque centrale qui pourra à l'avenir être amenée à les échanger contre des titres.

Le bilan de la banque centrale se présente donc de la façon suivante :

Actif	Passif
Titres : 1000	Billets : 1000

⁵⁸ On parle parfois « d'emploi » pour désigner l'actif et de « ressources » pour le passif. Par convention, l'actif apparaît toujours à gauche et le passif à droite du bilan. On utilise ainsi également les expressions de « gauche » et « droite » du bilan pour désigner l'actif et le passif.

La quantité de monnaie en circulation correspond donc exactement à la quantité de titres rachetée par la banque centrale.

Cependant, notre description de la création monétaire reste très imparfaite parce qu'elle ne tient aucun compte de la monnaie scripturale qui représente pourtant la majeure partie de la monnaie en circulation dans les économies contemporaines. Il faut donc à présent intégrer le rôle des banques commerciales.

B- La création monétaire en présence de banques commerciales

On l'a vu, la monnaie ne se résume pas aux billets, tant s'en faut. La majeure partie de la monnaie en circulation aujourd'hui prend la forme de soldes sur des comptes bancaires. Il faut donc compléter notre description de la création monétaire en intégrant le rôle des banques et des dépôts.

Pour intégrer le rôle des banques, il faut comprendre comment les banquiers gagnent leur vie. Leur ressource principale vient des intérêts perçus sur les prêts accordés à leurs clients.

La question est alors de savoir d'où proviennent les fonds que les banquiers utilisent pour leurs prêts. La réponse est qu'ils proviennent des dépôts.

En effet, une banque sait que tous les fonds qui ont été déposés sur ses comptes ne vont pratiquement jamais être retirés en même temps. Elle peut donc en prêter une partie sans cesser de faire face aux demandes de retrait. Elle percevra donc des intérêts sur les fonds prêtés. Les banquiers ont donc intérêt à prêter une partie aussi importante que possible de leurs dépôts.

Cependant, si une banque prêtait toute la monnaie qui lui a été confiée, elle ne pourrait plus faire face aux retraits et se retrouverait en faillite. De même, si une partie suffisamment élevée de ses clients décidait de retirer leurs dépôts, la banque pourrait se retrouver dans l'incapacité d'y faire face.⁵⁹ C'est pourquoi une banque conserve toujours une partie de ses dépôts sous forme de monnaie. C'est ce qu'on appelle ses *réserves* ou sa *trésorerie*.

La législation complète souvent la prudence des banques en imposant un montant minimum de réserves, sous la forme d'un *coefficient de réserves obligatoires* ou *coefficient de trésorerie*. Ce coefficient donne aux banques la valeur minimum du ratio entre leurs réserves et leurs dépôts. Ainsi, la loi oblige les banques à respecter l'inégalité suivante :

⁵⁹ On observe ce genre de phénomène quand les clients d'une banque craignent une faillite de celle-ci. Ils se précipitent alors pour retirer leurs dépôts en espérant ne pas arriver trop tard. On parle de *ruée bancaire*. Une simple rumeur peut être à l'origine d'une ruée puis d'une faillite. Le système de Réserve Fédérale aux Etats-Unis a par exemple été créé suite à une ruée bancaire au début du XX^{ème} siècle.

$$\frac{\text{réserves}}{\text{dépôts}} \geq \text{coefficient de réserves obligatoires}$$

Pour comprendre comment le système bancaire va créer de la monnaie, imaginons une économie dans laquelle la monnaie peut prendre la forme de billets et de dépôts à vue, et où le coefficient de réserves obligatoires a été fixé à dix pourcents. Supposons par ailleurs qu'il existe trois banques (A, B et C) dans cette économie. Supposons enfin que les individus qui composent cette économie ne détiennent jamais de liquide et n'utilisent donc que de la monnaie scripturale pour leurs paiements.

Imaginons à présent que la banque centrale vient d'imprimer mille euros, qu'elle a utilisés pour acheter des obligations à une entreprise (opération d'open market), et voyons quel va être l'effet de cette création de monnaie sur la masse monétaire.

L'entreprise va utiliser ces mille euros pour payer l'un des ses fournisseurs. Ce fournisseur va alors déposer les mille euros auprès de la banque A. Le bilan de cette banque est alors affecté de la façon suivante :

Banque A	
Actif	Passif
Réserves : 1000	Dépôts : 1000

La banque A ne ferait aucun profit si elle en restait là. Elle va donc prêter une partie de ses réserves afin de percevoir un intérêt. Si elle souhaite maximiser les intérêts perçus, elle prêtera tout ce que la loi lui autorise, c'est-à-dire qu'elle respectera tout juste le coefficient de réserves obligatoires.

Banque A	
Actif	Passif
Réserves : 100	Dépôts : 1000
Crédits : 900	

On voit bien que la banque A respecte encore son ratio de réserves obligatoires, puisque ses réserves représentent encore 10% de ses dépôts. En revanche, elle a pu prêter 900€ en billets à l'un de ses clients.

Ce client va dépenser les 900€ auprès d'une entreprise qui est cliente de la banque B. Cette entreprise déposera les 900€ auprès de sa banque dont le bilan sera le suivant :

Les notes du cours ne sont transmises qu'à titre indicatif.

Banque B	
Actif	Passif
Réserves : 900	Dépôts : 900

Comme la banque B ne souhaite pas non plus laisser dormir des fonds en réserve, elle va prêter ce que la loi lui autorise, soit 90% des 900€ de dépôts.

Banque B	
Actif	Passif
Réserves : 90	Dépôts : 900
Crédits : 810	

Les 810 euros de crédit vont alors se retrouver dans les coffres de la banque C lorsqu'ils auront été dépensés.

Banque C	
Actif	Passif
Réserves : 810	Dépôts : 810

Comme la banque C va, elle aussi, prêter ses billets, son bilan deviendra :

Banque C	
Actif	Passif
Réserves : 81	Dépôts : 810
Crédits : 729	

L'une des deux autres banques retrouvera les 729€ sous forme de dépôts mais en prêtera 90%, et le mécanisme se poursuivra indéfiniment.

Puisque la masse monétaire est composée des billets et des dépôts, on peut donc dire que l'augmentation totale du volume des dépôts sera de :

$$\Delta M = 1000 + 0,9 \times 1000 + 0,9^2 \times 1000 + 0,9^3 \times 1000 + 0,9^4 \times 1000 \dots + 0,9^n \times 1000$$

$$\Delta M = 1000 + 900 + 810 + 729 + 656,1 + \dots$$

On montre facilement que cette somme tend vers 10000 lorsque n tend vers l'infini.⁶⁰ L'augmentation du montant des dépôts est donc dix fois supérieure à celle du montant des billets. Comme les dépôts sont une partie de la masse monétaire, nous obtenons une première intuition importante et contraire au sens commun : les banques commerciales créent de la monnaie.

⁶⁰ Une simulation est proposée en annexe.

On peut se demander si on aurait pu prévoir le rapport de un à dix observé entre la quantité de billets imprimés et l'augmentation de la quantité de dépôts. La réponse est oui. Ce rapport correspond précisément à l'inverse du coefficient de réserves obligatoires. Pour s'en convaincre, il suffit de raisonner au niveau agrégé en représentant le bilan consolidé de toutes les banques.

Au départ, des réserves égales au dépôt apparaissent :

Système bancaire	
Actif	Passif
Réserves : 1000	Dépôts : 1000

Cependant, les banques sont autorisées à accorder des crédits qui se retrouvent sous forme de dépôts dans d'autres banques, c'est-à-dire au passif du système bancaire.

Globalement, puisque le coefficient de réserves obligatoires est de dix pourcents, le montant des dépôts pourra s'élever à dix fois celui des réserves, soit 10000 euros. Comme il y a déjà eu un dépôt de mille euro, le système bancaire peut donc accorder 9000 euros de crédits. Le bilan consolidé du système bancaire sera donc :

Système bancaire	
Actif	Passif
Réserves : 1000	Dépôts : 1000
Crédits : 9000	Dépôts : 9000

Dans notre économie simplifiée, l'augmentation des dépôts est donc égale à la quantité de billets créée divisée par le coefficient de réserves obligatoire. Plus généralement on peut dire que le volume des dépôts (D) est égal à celui des réserves (R) divisé par le coefficient de réserves obligatoires (θ), puisque tous les billets en circulation ont été créés à un moment ou un autre.

$$D = R / \theta \quad (1)$$

Cette expression est très simple mais reste trop irréaliste. En effet, elle repose sur l'hypothèse très forte que les agents ne conservent jamais de billets. De plus, elle limite l'activité de la banque centrale à l'impression de monnaie. Pourtant, les banques centrales accordent également des crédits et ces crédits peuvent aussi servir de réserves. On appelle d'ailleurs *monnaie banque centrale* ou *base monétaire* la somme des billets et des crédits accordés par la banque centrale.

Pour rendre notre raisonnement un peu moins abstrait, supposons qu'une fraction c_b de la masse monétaire est détenue sous forme de billets, le complément étant conservé sous

forme de dépôts. Si M représente la masse monétaire et E , pour espèces, la quantité émise de billets, on peut écrire :

$$E = c_b \cdot M \quad (2a)$$

$$D = (1 - c_b)M \quad (2b)$$

Par ailleurs, la monnaie banque centrale peut être utilisée soit par les ménages sous forme de billets, soit par les banques commerciales sous forme de réserves. Si on note H la monnaie banque centrale,⁶¹ on peut alors écrire :

$$H = E + R \quad (3)$$

Or on sait que les réserves sont proportionnelles aux dépôts. On peut donc réécrire l'expression de la quantité de monnaie banque centrale :

$$H = E + \theta D \quad (4)$$

On peut ensuite remplacer E et D par leur valeur et exprimer la masse monétaire en fonction de la base monétaire :

$$H = c_b \cdot M + \theta(1 - c_b)M$$

$$H = [c_b + \theta(1 - c_b)]M$$

$$\Rightarrow \boxed{M = \frac{1}{c_b + \theta(1 - c_b)} H} \quad (5)$$

Comme c_b et θ sont des paramètres, l'expression ci-dessus montre que la masse monétaire est proportionnelle à la base monétaire. De plus, le dénominateur est inférieur à un. Par conséquent, le coefficient de proportionnalité est sans ambiguïté supérieur à un. C'est pourquoi on l'appelle le *multiplieur monétaire*. Il signifie qu'une augmentation donnée de la base monétaire se traduit par une augmentation supérieure de la masse monétaire.

L'une des implications importantes du multiplieur est qu'une part importante de la monnaie est créée par les banques. Ainsi, la banque centrale ne détient pas le monopole de la création monétaire, ce qui est assez contre-intuitif.⁶²

⁶¹ La monnaie banque centrale est appelée « high-powered money » en anglais.

⁶² Le Dabe, le personnage interprété par Jean Gabin dans « Le cave se rebiffe », exagérait donc en déclarant « Dis-toi bien qu'en matière de monnaie, les Etats ont tous les droits et les particuliers aucun. » Il est vrai qu'il parlait là de fausse monnaie :

<http://www.youtube.com/watch?v=pBhE65uZrJc&playnext=1&list=PLA776A32114BEE2BF&index=38>.

III- Les moyens de contrôle de l'offre de monnaie

Nous venons de constater que c'était les banques commerciales qui créaient la majeure partie de la monnaie en circulation. Cela ne signifie pas pour autant que ce sont elles qui contrôlent la quantité de monnaie en circulation. En effet, la banque centrale dispose d'une batterie d'instruments qui lui permettent de contrôler la quantité de monnaie en circulation.

Certains de ces instruments peuvent se déduire de la formule du multiplicateur. Il s'agit des opérations d'open market (1), qui jouent sur l'offre de base monétaire, et du coefficient de réserves obligatoires (2), qui détermine la quantité de dépôts que les banques commerciales peuvent accorder pour une base monétaire donnée. Cependant, il existe un deuxième type d'instrument qui permet d'ajuster la base monétaire. Ce sont les facilités permanentes (3).

1) Les opérations d'open market

Le principe des opérations d'open market est d'échanger de la monnaie banque centrale aux banques commerciales contre des titres. Les banques s'y livrent lorsqu'elles ont besoin de liquidités.

Ces opérations s'apparentent à un prêt, à la différence qu'elles supposent que les emprunteurs déposent une garantie. On dit qu'il ne s'agit pas d'un *prêt en blanc*. Comme tout prêt, ces opérations donnent lieu au versement d'un intérêt.

Pour encourager les banques à avoir recours à ces prêts, la banque centrale peut ajuster son taux d'intérêt. Plus ce taux d'intérêt sera bas, plus les banques commerciales seront incitées à demander des liquidités à la banque centrale qui pourra alors écouler une quantité plus importante de monnaie.

On voit donc que la banque centrale dispose d'un instrument supplémentaire de contrôle de la base monétaire : le taux d'intérêt sur les opérations de refinancement des banques commerciales.

Dans la zone euro, ce taux d'intérêt est le *taux directeur* principal. On l'appelle le *Refi*, pour refinancement.⁶³ Vous comprenez à présent pourquoi c'est l'évolution de ce taux qui est utilisée pour commenter la politique monétaire de la BCE. C'est ce taux qui va déterminer la base monétaire de la zone euro.

⁶³ En anglais, on parle du taux « repo ».

2) Les réserves obligatoires

Si on reprend la formule du multiplicateur monétaire (5) et qu'on suppose que la base monétaire est donnée, la masse monétaire devient une fonction du coefficient de réserves obligatoires (θ). Plus précisément, le multiplicateur est une fonction décroissante du coefficient de réserves obligatoires.

Par conséquent, la banque centrale peut ajuster la masse monétaire en circulation en manipulant le coefficient de réserves obligatoires. Si elle réduit ce coefficient, les banques pourront accorder plus de crédits avec les mêmes réserves. Si, au contraire, la banque centrale augmente le coefficient de réserves obligatoires, les banques seront contraintes d'accorder moins de prêts. Les dépôts augmenteront dans le premier cas alors qu'ils diminueront dans le second.

Au départ, la justification de l'imposition d'un coefficient de réserves obligatoires n'avait pour objet que d'assurer la viabilité des banques. Il s'agissait d'une réglementation *prudentielle*. C'est à partir des années 1960 et 1970 que ce coefficient a été utilisé comme instrument de politique monétaire. Certains pays, comme la France, qui n'en imposaient pas jusque là, l'ont introduit à la même époque.

Pourtant, on s'est aperçu que si cet instrument était puissant, il ne permettait pas de régler finement l'offre de monnaie. Il a donc perdu de sa popularité dans les années 1980, à cause de son manque de précision. Il a même été purement et simplement supprimé par la Banque d'Angleterre.

Aujourd'hui le coefficient de réserves obligatoires est fixé à 2% dans la zone euro et n'a pas été modifié depuis le lancement de la monnaie unique. Il est utilisé pour son effet structurel sur la demande de monnaie banque centrale mais pas de façon conjoncturelle.

3) Les facilités permanentes

Les *facilités permanentes* sont des possibilités supplémentaires de gérer leurs liquidités offertes aux banques commerciales par la banque centrale. Il s'agit schématiquement de prêts et de dépôts.

A la différence des opérations d'open market, les facilités permanentes ne passent pas par le marché mais par des relations bilatérales entre une banque commerciale et la banque centrale. Ici, ce sont les banques commerciales qui en ont l'initiative.

En général, ces opérations sont conçues pour n'être utilisées que de façon exceptionnelle et à très court terme (une journée) par les banques commerciales. Les prêts sont donc très chers et les dépôts très mal rémunérés. C'est pourquoi les taux d'intérêt qui leur

sont appliqués sont les taux directeurs *plafond* et *plancher* de la banque centrale. Ils déterminent l'intervalle dans lequel le taux d'intérêt du marché va fluctuer. Les facilités permanentes permettent un réglage encore plus fin de la politique monétaire que les opérations d'open market.

Dans la zone euro, les facilités permanentes sont appelées *facilités marginales*. Le taux d'intérêt pratiqué sur les facilités marginales de prêts est le taux directeur plafond de la zone et le taux d'intérêt des facilités marginales de dépôts en constitue le taux d'intérêt plancher. Voilà pourquoi il y a trois taux directeurs dans la zone euro au lieu d'un seul.

Les banques centrales disposent donc d'une batterie d'instruments qui leur permettent d'influencer la quantité de monnaie en circulation, même si elles n'en créent qu'une faible partie. Dans la suite du cours, nous supposerons donc que la quantité de monnaie offerte est déterminée directement par la banque centrale, en faisant abstraction de la méthode utilisée.

Puisque nous avons décrit le fonctionnement de l'offre de monnaie, il convient à présent de se tourner vers sa demande.

Section 2 : La demande de monnaie

Lorsqu'on parle de demande de monnaie, il faut se méfier des contresens. En effet, par demande de monnaie, on n'entend pas demande de revenu ou demande de richesse. Nous souhaiterions tous disposer d'un revenu ou d'une richesse infinis.

Ce qu'on appelle demande de monnaie correspond à la quantité de liquidité que souhaitent détenir les agents. Elle dépend donc de la répartition des avoirs entre différents actifs. Pour comprendre cette répartition, il faut comprendre pourquoi un agent va souhaiter détenir un actif non rémunéré, la monnaie, plutôt que des actifs rémunérés. La réponse à ce paradoxe apparent tient au fait que la monnaie rend des services.

Longtemps on n'a considéré que la fonction d'intermédiaire des échanges de la monnaie. Dans une économie monétaire, les agents sont en effet tenus de détenir de la monnaie pour réaliser leurs transactions. On peut ainsi obtenir une première théorie de la demande de monnaie, en supposant qu'elle est une fonction du nombre de transactions réalisées (I).

En tenant compte des autres raisons pour lesquelles les agents détiennent de la monnaie, on peut raffiner les déterminants de la demande de monnaie et souligner le rôle du taux d'intérêt (II).

Avant de se lancer dans la description des déterminants de la demande de monnaie, il faut préciser quels sont les actifs qui peuvent être détenus à sa place. En réalité, n'importe quel actif est envisageable. Cependant, pour simplifier notre raisonnement, nous allons supposer que le seul type d'actif qui puisse être détenu à la place de la monnaie est les *obligations*.

Les obligations sont des *titres de créance négociables* qui assurent le versement d'un intérêt à l'échéance. Ce sont donc des actifs qui proposent une rémunération sans risque et qui peuvent être achetés et revendus. Ces caractéristiques seront importantes au moment de déterminer l'impact du taux d'intérêt sur la demande de monnaie. Le taux d'intérêt dont nous parlerons sera donc le taux d'intérêt sur les obligations.⁶⁴

I- La théorie quantitative de la monnaie

La *théorie quantitative de la monnaie* (TQM) a été développée par les économistes classiques et néo-classiques, mais on en trouve déjà l'intuition chez les mercantilistes. Elle part d'une simple relation comptable mais fournit également une théorie de la demande de monnaie (A). Sous certaines conditions, elle peut être utilisée pour comprendre l'évolution des prix (B).

A- La quantité de monnaie nécessaire aux transactions

Dans une économie monétaire, il est nécessaire de détenir de la monnaie pour réaliser la moindre transaction. Cependant, puisque la monnaie ne rapporte rien alors que les titres proposent une rémunération positive, les agents ne détiendront que le strict minimum de ce qui est nécessaire aux transactions. Pour les classiques et les néo-classiques, toute encaisse monétaire détenue en plus de ce qui est indispensable pour les transactions serait irrationnelle puisqu'elle impliquerait la perte inutile d'un intérêt.

Par conséquent, la demande de monnaie sera proportionnelle à la valeur des transactions exprimées en unités monétaires (euros, dollars...). Si on note T le nombre de

⁶⁴ On trouve une définition précise des obligations et des actions dans la bouche du personnage de Mme Caroline dans l'Argent d'Emile Zola : « [...] un obligataire est un simple prêteur qui touche tant pour cent sur son prêt, sans être intéressé dans les bénéfices, tandis qu'un actionnaire est un associé courant la chance des bénéfices et des pertes... » (E. Zola, *L'argent*, 1891. Réédition Folio classique, 2008, p.163).

transactions et P le niveau général des prix, qui mesure le prix moyen des transactions, la quantité de monnaie demandée pourra s'écrire :

$$M^d = \phi.P.T \quad (6)$$

k est un paramètre positif dont la signification apparaît en manipulant l'expression (6).

$$\phi = M^d / PT \quad (7)$$

D'après l'expression ci-dessus, ϕ mesure le nombre moyen d'unités monétaires nécessaires pour réaliser une transaction d'une valeur d'une unité monétaire.

ex : Si au cours d'une année on a utilisé 10 milliards d'euros en billets pour les transactions mais que la valeur des transactions s'est établie à 20 milliards d'euros, ϕ sera égal à $\frac{1}{2}$. Cela signifie qu'il aura suffi de cinquante centimes d'espèces pour réaliser un euro de transaction.

Par conséquent, chaque euro aura été utilisé deux fois au cours de la période. Le nombre moyen d'utilisations de chaque unité monétaire est appelé *vitesse de circulation de la monnaie*. Le paramètre k est donc l'inverse de la vitesse de circulation de la monnaie, et vice versa : $V \equiv 1/\phi$.

On peut alors réécrire la demande de monnaie :

$$M^d = P.T / V \quad (8)$$

L'équation (8) est appelée « équation des échanges ». Bien que son intuition soit apparue très tôt, c'est à Irving Fisher, que l'on doit sa version moderne.⁶⁵

Cette expression est difficile à utiliser empiriquement. En effet, on ne connaît pas le nombre de transactions. En revanche, il est raisonnable de considérer que le nombre de transactions est proportionnel au volume de la production. C'est pourquoi on remplace le nombre de transactions par la production :

$$M^d = P.Y / V \quad (9)$$

Il apparaît alors que la demande de monnaie est une fonction croissante de la production et des prix, et décroissante de la vitesse de circulation de la monnaie. Sous cette forme, l'équation des échanges est appelée « équation de Cambridge ».

⁶⁵ I. Fisher *The purchasing power of money*, MacMillan, 1911. I. Fisher « The equation of exchange, 1896-1911 », *American Economic Review*, vol. 1 n°2, p.296-305, juin 1911. En fait, on écrit en général l'équation des échanges sous la forme $MV = PT$, qui est tout à fait équivalente.

B- La relation entre la masse monétaire et les prix

La demande de monnaie telle que nous l'avons définie est tautologique. En quelque sorte, elle indique que la quantité de monnaie nécessaire pour les transactions sera égale à... la quantité de monnaie nécessaire pour les transactions. Cependant, et moyennant quelques hypothèses complémentaires, elle peut fournir une théorie de l'évolution des prix. Elle permet en effet d'établir la *théorie quantitative de la monnaie* (ou *TQM*).

On peut considérer que la vitesse de circulation est un paramètre stable. En effet, il est déterminé par des facteurs structurels tels que les habitudes de paiements ou le niveau de développement du système bancaire. Cela ne signifie pas forcément que la vitesse de circulation de la monnaie est constante, mais qu'elle ne dépend pas de l'évolution des autres variables qui apparaissent dans l'équation de Cambridge, et surtout pas de celle de la masse monétaire.

Si on considère de plus que l'économie utilise toutes ses capacités de production et qu'il n'y a pas de sous emploi, le revenu sera lui aussi un paramètre de l'équation.

Dans ce cas, les deux variables qui pourront s'ajuster pour assurer l'égalité des deux membres de l'équation de Cambridge sont la masse monétaire et les prix. Comme la masse monétaire est contrôlée par la banque centrale, ce sont les prix qui s'ajusteront pour maintenir l'égalité. On peut donc exprimer les prix en fonction de la masse monétaire :

$$P = MV/Y \quad (10a)$$

La théorie quantitative de la monnaie implique donc que les prix sont proportionnels à la masse monétaire. Plus précisément, il existe un lien de causalité entre la masse monétaire et le niveau des prix. Toute augmentation de la masse monétaire provoque une augmentation des prix, et ne provoque que cela. La TQM implique donc une séparation entre les variables nominales et réelles. On appelle *dichotomie entre la sphère nominale et la sphère réelle* cette séparation. C'est cette dichotomie que les auteurs classiques considéraient lorsqu'ils affirmaient que la monnaie n'est qu'un voile.

On a par ailleurs pris l'habitude d'exprimer la TQM en utilisant non le niveau des variables mais leur taux de variation. Le taux d'inflation devient alors une fonction du taux de croissance de la masse monétaire. Si on utilise une lettre minuscule pour désigner les taux de variation d'une variable, la TQM s'écrit :

$$p = m + v - y \quad (10b)$$

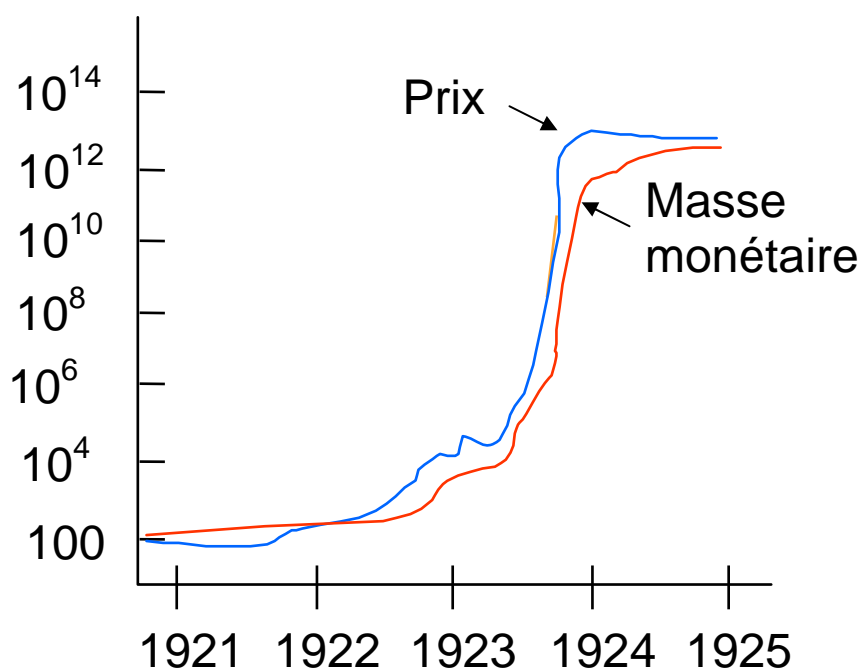
Nous avons vu dans le chapitre précédent que l'économie pouvait se trouver en situation de sous emploi à court terme à cause de la rigidité des prix. Nous verrons en effet dans la suite du cours que la politique monétaire peut alors influencer le niveau de production. La TQM n'est en principe pertinente que pour réfléchir à la relation de long terme entre le niveau de prix et la masse monétaire puisqu'elle repose sur l'hypothèse de plein emploi.

Elle est par exemple utilisable pour expliquer pourquoi la découverte du Nouveau Monde a été suivie par un siècle d'augmentation soutenue des prix en Europe et surtout en Espagne. La raison en est que l'or rapporté des Amériques est venu gonfler la masse monétaire qui était alors essentiellement métallique. La TQM implique que ce gonflement de la masse monétaire ne pouvait se traduire que par une augmentation des prix.

Pourtant, il existe des situations dans lesquelles la TQM est utilisable même à court terme. Ce sont les situations *d'hyperinflation*. On parle d'hyperinflation lorsque le taux d'inflation atteint 50% par mois.

Pendant les périodes d'hyperinflation, il est extrêmement coûteux de fixer les prix. Par conséquent, tous les agents révisent leurs prix en permanence et les prix peuvent être considérés comme flexibles même à court terme. La TQM est donc applicable.

Graphique 1 :
Evolution de la masse monétaire et des prix en Allemagne entre 1922 et 1924
Indice 100 en janvier 1921



Le cas d'hyperinflation le plus marquant est celui de l'hyperinflation allemande qui s'est déroulé au cours de la première moitié des années 1920. Entre 1922 et 1924, les prix y

ont été multipliés par dix milliards (10^{10}). Ainsi, le prix d'un journal est passé de 30 pfennigs (0,3 marks) en janvier 1921 à 70 millions de marks le 17 novembre 1923. Il n'était encore (que) de 15 millions de marks le 9 novembre. Pendant cette période, il était devenu si coûteux de fixer le moindre prix qu'on n'imprimait plus la moindre étiquette. Les salaires étaient même versés deux fois par jour pour éviter qu'ils soient trop érodés par l'inflation. On peut donc considérer que les prix étaient flexibles.

Le graphique 1 décrit l'évolution de la masse monétaire et des prix entre 1922 et 1924. On constate que l'évolution des prix suit de très près celle de l'offre de monnaie. C'est précisément ce que prévoit la TQM.

Ainsi, c'est parce que les autorités allemandes se sont lancées dans une politique de création monétaire très rapide que l'hyperinflation a démarré. On dit familièrement qu'elles ont fait tourner la planche à billets. En l'occurrence, l'expression n'est pas galvaudée puisque des billets de dix milliards et cent milliards de marks ont été émis à cette époque.⁶⁶

La raison de cette émission massive de billets est que le gouvernement allemand a essayé de financer son budget, et en particulier les réparations de guerre imposées à l'issue de la Première guerre mondiale, en créant de la monnaie. S'il a finalement réussi à stopper l'inflation, c'est grâce à une réduction drastique des dépenses accompagnée d'une réforme monétaire.

S'il peut exister des causes réelles à l'inflation, on peut affirmer que l'hyperinflation trouve toujours son origine dans la création excessive de monnaie.

II- La préférence pour la liquidité

Jusqu'à Keynes, l'analyse de la demande de monnaie est restée en gros focalisée sur son rôle lors des transactions. C'est à Keynes qu'on doit d'avoir considéré qu'il existait d'autres motifs de détention de monnaie. Les agents expriment alors une *préférence pour la liquidité*.

Il existe trois raisons pour lesquelles les agents peuvent préférer détenir de la liquidité : le motif de transaction (A), le motif de précaution (B) et le motif de spéculation (C). C'est en tenant compte de ces trois motifs qu'on peut définir une fonction de demande de monnaie utilisable dans les raisonnements de court terme (D).

⁶⁶ L'un de ces billets est reproduit en annexe. Si des billets de cent milliards d'euros étaient imprimés aujourd'hui, la totalité de l'agrégat M3 de la zone euro pourrait tenir en une liasse de 95 billets. Pourtant, en Allemagne en 1924, il fallait une brouette de billets pour faire ses courses.

A- Le motif de transaction

Keynes ne se distingue pas des classiques en considérant que les agents détiennent de la monnaie parce qu'ils doivent en utiliser pour leurs transactions. C'est ce qu'on appelle le *motif de transaction*.

Ce motif de détention de la monnaie s'impose à la fois aux ménages et aux entreprises. Keynes parle du motif de revenu pour les premiers et du motif d'entreprise pour les secondes. Ces deux demandes de liquidité s'additionnent pour former la demande de monnaie à des fins de transaction.

Comme le nombre de transactions augmente avec le volume de production, la demande de monnaie à des fins de transaction sera une fonction croissante du revenu.

B- Le motif de précaution

Le *motif de précaution* est défini par Keynes dans le chapitre XV de la Théorie Générale :

« Le souci de parer aux éventualités qui exigent des dépenses inopinées, l'espoir de profiter d'occasions imprévues pour réaliser des achats avantageux, et enfin le désir de conserver une richesse d'une valeur immuable pour faire face à une obligation future stipulée en monnaie sont autant de nouveaux motifs à conserver de l'argent liquide ».

On conserve donc de la monnaie parce qu'on ne peut prévoir toutes les transactions qu'on va être amené à effectuer. Détenir de la liquidité a donc une utilité puisque cela permet de profiter de bonnes affaires imprévues (acheter un vélo d'occasion, profiter des soldes...) et de faire face à une obligation inopinée (mettre de la monnaie dans un parcmètre, payer une rançon...).

Ce motif de détention de monnaie apparaît dès qu'on tient compte de l'incertitude dans laquelle se tient l'activité économique.

Le motif de précaution fait apparaître une nouvelle variable susceptible d'affecter la demande de monnaie. Il s'agit du taux d'intérêt nominal.

En effet, puisque la détention de monnaie a une utilité, les agents détiendraient toutes leurs encaisses sous forme liquide si cela n'avait pas un coût. Ce coût est le coût d'opportunité de la liquidité, soit le taux d'intérêt. A chaque fois qu'un agent décide de détenir un euro supplémentaire sous forme liquide, il renonce aux intérêts que lui aurait rapportés cet euro s'il avait été conservé sous forme de titre.

Les agents doivent donc arbitrer entre l'utilité de la liquidité et son coût d'opportunité. Par conséquent, plus le taux d'intérêt sera élevé moins les agents conserveront de monnaie. La demande de monnaie de précaution est donc une fonction décroissante du taux d'intérêt sur les titres.

On peut faire remarquer que l'importance de la demande de monnaie à des fins de transaction et de précaution dépend de la facilité avec laquelle on peut se procurer de la liquidité (les découverts sont-ils autorisés ?, les transferts entre comptes sont-ils faciles ?, les placements sont-ils liquides ? etc.). L'évolution du système financier va donc affecter la demande de monnaie. Il ne s'agit cependant pas d'un phénomène conjoncturel et il n'est pas nécessaire de l'intégrer de façon systématique à notre raisonnement.

C- Le motif de spéculation

La spéculation consiste à acheter un actif pour le revendre en espérant réaliser une plus-value. Un spéculateur cherche donc à acheter des titres au prix le plus faible pour les revendre au prix le plus élevé. La grande innovation qu'apporte Keynes à l'analyse de la demande de monnaie est de montrer que la spéculation sur le marché des titres implique un motif supplémentaire de détenir de la monnaie. Ce *motif de spéculation* amène à préciser la relation entre la demande de monnaie et le taux d'intérêt. Pour cela, il faut d'abord comprendre qu'il existe une relation précise entre le prix des titres et le taux d'intérêt courant.

La particularité d'une obligation est d'être un titre de créance négociable. On achète donc le titre en sachant qu'on obtiendra en échange une somme donnée à l'échéance. On sait cependant qu'on peut très bien la revendre avant l'échéance. Dans ce cas, c'est l'acheteur de cette obligation qui bénéficiera du versement de la somme à l'échéance. Il s'agit à présent de déterminer comment va évoluer le prix de revente de l'obligation.

Considérons une obligation émise par l'entreprise A, le 1^{er} janvier à dix heures, au taux d'intérêt en vigueur sur le marché à ce moment là, soit 10%. Chaque obligation est vendue initialement au prix de 100€ C'est la *valeur faciale* de cette obligation. Sa date d'échéance est fixée au 31 décembre.

L'entreprise A émet donc des titres qui garantissent à leur porteur le versement d'une somme de $100+10 = 110$ € le 31 décembre.

Une minute plus tard, le taux d'intérêt est passé à 20%. C'est le moment que choisit l'entreprise B pour émettre des obligations dont les caractéristiques sont les mêmes que celles

de l'entreprise A. Leur valeur faciale est donc de 100€ et leur échéance est fixée au 31 décembre. La différence est que l'entreprise B ne peut vendre ses obligations que s'il elle propose au moins le même rendement que le marché, soit 20%.

Par conséquent, les acheteurs d'une obligation B obtiendront $100+20 = 120€$ le 31 décembre.

Les détenteurs d'obligations A ont alors intérêt à les revendre pour acheter des obligations B qui offrent un meilleur rendement. Le prix des obligations A va donc baisser. Mais la baisse ne se poursuit pas indéfiniment. Au fur et à mesure que le prix des obligations A baisse, leur rendement augmente. A l'équilibre, il faut que le rendement des obligations A et B soit le même. C'est la condition d'*arbitrage*.

Chaque euro consacré à l'achat d'obligations B rapporte $(100+20)/100 = 1,2$ euros le 31 décembre.

Si le prix de revente de l'obligation A est P_A , chaque euro consacré à l'achat d'obligations A va rapporter $(100+10)/P_A$.

L'arbitrage assure que les deux rendements soient égaux : $110/P_A = 1,2$.

Le prix de l'obligation A doit donc passer à $110/1,2 \cong 91,66€$

On constate que l'augmentation du taux d'intérêt s'est traduite par une diminution du prix de l'obligation A.

On peut généraliser cette relation. Soit VF_A la valeur faciale d'une obligation et i_A le taux d'intérêt au moment de son émission. Soit i le taux d'intérêt courant sur les obligations émises par la suite mais de même échéance.

Si P_A est le prix de revente de l'obligation A, son rendement va être égal à :

$$rdt = \frac{VF(1 + i_A) - P_A}{P_A} \quad (11)$$

Puisque le rendement de l'obligation A doit en permanence être égal à celui des obligations émises par la suite, on peut écrire $rdt = i$ et résoudre l'équation pour obtenir l'expression du prix auquel l'obligation se négocie :

$$P_A = \frac{1 + i_A}{1 + i} VF \quad (12)$$

On retrouve une relation négative entre le taux d'intérêt courant et le prix de l'obligation A.⁶⁷ Cette relation va nous permettre de déterminer l'influence du taux d'intérêt

⁶⁷ Cette expression est en fait une simplification puisque nous supposons que toutes les obligations émises après l'obligation A ont la même échéance. Vous verrez dans vos cours de finance des relations plus réalistes. Dans tous les cas, la relation entre le taux d'intérêt et le prix des obligations émises par le passé subsistera.

sur la demande de monnaie à des fins de spéculation et donc la forme de la fonction de demande de monnaie à des fins de spéculation.

Si le taux d'intérêt est élevé, le prix des titres sera bas. Une forte proportion des spéculateurs anticipera donc une augmentation du prix des titres. Un grand nombre de spéculateurs cherchera donc à acheter des titres, c'est-à-dire à se débarrasser de leur monnaie. La demande de monnaie sera donc faible.

Si le taux d'intérêt est bas, le prix des titres sera élevé. Une forte proportion des spéculateurs anticipera donc une baisse du prix des titres. Un grand nombre de spéculateurs cherchera donc à vendre des titres, c'est-à-dire à les échanger contre de la monnaie. La demande de monnaie sera donc forte.

On voit donc qu'il existe une deuxième raison pour laquelle la relation entre demande de monnaie et taux d'intérêt est décroissante. Cette raison dépend du comportement des spéculateurs.

L'impact du taux d'intérêt sur la demande de monnaie à des fins de spéculation est cependant plus complexe que ce que suggère notre raisonnement pour l'instant. En effet, le prix des titres ne peut augmenter indéfiniment. Les financiers disent que « les arbres ne montent pas au ciel ».

Par conséquent, une fois que le taux d'intérêt sera devenu suffisamment faible, le prix des titres sera tellement élevé que tous les spéculateurs anticiperont un retournement de tendance au même moment. Ainsi, ils chercheront tous à vendre leurs titres. La demande de monnaie sera alors très importante et infiniment élastique au taux d'intérêt.

Cette situation est qualifiée de *trappe à liquidité* parce que les spéculateurs absorbent toute la liquidité qui est mise sur le marché. Le taux d'intérêt atteint donc une valeur minimum.

Une telle situation est plus qu'une simple curiosité théorique. En effet, on peut considérer que le Japon se trouve en situation de *trappe à liquidité* depuis le début des années 1990. Les taux d'intérêt y sont si bas que la banque centrale ne parvient plus à les faire baisser. Nous verrons que cela a des implications très fortes sur la politique monétaire.

Par ailleurs, si le taux d'intérêt devient très élevé, le prix des titres sera si bas que tous les spéculateurs anticiperont son augmentation. Ils chercheront alors à échanger toutes leurs encaisses monétaires contre des titres. La demande de monnaie sera alors très faible voire nulle. Une augmentation supplémentaire du taux d'intérêt n'aura donc plus d'effet sur la demande de monnaie qui sera donc totalement inélastique au taux d'intérêt.

On voit donc que la relation entre la demande de monnaie à des fins de spéculation et le taux d'intérêt est fortement non linéaire. Cela se répercutera sur la demande totale de monnaie.

D- La fonction de demande de liquidité

La demande totale de liquidité est la somme de la demande de monnaie à des fins de transaction, de la demande de monnaie de précaution et de la demande de monnaie de spéculation. La première est proportionnelle au revenu nominal, les deux autres diminuent lorsque le taux d'intérêt augmente.

Une façon simple de tenir compte de tous les déterminants de la demande de monnaie est la suivante :

$$M^d = L(i)PY \quad (13a)$$

Où L est la somme de la demande de monnaie de précaution et de la demande de monnaie de spéculation. C'est une fonction décroissante de i , le taux d'intérêt nominal.

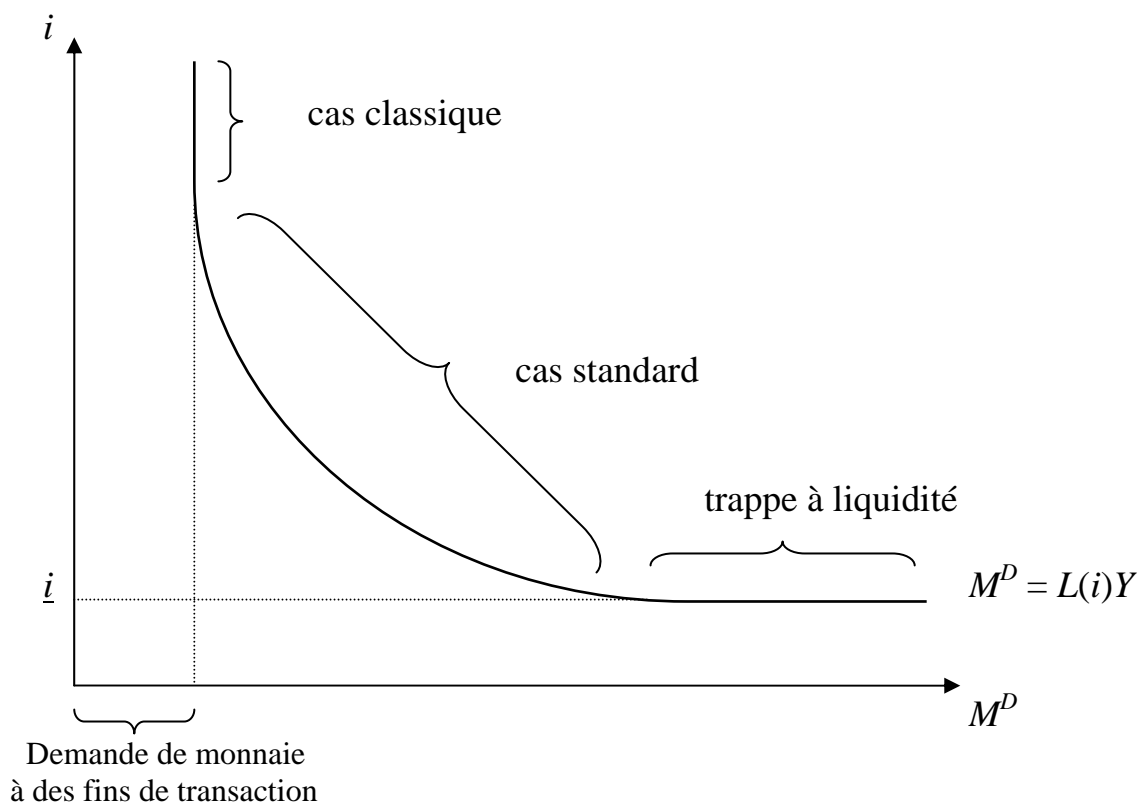
Traditionnellement on exprime la demande de monnaie en termes réels. On parle alors de *demande d'encaisses réelles*. Cette demande d'encaisses réelles représente intuitivement le « pouvoir d'achat » de la quantité de monnaie détenue, c'est-à-dire la quantité de biens que les encaisses monétaires permettraient d'acquérir instantanément.

On obtient facilement la fonction de demande d'encaisses réelles en divisant les deux membres de l'expression (13a) par le niveau des prix P :

$$M^D \equiv M^d/P = L(i)Y \quad (13b)$$

La demande d'encaisses réelles est donc une fonction du taux d'intérêt et du revenu national exprimé en termes réels. Compte tenu de nos hypothèses, on peut représenter graphiquement la demande d'encaisses réelles :

Graphique 2 :
La demande d'encaisses réelles



On voit que la demande passe par trois phases. Lorsque le taux d'intérêt est très faible, on assiste à une situation de trappe à liquidité. La demande de monnaie est alors infiniment élastique au taux d'intérêt et le taux d'intérêt atteint une valeur plancher \bar{i} .

Lorsque le taux d'intérêt est très élevé, la demande de monnaie à des fins de spéculation et la demande de monnaie de précaution sont presque nulles. La demande de monnaie se résume alors à la demande de monnaie à des fins de transaction, qui ne dépend que du revenu. L'élasticité de la demande de monnaie au taux d'intérêt est alors nulle.

Entre ces deux situations, la demande de monnaie est une fonction décroissante du taux d'intérêt. On peut parler du cas standard.

NB : Traditionnellement, on note i le taux d'intérêt nominal et r le taux d'intérêt réel. Le taux d'intérêt nominal mesure le nombre d'unités monétaires reçues en intérêt pour un prêt initial d'un euro. Le taux d'intérêt réel corrige l'impact de l'inflation. Il mesure donc la quantité de biens reçue en intérêts pour l'équivalent d'une unité de biens prêtés.

Comme le coût d'opportunité de la détention de monnaie est le taux d'intérêt nominal, c'est ce taux que les agents prennent en compte dans leurs décisions de détenir plus ou moins d'encaisses. C'est pourquoi nous utilisons le taux d'intérêt nominal dans notre raisonnement.

Cependant, beaucoup de manuels notent r le taux d'intérêt lorsqu'ils présentent le marché de la monnaie et le modèle IS-LM. Tant que les prix sont fixes, comme c'est le cas ici, cette différence n'a aucune importance puisque le taux d'intérêt nominal et le taux d'intérêt réel sont alors égaux.

Section 3 : La détermination du taux d'intérêt

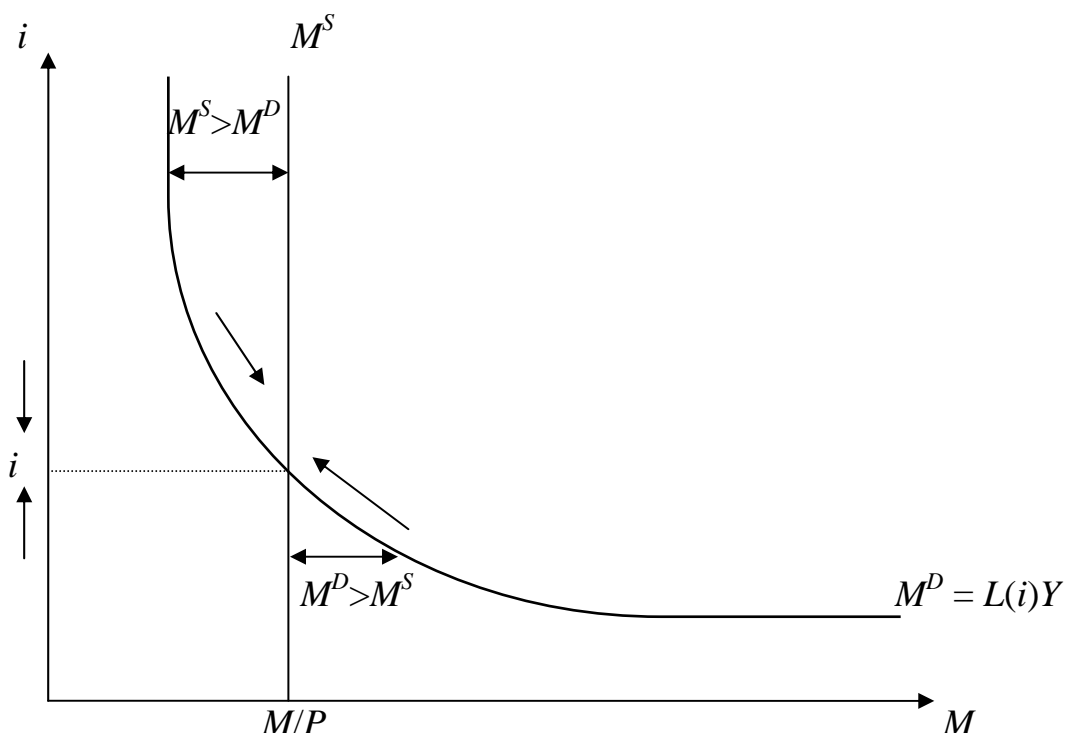
Puisque nous avons à présent défini à la fois l'offre et la demande de monnaie, nous pouvons nous intéresser à l'équilibre sur le marché de la monnaie (I). C'est lui qui va déterminer le taux d'intérêt. On pourra alors se demander comment une variation du revenu (II) ou de la masse monétaire (III) affecteront le taux d'intérêt.

I- L'équilibre sur le marché de la monnaie

L'équilibre sur le marché de la monnaie est atteint lorsque l'offre de monnaie est égale à la demande. Comme c'est le taux d'intérêt qui est déterminé sur ce marché, c'est lui qui va s'ajuster pour réaliser l'équilibre. Pour décrire comment on atteint l'équilibre, on va donc représenter l'offre et la demande de monnaie en fonction du taux d'intérêt.

Nous avons décrit la relation entre le taux d'intérêt et la demande d'encaisses réelles dans la section précédente. Auparavant, nous avons vu que l'offre de monnaie était contrôlée par la banque centrale. L'offre d'encaisses réelles, qui est égale à la masse monétaire divisée par le niveau des prix (M/P), est donc exogène au marché de la monnaie. Elle est par conséquent représentée par une droite verticale dans le repère (M, i) dans lequel nous avons déjà tracé la demande de monnaie.

Graphique 3 :
L'équilibre sur le marché de la monnaie



Compte tenu de nos hypothèses, il n'existe qu'une seule valeur du taux d'intérêt pour laquelle la demande d'encaisses réelles est égale à l'offre d'encaisses réelles. En effet, le taux d'intérêt d'équilibre est déterminé par la seule intersection entre la courbe de demande de monnaie et la droite qui représente l'offre de monnaie.

On peut de plus constater que l'équilibre est stable. Si le taux d'intérêt est supérieur au taux d'intérêt d'équilibre, l'offre d'encaisses réelles est supérieure à la demande. Le taux d'intérêt diminue donc pour que la demande puisse absorber cette offre excédentaire.

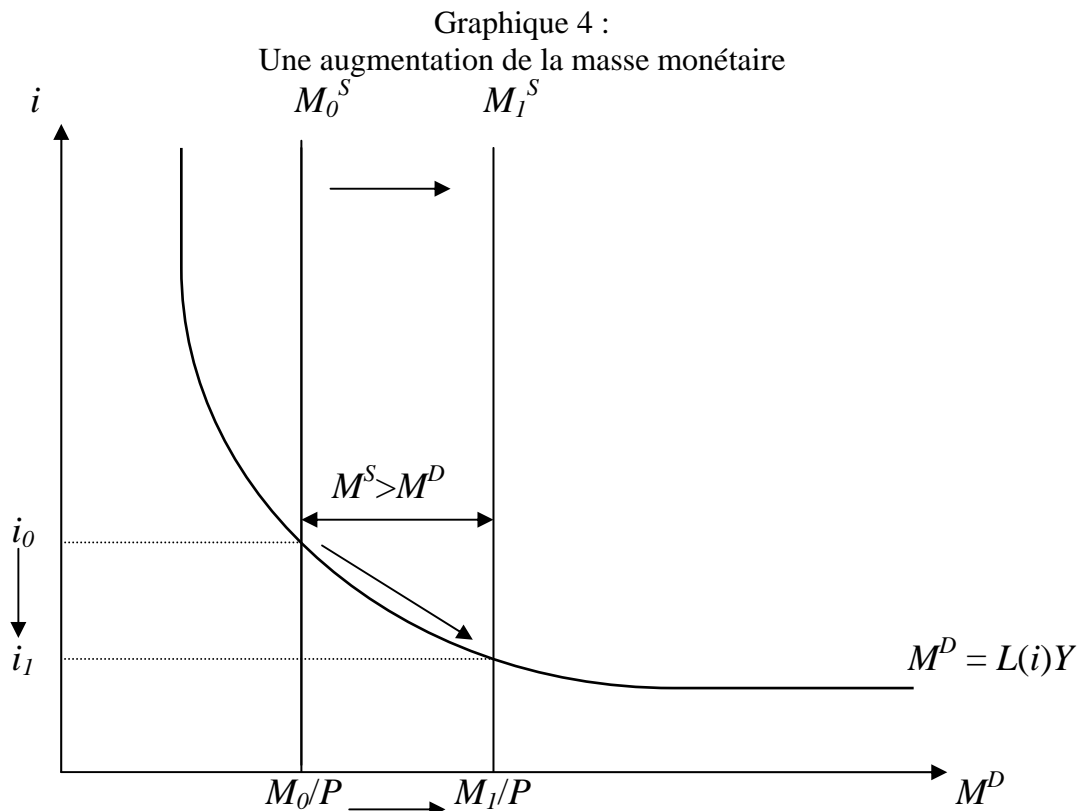
Si le taux d'intérêt est inférieur au taux d'intérêt d'équilibre, la demande d'encaisses réelles est supérieure à l'offre. Le taux d'intérêt augmente donc pour réduire la demande.

Par conséquent, le taux d'intérêt retourne systématiquement vers son niveau d'équilibre en cas de déséquilibre.

II- L'impact d'une augmentation de l'offre de monnaie

A présent que nous avons décrit la détermination du taux d'intérêt, nous pouvons réfléchir à la façon dont des perturbations exogènes au marché de la monnaie vont l'affecter. Commençons par réfléchir à l'impact d'une augmentation de l'offre de monnaie. Une telle

augmentation de l'offre de monnaie est décidée par la banque centrale. Il s'agit donc d'une décision de politique monétaire.



L'augmentation de la masse monétaire se traduit par une augmentation de l'offre d'encaisses réelles. Elle passe de M_0/P à M_1/P . Dans notre graphique, la droite qui représente l'offre d'encaisses réelles se déplace donc vers la droite.

Le graphique ci-dessus montre que l'augmentation de l'offre de monnaie provoque une offre excédentaire d'encaisses réelles. Pour résorber cette offre d'encaisses réelles, il faut que la demande de monnaie augmente. Le taux d'intérêt va donc diminuer. On atteindra donc un nouvel équilibre, caractérisé par un taux d'intérêt plus faible que le taux d'intérêt initial ($i_1 < i_0$).

On peut donc conclure qu'une augmentation de l'offre de monnaie se traduit par une diminution du taux d'intérêt. Or nous avons vu dans le chapitre précédent que l'investissement était une fonction décroissante du taux d'intérêt. Par conséquent, si le taux d'intérêt diminue, l'investissement va augmenter, ce qui va augmenter la demande globale et éventuellement relancer l'activité. C'est pourquoi une politique d'augmentation de l'offre de monnaie est qualifiée de *politique monétaire expansionniste*.

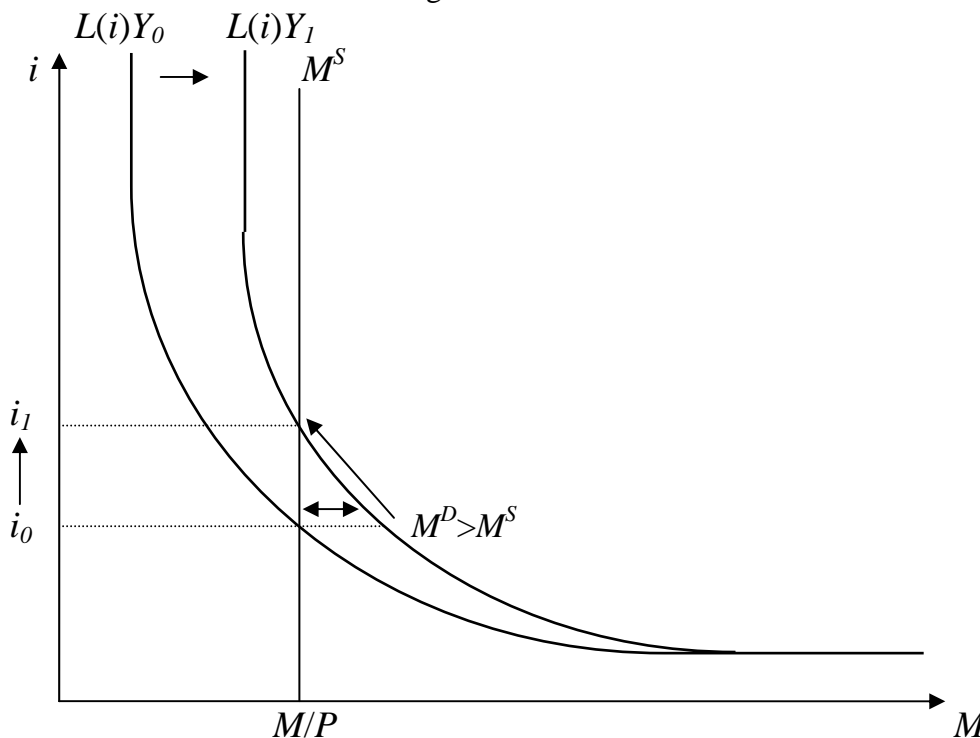
On peut lire le graphique à l'envers et supposer que la banque centrale réduit l'offre de monnaie. Dans ce cas, l'offre d'encaisses réelles passe de M_1/P à M_0/P . Il s'ensuit que le taux d'intérêt augmente pour passer de i_1 à i_0 . Cette augmentation aura donc un effet restrictif sur l'investissement, la demande et l'activité. On parle donc de *politique monétaire restrictive*.

Ce résultat n'est cependant pas valable lorsque l'économie est en situation de trappe à liquidité. Dans ce cas, la droite d'offre de monnaie coupe la courbe de demande de monnaie dans sa partie horizontale. L'augmentation de l'offre de monnaie est donc sans effet sur le taux d'intérêt, qui ne peut pas passer en dessous de sa valeur minimum.

III- L'impact d'une augmentation du revenu

La politique monétaire n'est pas la seule source d'évolution sur le marché de la monnaie. Dans la mesure où la demande d'encaisses réelles est une fonction du revenu, toute augmentation de l'activité augmentera également la demande d'encaisses réelles.

Graphique 5 :
Une augmentation du revenu



Dans le graphique ci-dessus, l'augmentation du revenu se traduit par un déplacement de la courbe de demande de monnaie vers le nord-est. Ce déplacement signifie que la demande de monnaie est supérieure à son niveau initial quel que soit le taux d'intérêt

considéré. Cependant, la partie horizontale de la demande de monnaie, qui correspond à la situation de trappe à liquidité n'est pas affectée.

L'augmentation du revenu se traduit alors instantanément par une demande excédentaire d'encaisses réelles. Pour résorber cette demande excédentaire, le taux d'intérêt augmente. La demande de monnaie de précaution et à des fins de spéculation diminue donc et on retrouve l'équilibre.

On constate donc qu'une augmentation du revenu se traduit par une augmentation du taux d'intérêt ($i_1 > i_0$). A l'inverse, si on lisait le graphique à l'envers, on observerait qu'une diminution du revenu se traduirait par une diminution du taux d'intérêt qui passerait alors de i_1 à i_0 .

Ici encore, ce résultat n'est pas valable lorsque l'économie est en situation de trappe à liquidité. Dans ce cas, la droite d'offre de monnaie coupe la courbe de demande de monnaie dans sa partie horizontale. L'augmentation de la demande de monnaie est donc sans effet sur le taux d'intérêt.

Conclusion

Après avoir étudié les facteurs qui déterminent l'offre et la demande de monnaie, nous avons décrit l'équilibre sur le marché de la monnaie. Il apparaît que le marché de la monnaie subit l'influence du marché des biens, puisque le revenu détermine la demande de monnaie à des fins de transaction. A l'inverse, comme le taux d'intérêt affecte l'investissement, le marché de la monnaie va influencer le marché des biens.

Si on souhaite décrire correctement le fonctionnement de l'économie, il est donc nécessaire de tenir compte conjointement de l'équilibre sur le marché de la monnaie et sur le marché des biens et de leurs interactions. C'est précisément ce que permet le modèle IS-LM que nous allons étudier dans le prochain chapitre.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 5.

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 9.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 18.

Pour les passionnés

Lavigne A. et J.P. Pollin : *Les théories de la monnaie*, Repères, La Découverte, Paris, 1997.

Mishkin F. : *Monnaie, banques et marchés financiers*, Pearson Education France, Paris, 2004
(chapitres 3, 15, 16 et 21)

Annexe

Document A1 : Evolution de la masse monétaire suite à un dépôt initial de 1000€

Etape	Dépôts	Total
1	1000	1000
2	900	1900
3	810	2710
4	729	3439
5	656,1	4095,1
10	387,42	6513,21
20	135,09	8784,23
30	47,10	9576,09
40	16,42	9852,19
50	5,73	9948,46
100	0,0295	9999,73

Document A2 : Un billet de cent milliards de marks



Le 15 novembre 1924, ce billet permettait d'acheter deux verres de bière.

(<http://www.joelscoins.com/exhibger2.htm>)

Les notes du cours ne sont transmises qu'à titre indicatif.

Chapitre 6 : Le modèle IS-LM

Introduction : 74 ans et toutes ses dents

Le modèle IS-LM a été proposé en 1937 par John Hicks (prix Nobel 1972)⁶⁸. Ce dernier l'a construit pour formaliser ce qu'il considérait être le principal apport de Keynes, c'est-à-dire l'analyse de l'interdépendance des marchés des biens et de la monnaie. Il a été développé par la suite par Alvin Hansen dans deux ouvrages parus en 1949 et 1953. C'est pourquoi on parle aussi du diagramme de Hicks et Hansen.

Depuis 1937, les connaissances en macroéconomie se sont étoffées et des théories plus complètes et rigoureuses du fonctionnement de l'économie ont été développées. On continue pourtant d'étudier le modèle IS-LM. La première raison de la longévité de ce modèle est qu'il fournit encore la logique de base des modèles macroéconométriques qui sont utilisés à des fins de prévision.

Par ailleurs, le modèle IS-LM décrit les principaux mécanismes à l'œuvre lorsqu'une politique de gestion de la demande globale est appliquée. De plus, il se prête à un grand nombre d'extensions qui étendent son pouvoir explicatif. Malgré sa simplicité, il est donc un instrument d'analyse de la politique macroéconomique puissant et flexible.

Enfin, le modèle IS-LM conserve des qualités pédagogiques certaines. Il permet de traiter très simplement des phénomènes relativement complexes. Il permet de confronter les arguments des différents courants dans un cadre unique et d'isoler les variables qu'il est nécessaire d'observer empiriquement. C'est pourquoi il est toujours utilisé en premier cycle et que la plupart des économistes en activité aujourd'hui ont appris à l'utiliser.

Le modèle IS-LM est construit pour décrire l'interdépendance entre le marché des biens et celui de la monnaie. Il est donc constitué de deux blocs principaux : la courbe IS qui décrit l'équilibre sur le marché des biens et la courbe LM qui décrit l'équilibre sur le marché de la monnaie. La construction de ces deux blocs et la façon dont ils permettent de décrire l'équilibre de l'économie feront l'objet de la première section de ce chapitre. Dans la deuxième section, nous verrons comment le modèle permet d'analyser les effets des politiques de gestion de la demande.

⁶⁸ J.R. Hicks « Mr. Keynes and the "Classics": a suggested interpretation », *Econometrica*, vol 5 n°2, p. 147-159, avril 1937. A. Hansen, *Monetary theory and fiscal policy*, New York, McGraw-Hill, 1949, et *A guide to Keynes*, New York, McGraw-Hill, 1953.

Section 1 : Construction du diagramme IS-LM

Comme son nom l'indique, le diagramme IS-LM est constitué de deux courbes : IS et LM. Nous allons d'abord étudier la construction et les caractéristiques de ces deux courbes (I et II) avant de déterminer l'équilibre global de l'économie (III).

I- La courbe IS

La courbe IS représente l'équilibre sur le marché des biens. Nous allons l'obtenir à partir de l'égalité entre la demande et la production de biens, mais on pourrait tout aussi bien partir de l'égalité entre l'épargne et l'investissement, d'où son nom. I et S sont en effet les initiales de « investment » et « savings ».

Nous allons commencer par décrire le principe général de construction de la courbe IS (A) avant d'analyser les déterminants de sa forme et de sa position (B), ce qui sera utile lorsque nous utiliserons le diagramme pour comprendre le fonctionnement des politiques économiques.

A- Pourquoi IS est décroissante

La courbe IS regroupe *toutes les combinaisons du revenu et du taux d'intérêt qui assurent l'équilibre sur le marché des biens*. Plus précisément, elle représente l'ensemble des valeurs d'équilibre du revenu pour chaque valeur possible du taux d'intérêt. En d'autres termes, elle décrit le revenu déterminé de façon endogène sur le marché des biens pour chaque valeur du taux d'intérêt, qui est exogène à ce marché.

Dans le chapitre consacré au marché des biens (chapitre 4), nous avons étudié l'équation qui définit l'équilibre sur le marché des biens. Nous avons supposé que l'investissement était exogène. Cette hypothèse n'est plus tenable si on étudie l'équilibre de toute l'économie au lieu de se concentrer sur le marché des biens. Si on rappelle que l'investissement dépend du taux d'intérêt, on obtient l'équation de la courbe IS :

$$Y = C(Y) + I(i) + G \quad (1)$$

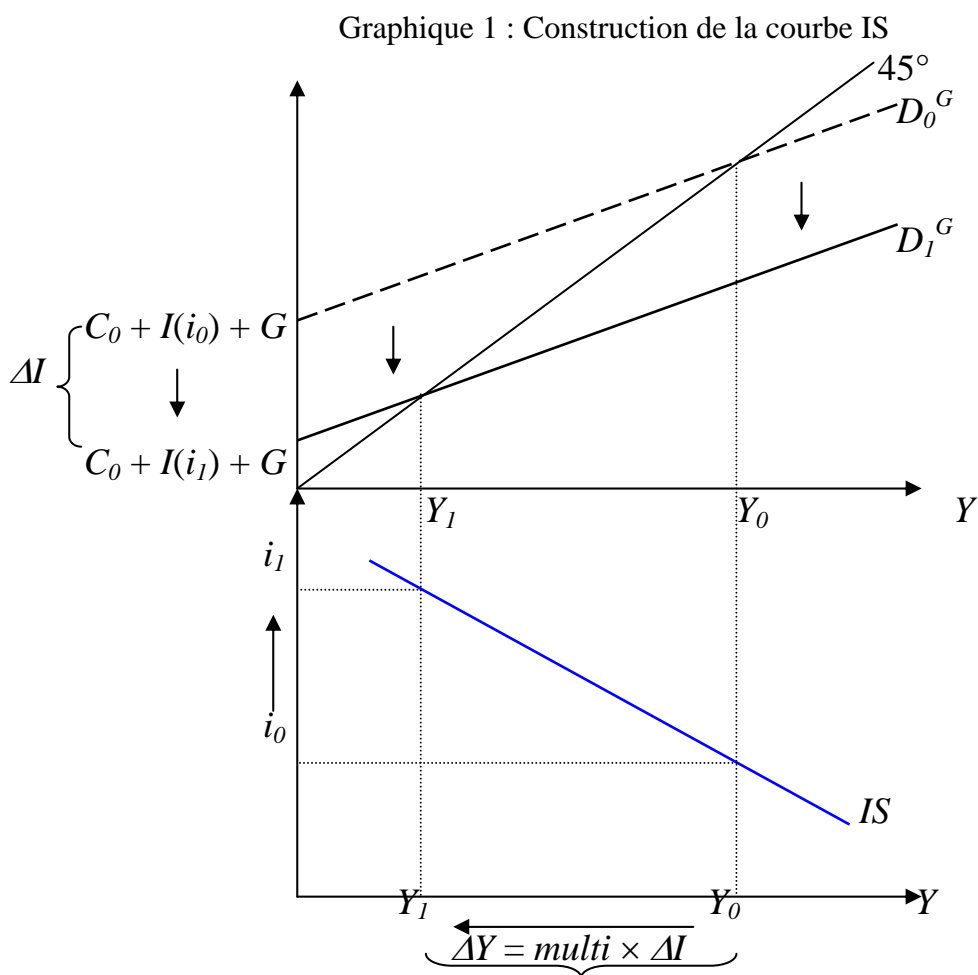
Cette équation permet de définir IS de façon algébrique, mais l'intuition de ce que représente cette courbe peut être saisie graphiquement.⁶⁹ Pour construire la courbe IS, il faut

⁶⁹ A titre d'exercice, vous pouvez construire la courbe IS et vérifier les déterminants de sa forme et de sa position en supposant que la consommation est décrite par une fonction de consommation keynésienne $C = c.Y + C_0$ et que l'investissement est une fonction linéaire du taux d'intérêt : $I = I_0 - z.i$.

repartir de l'équilibre sur le marché des biens, c'est-à-dire du diagramme à 45 degrés et y représenter l'impact d'une variation du taux d'intérêt. La courbe IS se représente cependant dans un autre repère que le diagramme à 45 degrés.

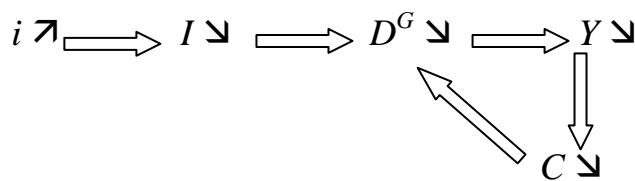
Le graphique 1 est donc constitué de deux repères superposés. Celui du dessus permet de tracer le diagramme à 45 degrés. Celui du dessous est consacré à la courbe IS qu'il s'agit de déterminer. Ce repère mesure le revenu en abscisses et le taux d'intérêt en ordonnées.

Supposons que le taux d'intérêt augmente d'un montant arbitraire pour passer de i_0 à i_1 ($i_1 > i_0$). Cette variation exogène du taux d'intérêt va inciter les entreprises à réduire leurs investissements, ce qui réduit la demande globale et déplace ainsi la courbe de demande globale vers le bas. On constate alors que le revenu d'équilibre a diminué ($Y_1 < Y_0$).



Si on reporte les valeurs du taux d'intérêt et du revenu d'équilibre dans le repère du dessous, et qu'on les relie, on constate que **la courbe IS est décroissante**.

La logique de construction de la courbe IS est résumée par l'enchaînement ci-dessous.



L'augmentation du taux d'intérêt réduit l'investissement, ce qui déprime la demande globale. La production s'ajuste à la demande, ce qui réduit le revenu puis la consommation. Il s'enclenche alors un mécanisme multiplicateur qui amplifie la réduction initiale de l'investissement.

On peut donc conclure que la diminution de la production va être égale à la variation de l'investissement multipliée par le multiplicateur de dépenses. ($\Delta Y = multi \times \Delta I$).

B- Forme et position de la courbe IS

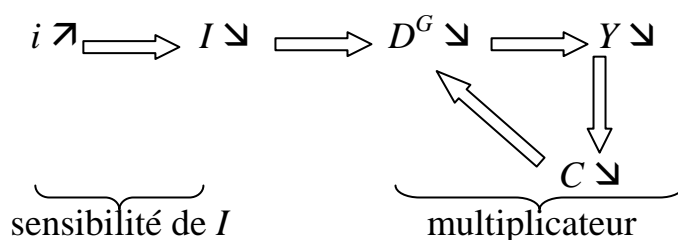
Même si nous la représentons par une droite, la courbe IS est bien une courbe qui peut prendre n'importe quelle forme. La seule de ses caractéristiques dont nous pouvons être sûrs est qu'elle est décroissante. Nous l'avons représentée par une droite et continuerons de le faire parce que c'est plus simple et aussi parce que nous avons supposé que la consommation était une fonction linéaire du revenu.

Il faut cependant garder à l'esprit qu'il ne s'agit là que d'une simplification dont les vertus sont essentiellement pédagogiques, même si nos conclusions resteraient les mêmes si nous utilisions des courbes IS de forme différente. Cela ne nous empêche pas de réfléchir à la pente de la courbe IS (1), ni à sa position dans le repère (2).

1) Forme de la courbe IS

La pente de la courbe IS représente la sensibilité du revenu d'équilibre sur le marché des biens aux variations du taux d'intérêt. Il s'agit donc d'un paramètre essentiel qui mesure la puissance du canal de transmission entre le marché de la monnaie et celui des biens.

Pour isoler les paramètres qui déterminent la pente de la courbe IS, il suffit de reprendre l'enchaînement des effets qui transforment une variation du taux d'intérêt en une variation du revenu.



Les notes du cours ne sont transmises qu'à titre indicatif.

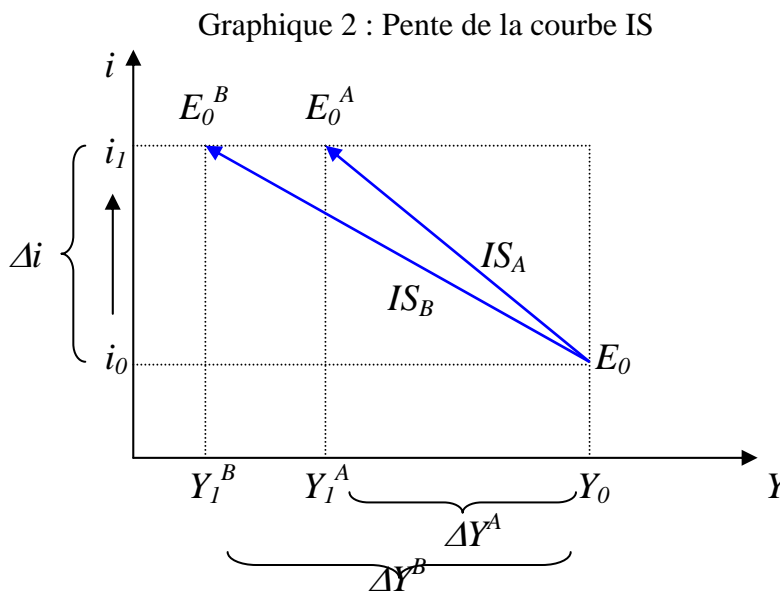
Le graphique suggère qu'on peut diviser l'effet total du taux d'intérêt en deux effets élémentaires : un effet direct du taux d'intérêt sur l'investissement et un effet induit qui dépend de la valeur du multiplicateur.

Par conséquent, l'effet total sera d'autant plus fort que l'investissement sera sensible aux variations du taux d'intérêt.

De même, l'effet total sera d'autant plus fort que l'effet multiplicateur sera puissant. Comme nous avons vu que l'effet multiplicateur dépend de la propension marginale à consommer, nous pouvons affirmer que l'effet total sera d'autant plus fort que la propension marginale à consommer sera élevée.

Si ces deux effets sont plus puissants, une même augmentation du taux d'intérêt aura un effet plus important. La courbe IS sera donc plus plate.

Pour s'en convaincre, considérons deux économies (A et B) qui connaissent au départ le même revenu et le même taux d'intérêt (Y_0, i_0) et supposons que le taux d'intérêt augmente de la même façon dans les deux économies ($i_1 > i_0$). Supposons que la propension marginale à consommer soit plus élevée et que l'investissement soit plus sensible au taux d'intérêt dans la deuxième économie que dans la première. Cela signifie que l'augmentation du taux d'intérêt déprimera plus le revenu dans l'économie B que dans l'économie A. Cela se traduit dans le graphique 2.



Les notes du cours ne sont transmises qu'à titre indicatif.

Il apparaît donc que puisque le revenu de l'économie B est plus affecté que celui de l'économie A par la même augmentation du taux d'intérêt, la courbe IS de l'économie B sera moins raide que celle de l'économie A.

En résumé, on peut dire qu'une sensibilité plus importante de l'investissement aux variations du taux d'intérêt et/ou une propension marginale à consommer plus élevée se traduiront par une courbe IS moins raide.

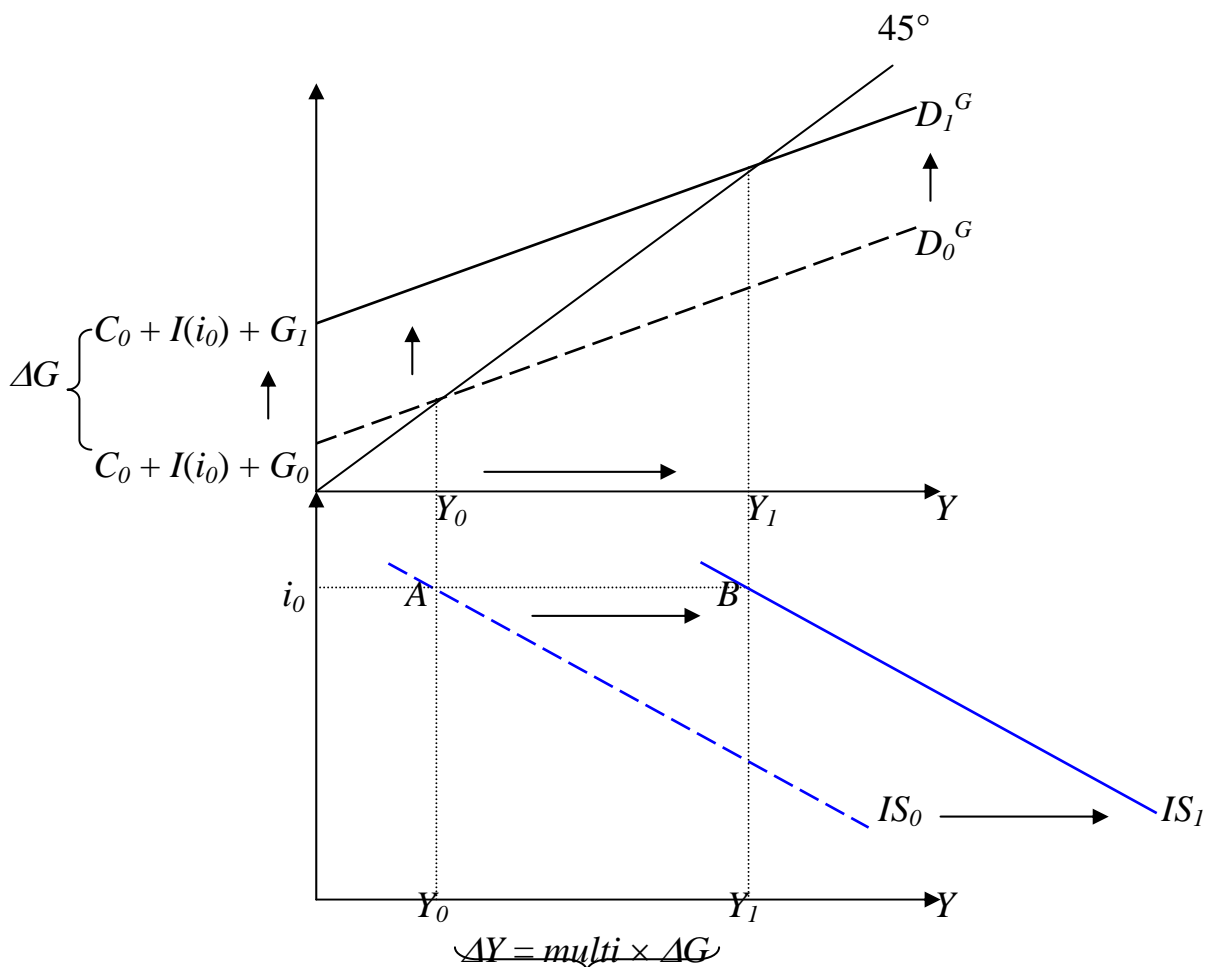
Cette conclusion est importante parce qu'elle indique deux paramètres qu'il serait utile d'estimer empiriquement pour prévoir l'impact d'une diminution des taux d'intérêts.

2) Position de la courbe IS

La position de la courbe IS dans le repère (Y, i) se modifie dès que l'un des paramètres qui déterminent la position de la courbe de demande globale se modifie. Dans le cas linéaire sur lequel nous nous concentrons, les déplacements de la courbe D^G peuvent être dus à une modification de la consommation incompressible (C_0) ou des dépenses publiques (G). Toute évolution du comportement des investisseurs qui les amènerait à modifier le volume de leurs investissements indépendamment de la valeur du taux d'intérêt aurait le même effet.

Pour être plus concret, supposons que les dépenses publiques augmentent ($G_1 > G_0$) alors que le taux d'intérêt reste constant (i_0). La demande globale augmente alors et la courbe de demande globale se déplace vers le haut. On constate alors que le revenu d'équilibre a augmenté ($Y_1 > Y_0$) grâce à l'effet multiplicateur.

Graphique 3 : Déplacement de la courbe IS



Lorsqu'on reporte cette augmentation dans le repère inférieur, on passe du point A au point B. Comme le choix du taux d'intérêt initial était arbitraire, on pourrait recommencer le même exercice pour n'importe quelle valeur du taux d'intérêt. Dans tous les cas l'augmentation du revenu serait rigoureusement la même puisqu'elle est égale à l'augmentation des dépenses multipliée par le multiplicateur. La nouvelle courbe IS est par conséquent parallèle à l'ancienne. **Une augmentation des dépenses publiques se traduit donc par une translation de la courbe IS vers la droite.**

On peut aussi lire le graphique à l'envers pour décrire les effets d'une réduction des dépenses publiques. **Une diminution des dépenses publiques se traduit donc par une translation de la courbe IS vers la gauche.** Ces résultats seront utiles au moment d'étudier les effets de la politique budgétaire.

II- La courbe LM

La courbe LM représente l'équilibre sur le marché de la monnaie. Nous allons l'obtenir à partir de l'égalité entre la demande de liquidité et l'offre de monnaie. Les initiales LM signifient justement « liquidity » et « money ».

Comme pour la courbe IS, nous allons commencer par décrire le principe général de construction de la courbe LM (A) avant d'analyser les déterminants de sa forme et sa position (B).

A- Pourquoi LM est croissante

La courbe LM regroupe *toutes les combinaisons du revenu et du taux d'intérêt qui assurent l'équilibre sur le marché de la monnaie*. Plus précisément, elle représente l'ensemble des valeurs d'équilibre du taux d'intérêt pour chaque valeur possible du revenu. En d'autres termes, elle décrit le taux d'intérêt déterminé de façon endogène sur le marché de la monnaie pour chaque valeur du revenu qui est exogène à ce marché.

La courbe LM est tracée dans le même repère (Y, i) que la courbe IS. Il s'agit donc de reporter le taux d'intérêt déterminé sur le marché de la monnaie pour chacune des valeurs possibles du revenu. La construction de la courbe LM part donc de l'équilibre sur ce marché de la monnaie. Algébriquement, on pourrait l'étudier grâce à la condition d'équilibre de ce marché :

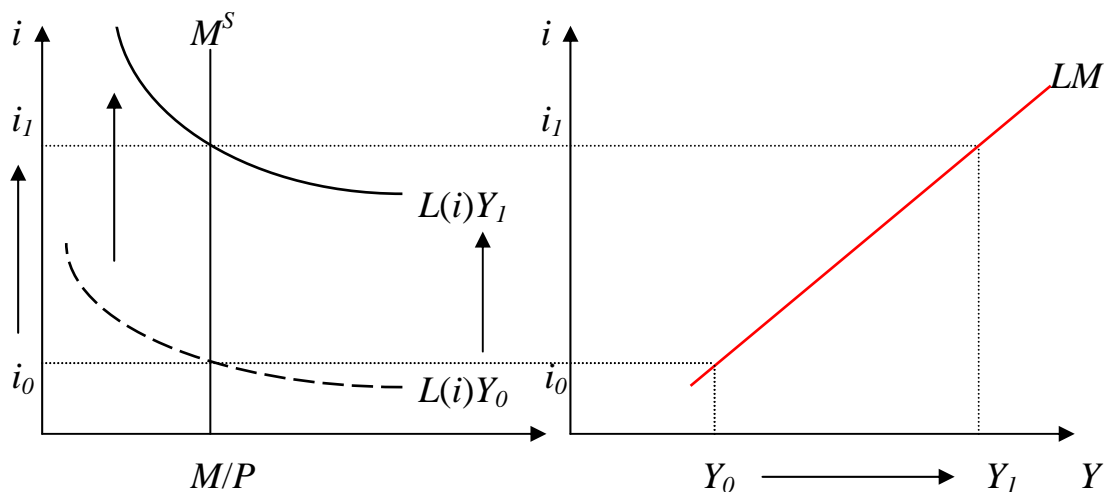
$$M/P = L(i)Y \quad (2)$$

Cette équation permet de définir LM de façon algébrique, mais l'intuition de ce que représente cette courbe peut être saisie graphiquement.⁷⁰

Le graphique 4 est constitué de deux repères disposés l'un à côté de l'autre. Celui de gauche permet de représenter l'équilibre sur le marché de la monnaie. Celui de droite est consacré à la courbe LM qu'il s'agit de tracer. Ce repère mesure le revenu en abscisses et le taux d'intérêt en ordonnées.

⁷⁰ A titre d'exercice, vous pouvez construire la courbe LM et vérifier les déterminants de sa forme et de sa position de façon simple en supposant que la demande de monnaie est une fonction linéaire du revenu et du taux d'intérêt : $L^d = L_0 + g.Y - hi$.

Graphique 4 : Construction de la courbe LM



Supposons que le revenu augmente d'un montant arbitraire pour passer de Y_0 à Y_1 ($Y_1 > Y_0$). Cette variation exogène du revenu va augmenter la demande de monnaie à des fins de transaction, ce qui déplace la courbe de demande vers le haut. On constate alors que le taux d'intérêt d'équilibre a augmenté ($i_1 > i_0$).

Si on reporte les valeurs du taux d'intérêt et du revenu d'équilibre dans le repère de droite, et qu'on les relie, on constate que **la courbe LM est croissante**.

La logique de construction de la courbe LM est résumée par l'enchaînement ci-dessous.

$$Y \nearrow \implies M^D \nearrow \implies i \nearrow \implies M^D \searrow$$

L'augmentation du revenu augmente les besoins de monnaie pour les transactions. Il apparaît alors une demande excédentaire de liquidité qui provoque à son tour une augmentation du taux d'intérêt. Comme le taux d'intérêt augmente, la demande de monnaie de précaution et à des fins de spéculation diminue. On retourne ainsi à l'équilibre sur le marché de la monnaie.

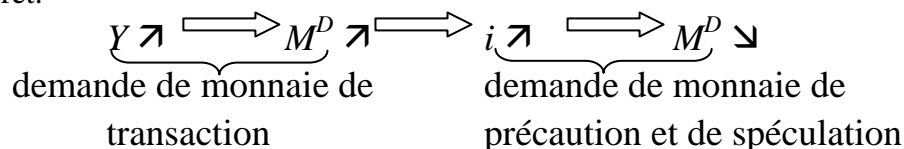
B- Forme et position de la courbe LM

Comme nous l'avons fait pour la courbe IS, nous allons d'abord examiner les déterminants de la pente de la courbe LM (1), puis de sa position dans le repère (2).

1) Forme de la courbe LM

La pente de la courbe LM représente la sensibilité du taux d'intérêt d'équilibre sur le marché de la monnaie aux variations du revenu. Il s'agit donc d'un paramètre essentiel qui mesure la puissance du canal de transmission entre le marché des biens et celui de la monnaie.

Pour isoler les paramètres qui déterminent la pente de la courbe LM, il suffit de reprendre l'enchaînement des effets qui transforment une variation du revenu en une variation du taux d'intérêt.



L'enchaînement suggère qu'on peut diviser l'effet total du revenu en deux effets élémentaires : un effet du revenu sur la demande de monnaie à des fins de transaction et un effet indirect sur la demande de monnaie de précaution et à des fins de spéculation.

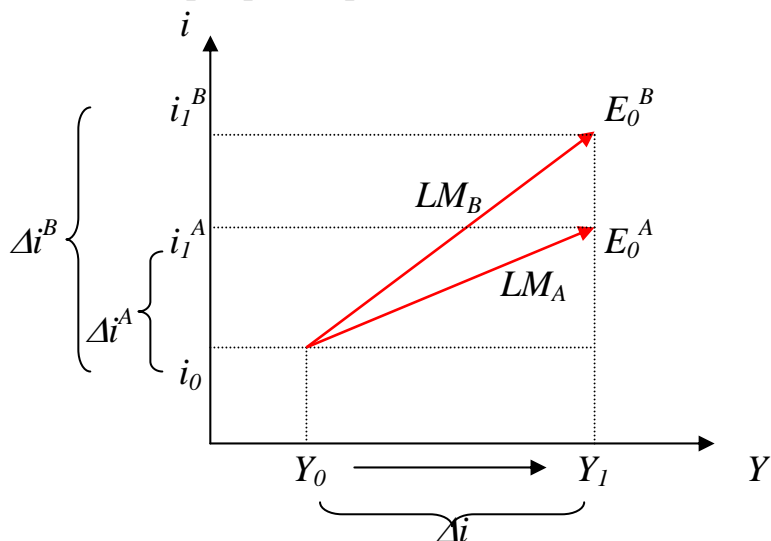
Plus la demande de monnaie à des fins de transaction sera sensible aux variations du revenu, plus l'augmentation de la demande de monnaie sera forte. Il faudra alors une forte augmentation du taux d'intérêt pour rétablir l'équilibre sur le marché de la monnaie, en réduisant la demande de monnaie aux fins de précaution et de spéculation.

De même, si la demande de monnaie de précaution et à des fins de spéculation est peu sensible aux variations du taux d'intérêt, il faudra aussi une forte augmentation du taux d'intérêt pour réduire la demande.

Par conséquent, on peut dire que plus la demande de monnaie sera sensible aux variations du revenu et peu sensible aux variations du taux d'intérêt plus une même augmentation du revenu provoquera une augmentation importante du taux d'intérêt. La courbe LM sera donc d'autant plus raide que la demande de monnaie sera sensible aux variations du revenu et peu sensible aux variations du taux d'intérêt.

Pour s'en persuader, considérons à nouveau deux économies (A et B) qui connaissent au départ le même revenu et le même taux d'intérêt (Y_0, i_0), et supposons que le revenu augmente de la même façon dans les deux économies ($Y_1 > Y_0$). Supposons que la demande de monnaie soit plus sensible aux variations du revenu et moins sensible aux variations du taux d'intérêt dans l'économie B que dans l'économie A. Cela se représente facilement dans le graphique 5.

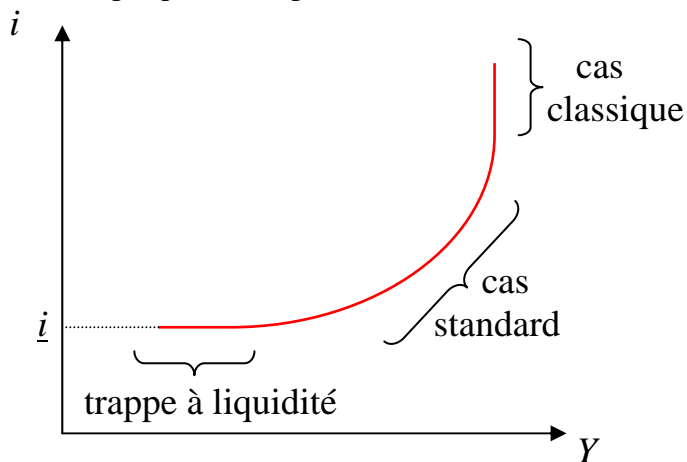
Graphique 5 : pente de la courbe LM



Il apparaît donc que puisque le taux d'intérêt de l'économie B est plus affecté que celui de l'économie A par la même augmentation du revenu, la courbe LM de l'économie B sera plus raide que celle de l'économie A.

Nous avons souligné dans le chapitre précédent que la relation entre la demande de monnaie et le taux d'intérêt était non linéaire. Plus précisément, la demande devient infiniment sensible au taux d'intérêt lorsque le taux d'intérêt est très faible à cause de la trappe à liquidité. Elle y devient insensible lorsque le taux d'intérêt est très élevé car il ne subsiste plus que la demande de monnaie de transaction. D'après la relation que nous venons d'analyser entre la pente de la courbe LM et la sensibilité au taux d'intérêt de la demande de monnaie, cela implique que la pente de la courbe LM ne sera pas constante.

Graphique 6 : les phases de la courbe LM



Les notes du cours ne sont transmises qu'à titre indicatif.

Au contraire, la pente de la courbe LM sera très faible pour de faibles valeurs du taux d'intérêt et très élevée pour des valeurs élevées du taux d'intérêt. Comme la demande de monnaie, la courbe LM passera par plusieurs phases.

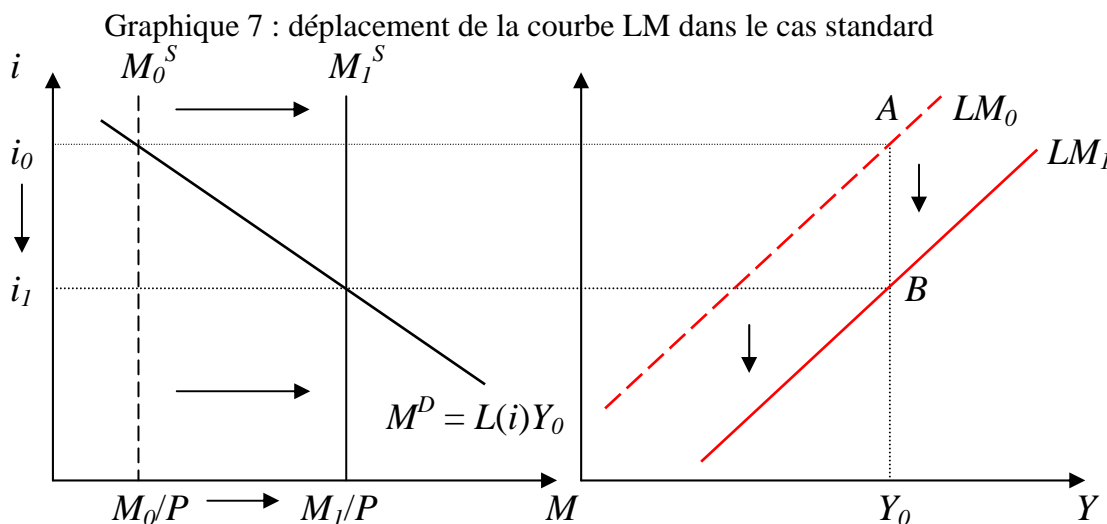
Pour de faibles valeurs du taux d'intérêt et du revenu, la courbe LM est horizontale. C'est la situation de trappe à liquidité. Pour des valeurs plus élevées du taux d'intérêt et du revenu, LM est verticale. C'est le cas classique. Entre ces deux extrêmes, la courbe LM est croissante.

Il se peut que nous représentions la courbe LM par une droite dans la suite du cours. Plus encore que pour la courbe IS, cela doit être interprété comme une approximation. Il s'agit de plus d'un cas particulier.

2) Position de la courbe LM

Le principal paramètre qui détermine la position de la courbe LM dans le repère (Y, i) est l'offre de monnaie.

Supposons ainsi que la banque centrale augmente la masse monétaire ($M_1 > M_0$) alors que le revenu reste constant (Y_0). La droite d'offre d'enchasses réelles se déplace alors vers la droite. On constate alors l'apparition d'une offre excédentaire d'enchasses réelles. Le marché de la monnaie ne retrouvera son équilibre que si le taux d'intérêt diminue afin d'inciter les agents à augmenter leurs enchasses ($i_1 < i_0$).



Lorsqu'on reporte cette augmentation dans le repère de droite, on passe du point A au point B. Comme le choix du revenu initial était arbitraire, on pourrait recommencer le même exercice pour n'importe quelle valeur du revenu. La nouvelle courbe LM est par conséquent

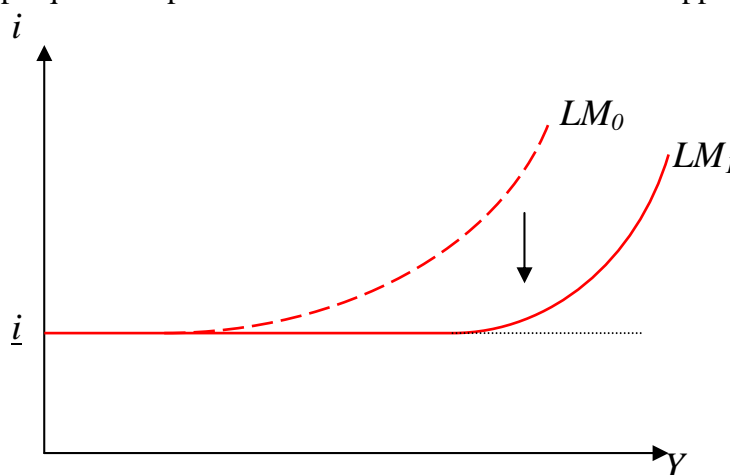
parallèle à l'ancienne. **Une augmentation de la masse monétaire se traduit donc par une translation de la courbe LM vers le bas.**

On peut aussi lire le graphique à l'envers pour décrire les effets d'une contraction de la masse monétaire. **Une contraction de la masse monétaire se traduit donc par une translation de la courbe LM vers le haut.**

Ces résultats ne sont cependant pas vrais tout au long de la courbe LM. En effet, si le taux d'intérêt initial est trop faible, l'économie se trouvera en situation de trappe à liquidité. L'offre supplémentaire de monnaie sera donc absorbée par les spéculateurs sans pour autant diminuer le taux d'intérêt. On peut donc dire qu'une augmentation de l'offre de monnaie n'affectera pas la partie horizontale de la courbe LM.

Le graphique ci-dessous représente l'effet d'une augmentation de l'offre de monnaie sur la courbe LM de façon plus complète.

Graphique 8 : déplacement de la courbe LM avec une trappe à liquidité



S'il existe une trappe à liquidité, ce n'est que la partie croissante de la courbe LM qui se déplace vers le bas. La partie horizontale, qui correspond à la trappe à liquidité, ne se modifie pas.

III- L'équilibre global

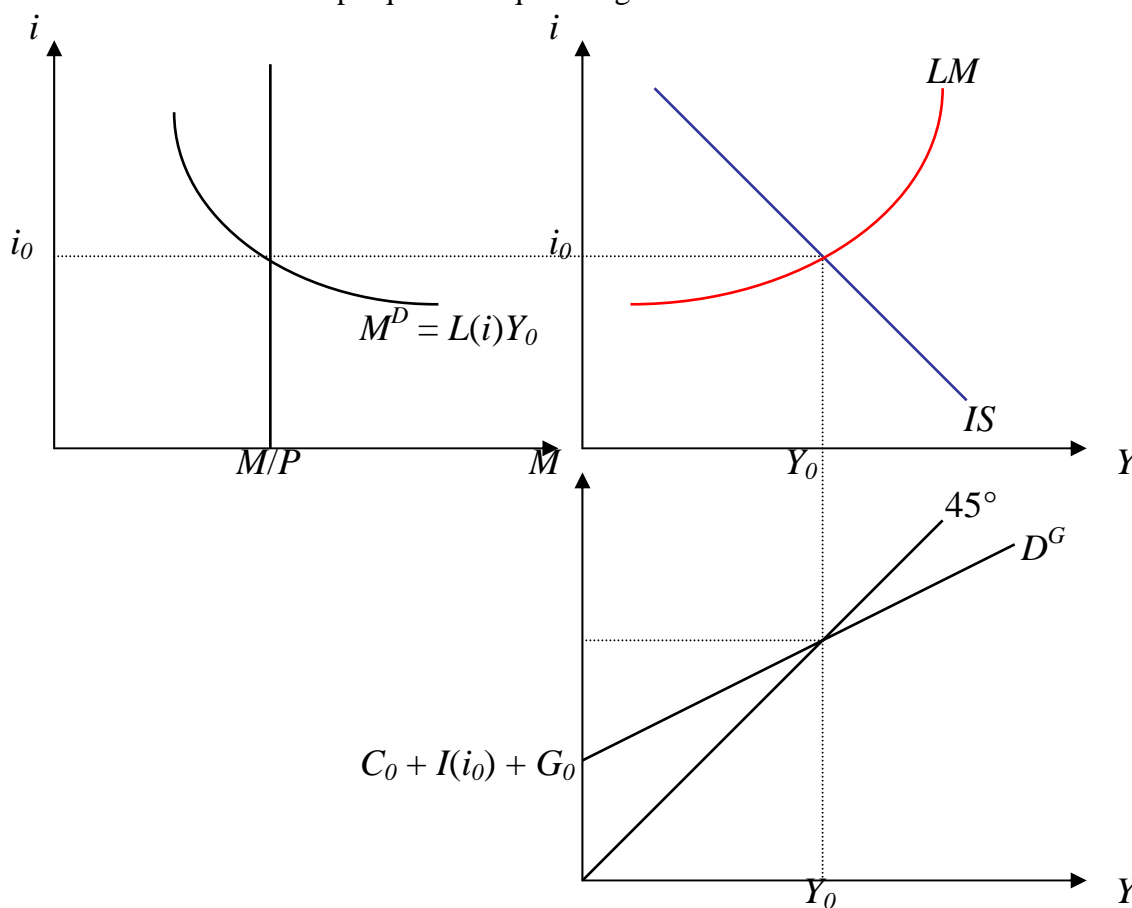
Nous avons à présent décrit séparément l'équilibre du marché de la monnaie et de celui des biens. Pour respecter la contrainte de bouclage, il est nécessaire de déterminer à

présent l'équilibre conjoint des deux marchés. C'est ce qu'on appelle *l'équilibre global* de l'économie.⁷¹

Cet équilibre global s'obtient en représentant les courbes IS et LM dans le même graphique, ce qui est possible puisqu'elles sont toutes les deux tracées dans le même repère (Y, i) . Rappelons que la première rassemble tous les couples (Y, i) qui assurent l'équilibre sur le marché des biens, alors que la seconde rassemble tous les couples (Y, i) qui assurent l'équilibre sur le marché de la monnaie.

Comme l'équilibre global de l'économie suppose que les deux marchés soient en équilibre, il sera représenté par un point qui appartient aux deux courbes à la fois. L'équilibre global correspond donc à l'intersection entre la courbe IS et la courbe LM.⁷²

Graphique 9 : l'équilibre global de l'économie



⁷¹ On peut se demander s'il ne faudrait pas aussi prendre en compte le troisième marché que nous avons évoqué, à savoir le marché des titres. Le faire serait inutile. En effet, la loi de Walras stipule que si deux marchés sur trois sont en équilibre alors le troisième sera forcément aussi en équilibre.

⁷² Algébriquement, l'équilibre global s'obtient en résolvant le système d'équations formé par l'équation de IS et celle de LM. Ce système est constitué de deux équations et a deux inconnues. On peut donc le résoudre. A titre d'exercice, vous pouvez essayer de retrouver les résultats que nous allons obtenir de façon graphique en utilisant les équations des deux courbes données plus haut.

Le graphique 9 montre qu'au point d'intersection entre les courbes IS et LM, le marché de la monnaie et celui des biens sont conjointement en équilibre. On a donc déterminé à la fois le revenu et le taux d'intérêt d'équilibre (Y_0, i_0). Comme l'équilibre sur ces deux marchés est stable, l'équilibre global sera stable lui aussi. Il ne se modifiera que si l'un ou l'autre des paramètres du modèle est modifié.

L'intérêt du modèle IS-LM est qu'il synthétise l'équilibre de deux marchés en un seul diagramme. Dorénavant, nous n'utiliserons plus que ce diagramme pour décrire l'équilibre de l'économie. Il faudra cependant garder à l'esprit ce que représentent les deux courbes et comment elles ont été construites.

Enfin, rien ne permet d'affirmer que le revenu d'équilibre correspond au plein emploi. La prochaine section vérifie dans quelles circonstances les politiques économiques permettent de s'en rapprocher.

Section 2 : La politique économique dans le modèle IS-LM

Le diagramme IS-LM permet de traiter facilement deux types de politiques de demande : la politique monétaire (I) et la politique budgétaire (II). Ces deux politiques peuvent se compléter ou se contrecarrer. Il est donc intéressant d'analyser le *policy-mix* dans son ensemble (III).

I- La politique monétaire

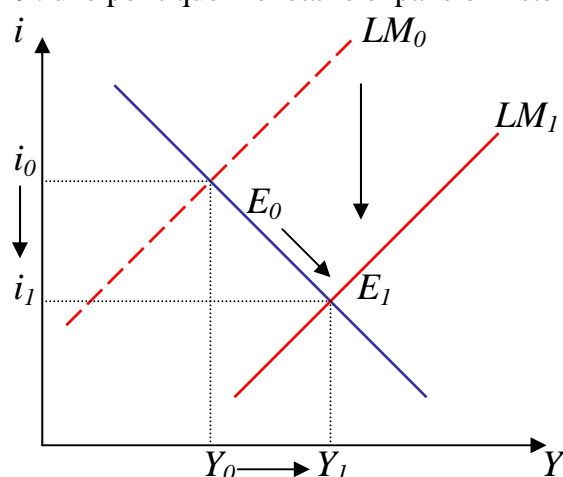
La politique monétaire consiste à manipuler l'offre de monnaie dans l'espoir d'affecter le revenu. Elle peut tout aussi bien consister dans une expansion de la masse monétaire (politique monétaire expansionniste) que dans une contraction de la masse monétaire (politique monétaire restrictive). Dans cette section nous allons nous intéresser aux effets d'une politique monétaire expansionniste. Les effets d'une politique monétaire restrictive pourront s'obtenir par symétrie.

Après avoir décrit le principe de la politique monétaire (A), nous verrons les facteurs qui déterminent son efficacité (B).

A- Principe de la politique monétaire

Une politique monétaire expansionniste consiste à augmenter l'offre de monnaie dans l'espoir de faire baisser les taux d'intérêt et de relancer l'activité.

Graphique 10 : une politique monétaire expansionniste dans le cas standard



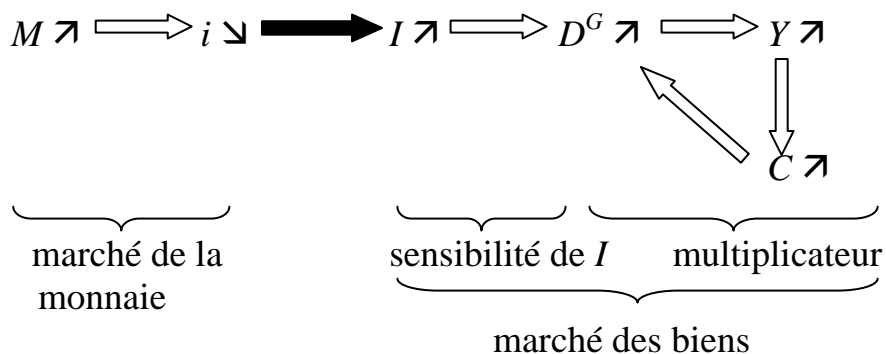
Considérons le cas standard, dans lequel la courbe LM est croissante et la courbe IS décroissante. Dans ce cas, l'augmentation de la masse monétaire provoque une offre excédentaire d'encaisses réelles qui ne se résorbe qu'à condition que le taux d'intérêt diminue. Nous avons vu que cela se traduisait par une translation vers le bas de la courbe LM. Le diagramme IS-LM permet d'en visualiser toutes les conséquences.

Le graphique 10 montre qu'**une politique monétaire expansionniste provoque à la fois une diminution du taux d'intérêt ($i_1 < i_0$) et une augmentation du revenu ($Y_1 > Y_0$)**.

Ce résultat est fondamental et il faut en comprendre les mécanismes, ce que le diagramme ne permet pas à lui seul. Il faut donc reprendre les différentes étapes qui ont permis à une augmentation de la masse monétaire d'affecter le revenu.⁷³

L'augmentation de la masse monétaire crée une offre excédentaire de monnaie qui provoque une diminution du taux d'intérêt. L'effet de la politique est alors transmis au marché des biens (flèche noire) par l'investissement. La demande globale augmente, ce qui incite les entreprises à produire davantage. Le revenu augmente, ce qui relance la consommation et enclenche un effet multiplicateur.

⁷³ Ce résultat ne va pas de soi. A l'époque où il a été obtenu, on considérait qu'il existait une dichotomie entre la sphère réelle et la sphère monétaire.



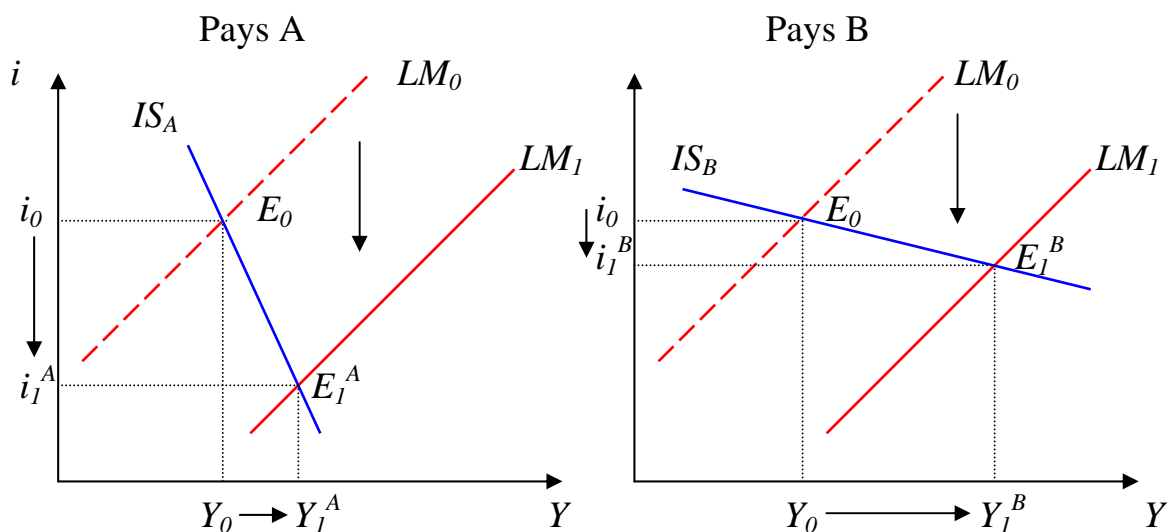
C'est donc le taux d'intérêt qui constitue le canal de transmission entre la sphère monétaire et la sphère réelle. Il s'agit à présent de déterminer ce qui affecte l'efficacité de ce canal.

B- Efficacité de la politique monétaire

L'efficacité de la politique monétaire se mesure par l'augmentation du revenu que permet l'augmentation de la masse monétaire. Ce que montre notre raisonnement, c'est que la politique monétaire aura un effet plus ou moins important sur le revenu en fonction des facteurs qui affectent le canal de transmission entre les deux marchés. Il s'agit là d'abord des facteurs qui déterminent la pente de la courbe IS, mais le fonctionnement du marché de la monnaie joue aussi un rôle.

Pour visualiser ce résultat, comparons deux pays qui partent de la même situation, avec la même courbe LM, et qui mettent en œuvre la même politique monétaire expansionniste. Supposons que la courbe IS du pays B est moins raide que celle du pays A. Nous savons que cela peut être dû soit au fait que l'investissement est plus sensible au taux d'intérêt et où que la propension marginale à consommer est plus élevée dans le pays B que dans le pays A.

Graphique 11 : L'efficacité de la politique monétaire



Le graphique 11 permet deux constats. D'abord, plus la courbe IS est raide moins la politique monétaire est efficace. L'augmentation du revenu est en effet plus faible dans le pays A que dans le pays B. Ce résultat était prévisible puisque la pente de la courbe IS de ce pays indique que l'investissement y est très sensible au taux d'intérêt et/ou que le multiplicateur y est puissant.

On peut donc conclure que **la politique monétaire est d'autant plus efficace que l'investissement est sensible au taux d'intérêt et la propension marginale à consommer élevée.**

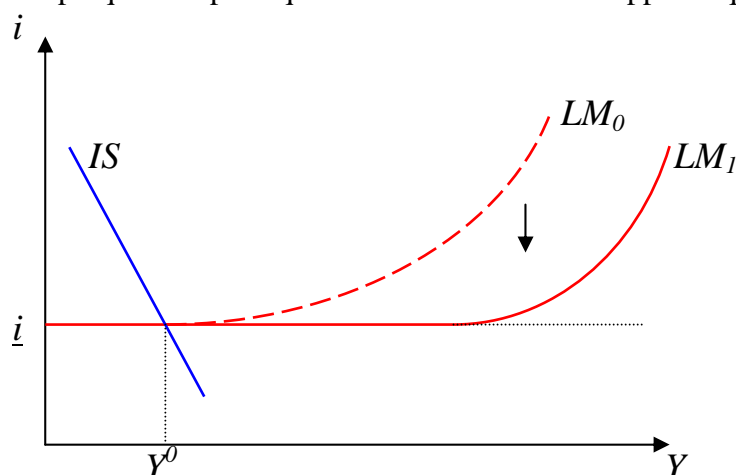
On constate aussi que le taux d'intérêt a moins diminué dans le pays B que dans le pays A. Ce résultat peut sembler plus contre intuitif a priori. Il s'explique cependant si on réalise que le raisonnement qui nous a permis de prévoir l'effet d'une augmentation de l'offre de monnaie sur le revenu est incomplet. En effet, il ne tient pas compte de l'impact de l'augmentation du revenu sur la demande de monnaie de transaction.

Si on tient compte de ce dernier effet, on réalise que la demande de monnaie doit augmenter, ce qui compense un peu l'augmentation de la masse monétaire. Le taux d'intérêt va donc être plus élevé que ce que nous aurions prévu en nous concentrant simplement sur le marché des biens.

Comme le revenu augmente plus dans le pays B que dans le pays A, l'augmentation de la demande de monnaie y sera forcément plus importante. Le taux d'intérêt y diminuera donc moins.

Ces conclusions sont valables dans le cas standard. Cependant, si on veut déterminer les conditions d'efficacité de la politique monétaire de façon exhaustive, on doit tenir compte d'une situation particulière que nous avons déjà évoquée : la trappe à liquidité. Le graphique 12 permet de visualiser l'effet d'une politique monétaire lorsque l'économie se trouve en situation de trappe à liquidité. Dans ce cas, la courbe IS coupe initialement la courbe LM dans sa portion horizontale.

Graphique 12 : politique monétaire en cas de trappe à liquidité



L'augmentation de la masse monétaire déplace la portion croissante de la courbe LM vers le bas mais n'affecte pas sa portion horizontale. Par conséquent, l'équilibre de l'économie n'est pas modifié.

On peut donc conclure que **la politique monétaire est inefficace en situation de trappe à liquidité.**

Intuitivement, lorsque l'économie est en situation de trappe à liquidité, les spéculateurs absorbent instantanément toute la liquidité que la banque centrale peut créer. Par conséquent, le taux d'intérêt reste constant ce qui empêche la politique monétaire d'affecter le marché des biens. Les effets de la politique monétaire restent donc circonscrits au seul marché de la monnaie.

II- La politique budgétaire

La politique budgétaire consiste à manipuler les dépenses publiques pour affecter la demande globale. Elle peut tout aussi bien consister dans une augmentation des dépenses (politique budgétaire expansionniste) qu'en leur réduction (politique budgétaire restrictive).

Comme pour la politique monétaire, nous n'allons nous intéresser qu'aux effets d'une politique budgétaire expansionniste dans cette section. Les effets d'une politique budgétaire restrictive pourront s'obtenir par symétrie.

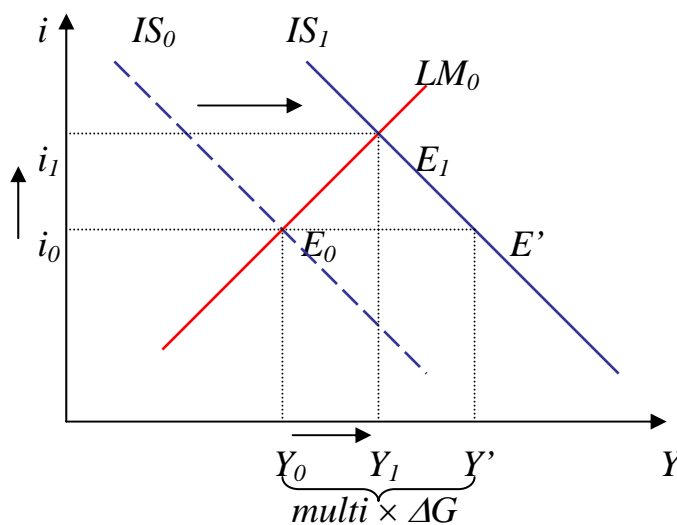
Après avoir décrit le principe de la politique budgétaire (A), nous verrons les facteurs qui déterminent son efficacité (B).

A- Principe de la politique budgétaire

Une politique budgétaire expansionniste consiste à augmenter les dépenses publiques dans l'espoir d'augmenter la demande et relancer l'activité.

Considérons le cas standard, dans lequel la courbe LM est croissante et la courbe IS décroissante. Dans ce cas, l'augmentation des dépenses relance la demande globale, ce qui provoque une translation de la courbe IS vers la droite, l'ampleur de ce déplacement correspondant à l'effet multiplicateur. C'est ce que montre le graphique ci-dessous :

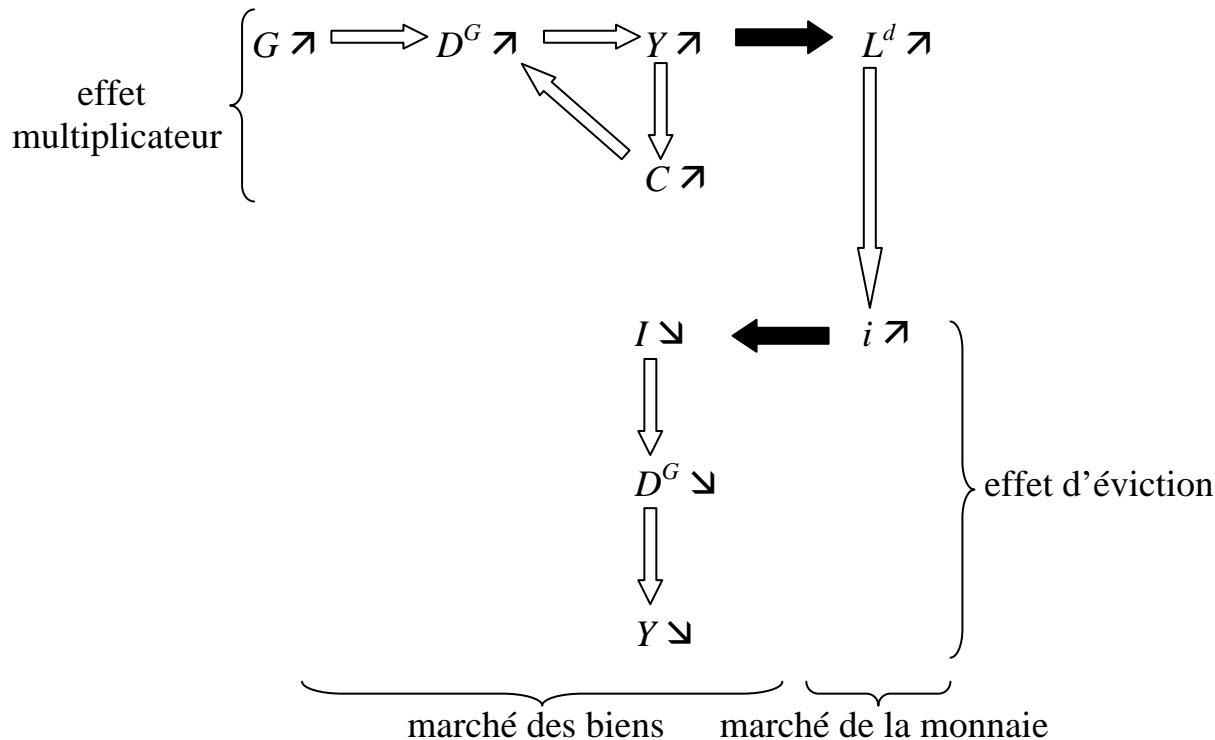
Graphique 13 : une politique budgétaire expansionniste dans le cas standard



Le graphique 13 montre qu'une **politique budgétaire expansionniste provoque une augmentation à la fois du revenu ($Y_1 > Y_0$) et du taux d'intérêt ($i_1 > i_0$).**

On peut cependant aussi remarquer que **l'augmentation du revenu est inférieure à celle que l'effet multiplicateur aurait laissé prévoir ($Y_1 < Y'$).**

Ce résultat est fondamental mais il faut en comprendre les mécanismes, ce que le diagramme ne permet pas à lui seul. Il faut donc reprendre les différentes étapes qui ont permis à une augmentation des dépenses publiques d'augmenter le revenu et le taux d'intérêt.



L'augmentation des dépenses publiques augmente la demande globale. La production va donc augmenter pour satisfaire la demande supplémentaire. Comme la production augmente, le revenu augmente lui aussi, ce qui se répercute sur la consommation et enclenche un effet multiplicateur.

Ces effets pouvaient déjà être analysés dans le cadre du diagramme à 45 degrés. La nouveauté ici est que l'augmentation du revenu va se répercuter sur le marché de la monnaie (flèche noire) parce que la demande de monnaie de transaction augmente. Il apparaît alors une demande excédentaire de monnaie qui ne peut se résorber que grâce à une augmentation du taux d'intérêt.

Cette augmentation du taux d'intérêt va à son tour se répercuter sur le marché des biens en déprimant l'investissement (deuxième flèche noire). L'augmentation de la demande publique est ainsi en partie compensée (on dit aussi évincée) par la diminution de l'investissement. L'effet total de la politique budgétaire est donc inférieur à l'effet multiplicateur.

L'effet négatif de l'augmentation des dépenses publiques sur l'investissement privé est appelé *effet d'éviction*.

Ici non plus, il n'y a pas de dichotomie entre sphère réelle et sphère monétaire. La demande de monnaie de transaction est affectée par l'augmentation du revenu. En retour, l'investissement est affecté par l'augmentation du taux d'intérêt.

Il s'agit à présent de déterminer de quoi dépend l'importance de ces effets.

B- Efficacité de la politique budgétaire

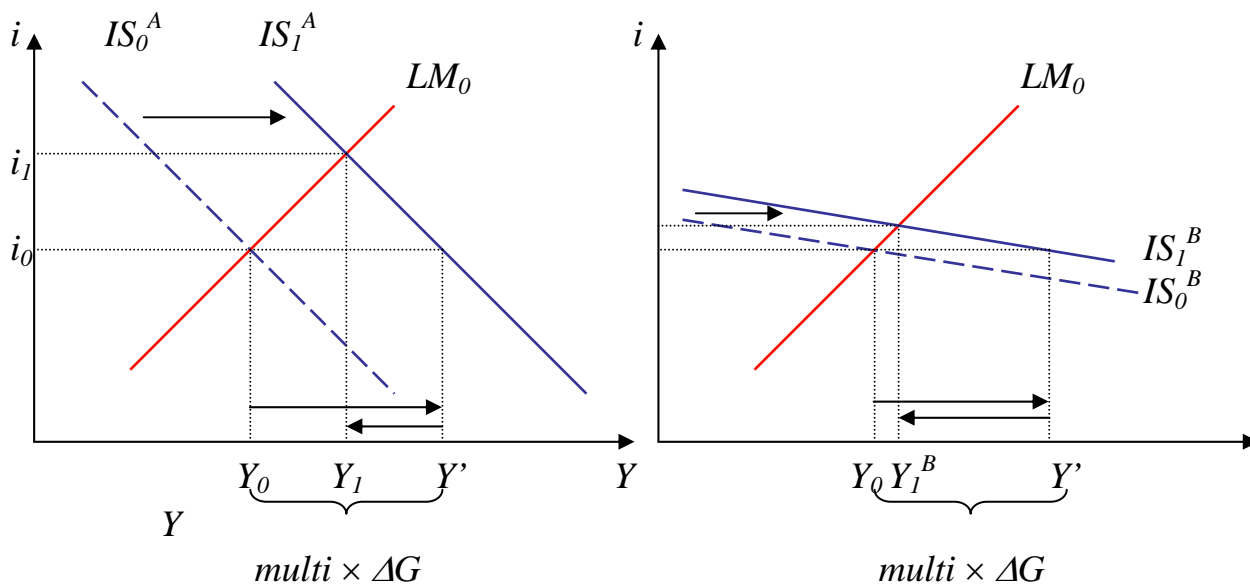
L'efficacité de la politique budgétaire se mesure par l'augmentation du revenu que permet l'augmentation des dépenses. Ce que montre notre raisonnement, c'est qu'il y a au moins deux effets dont l'ampleur va déterminer l'effet total de la politique budgétaire : l'effet multiplicateur et l'effet d'éviction. Comme nous avons déjà étudié l'effet multiplicateur dans le chapitre consacré au marché des biens, nous allons nous concentrer sur l'effet d'éviction. Ce dernier va dépendre à la fois de facteurs qui ont trait au marché des biens et de la monnaie.

Nous venons de voir que l'effet d'éviction dépend de la réaction de l'investissement au taux d'intérêt. C'est donc la pente de la courbe IS qui est en jeu ici. Si l'investissement est très sensible au taux d'intérêt, la courbe IS sera très plate.

Le graphique ci-dessous compare l'effet de la politique budgétaire dans deux économies qui se trouvent au départ dans la même situation (Y_0, i_0). Tous leurs paramètres sont identiques, sauf la sensibilité de l'investissement au taux d'intérêt. L'investissement est plus sensible au taux d'intérêt dans l'économie B que dans l'économie A.

Dans les deux quadrants, l'accroissement des dépenses publiques se traduit par une translation de la courbe IS vers la droite. Mesuré sur l'axe des abscisses, ce déplacement mesure l'effet multiplicateur, qui est le même dans les deux graphiques.

Graphique 14 : L'efficacité de la politique budgétaire en fonction de la sensibilité de l'investissement au taux d'intérêt



Les deux économies ne se distinguent que par la pente de la courbe IS qui est due à une sensibilité différente de l'investissement au taux d'intérêt. On constate que l'effet d'éviction est plus important dans l'économie B que dans l'économie A. L'augmentation du revenu est donc plus faible dans l'économie B que dans l'économie A.

On peut donc conclure que **la politique budgétaire est d'autant plus efficace que la valeur du multiplicateur est élevée et que l'investissement est peu sensible au taux d'intérêt.**

Si la forme de la courbe IS peut déterminer l'efficacité de la politique budgétaire, c'est surtout sur le rôle du marché de la monnaie que les débats ont porté. En effet, c'est l'augmentation du taux d'intérêt qui détermine l'ampleur de l'effet d'éviction. Par conséquent, c'est la forme de la courbe LM qui joue le rôle principal.

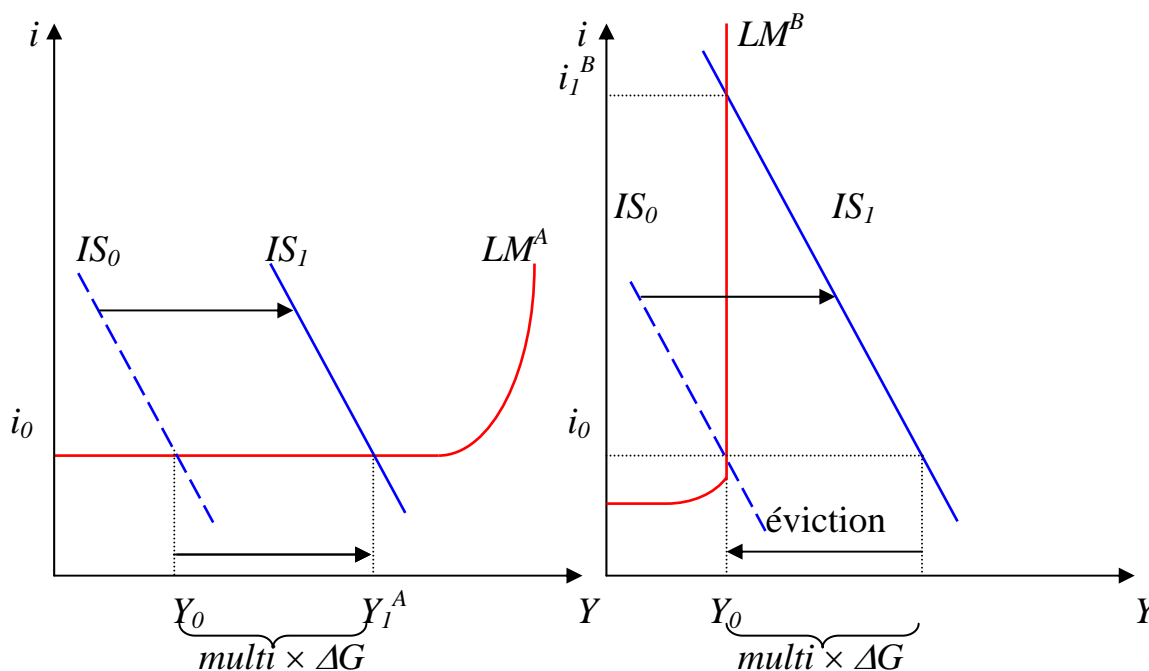
Les deux paramètres qui déterminent la pente de la courbe LM sont la sensibilité au revenu de la demande de monnaie de transaction et la sensibilité au taux d'intérêt de la demande de monnaie de précaution et à des fins de spéculation. En effet, plus la demande de monnaie de transaction sera sensible au revenu, plus l'augmentation initiale du revenu provoquera une demande excédentaire de monnaie. L'augmentation du taux d'intérêt nécessaire pour résorber cette demande excédentaire sera donc plus importante, ce qui augmentera l'effet d'éviction.

Bien que la sensibilité de la demande de monnaie de transaction puisse influencer l'efficacité de la politique budgétaire, c'est sur le rôle de la sensibilité de la demande de monnaie au taux d'intérêt que les débats ont porté. En effet, une des principales différences entre les théories keynésienne et classique porte sur les déterminants de la demande de monnaie.

Pour les classiques, la monnaie n'est détenue que pour réaliser des transactions. Sa demande est donc insensible au taux d'intérêt. L'apport de la théorie keynésienne est d'avoir suggéré qu'il existait d'autres motifs de détention de la monnaie et que la demande de monnaie était donc sensible au taux d'intérêt.

On peut donc présenter l'opposition entre les théories keynésienne et classique comme le résultat d'un désaccord portant sur la pente de la courbe LM. Pour les classiques, cette courbe est verticale. Pour les keynésiens, elle est croissante, voire horizontale lorsque l'économie est en situation de trappe à liquidité. Comme le cas d'une courbe LM croissante a déjà été présenté, nous allons nous concentrer sur les deux cas extrêmes.

Graphique 15 : effet de la politique budgétaire en cas de trappe à liquidité et dans le cas classique



Le graphique ci-dessus présente l'effet d'une politique budgétaire de même ampleur dans deux économies qui partent de la même position mais qui se distinguent par la situation de leur marché de la monnaie. L'économie A est confrontée à une trappe à liquidité alors que l'économie B se trouve en situation classique.

Dans l'économie A, on constate que la politique budgétaire provoque une augmentation du revenu correspondant à l'effet multiplicateur et que le taux d'intérêt reste constant. Dans cette économie, il n'y a donc pas d'effet d'éviction et l'efficacité de la politique budgétaire est maximale.

En effet, comme l'économie est en situation de trappe à liquidité, le taux d'intérêt a atteint sa valeur minimum et la demande de monnaie est infiniment sensible au taux d'intérêt. L'augmentation de la demande de monnaie sera résorbée sans que le taux d'intérêt ait besoin de réagir. Comme le taux d'intérêt est constant, l'effet d'éviction sera nul. L'effet total de la politique budgétaire sera donc égal à l'effet multiplicateur. On peut donc conclure que **l'efficacité de la politique budgétaire est maximale lorsque l'économie est en situation de trappe à liquidité.**

La situation est toute différente dans l'économie B où la demande de monnaie est totalement insensible au taux d'intérêt. Dans cette économie, l'effet d'éviction compense totalement l'effet multiplicateur.

Dans cette économie, la politique budgétaire provoque une relance temporaire grâce à l'effet multiplicateur. Comme le revenu augmente, la demande de monnaie de transaction augmente, ce qui entraîne une demande excédentaire de monnaie. Le taux d'intérêt se met alors à augmenter, mais cette augmentation n'affecte pas la demande de monnaie.

Elle affecte en revanche l'investissement, qui se met à décroître. Comme l'investissement diminue, la demande globale diminue, ce qui réduit le revenu. La demande de monnaie de transaction diminue alors, ce qui réduit la demande excédentaire de monnaie. Ce mécanisme se poursuit tant qu'il subsiste une demande excédentaire de monnaie.

Le marché de la monnaie ne retourne à l'équilibre que lorsque la demande de monnaie de transaction a retrouvé son niveau initial. Cela signifie que le revenu a également retrouvé son niveau initial. Par conséquent, la diminution de l'investissement a intégralement compensé l'augmentation initiale des dépenses publiques. On dit que les dépenses publiques ont évincé l'investissement privé. Le seul effet de la relance budgétaire est d'avoir fait augmenter le taux d'intérêt et réduit l'investissement. On peut donc conclure que la **politique budgétaire est totalement inefficace lorsque la demande de monnaie est insensible au taux d'intérêt.**

En comparant ces deux situations et en les complétant par le cas standard, on peut donc affirmer que **la politique budgétaire est d'autant plus efficace que la demande de monnaie est sensible au taux d'intérêt.**

III- Le policy-mix

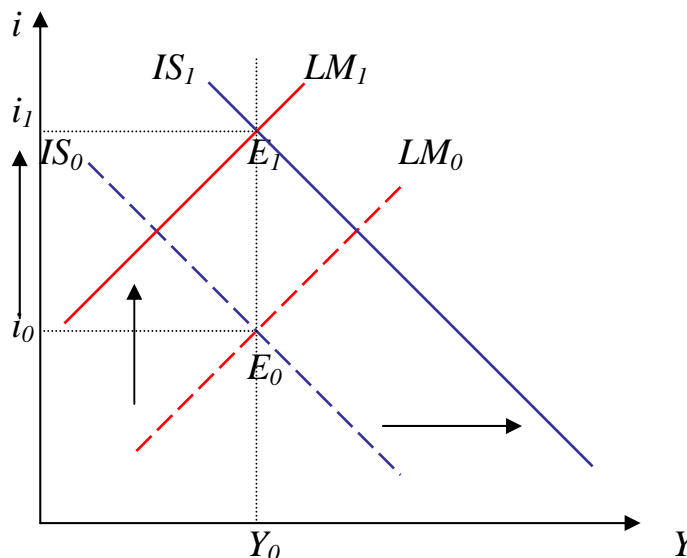
Dans les deux paragraphes précédents, nous avons étudié séparément l’effet des politiques budgétaire et monétaire. Pourtant les deux instruments peuvent être utilisés conjointement. Le *policy-mix* est le dosage des deux politiques. Nous allons voir que ces deux politiques peuvent tout autant se renforcer que se neutraliser. Pour ce faire, nous allons nous concentrer sur le cas standard et n’étudier que deux configurations. Dans la première (A), la politique budgétaire est expansionniste mais la politique monétaire est restrictive. Dans la seconde (B), la politique budgétaire est toujours expansionniste, mais elle est accompagnée d’une politique monétaire expansionniste.

Le policy-mix ne se résume pas à ces deux configurations. Vous pourrez en étudier d’autres à titre d’exercice.

A- Politique budgétaire expansionniste accompagnée d’une politique monétaire restrictive

Nous avons vu qu’une politique budgétaire expansionniste se traduisait par un déplacement vers la droite de la courbe IS alors qu’une politique monétaire restrictive déplaçait la courbe LM vers le haut. Le policy-mix que nous étudions ici est donc représenté par ces deux effets.

Graphique 16 : une politique budgétaire expansionniste accompagnée d’une politique monétaire restrictive



Les notes du cours ne sont transmises qu’à titre indicatif.

Le graphique 16 montre que la politique monétaire accentue l'effet d'éviction. Le taux d'intérêt augmente encore plus et le revenu encore moins qu'en présence d'une politique monétaire neutre. Dans le cas qui est représenté, la politique monétaire contrecarre exactement la politique budgétaire, si bien que le revenu reste constant.

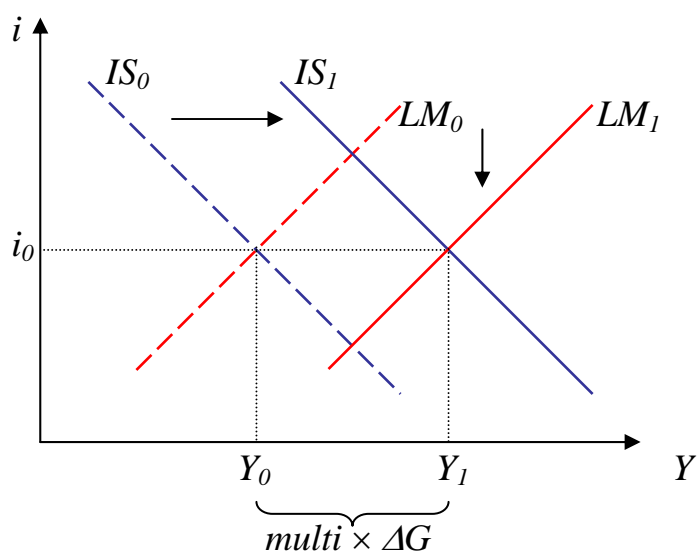
Il s'agit ici d'un cas particulier. Si la politique monétaire était moins restrictive, le revenu pourrait augmenter légèrement. En revanche si la politique monétaire était plus restrictive, le revenu pourrait même diminuer.

Dans tous les cas, un tel policy-mix est incohérent. Il n'a en effet aucun impact sur le revenu mais implique une augmentation inutile des dépenses et du taux d'intérêt. La situation initiale serait donc préférable puisque les dépenses seraient inférieures, ce qui permettrait de ne pas augmenter la dette publique sans que le revenu soit différent.

B- Politique budgétaire expansionniste accompagnée d'une politique monétaire expansionniste

Comme précédemment, la politique budgétaire est expansionniste et déplace donc la courbe IS vers la droite. En revanche, la politique monétaire est à présent elle aussi expansionniste. Par conséquent, elle déplace la courbe LM vers le bas.

Graphique 17 : une politique budgétaire expansionniste accompagnée d'une politique monétaire expansionniste



Le graphique 17 montre que dans ces conditions que ces deux effets se complètent et résultent en une augmentation du revenu, accompagnée d'une hausse limitée du taux d'intérêt.

Les notes du cours ne sont transmises qu'à titre indicatif.

Le graphique montre que la politique monétaire stabilise exactement le taux d'intérêt. L'effet d'éviction est ainsi évité et l'augmentation du revenu correspond exactement à l'effet multiplicateur.

Comme précédemment, il s'agit là en fait d'un cas extrême. Plus généralement on peut dire que lorsque la politique monétaire est expansionniste, elle augmente l'efficacité de la politique budgétaire. L'augmentation du revenu est supérieure, et celle du taux d'intérêt inférieure, à celle qui aurait été constatée si la politique monétaire était restée passive.

Si on avait inversé les rôles et étudié une politique monétaire expansionniste confrontée à une politique budgétaire soit expansionniste, soit restrictive, on aurait obtenu des résultats comparables.

L'étude du policy-mix permet donc de conclure que **l'efficacité de la politique budgétaire dépend de la politique monétaire mise en œuvre et vice versa.**

Ce résultat est important car il souligne l'importance de la coordination des politiques budgétaire et monétaire. Il amène donc à réfléchir aux institutions qui permettent de faciliter cette coordination. On doit se demander s'il faut confier les deux instruments de politique macroéconomique à la même institution (gouvernement, agence indépendante etc.) ou les séparer.

Cette question se pose à l'heure actuelle dans l'Union européenne. La Banque centrale européenne est indépendante du pouvoir politique et gérée de façon fédérale alors que les politiques budgétaires restent largement aux mains des gouvernements nationaux. Certains considèrent que la coordination entre la BCE et les gouvernements nationaux devrait être renforcée alors que d'autres pensent que l'absence de contacts soutenus est le gage de l'indépendance de la banque.

Conclusion

Le diagramme IS-LM décrit une économie où les ajustements se font par les quantités, parce que les prix sont rigides et qu'il existe un sous-emploi des ressources. On constate alors une interdépendance des sphères réelle et monétaire.

Dans ce cas, les politiques monétaire et budgétaire peuvent en principe contribuer à amortir les fluctuations conjoncturelles. Cependant, leur efficacité dépendra d'un certain nombre de paramètres qui affectent la forme des courbes IS et LM. La question de

l'instrument à utiliser devient alors empirique : il s'agit en particulier de mesurer notamment le multiplicateur des dépenses, la sensibilité de l'investissement au taux d'intérêt, ou encore la sensibilité de la demande de monnaie au taux d'intérêt.

Cependant, le modèle IS-LM, même s'il permet d'obtenir certains résultats fondamentaux, reste très incomplet. D'abord, son application est limitée aux situations de sous emploi. De plus, il ne permet pas d'étudier l'impact des politiques économiques sur l'inflation, ne serait-ce que parce qu'il repose sur l'hypothèse de fixité des prix, ce qui restreint son application au court terme.

Pour pouvoir dépasser ces limites, il est nécessaire de tenir compte des capacités de production, et de l'évolution des coûts. Il faut donc introduire l'offre dans notre raisonnement. C'est ce que le prochain chapitre permet de faire.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2004 : chapitre 6.

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 10.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 10 et chapitre 11 (section 11.1 uniquement).

Pour les passionnés

Benassy-Quéré A. et B. Cœuré : *Economie de l'euro*, Repères, La Découverte, Paris, 2002 : chapitres 2, 3 et 4, pour une description du policy-mix dans la zone euro.

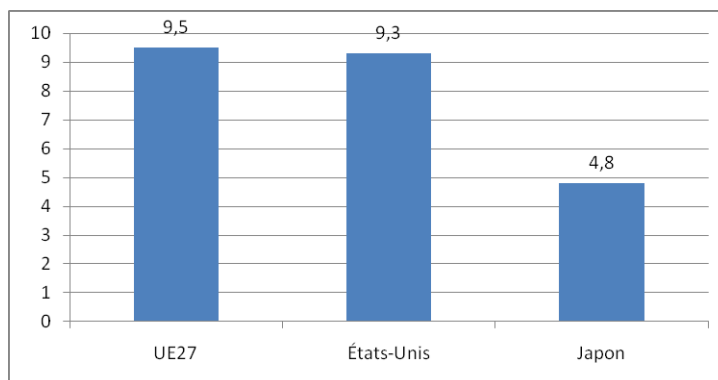
Chapitre 7 : Le marché du travail

Introduction : Les différences internationales de taux de chômage

Il y a au moins deux raisons de s'intéresser au marché du travail. La première est qu'on ne peut concevoir de réfléchir à l'évolution des prix et de la production à court et moyen terme sans tenir compte des capacités de production de l'économie. Comme le stock de capital met du temps à s'ajuster aux variations de la demande, c'est l'emploi qui est la variable d'ajustement de l'offre à court terme.

La deuxième raison tient à l'importance qu'a prise la question du chômage dans les économies occidentales en général et européennes en particulier. Après trois décennies de plein emploi ou de quasi-plein emploi, la plupart des économies européennes ont vu leur taux de chômage augmenter. Aujourd'hui, certaines ont retrouvé des niveaux de chômage très faibles alors que d'autres peinent à réduire le leur. C'est ce que montre le graphique ci-dessous :

Graphique 1 : taux de chômage dans l'UE, les Etats-Unis et le Japon en 2010



Source : Eurostat.

Le graphique montre qu'il existe de très fortes disparités entre les taux de chômage des pays européens, des Etats-Unis et du Japon. Il est globalement plus élevé dans l'Union européenne qu'aux Etats-Unis et au Japon. De plus, de fortes disparités existent au sein de l'Union. Le taux de chômage moyen y est en effet plus élevé parmi les douze nouveaux membres que parmi les quinze plus anciens.⁷⁴

⁷⁴ L'agrégation par groupe est cependant trompeuse. Un aperçu détaillé du taux de chômage des vingt-cinq pays membres est reproduit en annexe. Les taux de chômage des pays membres en 2010 s'échelonnaient de 4,4 % aux Pays-Bas à 20,1 % en Espagne.

Compte tenu du coût social et humain du chômage, ces différences ne peuvent laisser insensible. L'un des objets de ce chapitre est de tenter de les expliquer. Il s'agit en partie de comprendre ce qui différencie l'Espagne où le taux de chômage dépasse vingt pourcents (20,6 %) et l'Allemagne, où le chômage est inférieur à sept pourcents (6,7 %) en 2010. Cependant, les explications que nous passerons en revue resteront très sommaires. Il s'agit ici de faire le strict minimum nécessaire à la construction d'un modèle macroéconomique. Les théories que nous allons étudier manquent donc probablement de subtilité, mais vous aurez l'occasion de davantage entrer dans le détail dans la suite de vos études.

L'une des simplifications les plus criantes consiste dans le simple fait de parler « du » marché du travail. Il s'agit bien là d'une hypothèse simplificatrice mais réductrice. Le travail est en effet un facteur hétérogène. Les travailleurs qualifiés et non qualifiés ne sont pas substituables. Au sein même de la catégorie des travailleurs qualifiés il existe des différences de qualification. Un médecin est fondamentalement différent d'un ingénieur agronome.

Il s'agit là en fait du même type de simplification que pour le marché des titres ou des biens. Nous allons supposer que le marché du travail est homogène parce que cela est suffisant pour traiter les questions qui vont nous occuper dans le cadre du cours. Si nous nous intéressions à des questions plus spécifiques, comme les écarts de salaire entre travailleurs qualifiés et non qualifiés ou la discrimination entre hommes et femmes etc., nous devrions évidemment utiliser des hypothèses plus réalistes.

Pour l'heure, nous allons nous contenter de la chimère d'un marché du travail homogène. Ce cadre permet déjà d'obtenir des intuitions importantes sur les causes possibles du chômage. Pour cela il faut comprendre les spécificités du marché du travail qui aboutissent à l'existence du chômage (section 2). Cela suppose d'avoir au préalable décrit comment fonctionnerait le marché du travail s'il était concurrentiel (section 1).

Section 1 : Description d'un marché du travail concurrentiel

Lorsqu'on étudie le marché du travail, on considère le travail comme un bien. Son prix est déterminé de façon à égaliser les quantités offertes et demandées. Les agents qui offrent le travail sont les travailleurs et ce sont les entreprises qui le demandent.

Une fois qu'on a déterminé l'offre (I) et la demande (II) de travail, on peut définir l'équilibre du marché et commencer à réfléchir aux causes du chômage (III).

I- L'offre de travail

L'offre de travail émane des travailleurs. Ce sont donc des agents particuliers qui offrent de leur temps en contrepartie d'un salaire. Comme le temps dont dispose chaque travailleur est limité, l'offre d'une quantité supplémentaire de travail ne peut se faire qu'en réduisant le temps consacré au loisir.⁷⁵

Il existe donc un arbitrage entre le revenu du travail et les loisirs.⁷⁶ Comme le salaire sert à consommer, la décision d'offrir du travail repose donc sur un arbitrage entre consommation et loisir. On peut donc utiliser les outils de la microéconomie standard pour déterminer l'offre de travail d'un consommateur. Nous pouvons donc supposer qu'il cherche à maximiser une fonction d'utilité dont les deux arguments sont la consommation et le temps de loisir.

La contrainte est constituée par le fait qu'une journée ne fait que vingt-quatre heures. Ces vingt-quatre heures doivent être réparties entre travail et loisir. Chaque heure consacrée au loisir doit être retranchée du temps de travail. Elle diminue donc le revenu du travail d'un montant équivalent au salaire horaire. Le taux de salaire horaire mesure donc le coût d'opportunité du loisir.

On distingue le *salaire nominal* et le *salaire réel*. Le salaire nominal (W) est mesuré en unités monétaires. Il se lit directement dans le contrat de travail. Le salaire réel mesure la quantité de biens que peut acheter le salaire versé pour une heure de travail. Il s'obtient en divisant le salaire nominal par le niveau des prix (W/P).

Comme on considère que les agents ne sont pas victimes d'illusion monétaire, c'est le salaire réel qui va déterminer leur comportement. En effet, un travailleur ne s'intéresse pas vraiment au nombre d'euros qu'il gagne mais à ce que ces euros lui permettent d'acheter. C'est donc son salaire réel qu'il va prendre en compte pour déterminer son arbitrage entre travail et loisir, même s'il négocie son salaire nominal avec son employeur.

Déterminer la fonction d'offre de travail du travailleur, consiste à déterminer comment il va modifier le temps qu'il consacre au travail lorsque son taux de salaire réel varie.

Supposons que le taux de salaire réel augmente. Cela signifie que le travailleur va pouvoir consommer davantage pour un temps de travail donné. Cela implique également que

⁷⁵ Par loisir, il faut entendre tout ce qui ne relève pas du travail rémunéré : passe-temps, repos, formation, production domestique etc.

⁷⁶ L'utilité marginale est l'opposé de la désutilité du travail. Supposer que le travail a une utilité marginale négative n'a rien d'étonnant lorsqu'on sait que le mot travail viendrait du latin « tripalium », qui signifie triple pal et désigne un instrument de torture constitué de trois pieux. Pour compenser l'épreuve du travail, on perçoit heureusement un salaire. Ce mot viendrait lui du mot sel, parce que les travailleurs étaient originellement rémunérés en recevant du sel.

le coût d'opportunité du loisir augmente. On peut dire qu'une augmentation du salaire réel se traduit par une augmentation du prix relatif du loisir par rapport à la consommation. Comme lors de l'arbitrage entre deux biens, deux effets vont entrer en jeu :

- un effet de revenu : puisque le salaire augmente, la valeur du temps que le travailleur peut affecter au travail augmente. Il est en quelque sorte « plus riche ». Si le loisir est un bien supérieur, le travailleur peut alors être incité à en prendre davantage et réduire ainsi son offre de travail.

- un effet de substitution : comme l'augmentation du salaire implique une augmentation du prix relatif du loisir par rapport à la consommation, le travailleur est incité à substituer de la consommation au loisir, ce qui augmente son offre de travail.

La relation entre le salaire réel et l'offre de travail est donc ambiguë. L'offre de travail peut soit augmenter soit diminuer lorsque le taux de salaire réel augmente. La solution est donc empirique. Elle dépend de plus de l'horizon considéré.

La tendance séculaire à la diminution du temps de travail suggère que l'effet de revenu l'emporte sur l'effet de substitution à long terme.⁷⁷ On considère en revanche que c'est l'effet de substitution qui l'emporte à court terme, parce que les travailleurs vont être disposés à travailler davantage lorsqu'ils considèrent que l'augmentation du salaire qui leur est proposée est temporaire. De plus, les études statistiques montrent que l'élasticité de l'offre de travail au salaire est positive.⁷⁸

Comme nous nous intéressons ici au court terme, nous supposons donc que la courbe d'offre de travail est croissante.

On peut trouver curieux de supposer que le travailleur détermine librement le temps qu'il souhaite consacrer à son activité. En réalité, les emplois proposés ne laissent qu'exceptionnellement la possibilité au travailleur de déterminer son temps de travail.

Il faut cependant garder à l'esprit que nous raisonnons à l'échelle de l'économie dans son ensemble. L'offre de travail est donc constituée de la somme des offres de travail individuelles. C'est donc plus l'évolution du nombre de travailleurs que celle du nombre d'heures de travail offertes par chaque travailleur qui explique la réaction de l'offre de travail aux variations du salaire réel.

⁷⁷ Le tableau A1 reproduit en annexe montre qu'on travaillait en gros deux fois moins en 1987 qu'en 1870 dans les pays industrialisés. Sur la même période, le revenu par habitant a été multiplié par sept en moyenne dans les pays occidentaux.

⁷⁸ Une courte synthèse des grandes questions posées par l'estimation de l'élasticité de l'offre de travail est proposée par James J. Heckman "What have we learned about labor supply in the past twenty years?", *American Economic Review*, vol.83 n°2, p.116-121, mai 1993.

Lorsque le taux de salaire diminue, certains travailleurs peuvent décider de quitter le marché du travail parce qu'ils ne considèrent pas les emplois disponibles comme suffisamment rémunérateurs.

A l'inverse, l'augmentation du taux de salaire réel peut inciter certains travailleurs à revenir sur le marché du travail.

La courbe d'offre de travail est donc globalement croissante, même si chaque travailleur n'est confronté qu'à un choix binaire qui consiste à accepter ou refuser les emplois qu'on lui propose.

II- La demande de travail

La demande de travail est exprimée par les entreprises. Elles utilisent le travail comme facteur de production. Si on suppose que la quantité de capital est fixée à court terme, la production ne devient plus qu'une fonction de la quantité de travail utilisée :

$$Y = F(L) \quad (1)$$

L'entreprise va alors choisir la quantité de travail utilisée de façon à maximiser son profit. Ce profit est une fonction de la quantité de travail utilisée, du taux de salaire et du prix du bien :

$$\pi = P.F(L) - W.L \quad (2)$$

La quantité de travail qui maximise le profit est celle qui annule la dérivée du profit, soit :

$$\frac{\partial \pi}{\partial L} = P.Pm_L - W = 0 \quad (3)$$

On peut alors déduire une expression du salaire réel d'équilibre :

$$\boxed{\frac{W}{P} = Pm_L} \quad (4)$$

Ainsi, la maximisation du profit implique que le taux de salaire réel soit égal à la productivité marginale du travail. Ce résultat permet de déterminer la forme de la demande de travail dès lors qu'on connaît l'évolution de la productivité marginale du travail.

Comme on suppose que la productivité marginale du travail est décroissante, la demande de travail sera une fonction décroissante du salaire réel. En effet, si le taux de salaire réel est inférieur à la productivité marginale du travail, les entreprises peuvent augmenter leurs profits en embauchant des travailleurs. En effet, chaque heure de travail leur rapporte

plus qu'elle ne leur coûte. La quantité de travail utilisée augmentera donc. La productivité marginale du travail diminuera alors, et ce jusqu'à ce que l'égalité soit retrouvée.

A l'inverse, si le taux de salaire réel est supérieur à la productivité marginale du travail, les entreprises seront incitées à réduire la quantité de travail qu'elles utilisent, puisque chaque heure de travail leur coûte plus cher qu'elle ne leur rapporte.

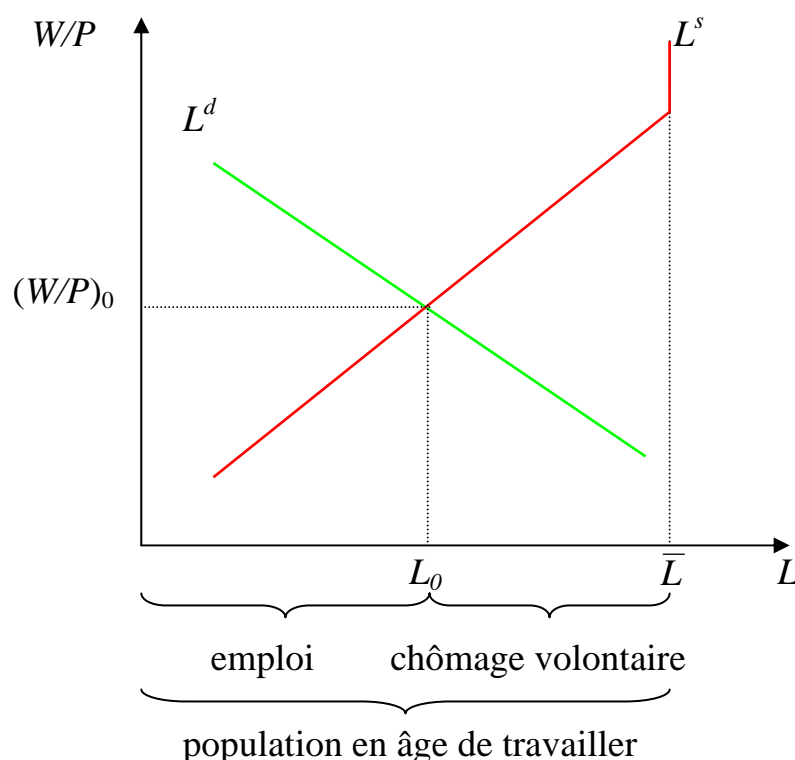
Par conséquent, le fait que la productivité marginale du travail soit décroissante implique que la demande de travail soit une fonction décroissante du taux de salaire réel.

III- L'équilibre du marché du travail

Nous avons à présent défini l'offre et la demande de travail. Si le marché est organisé de façon concurrentielle, nous pouvons alors déterminer le taux de salaire réel d'équilibre. C'est celui qui permet d'assurer l'égalité entre la quantité offerte et la quantité demandée de travail. Si on considère que l'offre et la demande de travail ont des formes standards, le graphique ci-dessous représente l'équilibre sur le marché du travail.

La courbe de demande de travail est décroissante. La courbe d'offre de travail est croissante. Au-delà d'une certaine quantité \bar{L} , elle devient verticale. Cette forme particulière traduit le fait que l'offre de travail butte sur une limite supérieure lorsque toute la population en âge de travailler est employée.

Graphique 2 : L'équilibre sur un marché du travail concurrentiel



Le graphique suggère que si le marché du travail fonctionnait de façon concurrentielle et que le taux de salaire était fixé de façon à égaliser l'offre et la demande de travail, la population en âge de travailler se répartirait entre travailleurs et chômeurs volontaires.

Le chômage volontaire inclut tous les actifs qui ne travaillent pas parce qu'ils préfèrent ne pas exercer une activité rémunérée au taux de salaire en vigueur. En effet, si le taux de salaire augmentait, certains chômeurs volontaires pourraient alors souhaiter travailler.

L'expression de chômage volontaire peut sembler choquante et elle l'est en partie, mais elle perd de sa charge agressive lorsqu'elle est définie correctement. En effet, elle ne signifie pas que les personnes qui n'ont pas d'emploi sur un marché concurrentiel le font de gaieté de cœur ou par fainéantise.

Un exemple typique est celui d'une mère de famille qui a des enfants en bas âge. Pour elle, travailler impliquerait de faire appel à une garderie ou d'employer une nourrice. Si le taux de salaire qu'elle peut obtenir en travaillant est trop faible, elle pourrait perdre de l'argent en travaillant. Elle a donc intérêt à ne pas travailler.

On peut se demander si la seule forme de chômage possible sur un marché du travail concurrentiel est le chômage volontaire. La réponse est non. Il existe deux raisons fondamentales pour lesquelles il subsiste du chômage involontaire même sur un marché du travail en équilibre concurrentiel.

La première raison tient à l'inadéquation entre les caractéristiques des chômeurs et celles que recherchent les employeurs. Si la sidérurgie débauche mais que le secteur informatique cherche à embaucher, il est peu probable que les qualifications des anciens sidérurgistes correspondent à celles que recherchent les entreprises du secteur informatique. Les anciens sidérurgistes resteront donc au chômage, alors qu'il existe une demande insatisfaite d'informaticiens. La résorption de ce type de chômage passe donc par l'acquisition par les chômeurs de qualifications nouvelles. Le même raisonnement peut s'appliquer lorsque l'inadéquation est géographique et que les chômeurs sont éloignés des emplois.

La deuxième source de chômage à l'équilibre est appelée *chômage frictionnel*. Elle correspond à la période de chômage temporaire que traversent les travailleurs qui ont quitté un emploi avant d'en trouver un autre. L'existence du chômage frictionnel est principalement due à la lenteur avec laquelle l'information circule. Même si une entreprise cherche à embaucher un employé dont les qualifications correspondent exactement à celles d'un travailleur qui vient de quitter son emploi précédent, il faut que l'entreprise apprenne l'existence du travailleur et vice versa. On dit que le travailleur et l'entreprise doivent s'apparier. Le travailleur restera donc au chômage le temps de réaliser son *appariement* avec l'entreprise qui cherche à pourvoir un poste correspondant à ses qualifications.

Il existera donc toujours un stock de chômage frictionnel dont la taille dépendra des flux de nouveaux chômeurs et de chômeurs retrouvant un emploi. Ce stock aura tendance à augmenter pendant les périodes d'augmentation du nombre de licenciements et/ou de ralentissement des reprises d'emploi. Une amélioration du système de recherche d'emploi permettra également de le réduire. Les organismes de placement, les agences pour l'emploi et les cabinets de recrutement trouvent entre autres la justification de leur existence dans le fait qu'ils permettent de réduire les coûts d'information et de recherche d'emploi.

Les deux types de chômage à l'équilibre reposent sur des mécanismes différents mais impliquent tous les deux qu'il existe simultanément des chômeurs et des emplois vacants, puisque la quantité de travail offerte est égale à la quantité de travail demandée. Ces deux types de chômage n'apparaissent donc pas explicitement dans le graphique qui représente

l'équilibre d'un marché du travail concurrentiel.⁷⁹ On les désigne par l'expression paradoxale de *chômage de plein emploi*.

Section 2 : La possibilité du sous-emploi

En plus du chômage volontaire et des deux formes de chômage de plein emploi, il existe des formes involontaires de chômage qui ne peuvent être dues qu'à une offre excédentaire de travail. En effet, il arrive que le nombre de chômeurs soit supérieur au nombre d'emplois vacants. De plus, le taux de chômage moyen peut atteindre des niveaux trop élevés pour être compatibles avec le plein emploi. Il doit donc exister un chômage involontaire (II).

S'il le chômage involontaire est possible, c'est qu'il existe des mécanismes qui empêchent d'atteindre l'équilibre sur le marché du travail. Il faut donc d'abord comprendre en quoi consistent ces mécanismes, qui distinguent le marché du travail des autres marchés (I).

I- Les rigidités spécifiques au marché du travail

Lorsqu'on évoque les sources de rigidités sur le marché du travail, on cite souvent en priorité des facteurs législatifs imposés de façon exogène par l'Etat (1). Cependant, une telle vision est réductrice. En effet, la spécificité de l'organisation du marché du travail peut inciter à la fois les offreurs (2) et les demandeurs (3) à créer des rigidités à la baisse du salaire réel.

A- Les rigidités exogènes au marché du travail

La législation sociale est une source directe ou indirecte de rigidité à la baisse du salaire réel. Par définition, l'imposition d'un *salaire minimum* empêche le salaire de baisser en dessous d'une certaine limite. Une telle mesure augmente la rémunération des travailleurs les moins bien rémunérés, dont la productivité marginale est aussi souvent la plus faible.

⁷⁹ En revanche, on peut penser qu'il existe une relation décroissante entre le nombre d'emplois vacants et le taux de chômage, puisque la durée du chômage frictionnel devrait se réduire lorsqu'il y a plus d'emplois proposés donc plus de chances pour les chômeurs d'en trouver un qui leur corresponde. Cette relation est traduite par la *courbe de Beveridge* du nom de William Henry Beveridge, l'économiste et homme politique britannique à qui le Royaume-Uni doit son système de protection sociale. C'est à lui que l'on doit la définition du plein emploi comme une situation où le nombre d'emplois vacants est égal au nombre de chômeurs, dans son livre *Full Employment in a Free Society*, George Allen and Unwin, Londres, 1944.

La contrepartie est que les entreprises peuvent être incitées à moins embaucher ces travailleurs si le salaire minimum dépasse leur productivité marginale. Un chômage apparaît ainsi dans le segment des travailleurs à bas salaire.

Ce raisonnement ne s'applique que si on considère que les employeurs ne disposent d'aucun pouvoir de marché. Supposons au contraire que le marché du travail soit en situation de monopsonne ou d'oligopsonne. Dans ce cas, les employeurs peuvent utiliser leur pouvoir de marché en réduisant leur demande de travail afin de peser sur les salaires. Ils maximisent ainsi leur profit en créant du chômage. Dans une telle situation, imposer un salaire minimum peut augmenter l'emploi parce qu'il devient optimal pour les entreprises d'augmenter leur demande de travail.

Une telle possibilité est très discutée, mais il existe des observations troublantes. Dans les années 90, le salaire minimum a augmenté aux Etats-Unis, en 1990 et 1991, après presque une décennie de stagnation. Deux économistes américains ont étudié l'impact de ces deux mesures sur l'emploi dans l'industrie de la restauration rapide au Texas.⁸⁰ Contre toute attente, de nombreux restaurants ont augmenté leurs effectifs suite à cette mesure. Les restaurants qui ont été les plus touchés, parce que leurs salaires initiaux étaient les plus faibles, sont même ceux qui ont le plus augmenté leurs effectifs. Cette observation peut s'expliquer par le pouvoir de monopsonne des employeurs sur le marché de la restauration rapide, qui sont confrontés à une main-d'œuvre peu qualifiée qui dispose de peu d'alternatives en dehors de ces emplois.

La législation peut provoquer des rigidités de façon plus indirecte. En effet, la possibilité de bénéficier d'un revenu de remplacement sous forme d'assurance chômage peut inciter les travailleurs à augmenter leur salaire de réservation. Le risque est alors celui de l'apparition de situations de *trappe à inactivité*.

Ces situations apparaissent lorsqu'un chômeur perd ses revenus de transfert lorsqu'il retrouve un emploi (allocation chômage, aide au logement, aides sociales etc.). Son gain net en termes de revenu peut alors s'avérer très faible voire négatif. Dans ce cas, le chômeur n'est pas incité à reprendre un emploi.

L'existence de trappes à inactivité semble corroborée par le fait que l'augmentation du revenu des chômeurs qui retrouvent un emploi est largement inférieure à leur salaire dans les pays industrialisés.⁸¹ Cela ne signifie pas qu'il faille supprimer les prestations sociales aux

⁸⁰ L. Katz et A. Krueger dans « The effect of minimum wage on the fast-food industry », *Industrial and Labor Relations Review*, vol 46, p. 6-21, octobre 1992.

⁸¹ Comme la perte des transferts sociaux peut être interprétée comme une forme d'imposition, l'effet des transferts sur l'incitation à reprendre un emploi se mesure par le taux marginal effectif d'imposition, c'est-à-dire

chômeurs mais qu'il est nécessaire de réfléchir à des systèmes de protection sociale dont les effets désincitatifs soient aussi limités que possible (ex : crédit d'impôt, prestations dégressives plutôt que forfaitaires etc.).

B- Rigidités créées du côté de l'offre de travail

Si les travailleurs restaient isolés, l'offre de travail serait atomistique et les travailleurs de simples preneurs de prix sur le marché du travail. Certes, certains travailleurs peuvent négocier directement leur salaire mais ils sont relativement rares. Ce sont les travailleurs les plus qualifiés et dont les qualifications sont les plus recherchées qui peuvent se permettre de négocier leur contrat de travail, y compris leur salaire. C'est le cas par exemple des chefs d'entreprise reconnus ou encore des sportifs de haut niveau. Une telle possibilité ne s'offre cependant pas aux employés non qualifiés qu'on peut trouver par exemple dans la restauration rapide.

Si les travailleurs sont individuellement preneurs de prix, ils peuvent agir collectivement grâce aux organisations syndicales. Les syndicats en organisant les revendications de milliers de travailleurs peuvent obtenir un pouvoir de négociation important et revendiquer des salaires plus élevés que sur un marché concurrentiel. Certains y voient donc une source importante de rigidité des salaires.

Une raison supplémentaire pour laquelle l'existence de syndicats peut se traduire par un salaire supérieur au salaire de plein-emploi et provoquer du chômage tient au fait que le syndicat ne défend en principe que les intérêts des travailleurs et pas ceux des chômeurs. Comme les travailleurs qui ont un emploi, ont intérêt à obtenir des salaires élevés, les revendications des syndicats tendront à faire augmenter les salaires. Si ces revendications aboutissent, elles réduiront les chances des chômeurs d'être embauchés. Il apparaît donc un conflit entre les *insiders* et les *outsiders*.

L'explication du chômage par le conflit insiders-outsiders ne nécessite pas l'existence de syndicats. Il suffit en effet que les entreprises subissent un coût en cas de licenciement des

la part de l'augmentation du revenu qui est prélevée par l'Etat, soit par impôt soit en supprimant des transferts. On a calculé qu'en France le taux marginal effectif pouvait atteindre 90 à 100% pour un chômeur quittant les minima sociaux pour retrouver un emploi à bas salaire. T. Piketty *Economie des inégalités*, Repères, La Découverte, Paris, 2001.

insiders. Ces derniers disposent alors d'un pouvoir de négociation qui leur permet de revendiquer des augmentations salariales.⁸²

Il faut de plus se garder de penser que l'existence de syndicats puissants implique nécessairement un taux de chômage élevé. La relation n'est en fait pas du tout claire empiriquement. Par exemple, la Suède est un des pays où le taux de syndicalisation est le plus élevé (80,1% en 2004, 1^e trimestre). Pourtant, le taux de chômage en Suède reste l'un des plus bas de l'OCDE (6,3% en 2004).

C- Rigidités créées du côté de la demande de travail

A priori, il semble irrationnel pour le demandeur d'un bien de créer des rigidités à la baisse du prix de ce bien.⁸³ Ce comportement paradoxal pourrait cependant être à l'origine de la rigidité des salaires. Il s'expliquerait par la spécificité de la relation entre un employeur et ses employés et serait tout à fait rationnel.

La première spécificité du marché du travail est qu'il repose la plupart du temps sur des relations de long terme entre le travailleur et son employeur. Un contrat de travail n'est en effet pas révisé quotidiennement. Il stipule en général la mission du travailleur, en des termes plus ou moins précis, et sa rémunération.

La conclusion d'un tel contrat peut s'interpréter comme un *contrat implicite* d'assurance. En effet, le salarié sait que sa rémunération est indépendante des aléas économiques. C'est son employeur qui supporte le risque économique. En contrepartie, le travailleur accepterait une rémunération moyenne légèrement plus faible que s'il acceptait de voir son salaire fluctuer.

La conséquence de cette assurance implicite est que le salaire sera certes inférieur à ce qu'il aurait pu être en période de conjoncture favorable mais qu'il sera supérieur à son niveau d'équilibre lorsque la conjoncture sera défavorable. Cette *poignée de main invisible* crée donc une rigidité du salaire à la baisse.⁸⁴

⁸² Cette théorie trouve ses origines dans les travaux de Assar Lindbeck et Dennis J. Snower qui sont présentés de façon synthétique dans leur article "Wage Setting, Unemployment, and Insider-Outsider Relations" *American Economic Review*, vol 76, n°2, pp. 235-239, May, 1986.

⁸³ Par exemple, on imagine mal les automobilistes et les chauffeurs routiers revendiquer un prix minimum du carburant.

⁸⁴ L'assurance est cependant imparfaite. Lorsque la situation de l'entreprise se dégrade trop, les travailleurs peuvent être licenciés. Ils supportent alors pleinement le risque économique.

Une explication plus récente de la rigidité à la baisse du salaire réel repose sur la *théorie du salaire d'efficience*.⁸⁵ Cette théorie regroupe quatre arguments principaux qui impliquent que les entreprises maximisent leurs profits en proposant des salaires plus élevés que leur niveau de plein emploi parce que cela augmente la productivité moyenne de leurs employés.

L'argument le plus intuitif est principalement applicable dans les pays où les salaires sont si bas qu'ils déterminent l'état de santé des travailleurs. Un employeur qui verse des salaires plus élevés sait alors que ses travailleurs seront plus productifs parce qu'ils seront en meilleure santé.

Le deuxième argument repose sur l'idée qu'il est coûteux pour une entreprise de former ses nouveaux salariés. Elle a donc intérêt à inciter ses employés à rester. La meilleure façon de le faire consiste à leur verser des salaires supérieurs à ceux qu'ils trouveraient sur le marché.

Le troisième argument suppose que l'offre de travail n'est pas homogène. On y trouve des travailleurs plus ou moins sérieux, plus ou moins efficaces etc. Cependant un employeur est incapable de les distinguer au moment de l'embauche. Il sait cependant que les travailleurs les plus productifs seront incités à aller voir ailleurs afin d'obtenir une meilleure rémunération. Si l'employeur se contente de verser le salaire d'équilibre du marché, il sait par conséquent qu'il ne gardera à terme que les travailleurs les moins efficaces. Ce phénomène de *sélection adverse*, ou *antisélection*, se produit en raison de l'asymétrie d'information entre l'employeur et l'employé. Afin d'éviter le départ de ses meilleurs employés, l'entreprise a intérêt à verser des salaires supérieurs au niveau d'équilibre du marché.

Le dernier argument est aussi une conséquence de l'asymétrie d'information entre les employeurs et les employés mais porte cette fois sur l'effort fourni par ces derniers. L'employeur ne peut en effet pas contrôler en permanence que les travailleurs fournissent l'effort qu'il attend d'eux. Il se pose donc un problème d'*aléa moral*. L'employeur peut toutefois organiser des contrôles et licencier les employés surpris en train de « tirer au flanc ». Cependant la sanction est peu dissuasive si l'employé licencié sait qu'il va retrouver un emploi proposant le même salaire que celui qu'il vient de perdre. L'entreprise va donc avoir intérêt à verser des salaires supérieurs à ceux du marché afin de garantir le caractère dissuasif

⁸⁵ Il est difficile d'associer un seul auteur à ces théories. On trouvera en revanche une synthèse très lisible sur les principales contributions à l'origine des théories des salaires d'efficience dans l'article de Janet L. Yellen "Efficiency wage models of unemployment", *American Economic Review*, vol 74 n°2, p. 200-205, mai 1984.

du licenciement et s'assurer de la motivation de ses salariés. Ce *modèle du tire-au-flanc* fournit ainsi une dernière explication de la rigidité à la baisse des salaires.

Quelle que soit la logique des salaires d'efficience, ils impliquent tous que les salaires versés par les entreprises vont être supérieurs à ceux qui permettraient d'assurer le plein emploi. Les employeurs ne cherchent pas à créer du chômage mais à augmenter la productivité de leurs travailleurs. Le chômage apparaît alors comme un effet pervers de cette stratégie.

Cette sous-section suggère qu'il existe une batterie des raisons pour lesquelles les salaires nominaux peuvent être rigides à la baisse. Ces explications ne sont pas forcément incompatibles. Il se peut que l'une soit pertinente dans un secteur et une autre dans un autre. De plus, les institutions du marché du travail peuvent être très différentes d'un pays à l'autre. Par conséquent, lorsqu'on cherche à expliquer les différences de situation entre pays, il convient d'éviter les explications mono-causales. Les spécialistes du marché du travail insistent sur l'importance des *complémentarités institutionnelles*, pour expliquer le fonctionnement du marché du travail.

Quoi qu'il en soit, les explications que nous avons vues aboutissent à un même résultat : la rigidité à la baisse du salaire réel. Dans la section suivante, nous allons voir quelles peuvent être les conséquences de cette rigidité.

II- Le chômage involontaire

S'il existe des rigidités à la baisse du salaire réel, il peut aussi exister une offre excédentaire de travail. Cette offre excédentaire détermine le chômage involontaire (A). Cependant, le taux de chômage peut fluctuer. Il convient donc de distinguer, au sein du chômage, une composante structurelle et une composante conjoncturelle (B).

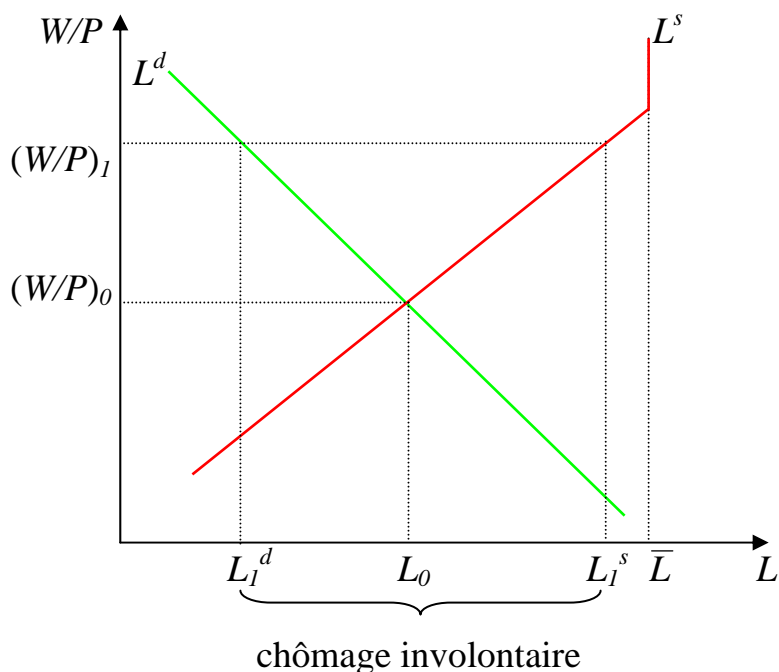
A- Conséquence d'une rigidité à la baisse du salaire réel

Reprenons le graphique qui représente l'offre et la demande de travail. Ce graphique montre qu'il existe une valeur du salaire réel $(W/P)_0$ pour laquelle la quantité offerte de travail correspond exactement à la quantité demandée. Si on fait abstraction du chômage frictionnel et de l'inadéquation entre les qualifications offertes et recherchées, il n'existe donc que du chômage volontaire.

Si on veut expliquer, l'existence d'un chômage involontaire, il faut donc que le taux de salaire réel soit différent de $(W/P)_0$. Plus précisément, il faut que le salaire en vigueur soit supérieur au salaire d'équilibre. On se retrouve alors présence d'une offre excédentaire de travail.

Le graphique 3 montre que si le salaire réel est fixé à un niveau supérieur à celui qui équilibre le marché du travail, il apparaît une offre excédentaire de travail, c'est-à-dire du chômage involontaire ($U_{inv} = L_1^s - L_1^d$). En d'autres termes, certains travailleurs souhaitent offrir leur travail au taux de salaire réel en vigueur mais ne trouvent pas d'emploi. Le chômage total sera donc composé du chômage volontaire et du chômage involontaire.

Graphique 3 : rigidité à la baisse du salaire réel



Ce volume de chômage involontaire s'explique d'une part par la diminution de la demande de travail ($L_1^d < L_0$) et d'autre part par l'augmentation de l'offre de travail ($L_1^s > L_0$). En effet, les entreprises sont incitées à réduire la quantité de travail qu'elles utilisent, afin de faire augmenter la productivité marginale du travail. Par ailleurs, l'augmentation du salaire réel incite certains chômeurs à offrir leur travail.

Si le salaire réel était flexible, il s'adapterait à l'offre excédentaire de travail en diminuant. Le marché convergerait alors vers l'équilibre. Si ce n'est pas le cas, c'est qu'il

existe sur le marché du travail des phénomènes qui maintiennent le taux de salaire réel au-dessus de son niveau d'équilibre.

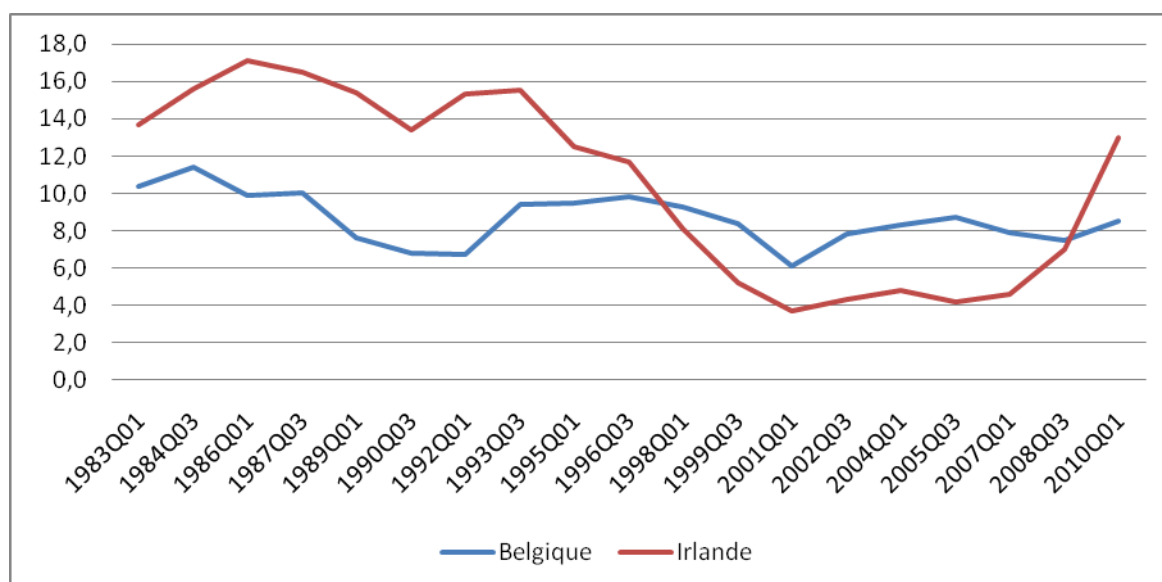
B- La distinction entre chômage structurel et chômage conjoncturel

S'il existe des déterminants structurels du chômage, on observe également des fluctuations du taux de chômage autour de son niveau structurel. C'est ce que montre le graphique ci-dessous qui représente l'évolution sur une vingtaine d'années du taux de chômage de la Belgique et de l'Irlande.

L'évolution du chômage en Irlande semble faire apparaître deux périodes. Au cours de la première, qui correspond en gros aux dix premières années de la période d'observation, le taux de chômage irlandais fluctue autour d'une valeur moyenne comprise entre 14 et 15 %. Il atteint même les 17,2 % au premier trimestre 1986. Le taux de chômage se met alors à baisser sensiblement pour sembler atteindre un nouveau plateau situé autour de 4 %.

Le passage du premier plateau au second illustre bien la possibilité d'une évolution du chômage structurel. La politique de l'Irlande sur le marché du travail et le développement de son économie dans les années 90 ont permis de réduire durablement son chômage structurel.

Graphique 7 : Evolution du taux de chômage
(1983-2010)



Source : Eurostat.

Comparée à l'évolution du chômage en Irlande, celle du chômage belge semble beaucoup plus stable. On ne constate pas de tendance très nette sur la période. Le taux de

chômage de la Belgique a fluctué entre 6 % au deuxième trimestre 1991 et 11,3 % au dernier trimestre 1983. Il semble fluctuer autour d'une moyenne en gros constante d'à peu près 8,5 %.

Ces fluctuations du taux de chômage belge autour d'une valeur apparemment stable illustrent le fait qu'il existe des raisons conjoncturelles aux variations du taux de chômage. On peut aussi remarquer que le taux de chômage irlandais fluctue lui aussi d'un trimestre sur l'autre. C'est le cas autour des deux plateaux que nous avons identifiés.

Les explications de la rigidité des salaires vues dans le I permettent d'expliquer les différences internationales du taux de chômage ou encore le passage d'un palier à l'autre, mais on voit mal comment elles peuvent rendre compte des fluctuations trimestrielles. Par exemple, si on pense que les employeurs ont intérêt à verser des salaires plus élevés que le salaire d'équilibre, par exemple pour fidéliser leurs salariés, on voit mal ce qui pourrait les amener à faire varier l'emploi d'un mois à l'autre.

Il existe une variété d'explications aux fluctuations de l'emploi et du chômage. Leur point commun est de tenir compte des aléas qui vont affecter la demande de travail dans le court terme, et de façon imprévue. Le prix des matières premières fluctue, le moral des consommateurs peut les amener à consommer plus ou moins, l'optimisme des entreprises se traduit par des investissements plus ou moins importants, la politique économique peut connaître des inflexions imprévues etc. On parle de *chocs conjoncturels* pour désigner ces phénomènes imprévisibles. Tous ces chocs peuvent se produire à des intervalles très courts et affecter soit les quantités de biens demandées aux entreprises, soit les prix, ce qui va se répercuter sur le salaire réel si le salaire nominal ne réagit pas assez. In fine, ils vont se traduire par des fluctuations de la demande de travail, donc de l'emploi. On voit ici que l'évolution du marché du travail ne peut être dissociée de celle du marché des biens.

Dans le chapitre suivant, nous proposerons plusieurs autres explications aux fluctuations de court terme de la production et de l'emploi. Toutes aboutiront à la conclusion que les variations non anticipées du niveau des prix peuvent affecter l'emploi et l'activité.

Conclusion

Le marché du travail est un marché particulier sur lequel il existe un taux de chômage structurel positif. Ce taux de chômage est probablement en large partie involontaire. Il inclut une composante frictionnelle et une composante due à l'inadéquation qualitative de l'offre et

de la demande de travail. Ces deux composantes existent même en situation de plein emploi. Elles correspondent alors exactement au nombre d'emplois vacants.

En plus du chômage de plein-emploi, le chômage est aussi dû à une offre excédentaire de travail. Cette offre excédentaire ne peut être résorbée à cause des rigidités qui empêchent le salaire réel de s'ajuster. Le niveau du chômage est alors déterminé par la législation, la nature des relations entre travailleurs et employeurs etc., qui peuvent prendre des formes très variées. Ce sont elles qui détermineront le chômage structurel.

A plus court terme, le taux de chômage peut fluctuer autour de son niveau structurel. Ces fluctuations sont alors dues à des chocs conjoncturels qui vont affecter à la fois le marché du travail et celui des biens.

Comme la production globale dépend de la quantité de travail utilisée par les entreprises, les fluctuations de court terme de l'emploi vont influencer les fluctuations de l'offre globale. C'est ce que nous allons étudier dans le prochain chapitre.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2004 : chapitre 7.

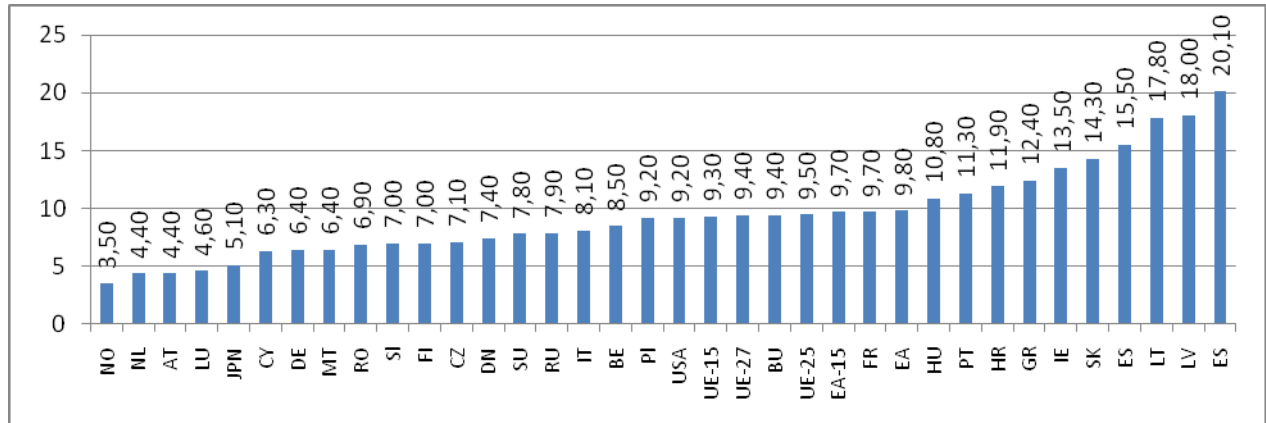
Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 10 et chapitre 6.

Pour les passionnés

Cahuc P. et A. Zylberberg, *Économie du travail : la formation des salaires et les déterminants du chômage*, De Boeck, Bruxelles, 1996.

Annexe

Graphique A1 : Taux de chômage dans l'Union européenne, aux Etats-Unis et au Japon en 2010



Source : Eurostat.

Tableau A1 : Evolution des heures effectivement travaillées par année et par personne (1870-1987)

	France	Royaume Uni	Etats-Unis	Japon
1870	2945	2984	2964	2945
1890	2770	2807	2789	2770
1913	2588	2624	2605	2588
1929	2297	2286	2342	2364
1938	1848	2267	2062	2361
1950	1926	1958	1867	2166
1960	1919	1913	1795	2318
1973	1771	1688	1717	2093
1987	1543	1557	1608	2020

Source : Maddison A. *Dynamic Forces in Capitalist Development: A Long-run Comparative View*, Oxford, Oxford University Press, 1991.

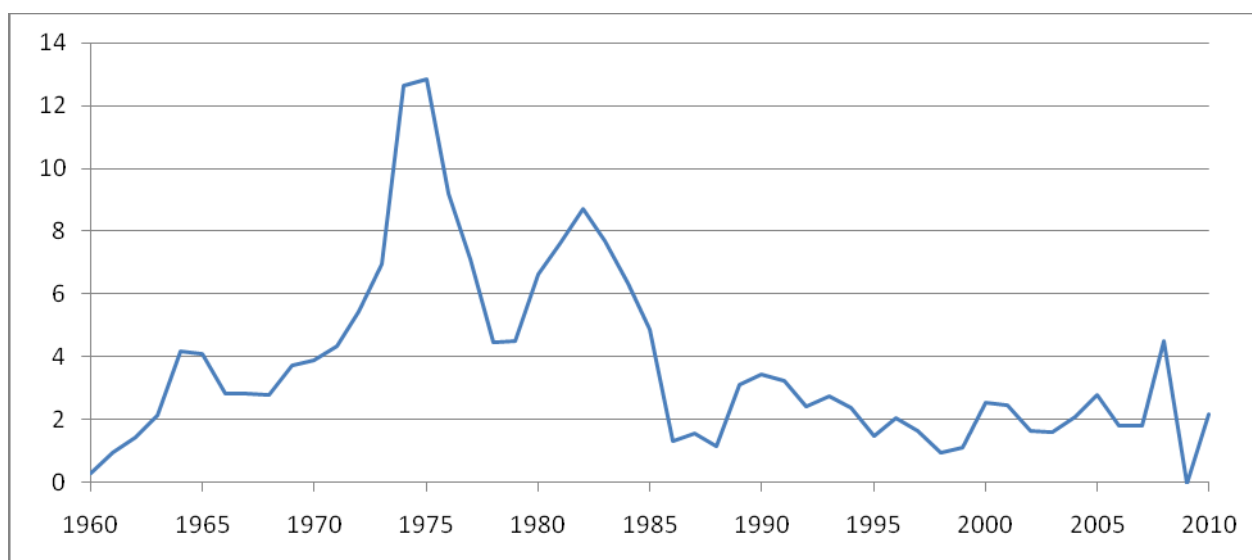
Chapitre 8 : L'offre agrégée et le modèle de la synthèse

Introduction : Pourquoi il faut compléter le modèle IS-LM

Tout comme la croissance, l'inflation connaît de fortes fluctuations. Le graphique 1 montre leur ampleur. Il permet de constater que certaines périodes sont globalement caractérisées par plus d'inflation que d'autres.

La période 1970-1985 a par exemple été marquée par une forte poussée inflationniste culminant à 12,8% en 1975 et 8,7% en 1982. L'inflation semble depuis avoir atteint un plateau situé autour de 2%.

Graphique 1 : L'inflation en Belgique
(1961-2010)



Source : Ministère de l'économie.

Nous avons vu qu'à long terme la théorie quantitative de la monnaie prévoyait que l'inflation devait être proportionnelle à l'augmentation de la masse monétaire. Même si la politique monétaire y joue certainement un rôle, cette théorie n'est pas applicable à court terme. A court terme en effet, certains prix sont rigides et la demande de monnaie peut être affectée par une foule de facteurs qui disparaissent à long terme.

Par ailleurs, l'évolution des prix peut être influencée à court terme par des phénomènes qui ne sont pas monétaires, comme les coûts de production. Ce n'est

certainement pas un hasard si les deux pics d'inflation observés dans le graphique 1 suivent de près deux chocs pétroliers.

Par conséquent, l'évolution de l'inflation dépend non seulement de celle de la politique monétaire, qui agit sur la demande, mais aussi de phénomènes, comme le prix des matières premières, qui agissent sur l'offre.

Le modèle IS-LM est donc doublement incapable de servir de base à une théorie de l'inflation. Non seulement il repose sur l'hypothèse de prix fixes, mais en plus il ne décrit que la détermination de la demande.

Pour dépasser ces deux limites, il est nécessaire de compléter le modèle IS-LM par une description de l'offre agrégée, ce qui permet de tenir compte des capacités de production et de l'évolution des coûts.

Dans ce chapitre, nous commencerons donc par construire un modèle qui tienne compte à la fois de l'offre et de la demande agrégée (section 1). Ce modèle est tout simplement appelé le modèle « demande agrégée-offre agrégée » ou DA-OA. C'est grâce à ce modèle que nous pourrions analyser les effets des politiques économiques et comprendre les déterminants des fluctuations des prix et de la production (section 2).

Section 1 : Construction et équilibre dans le modèle demande agrégée-offre agrégée

Comme son nom l'indique, le modèle DA-OA repose sur une courbe d'offre agrégée (I) et une courbe de demande agrégée (II). La confrontation de ces deux courbes permet de décrire l'équilibre global de l'économie (III).

I- L'offre agrégée

La fonction d'offre agrégée décrit la relation entre le niveau des prix et le produit national. Il existe une variété de théories qui prévoient une telle relation. Pour bien comprendre ce que représente la fonction d'offre agrégée, nous commencerons par voir comment l'hypothèse de rigidité des salaires permet d'en construire une (A). Cette théorie simple permettra de déterminer en détail les propriétés de cette fonction d'offre.

Nous pourrions alors présenter plus brièvement trois autres théories de l'offre agrégée plus sophistiquées (B). Si ces trois théories reposent sur des mécanismes différents de ceux du

modèle à salaires rigides, elles aboutissent toutes à la même forme de fonction d'offre agrégée et aux mêmes propriétés.

A- La rigidité des salaires comme fondement de l'offre agrégée

Dans le chapitre précédent, nous avons énuméré toute une série de raisons pour lesquelles le salaire nominal pouvait être rigide. Le modèle d'offre agrégée avec salaires rigides suppose précisément que les salaires nominaux ne peuvent pas être renégociés en permanence.

Cette hypothèse n'est pas irréaliste. Les contrats de travail fixent le salaire nominal et ne sont pas révisés en permanence. De plus, l'argument du contrat implicite entre l'employeur et le travailleur repose précisément sur l'intérêt pour le travailleur de disposer d'un salaire stable pour s'assurer contre les aléas de la conjoncture. Pour formaliser cette intuition, le modèle à salaire rigide suppose que les salaires sont négociés au début de chaque année et intangibles par la suite.⁸⁶ Le modèle repose donc sur une chronologie précise des négociations salariales :

- 1- les salaires nominaux sont négociés pour l'année entière,
- 2- au cours de l'année, les entreprises et les travailleurs observent l'évolution de l'économie en particulier des prix, ce qui détermine le salaire réel,
- 3- lorsque le salaire réel est enfin connu, les entreprises déterminent leurs embauches et leur production.

Pour analyser les conséquences de cette description, on peut choisir deux horizons. D'abord, on peut décrire ce qui se passe à partir du moment où les salaires ont été négociés. On décrit alors les caractéristiques de la fonction de production de court terme (1). Si on élargit l'horizon pour considérer toute l'année, voire plusieurs années, en supposant que le salaire nominal peut être renégocié, on aboutit à la fonction d'offre agrégée de long terme (2).

1) La fonction d'offre agrégée de court terme

Pour décrire la fonction de production agrégée de court terme, on doit reprendre chacune des trois étapes du modèle. La plus importante est la première. C'est celle où les agents déterminent le salaire nominal. Leur problème est qu'ils ne s'intéressent pas au salaire nominal en tant que tel mais au salaire réel. Ils doivent donc négocier le salaire nominal en essayant de prévoir l'évolution des prix.

⁸⁶ Ce modèle apparaît notamment dans l'article de J.A. Gray "Wage indexation: a macroeconomic approach", *Journal of Monetary Economics*, vol 73 n°2, p.221-235, avril 1976.

Pour décrire cette situation, nous allons supposer que lors du processus de fixation du salaire, les agents tombent d'accord sur un niveau du salaire réel. Leur problème est alors de choisir le salaire nominal qui permettra d'obtenir le salaire réel convenu. Ils doivent donc anticiper le niveau des prix de l'année. Le salaire nominal se déduit alors automatiquement du niveau anticipé des prix :

$$W = \left(\frac{W}{P} \right)_{cible} \times P^e$$

$$W = \omega \times P^e \quad (1)$$

Où P^e représente le niveau anticipé des prix pour l'année, au moment où les contrats sont signés. ω est le salaire réel cible.

On peut à présent passer à la deuxième étape de notre raisonnement. Si les prix évoluaient conformément aux anticipations, le salaire réel serait égal au salaire réel cible. Cependant, rien ne garantit que les anticipations des agents vont être validées. Il peut en effet se produire toute une série d'événements imprévus au cours de l'année, par exemple une modification du cours d'une matière première, une mauvaise récolte, un changement de la politique économique, un ouragan etc. Le niveau des prix peut donc être différent de celui qui était prévu.

Par conséquent, le taux de salaire réel effectivement en vigueur va être différent de son niveau anticipé. Il suffit pour cela de diviser le salaire nominal (1) par le niveau observé des prix :

$$\frac{W}{P} = \frac{\omega \times P^e}{P} = \omega \frac{P^e}{P} \quad (2)$$

L'expression (2) montre bien que le salaire réel serait égal à sa valeur cible si le niveau des prix était parfaitement anticipé ($P = P^e$). En revanche, si le niveau des prix a été surestimé ($P^e > P$), le salaire réel sera supérieur à sa valeur cible ($W/P > \omega$). A l'inverse, si le niveau des prix a été sous-estimé ($P^e < P$), le salaire réel sera inférieur à sa valeur cible ($W/P < \omega$).

La troisième et dernière étape de notre raisonnement consiste à décrire comment les agents vont réagir aux variations du salaire réel. Sans aucun doute, les travailleurs vont modifier leur offre et les entreprises leur demande. La question est de savoir qui va décider de la quantité de travail fournie.

Nous allons donc faire une dernière hypothèse. Nous supposons ainsi que ce sont les employeurs qui déterminent la quantité de travail qu'ils utilisent en fonction de la valeur du

salaires réels. Les négociations salariales de début d'année ne servent donc qu'à fixer le taux de salaire nominal mais pas la quantité de travail fournie. Les employés s'engagent à travailler autant que ce que leur employeur leur demandera par la suite. Ils savent cependant que toutes les heures travaillées seront rémunérées au taux de salaire négocié en début d'année.

Cette hypothèse est une simplification mais elle n'est pas extravagante dans une économie où il existe un chômage structurel. La demande est alors bien le côté court du marché du travail et c'est donc elle qui détermine l'emploi.

Il est à présent nécessaire de décrire la demande de travail. Or nous avons vu dans le chapitre précédent que les entreprises maximisent leur profit en ajustant leur demande de travail de telle façon que le salaire réel soit égal à la productivité marginale du travail :

$$\frac{W}{P} = Pm_L \quad (4, \text{ chapitre } 7)$$

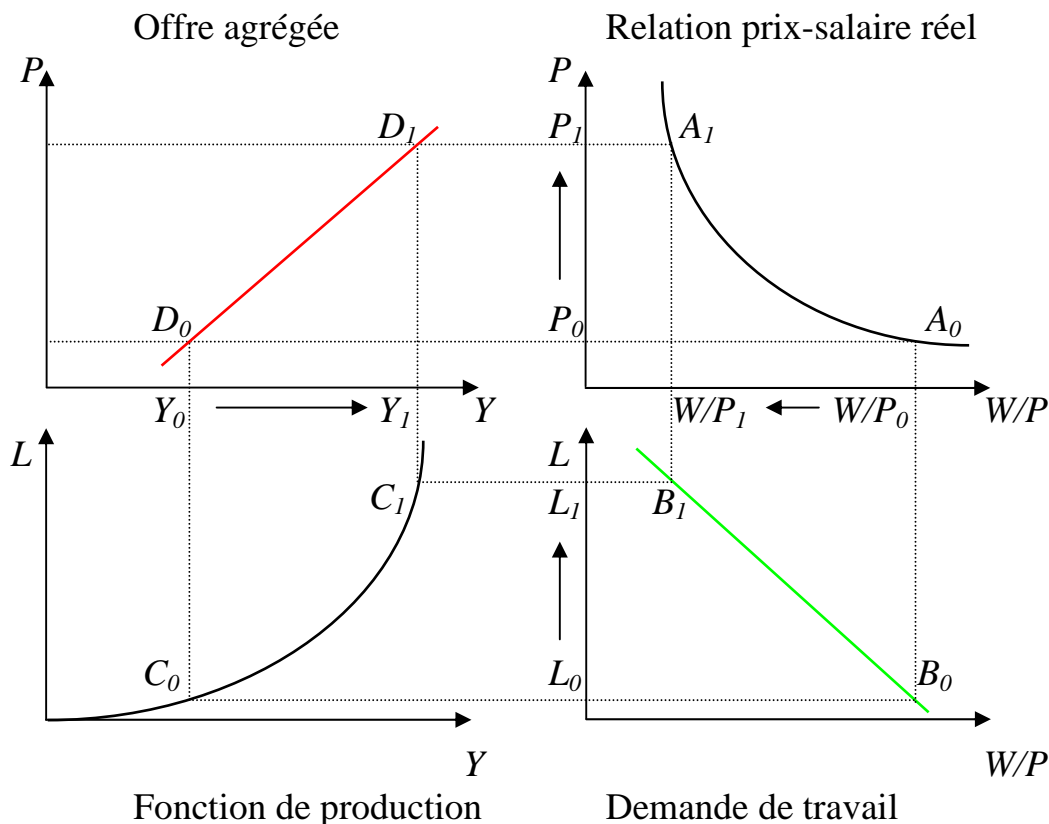
Si on suppose que la productivité marginale du travail est décroissante, l'égalité ci-dessus implique que les entreprises emploieront moins de travail si le salaire réel augmente. En effet, si le salaire réel augmente, il faut que la productivité marginale du travail augmente. Pour cela, il faut que la quantité de travail employée diminue.

A l'inverse, si le salaire réel diminue, les entreprises auront intérêt à augmenter leur demande de travail, ce qui en réduira la productivité marginale jusqu'à ce qu'elle soit à nouveau égale au salaire réel.

Ce comportement implique que si le niveau des prix est supérieur à son niveau anticipé, l'emploi et la production seront également supérieurs à leur niveau anticipé. On peut donc dire que **l'offre globale est une fonction croissante de l'inflation non anticipée.**

Par conséquent, pour un niveau donné du salaire nominal, c'est-à-dire pour une valeur donnée des anticipations, l'emploi et la production seront une fonction croissante des prix. Le modèle à salaire réel rigide prévoit donc que **l'offre agrégée est croissante à court terme.**

Graphique 2 : Construction de l'offre agrégée de court terme



Ce raisonnement est repris dans le graphique 2 qui comprend quatre quadrants. On veut tracer la courbe d'offre agrégée dans le quadrant nord-ouest. Le quadrant nord-est représente la relation entre le niveau des prix et le salaire réel, qui est mécanique à partir du moment où le salaire nominal est donné. Le quadrant sud-est représente la demande de travail, qui permet de déterminer la quantité de travail demandée par les entreprises pour chaque valeur du salaire réel. Enfin, le quadrant sud-ouest représente la fonction de production de l'économie qui relie la quantité produite à la quantité de travail employée.⁸⁷

Ce graphique permet de décrire comment une augmentation imprévue du niveau des prix va provoquer une augmentation de la production globale.

Pour ce faire, on commence la lecture du graphique par le quadrant nord-est. Ce quadrant montre que l'augmentation imprévue des prix réduit le salaire réel. Le quadrant sud-est permet alors de constater que la quantité de travail utilisée par les entreprises augmente.

⁸⁷ En toute rigueur, la courbe de demande de travail se déduit de la fonction de production. Comme le montre l'expression (4, chap.7), la demande de travail se déduit de la productivité marginale du travail. Par conséquent, la courbe de demande de travail représente la pente de la fonction de production. Elle ne se modifie que si la fonction de production se modifie, c'est-à-dire en cas de choc technologique.

Le quadrant sud-ouest décrit l'augmentation de la production que permet l'augmentation de la quantité de travail utilisée. Il n'y a plus qu'à reporter cette nouvelle quantité produite pour constater que la fonction d'offre agrégée de court terme est effectivement croissante.

Ces résultats décrivent les phénomènes qui se déroulent en cours d'année, c'est-à-dire tant que le salaire nominal reste fixé au niveau choisi en début d'année. Si on veut décrire l'offre sur un horizon plus long, il faut tenir compte du fait que les agents peuvent réviser les salaires nominaux, on décrit alors la fonction d'offre agrégée de long terme.

2) L'offre agrégée de long terme

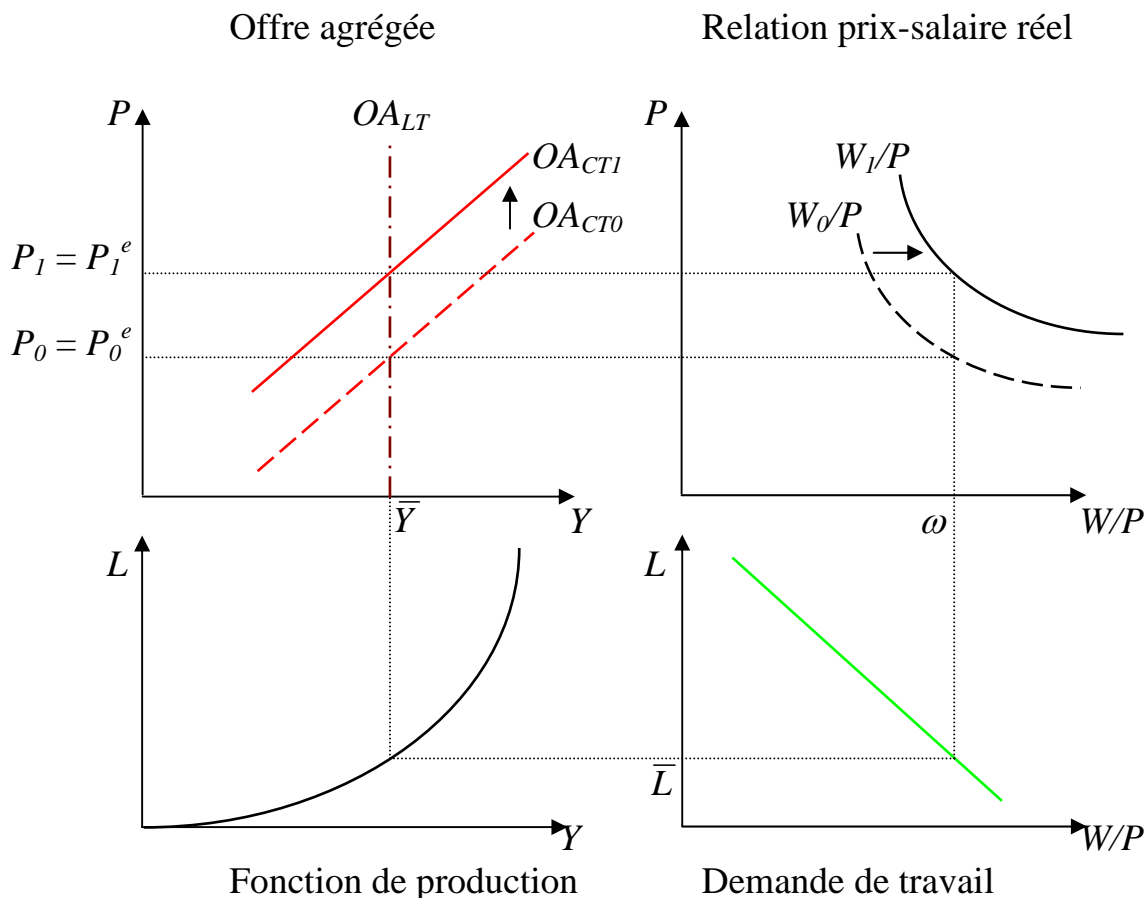
En début d'année, la négociation portera sur le salaire réel. Les employeurs et les travailleurs décident donc d'un salaire réel et en déduisent le salaire nominal correspondant, compte tenu du niveau anticipé des prix. Ils déterminent ainsi également les niveaux structurels de l'emploi et de la production. Ce sont ces niveaux d'emploi et de production qui prévaudront si le niveau observé des prix est bien égal à ce que les agents avaient anticipé. Nous les notons \bar{L} et \bar{Y} , pour souligner le fait qu'ils ne dépendent que de facteurs structurels.

Si les agents révisent à la hausse le niveau des prix qu'ils anticipent, ils augmenteront dans les mêmes proportions le salaire nominal. Le salaire réel négocié restera constant. En moyenne, et lorsque les anticipations seront vérifiées, le salaire réel, l'emploi et la production seront donc donnés par leur valeur structurelle. Par conséquent, **l'offre agrégée de long terme est indépendante du niveau des prix.**

Pour comprendre les raisons de ce résultat très important, il est utile de reprendre le graphique à quatre quadrants que nous avons utilisé pour décrire l'offre agrégée de court terme et de l'utiliser pour envisager les effets d'une augmentation anticipée des prix. Ce faisant, nous pourrions observer la relation entre l'offre agrégée à court et à long terme.

Supposons que l'économie se trouve au départ en situation d'équilibre de long terme, c'est-à-dire que les agents aient correctement anticipé le niveau initial des prix ($P_0^e = P_0$). Cela implique que le salaire réel est égal à sa valeur cible ($W_0/P_0 = \omega$). Supposons à présent que les agents anticipent correctement une augmentation des prix ($P_1^e = P_1 > P_0$).

Graphique 3 : L'offre agrégée de long terme



Comme les agents ont prévu l'augmentation des prix, ils vont ajuster le salaire nominal de façon à ce que le salaire réel reste égal à sa valeur cible ($W_1 = \omega \times P_1^e > W_0$). On voit que cet ajustement a pour effet de déplacer la relation entre le salaire réel et les prix, dans le quadrant nord-est. En revanche, les courbes représentées dans les deux quadrants du bas du graphique ne sont pas affectées, puisque ce sont des relations techniques. On peut en tirer deux conclusions :

- d'abord, la fonction d'offre agrégée de court terme se déplace vers le haut. Cela signifie qu'à production égale, le niveau des prix sera désormais toujours supérieur à ce qu'il était auparavant.

- ensuite, la production est la même avant et après l'augmentation des prix, tant que les prix anticipés sont conformes aux prix réalisés.

Ainsi, en l'absence d'erreurs d'anticipations, la production est indépendante du niveau des prix. Si on définit la fonction d'offre de long terme comme décrivant la quantité offerte

Les notes du cours ne sont transmises qu'à titre indicatif.

quand les agents ont correctement anticipé les prix, cela implique que **la courbe d'offre agrégée de long terme est verticale.**

On peut remarquer de plus que s'il existe une seule courbe d'offre agrégée de long terme, il existe une infinité de courbes d'offre agrégée de court terme. En effet, il correspond une courbe d'offre de court terme différente à chaque valeur du niveau anticipé des prix.

Cependant, on peut déterminer précisément le point d'intersection entre chaque courbe de court terme avec celle de long terme. En effet, nous savons que la production est égale à sa valeur de long terme quand l'erreur d'anticipation est nulle. Par conséquent, chaque courbe d'offre de court terme coupera la courbe de long terme au point d'ordonnée $P = P^e$. Ce résultat sera utile pour la suite du cours, notamment quand nous envisagerons la réaction de l'économie à un choc ou à une politique économique donnée.

Le modèle à salaire rigide suggère que seules des anticipations erronées permettent à la production courante d'être différente de sa valeur structurelle. Plus précisément, la production sera supérieure à sa valeur structurelle si les prix sont supérieurs à leur valeur anticipée, et inférieure dans le cas opposé.

Nous allons voir à présent qu'on peut retrouver ces résultats en partant de raisonnements complètement différents.

B- Trois autres fondements de l'offre agrégée

On peut envisager au moins trois autres théories qui permettent de construire une fonction d'offre agrégée. Toutes trois prédisent cependant que la fonction d'offre agrégée sera croissante à court terme et verticale à long terme et que les écarts du revenu par rapport à sa valeur de long terme dépendront des écarts entre l'inflation courante et l'inflation anticipée. Ces théories reposent respectivement sur l'information incomplète des producteurs (1), sur la rigidité des prix (2) et sur les négociations salariales (3).

1) L'information imparfaite

Certains macroéconomistes considèrent que l'hypothèse de rigidité des prix et des salaires n'est pas satisfaisante. Ils préfèrent supposer que les prix s'ajustent en permanence pour rétablir l'équilibre de tous les marchés. Cette hypothèse est incompatible avec les théories de l'offre globale que nous avons mentionnées jusqu'ici.

Elle n'amène cependant pas à rejeter l'idée que le produit global est une fonction de l'écart entre les niveaux constaté et anticipé des prix. On peut en effet retrouver ce résultat en supposant que l'information des producteurs est imparfaite.⁸⁸

Supposons qu'il existe une multitude de producteurs spécialisés chacun dans la production d'un seul bien mais consommant un grand nombre d'autres biens. La microéconomie nous enseigne que leurs décisions d'augmenter ou de réduire leur production dépendront de l'évolution du prix relatif du bien qu'ils produisent. Si le prix relatif de leur production augmente, ils augmenteront leur production. S'il diminue, ils la réduiront.

Afin d'évaluer le prix relatif de sa production chaque producteur a besoin de connaître non seulement l'évolution du prix nominal de sa production mais aussi celle de tous les biens qu'il consomme. Si on peut raisonnablement supposer que chaque producteur connaît l'évolution du prix de sa production, il est aussi raisonnable de supposer qu'il ne pourra connaître l'évolution des autres prix qu'avec imprécision. Son information est donc imparfaite.

Pour simplifier, supposons que chaque producteur n'observe que l'évolution du prix de son propre bien au moment où il doit choisir les quantités qu'il produit. Il n'apprendra que par la suite l'évolution du prix des autres biens.⁸⁹ Dans ces circonstances, quel peut être le comportement des producteurs ? Chaque producteur est contraint d'anticiper l'évolution des prix des biens qu'il ne produit pas sur la base de son information imparfaite. Il compare alors l'évolution de son propre prix à celle qu'il a anticipée pour les autres prix.

S'il constate que la variation du prix du bien qu'il produit est égale à la variation anticipée des autres prix, il en déduit que son prix relatif est resté constant. Il n'a donc aucune raison de modifier son comportement et sa production reste constante.

S'il observe que le prix du bien qu'il produit augmente, le producteur est confronté à une incertitude. Soit l'augmentation du prix nominal reflète une augmentation générale de tous les autres prix et le prix relatif reste donc constant. Soit l'augmentation correspond effectivement à une augmentation du prix relatif. Dans le premier cas, le producteur ne doit pas modifier sa production. Dans le second, il doit au contraire l'augmenter.

⁸⁸ Cette théorie a été initiée par Robert Lucas, le chef de file du courant des anticipations rationnelles. On en trouvera l'origine dans son article "Understanding business cycles", *Carnegie-Rochester Conferences on Public Policy*, supplément du *Journal of Monetary Economics* vol 5, p.7-29, 1977.

⁸⁹ Cette hypothèse est réaliste. Les processus de production sont en effet longs et un laps de temps non négligeable s'écoule entre le moment où les quantités à produire sont décidées et celui où les recettes sont dépensées. Un exemple typique est celui d'un agriculteur qui doit déterminer au moment des semailles les quantités qu'il souhaite produire, alors qu'il ne pourra dépenser ses profits que lorsque la récolte aura été vendue.

Dans le doute, il est rationnel pour le producteur d'adopter un comportement intermédiaire. S'il observe une augmentation non anticipée de son prix, il augmente sa production mais d'une quantité moindre que s'il était certain que son prix relatif avait effectivement augmenté.

Le résultat est toutefois que toute augmentation non anticipée du prix de son bien incitera le producteur à augmenter sa production. A l'inverse, si l'augmentation du prix est inférieure à celle qu'il avait anticipée, le producteur réduira sa production, parce que cette variation du prix pourrait correspondre à une diminution de son prix relatif.

Notre raisonnement, n'a pour l'instant concerné qu'un seul producteur à la fois, mais il faut considérer qu'il s'applique à tous les producteurs de l'économie en même temps. Par conséquent, si l'augmentation du niveau général des prix est supérieure à celle qui était anticipée, tous les producteurs verront augmenter le prix nominal de leur bien en même temps. Ils décideront tous d'augmenter leur production en même temps.

On peut donc conclure qu'une augmentation du niveau général des prix supérieure à ce qui était anticipé sera accompagnée d'une augmentation du produit national. A l'inverse, le même raisonnement nous amène à conclure qu'une augmentation du niveau général des prix inférieure aux anticipations sera accompagnée d'une diminution de la production.

Ainsi, le modèle avec information imparfaite permet de retrouver la même fonction d'offre agrégée du modèle précédent. La production est encore une fonction de l'écart entre le niveau des prix courant et son niveau anticipé. En l'absence d'erreur d'anticipation, la production est égale à son niveau structurel.

2) La rigidité des prix

Le modèle de rigidité des prix lève une dernière hypothèse sur laquelle reposaient les deux précédentes théories de l'offre agrégée, celle de l'ajustement permanent et uniforme de tous les prix. En effet, toutes les entreprises n'ajustent pas leurs prix en même temps. Si certaines peuvent facilement réviser leurs prix, d'autres sont contraintes de les fixer plus ou moins durablement.

La rigidité des prix peut s'expliquer par l'existence de catalogues, de contrats de long terme, ou de tout autre type de coûts de modification des prix (on parle de *coûts de menu*). Si une entreprise ne se trouve pas sur un marché en concurrence pure et parfaite, sa survie ne sera pas menacée si elle pratique un prix légèrement supérieur à celui de ses concurrentes puisqu'elle ne verra pas ses ventes s'effondrer.

Il est donc raisonnable de considérer qu'il existe dans l'économie certaines entreprises dont les prix sont flexibles et d'autres qui doivent s'engager pour une période donnée.

Lorsqu'un producteur choisit son prix, il doit tenir compte à la fois du niveau général des prix et de la demande. En effet, il doit non seulement tenir compte de l'évolution des prix de ses concurrents, mais aussi de ceux de ses fournisseurs et de celui des biens qu'il consomme. Comme précédemment, le producteur choisit son prix nominal mais son prix relatif dépend de l'évolution de tous les autres prix dans l'économie.

Par ailleurs, si la demande augmente et que le producteur la satisfait, il voit augmenter ses coûts marginaux. Il doit donc choisir un prix d'autant plus élevé que la demande est importante. Comme la demande dépend du revenu, on peut supposer que le prix choisi par chaque producteur dépend indirectement du niveau du revenu.

Par conséquent, le prix fixé par chaque producteur sera une fonction croissante du revenu et du niveau général des prix.

Cela est vrai pour les producteurs qui peuvent ajuster en permanence leur prix. Ceux qui doivent l'annoncer pour une période donnée sont contraints de fixer leur prix en fonction de la valeur anticipée du revenu et des prix.

Comme dans l'économie il existe à la fois des entreprises qui fixent leur prix et d'autres qui l'ajustent en permanence, le niveau général des prix sera une moyenne des prix choisis par les unes et les autres. Ce sera donc une fonction croissante à la fois du niveau anticipé des prix et de la production.

On peut utiliser cette expression pour exprimer la production en fonction des autres variables. Il suffit de la lire en sens inverse. On constate alors que l'offre agrégée est une fonction croissante du niveau des prix et des prix anticipés. La fonction d'offre agrégée a donc la même forme que celles que nous avons étudiées jusqu'à présent.⁹⁰

3) Les négociations salariales

Dans le chapitre précédent, nous avons vu que les relations sur le marché du travail sont spécifiques. En particulier, les négociations entre les employeurs et les syndicats y jouent un rôle particulier. En les intégrant dans un raisonnement macroéconomique, on peut proposer une théorie supplémentaire de l'offre agrégée. C'est ce que permet le *modèle WS-PS*, qui relie le marché du travail et celui des biens, développé au début des années 90.⁹¹ Ce modèle très souple repose sur une description simple de la formation des salaires nominaux et des prix.⁹²

⁹⁰ La théorie à prix rigides est notamment développée dans l'article de Julio Rotemberg "Monopolistic price adjustment and aggregate output", *Review of Economic Studies*, vol 49 n°4, p. 517-531, 1982.

⁹¹ Il trouve son origine dans l'ouvrage de R. Layard, S. Nickell et R. Jackman *Unemployment, macroeconomic performance and the labour market*, Oxford University Press, 1991.

⁹² En anglais, la formation des salaires se traduit par « wage setting », d'où WS, et celle des prix « price setting », d'où PS.

L'idée de base sur laquelle on peut entamer la réflexion sur la détermination des salaires est que le salaire évoluera en fonction du pouvoir de négociation des travailleurs et des employeurs. Or l'état de l'économie va influencer leur pouvoir de négociation.

Les négociations salariales ont lieu à intervalle plus ou moins régulier. Elles fixent alors les salaires nominaux pour une période donnée, même si des ajustements sont envisageables par la suite. Les négociations doivent donc se baser sur l'anticipation de l'évolution du niveau des prix.

En effet, c'est le salaire réel qui intéresse les agents. Ils doivent tenir compte de l'évolution prévue des prix pour le négocier. S'ils prévoient que les prix vont doubler au cours de l'année, la base de la négociation sera un doublement des salaires nominaux. Pour tenir compte de cette caractéristique des négociations sur le marché du travail, nous allons supposer qu'un cadre de négociation est adopté au début de chaque période, par exemple une année, sur la base du niveau anticipé des prix et qu'il s'impose pour toute la période en cours.

Cependant, les parties à la négociation vont par la suite essayer de modifier le salaire à leur avantage. Le résultat va dépendre de leur pouvoir de négociation. L'hypothèse centrale du modèle est que ce pouvoir de négociation va évoluer en fonction du niveau du chômage.

En effet, une augmentation du taux de chômage réduit le pouvoir de négociation des travailleurs au profit des employeurs. Cela va permettre à ces derniers d'obtenir une diminution du salaire. A l'inverse, lorsque l'économie se rapproche du plein emploi, les travailleurs disposent de plus de poids pour revendiquer des augmentations de salaire.⁹³

En résumé, le salaire nominal sera une fonction croissante du niveau anticipé des prix mais décroissante du taux de chômage. C'est la relation WS.

Pour compléter la description de la négociation, nous devons décrire sa chronologie. Nous supposons donc qu'elle se base sur les anticipations du niveau des prix formées au début de l'année. Ainsi nous supposons que le salaire peut être renégocié en permanence en fonction de l'évolution du taux de chômage et des caractéristiques institutionnelles de

⁹³ Cette représentation des négociations salariales n'est pas sans rappeler l'idée de Karl Marx selon laquelle les chômeurs constituaient une « armée industrielle de réserve » qui permettait aux capitalistes de limiter les salaires.

l'économie, mais que les anticipations du début de la période s'imposent comme base de négociation tout au long de la période. Une autre façon de traduire cette idée est de supposer que les anticipations se forment une fois pour toutes au début de l'année et ne sont révisées qu'une fois par an.

Parallèlement aux négociations salariales, les producteurs déterminent leurs prix. La microéconomie montre qu'un producteur fixe son prix en imposant une marge sur son coût marginal. Cette marge augmente avec le pouvoir de marché du producteur. Elle tend vers zéro lorsque le marché se rapproche d'une structure concurrentielle.

A court terme, le coût marginal est constitué essentiellement du travail. Ce résultat microéconomique montre donc que le prix pratiqué sera une fonction croissante du salaire nominal. Cette expression définit la relation PS.

Indirectement, la relation PS décrit comment les entreprises vont ajuster l'emploi aux fluctuations du salaire. Si le salaire réel dépasse la valeur donnée par PS, les entreprises réduiront leur demande de travail parce que le coût marginal du travail sera supérieur à la recette marginale qu'il permet d'obtenir.

Reprenons à présent la chronologie des négociations. En début de période, les employeurs et les travailleurs déterminent le salaire nominal qui servira de base aux négociations pour toute la période, sur la base de l'inflation qu'ils anticipent pour cette période. Ils déterminent ainsi l'emploi, la production et le chômage structurel.

Cependant, en cours d'année des événements imprévus surviennent. Supposons par exemple une augmentation inattendue de la demande de biens provoquée par une augmentation des dépenses publiques.

Pour satisfaire cette demande, les entreprises vont augmenter leur production, donc employer des travailleurs supplémentaires. L'emploi augmente donc et le chômage diminue. Le pouvoir de négociation des travailleurs augmente alors et ils obtiennent une augmentation du salaire nominal. Les producteurs voient alors leur coût marginal augmenter, ce qu'ils répercutent sur leurs prix.

In fine, on constate que la production a augmenté et a dépassé son niveau structurel. Parallèlement les prix ont augmenté et sont à présent supérieurs à leur niveau anticipé. On retrouve ainsi bien la relation habituelle de court terme entre prix et production agrégée.

A plus long terme, les agents corrigent leurs anticipations et le salaire nominal en intégrant l'augmentation des prix. Le salaire nominal retrouve donc sa valeur structurelle, de même que l'emploi et le chômage. On retrouve ainsi l'idée qu'une variation anticipée des prix n'affecte pas l'offre agrégée. En d'autres termes, l'offre agrégée de long terme est verticale.

Encadré 1 : Détermination algébrique de la courbe d'offre agrégée à partir du modèle WS-PS

On suppose que la relation WS prend la forme suivante :

$$W = P^e F(u) \quad (\text{E1})$$

Pour obtenir l'expression de la courbe PS, on suppose que la fonction de production est de la forme suivante : $Y = A.N$, où A est un paramètre positif et N mesure la quantité de travail utilisée. Pour éviter de traîner le paramètre A pendant le reste de nos calculs, nous allons de plus mesurer la production dans des unités telles que A soit égal à un. On dit qu'on *normalise* la production de façon à ce que A soit égal à un.⁹⁴ La fonction de production s'écrit alors simplement $Y = N$. Elle implique qu'il faut exactement une unité de travail pour produire une unité supplémentaire de bien. Le coût marginal est alors exactement égal au salaire nominal.

Si on baptise μ la marge appliquée par les producteurs sur le coût marginal, leur prix est alors donné par la relation PS, qui est de la forme suivante :

$$P = (1 + \mu)W \quad (\text{E2})$$

En combinant ces deux expressions, on peut aussi exprimer le niveau courant des prix en fonction du niveau anticipé des prix et du taux de chômage :

$$P = (1 + \mu) P^e F(u, z_1, \dots, z_n) \quad (\text{E3})$$

Cette expression implique indirectement une relation entre le niveau des prix et la quantité produite. Pour s'en assurer, il suffit de rappeler la définition du taux de chômage. Il s'agit du nombre de chômeurs rapporté à la population active (L). Par ailleurs, comme la population active se partage entre chômeurs et travailleurs ayant un emploi, le nombre de

⁹⁴ Il ne s'agit que d'une astuce qui permet d'alléger la présentation mais ne modifie pas les résultats. Imaginez que le produit soit des œufs et que la fonction de production s'écrive : $Y = 12N$ lorsque Y est mesuré en œufs. On peut tout aussi bien mesurer la production d'œufs en douzaines. La fonction de production s'écrit alors $Y = N$. Cela ne change rien à la quantité d'œufs produite. Par exemple, si $N = 1$, on aura produit douze œufs selon la première expression et une douzaine selon la seconde. Jusqu'à nouvel ordre, douze œufs font bien une douzaine.

chômeurs (U) est par définition égal à la différence entre la population active et le nombre de travailleurs ayant un emploi (N). On peut donc écrire :

$$u = \frac{U}{L} = \frac{L - N}{L} = 1 - \frac{N}{L} \quad (\text{E4})$$

En rappelant que la fonction de production relie la production nationale et le nombre de travailleurs employés, l'expression (E3) permet d'exprimer le taux de chômage en fonction de la production nationale. Nous avons supposé que la fonction de production agrégée était $Y = N$. Il suffit alors de remplacer N par sa valeur dans l'expression (E4) pour décrire la relation entre la production et le taux de chômage :

$$u = 1 - \frac{Y}{L} \quad (\text{E5})$$

On constate ainsi qu'une augmentation de la production se traduit par une diminution du taux de chômage. Si les entreprises désirent augmenter leur production, elles devront en effet embaucher, ce qui réduit le taux de chômage. On peut enfin exprimer le niveau des prix en fonction de la production, en combinant les expressions (E3) et (E5) :

$$P = (1 + \mu)P^e F\left(1 - \frac{Y}{L}, z_1, \dots, z_n\right) \quad (\text{E6})$$

Cette expression montre que toute augmentation de la production est accompagnée d'une augmentation des prix. La courbe d'offre agrégée de court terme est donc bien croissante.

L'expression (E6) permet de plus de déterminer l'offre agrégée de long terme. Supposons pour cela que le niveau des prix soit égal à sa valeur anticipée. La fonction d'offre agrégée peut alors se simplifier :

$$\begin{aligned} P &= P^e = (1 + \mu)P^e F\left(1 - \frac{Y}{L}, z_1, \dots, z_n\right) \\ \Rightarrow F\left(1 - \frac{Y}{L}, z_1, \dots, z_n\right) &= \frac{1}{1 + \mu} \\ \Leftrightarrow F(u, z_1, \dots, z_n) &= \frac{1}{1 + \mu} \end{aligned} \quad (\text{E7})$$

L'expression (E5) montre que lorsque le niveau des prix coïncide avec son niveau anticipé, le taux de chômage est égal à son niveau structurel déterminé dans le modèle WS-PS, qui est indépendant du niveau des prix. La courbe d'offre agrégée de long terme est donc bien verticale.

Nous disposons à présent de quatre théories qui prédisent toutes que la production est une fonction croissante des écarts entre les niveaux courant et anticipé des prix et qu'elle est égale à sa valeur structurelle lorsque cet écart est nul. Ce constat est rassurant dans la mesure où il suggère que la forme de la fonction agrégée est robuste.

En effet, les quatre théories que nous avons passées en revue ne sont pas incompatibles. Au contraire, elles reposent sur des imperfections de marché qui peuvent tout à fait coexister. En effet, on peut supposer à la fois que le salaire est rigide (modèle à salaire rigide), que l'information est imparfaite (modèle à information imparfaite), que le marché des biens n'est pas concurrentiel (modèle à rigidité des prix et modèle WS-PS), que certains prix sont rigides (modèle à rigidité des prix) et que le marché du travail n'est pas concurrentiel (modèle WS-PS).

Il faut garder à l'esprit que nous cherchons à décrire le fonctionnement d'une économie dans son ensemble. Or une économie est composée d'un grand nombre de secteurs qui fonctionnent de façon différente. Le marché du travail n'est pas homogène non plus. Dans ces conditions, la variété des théories de l'offre agrégée permet d'accorder une certaine confiance à nos résultats.⁹⁵

On pourrait donc conclure cette section par une seule expression de l'offre agrégée :

$$Y = \bar{Y} + F(P - P^e) \quad (3)$$

C'est cette expression que nous utiliserons pour analyser l'équilibre de l'économie. Il faudra auparavant avoir déterminé les caractéristiques de la demande agrégée.

II- La demande agrégée

Le fondement de la demande agrégée est fourni par le modèle IS-LM. Si ce modèle est un modèle à prix fixes, il permet cependant d'étudier les effets d'une variation exogène du niveau des prix, donc de construire la courbe de demande agrégée (A). On pourra alors l'utiliser pour étudier les effets des politiques économiques sur la position de la courbe (B).

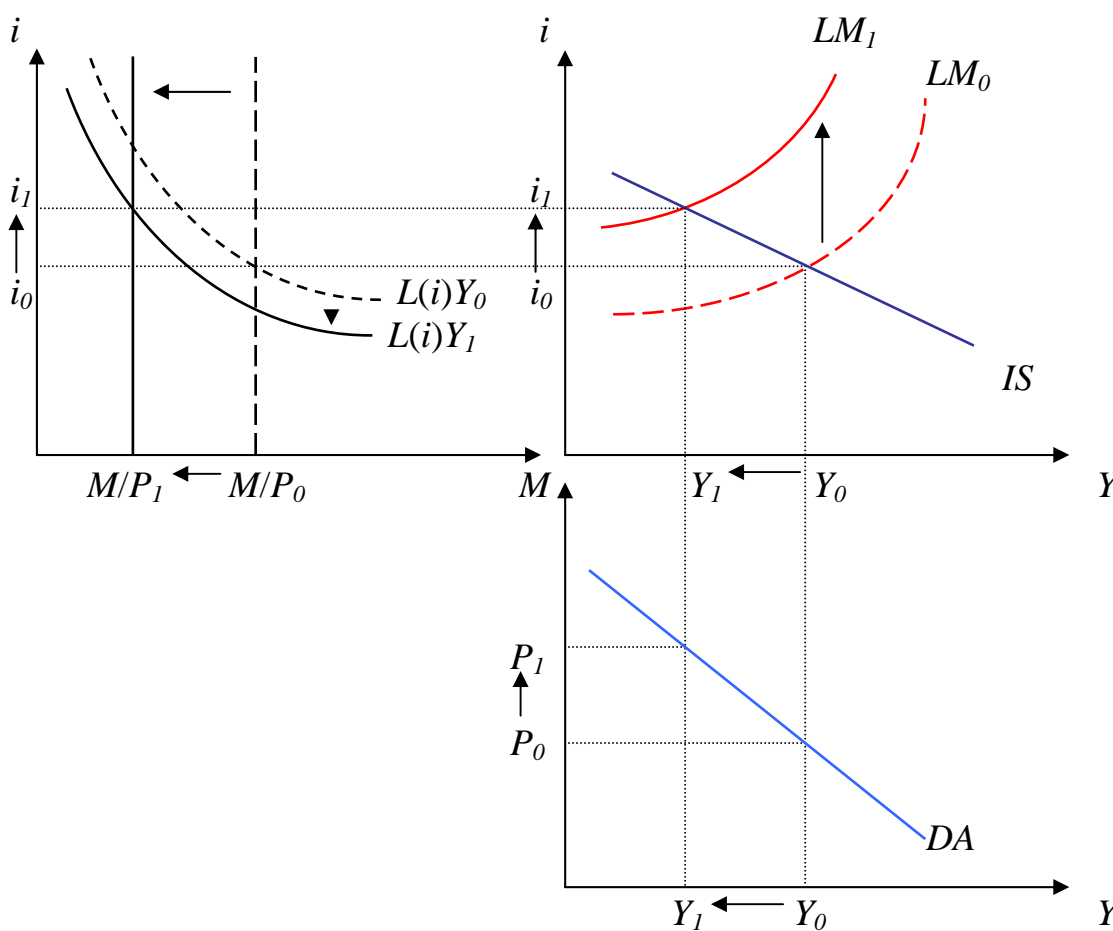
⁹⁵ Cet optimisme doit être cependant tempéré. Certes, les quatre théories aboutissent à la même forme d'offre agrégée. En revanche, leurs conclusions peuvent être opposées sur d'autres points. C'est le cas notamment de l'évolution cyclique du salaire réel. Le modèle WS-PS et le modèle à salaire rigide prévoient que le salaire réel diminue lorsque le revenu est supérieur à son niveau structurel. Les deux autres modèles prévoient l'inverse. A titre d'exercice essayez de comprendre pourquoi.

A- Construction de la courbe de demande agrégée

Lorsque l'on reprend la construction du modèle IS-LM, on constate que le niveau des prix n'intervient que dans la définition de l'offre d'encaisses réelles (M/P). Par conséquent, la relation entre le niveau des prix et la demande agrégée passera forcément par l'effet de la variation de prix sur l'offre d'encaisses réelles. Cet effet peut se représenter dans un graphique à trois quadrants (graphique 4).

L'augmentation du niveau des prix réduit l'offre d'encaisses réelles. On pourrait dire que le pouvoir d'achat de la quantité de liquidité en circulation diminue. Comme l'offre d'encaisses réelles diminue, une demande excédentaire de monnaie apparaît, ce qui provoque une augmentation du taux d'intérêt, quel que soit le revenu initial. La courbe LM se déplace donc vers le haut.

Graphique 4 : Construction de la courbe de demande agrégée



Les notes du cours ne sont transmises qu'à titre indicatif.

L'augmentation du taux d'intérêt comprime alors l'investissement, ce qui réduit la demande de biens et le revenu. On peut donc conclure qu'une augmentation du niveau des prix est accompagnée d'une diminution de la demande agrégée. Ainsi, **la courbe de demande agrégée est décroissante.**

On aurait pu obtenir les mêmes résultats en remarquant qu'une augmentation du niveau des prix dans le modèle IS-LM a le même effet qu'une politique monétaire restrictive. En effet, l'offre d'encaisses réelles diminue dans les deux cas. Il s'ensuit que la pente de la courbe de demande agrégée sera affectée par les mêmes facteurs que ceux qui déterminent l'efficacité de la politique monétaire.

Plus la politique monétaire sera efficace, plus l'effet négatif d'une augmentation des prix sur le revenu sera important, et la valeur absolue de la pente de la courbe de demande agrégée faible. Ainsi, en reprenant les résultats du chapitre consacré à la construction du diagramme IS-LM, on peut conclure que la valeur absolue de la pente de la demande agrégée sera :

- d'autant plus faible que la demande de monnaie est peu sensible au taux d'intérêt ;
- d'autant plus faible que l'investissement est sensible au taux d'intérêt et la propension marginale à consommer élevée.
- infinie en cas de trappe à liquidité.

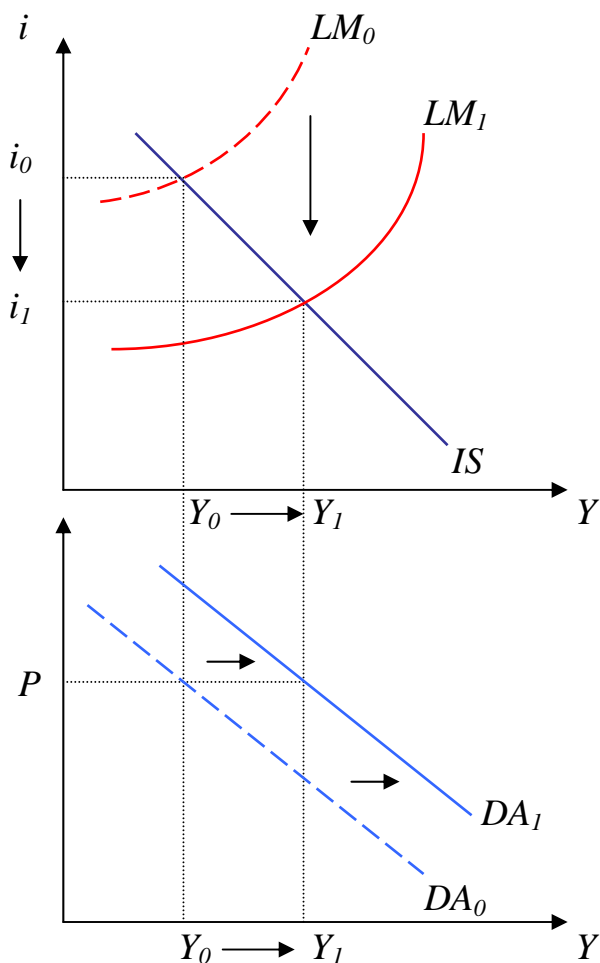
B- Position de la courbe de demande agrégée

Tous les facteurs susceptibles d'affecter le revenu d'équilibre pour un niveau donné des prix vont affecter la position de la courbe de demande agrégée. En d'autres termes, tous les paramètres qui déterminent la position des courbes IS et LM, à part le niveau des prix, vont déterminer la position de la courbe de demande agrégée. Nous allons nous concentrer sur les deux d'entre eux qui permettent d'analyser les effets des politiques économiques : l'offre de monnaie et les dépenses publiques.

1) Effet d'une augmentation de la masse monétaire sur la demande agrégée

L'augmentation de la masse monétaire pour un niveau donné des prix va se traduire par un déplacement vers le bas de la courbe LM. La conséquence dans le diagramme IS-LM en est une diminution du taux d'intérêt et une augmentation du revenu.

Graphique 5 : Effet d'une augmentation de l'offre de monnaie



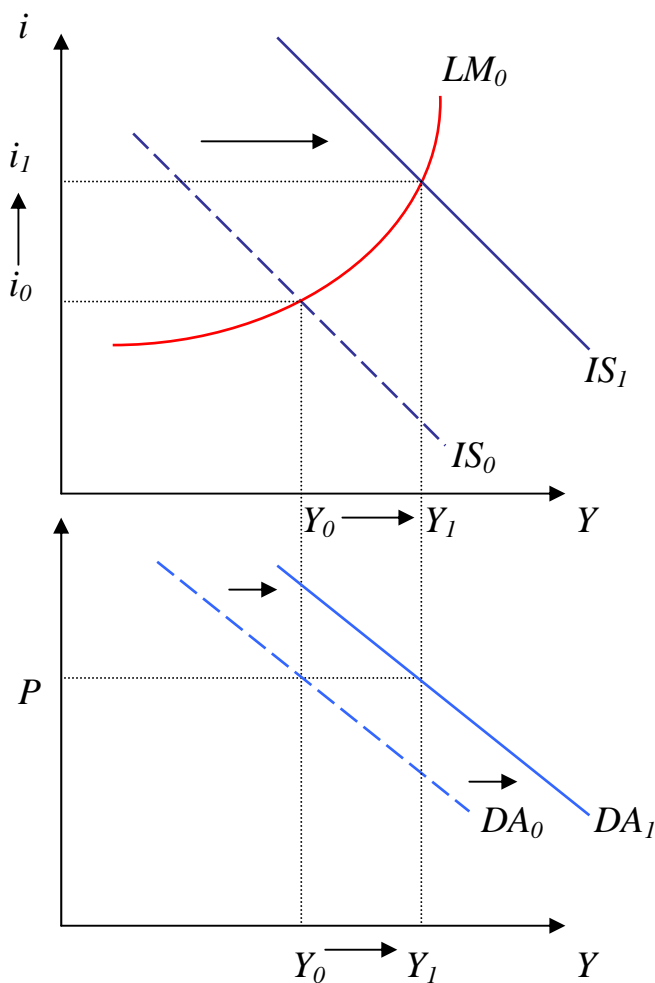
Ainsi, comme l'offre de monnaie augmente, le taux d'intérêt diminue, ce qui relance l'investissement. La demande globale augmente alors, ce qui augmente le revenu grâce au mécanisme multiplicateur.

Comme le revenu augmente quel que soit le niveau initial des prix, la demande agrégée se déplace vers la droite. Ainsi, **une politique monétaire expansionniste se traduit par un déplacement de la courbe de demande agrégée vers la droite.**

2) Effet d'une augmentation des dépenses publiques sur la demande agrégée

L'augmentation des dépenses publiques pour un niveau donné des prix va se traduire par un déplacement vers la droite de la courbe IS. La conséquence dans le diagramme IS-LM en est une augmentation du revenu et du taux d'intérêt.

Graphique 6 : Effet d'une augmentation des dépenses publiques



Ainsi, comme les dépenses publiques augmentent, la demande globale augmente, ce qui augmente la production grâce au mécanisme du multiplicateur.

Comme le revenu augmente quel que soit le niveau initial des prix, la demande agrégée se déplace vers la droite. Ainsi, **une politique budgétaire expansionniste se traduit par un déplacement de la courbe de demande agrégée vers la droite.**

III- L'équilibre global

L'équilibre global est par définition atteint lorsque l'offre agrégée est égale à la demande agrégée. Cela suppose donc implicitement que tous les marchés que nous avons décrits jusqu'à présent soient équilibrés. Ainsi, puisque l'offre agrégée dépend de l'équilibre du marché du travail, le marché du travail doit être équilibré. De plus, comme la demande

agrégée est déterminée grâce au schéma IS-LM, l'équilibre global implique également que le marché des biens et de la monnaie soient en équilibre.

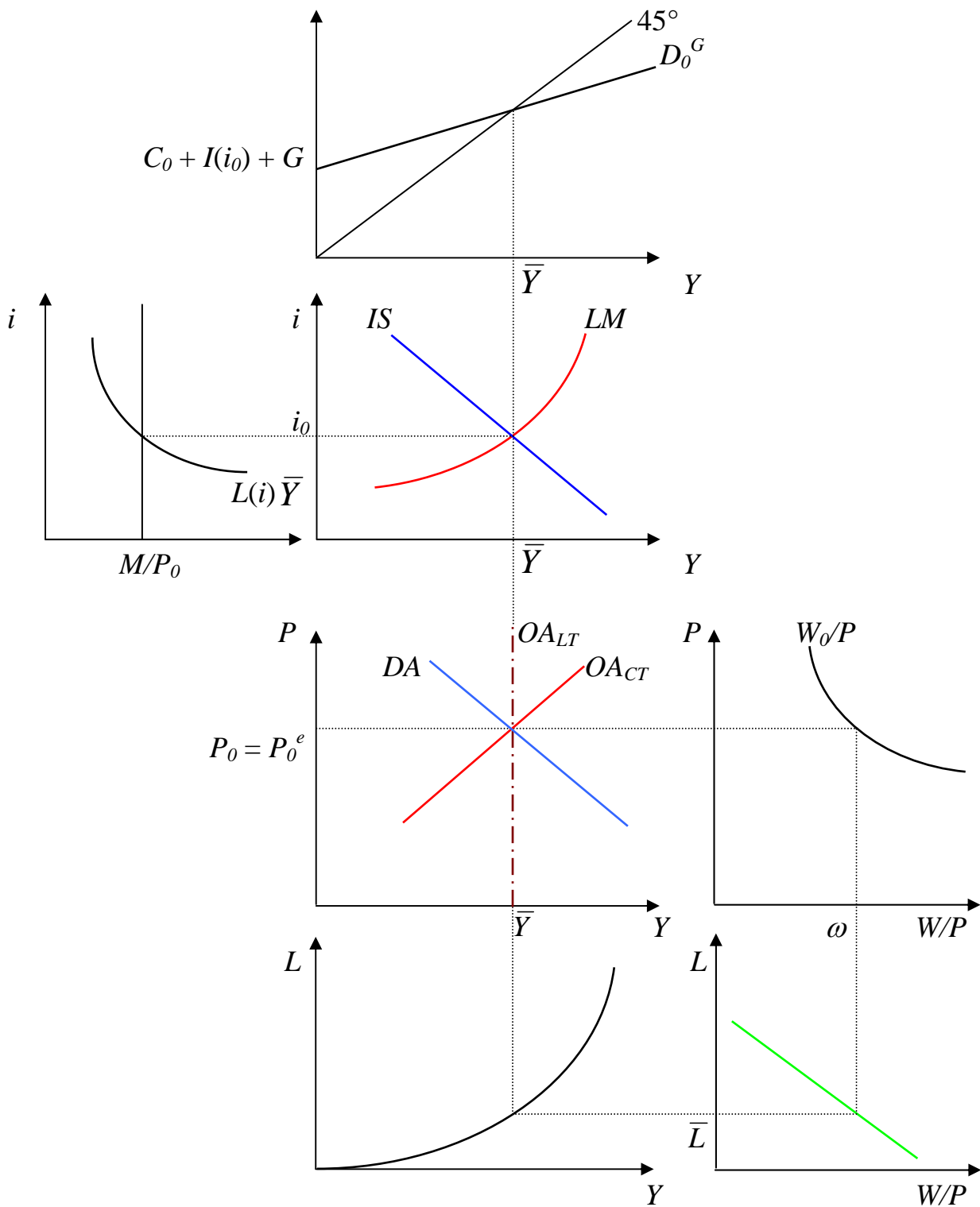
Le graphique 7 décrit cette situation. Pour ce faire, il reprend toutes les étapes de la construction de l'offre agrégée et de la demande agrégée.

Ainsi, les trois quadrants supérieurs décrivent comment la demande agrégée en étant dérivée du modèle IS-LM repose fondamentalement sur l'équilibre du marché des biens et du marché de la monnaie. Les trois quadrants situés tout en bas du schéma et à droite sur l'avant-dernière ligne montrent que l'offre globale est obtenue à partir du marché du travail, sur lequel la demande dépend des conditions techniques de production dans l'économie.

Cela permet de souligner *l'interdépendance des différents marchés*. On montre ainsi que l'égalité de l'offre agrégée et de la demande agrégée n'est possible que si tous les marchés décrits dans les chapitres précédents sont en équilibre en même temps.

Ce que montre également le graphique 7 est qu'il serait extrêmement fastidieux de devoir représenter l'équilibre de tous les marchés en même temps. C'est pourquoi dans la suite du cours nous allons nous contenter de représenter les courbes d'offre agrégée et de demande agrégée, qui fournissent un résumé de l'équilibre de tous les marchés de l'économie, comme dans le graphique 8.

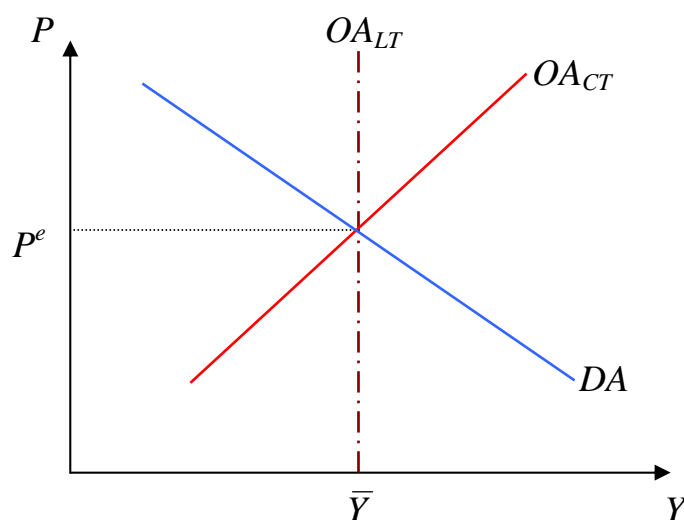
Graphique 7 : L'équilibre global



Les notes du cours ne sont transmises qu'à titre indicatif.

En revanche, lorsque nous utiliserons le modèle DA-OA pour décrire l'équilibre de l'économie ou les effets d'une politique économique, nous devons garder à l'esprit ce sur quoi la représentation de ces deux courbes repose. En d'autres termes, il sera nécessaire de garder à l'esprit l'interdépendance des différents marchés. Pour ce faire, il pourra être utile de se référer au graphique 7.

Graphique 8 : l'équilibre agrégé



Dans le modèle DA-OA, l'équilibre de l'économie est simplement déterminé par l'intersection entre la courbe de demande agrégée et la courbe d'offre agrégée. Encore faut-il préciser de quelle courbe d'offre il s'agit. En effet, les courbes d'offre à court et à long terme sont différentes.

Par conséquent, l'économie sera en situation d'équilibre de court terme lorsque la demande agrégée sera égale à l'offre agrégée de court terme. Cette situation est déterminée par l'intersection entre la courbe de demande agrégée et la courbe d'offre agrégée de court terme. C'est ce point d'intersection qui détermine la production et le niveau des prix de l'économie.

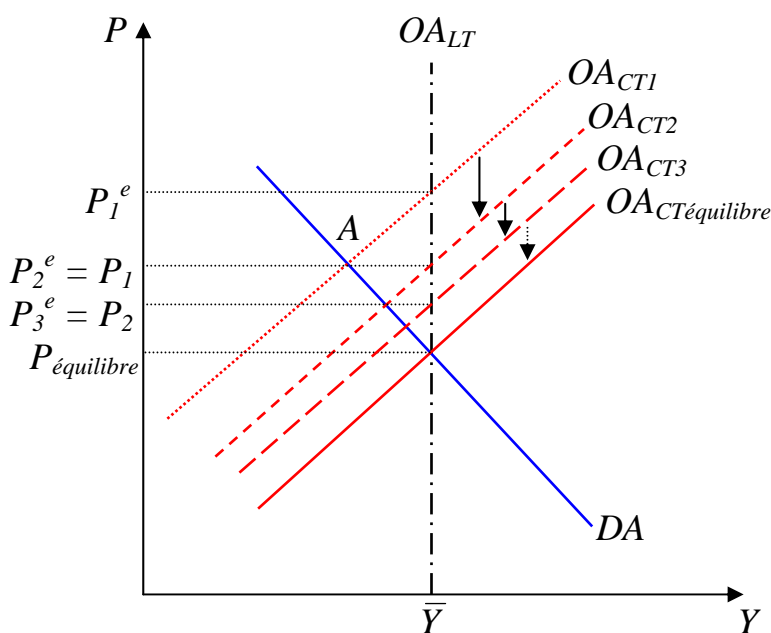
On pourra de plus affirmer que l'économie est également en situation d'équilibre de long terme lorsque la production d'équilibre correspondra à son niveau structurel. Ainsi, il faut que l'intersection entre la courbe d'offre de court terme et la courbe de demande agrégée se fasse sur la courbe d'offre de long terme. Puisque l'intersection entre les courbes d'offre de long et de court terme correspond au point dont l'ordonnée est égale au niveau anticipé des

prix, on peut conclure que l'équilibre de long terme de l'économie correspond à la situation où la production est égale à son niveau structurel et où le niveau anticipé des prix est égal à son niveau courant.

Il se peut que l'équilibre de court terme ne corresponde pas à celui de long terme. Dans ce cas, le niveau du prix courant est différent de son niveau anticipé, et la production diffère de son niveau structurel. Pour décrire le processus qui va ramener l'économie à l'équilibre, nous devons faire une hypothèse supplémentaire quant à la façon dont se forment les anticipations.

Nous allons ainsi supposer que les agents forment leurs anticipations en extrapolant les observations passées. Ainsi, nous supposons que le niveau anticipé des prix pour l'an prochain est égal au niveau des prix observé cette année. Cette forme d'anticipations est appelée *anticipations extrapolatives*.⁹⁶

Graphique 9 : le retour vers l'équilibre de long terme



⁹⁶ L'hypothèse d'anticipations extrapolatives n'est pas très satisfaisante d'un point de vue théorique parce qu'elle suppose une très grande naïveté de la part des agents. Les progrès de la macroéconomie au cours des cinquante dernières années sont dus en grande partie aux progrès réalisés dans la modélisation des anticipations. Nous n'utilisons les anticipations extrapolatives que parce qu'il s'agit d'un cas très simple mais discuterons d'autres formes d'anticipations dans la suite du cours. Nous pouvons également défendre cette hypothèse en avançant que, si on constate que l'économie converge vers l'équilibre avec des anticipations aussi frustes, des anticipations plus sophistiquées devraient a fortiori aussi ramener l'économie vers l'équilibre.

Supposons que les agents aient initialement surestimé le niveau des prix. La courbe OA_{CTI} correspond à cette situation. Le niveau anticipé des prix est donné par l'intersection entre la courbe d'offre agrégée de court terme (OA_{CTI}) et celle de long terme (OA_{LT}). Le niveau courant des prix est donné par l'intersection entre la courbe de demande (DA) et la courbe d'offre de court terme (OA_{CTI}). On voit bien que le revenu est inférieur à son niveau structurel mais aussi que le niveau anticipé des prix est supérieur à son niveau courant ($P_1^e > P_1$).

Les agents vont donc réviser leurs anticipations. Compte tenu de notre hypothèse d'anticipations extrapolatives, ils vont à présent anticiper $P_2^e = P_1$. Le prix anticipé va donc diminuer, ce qui va déplacer la courbe d'offre agrégée de court terme vers le bas. On peut tracer la nouvelle courbe en rappelant qu'elle est parallèle à la précédente et qu'elle coupe la courbe d'offre agrégée de long terme en un point dont l'ordonnée correspond au niveau des prix de la période précédente (P_1).

On constate alors que le revenu a augmenté alors que les prix ont diminué par rapport à la période précédente. Cependant le revenu reste inférieur à son niveau structurel et le niveau anticipé des prix reste supérieur au niveau réalisé des prix.

Par conséquent, les agents vont encore réviser à la baisse leurs anticipations. Le revenu va encore se rapprocher de son niveau d'équilibre et les anticipations de la valeur effective du niveau des prix. Cependant, les anticipations ne seront toujours pas validées.

Le même mécanisme va donc se poursuivre et les agents réviser leurs anticipations à la baisse jusqu'à ce que l'économie atteigne l'équilibre de long terme.⁹⁷ L'évolution des anticipations contribue donc à ramener l'économie vers sa situation d'équilibre de long terme. Ainsi, **le revenu tend à revenir vers sa valeur structurelle**. On peut donc dire qu'il existe un équilibre de long terme et une succession d'équilibres de court terme qui convergent vers l'équilibre de long terme.

Le temps que met l'économie pour retourner à son équilibre de long terme est une question centrale de la macroéconomie qui fait l'objet de débats très vifs. D'un côté, les économistes qui font confiance aux mécanismes du marché ont tendance à considérer que l'intervalle peut être très court, voire inexistant, et qu'il n'est donc pas pertinent d'en tenir compte.

De l'autre, les économistes pour qui il existe de nombreuses rigidités considèrent que le long terme est très éloigné, voire qu'il n'est jamais atteint. J.M. Keynes a ainsi déclaré

⁹⁷ En fait, l'économie tend de façon asymptotique vers sa situation d'équilibre de long terme sans jamais l'atteindre exactement. Ce résultat est une conséquence peu vraisemblable de l'hypothèse d'anticipations extrapolatives. Si les agents formaient leurs anticipations de façon moins naïve, ils devraient finir par se rendre compte que le niveau des prix tend vers une valeur donnée et anticiper directement cette valeur.

qu'« à long terme nous serons tous morts ». Dans ce cas, l'équilibre de long terme n'est que la situation vers laquelle l'économie converge, mais c'est l'équilibre de court terme qui est pertinent pour comprendre le comportement de l'économie.

Dans la suite du cours, nous nous contenterons de comparer l'équilibre de court terme et celui de long terme tel que nous venons de les déterminer sans considérer explicitement les étapes intermédiaires que nous venons de décrire.

Section 2 : La politique économique dans le modèle demande agrégée-offre agrégée

Décrire l'équilibre n'a pas d'intérêt en tant que tel si on ne l'utilise pas pour étudier les effets des politiques économiques. Nous allons donc utiliser le modèle DA-OA pour étudier les effets des politiques budgétaire (I) et monétaire (II). Il s'agit en fait de compléter l'analyse que permettait le modèle IS-LM en tenant compte de l'effet de ces politiques sur le niveau des prix et en distinguant leurs effets à court et long terme.

Une fois que les effets de ces deux politiques de gestion de la demande agrégée auront été étudiés dans l'absolu, on pourra affiner notre raisonnement en décrivant la réaction de l'économie à un choc de demande (III) et à un choc d'offre (IV).

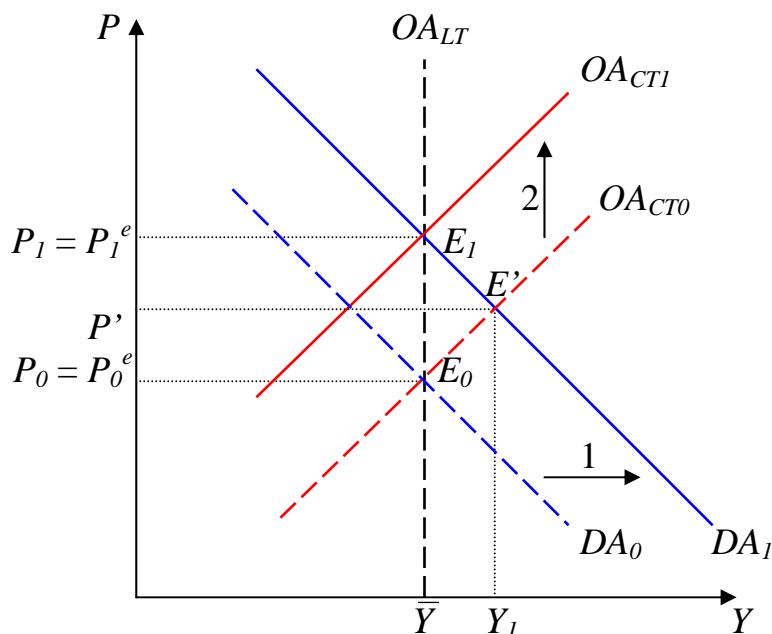
I- La politique budgétaire

La politique budgétaire affecte la demande agrégée. La base de l'analyse de ses effets dans le modèle DA-OA est donc fournie par le modèle IS-LM. Ces effets sont complétés par la réaction de l'offre et des prix.

Nous avons vu dans la section précédente qu'une augmentation des dépenses publiques augmentait la demande agrégée. Ce résultat nous permet d'étudier les effets d'une politique budgétaire expansionniste à court et long terme.

Graphiquement, la politique budgétaire déplace la courbe de demande agrégée vers la droite. DA_0 est remplacée par DA_1 . L'effet instantané en est de modifier l'équilibre de l'économie qui passe du point E_0 au point E' . Le niveau des prix et le revenu augmentent.

Graphique 10 : une politique budgétaire expansionniste



Cependant, au point E' , les anticipations sont erronées ($P' > P_0^e$). Le point E' ne peut donc pas représenter un équilibre de long terme. Les agents vont donc réviser leurs anticipations à la hausse, ce qui va déplacer vers le haut la courbe d'offre agrégée de court terme. Ainsi, les prix vont augmenter et le revenu se rapprocher de sa valeur structurelle.

In fine, l'économie aura convergé vers un nouvel équilibre, caractérisé par un niveau des prix plus élevé qu'auparavant et un revenu égal à sa valeur de long terme. Cet équilibre correspond au point E_1 .

Pour bien comprendre les conséquences d'une politique budgétaire dans le modèle DA-OA, un raisonnement graphique est insuffisant. Il faut en décomposer les mécanismes sous-jacents à court terme et long terme.

Il faut au préalable rappeler que l'économie est au départ en situation d'équilibre de court terme et long terme. La demande est alors exactement égale à la production structurelle. Cela signifie que les anticipations sont validées et que le marché du travail est en situation d'équilibre.

A court terme, l'augmentation des dépenses publiques augmente la demande de biens, ce qui augmente la demande agrégée grâce à un mécanisme multiplicateur. La production doit donc augmenter pour satisfaire la demande.

Selon le modèle de l'offre agrégée qu'on utilise, la réaction des producteurs diffère :

- d'après le modèle à rigidité des salaires, l'augmentation de la demande provoque d'abord une augmentation des prix qui réduit les salaires réels et incite les entreprises à embaucher pour augmenter la production.

- d'après le modèle à information imparfaite, l'augmentation de la demande provoque également une augmentation des prix qui est interprétée par les producteurs comme une augmentation de leurs prix relatifs. Ils embauchent donc pour augmenter leur production.

- d'après le modèle à prix rigides, les entreprises dont les prix sont flexibles augmentent leur production et leurs prix en réaction à l'augmentation de la demande.

- d'après le modèle WS-PS, l'augmentation de la production réduit le taux de chômage, ce qui augmente les salaires et les prix.

Quel que soit le modèle sous-jacent de l'offre agrégée, on constate que la production augmente, de même que les prix. On passe ainsi du point E_0 au point E' . Ainsi, on peut conclure que **la politique budgétaire est efficace à court terme mais qu'elle produit une augmentation des prix.**

L'équilibre de court terme n'est pas soutenable puisqu'il suppose que les anticipations sont erronées. Les agents vont donc finir par réviser leurs anticipations à la hausse. Cela va provoquer un déplacement vers le haut de la courbe d'offre agrégée de court terme tant que l'équilibre de long terme ne sera pas atteint.

Ici encore, le mécanisme qui va provoquer ce retour à l'équilibre de long terme va dépendre du modèle sous-jacent de l'offre agrégée :

- dans le modèle à salaire rigide, les agents privés se rendent compte que le salaire réel n'est pas celui qui équilibre le marché du travail. Ils ajustent donc le salaire nominal. Le salaire réel augmente donc, ce qui ramène l'emploi à son niveau structurel.

- dans le modèle à information imparfaite, les producteurs réalisent que l'augmentation de leur prix nominal ne correspond pas à une augmentation de leur prix relatif. Ils finissent donc par réduire leur production, ce qui augmente les prix à demande agrégée inchangée.

- dans le modèle avec rigidité des prix, les producteurs dont les prix étaient fixés les ajustent, ce qui provoque une augmentation des prix et leur permet de se rapprocher de leur production structurelle.

- dans le modèle WS-PS, les travailleurs révisent à la hausse le salaire nominal négocié. Les entreprises voient donc augmenter leurs coûts marginaux et le répercutent sur leurs prix.

Dans tous les cas, le niveau des prix augmente. Cela réduit les encaisses réelles donc la demande. L'économie se déplace alors le long de la courbe de demande agrégée.

Lorsque l'équilibre de long terme est atteint, la demande et la production ont retrouvé leur niveau initial. Seul le niveau des prix a augmenté. On peut donc conclure que **la politique budgétaire est inefficace à long terme.**

Cela ne signifie cependant pas que la politique budgétaire n'a aucun effet à long terme. Elle n'a certes aucun effet sur la quantité produite mais elle influence la composition de la demande globale. En effet, si la demande globale retrouve à long terme son niveau initial malgré l'augmentation des dépenses publiques, c'est qu'une autre de ses composantes a diminué.

Comme le revenu est retourné à son niveau initial, la consommation en a fait autant. C'est donc l'investissement qui s'est ajusté. En effet, l'augmentation des prix a réduit l'offre d'encaisses réelles, ce qui a provoqué une augmentation des taux d'intérêt. L'augmentation des prix s'est poursuivie jusqu'à ce que la demande globale retrouve son niveau initial, ce qui implique que le taux d'intérêt ait augmenté suffisamment pour que la diminution de l'investissement corresponde à l'augmentation des dépenses publiques. Ainsi, **à long terme, l'augmentation des dépenses publiques évince totalement l'investissement.**⁹⁸

Le même raisonnement peut être tenu quelle que soit la théorie sous-jacente de l'offre agrégée. Dans la suite du cours, nous ne commenterons donc pas systématiquement les quatre théories que nous avons passées en revue, parce que les mécanismes d'ajustement de l'offre et de la demande agrégée seront toujours les mêmes. Nous utiliserons prioritairement le modèle à salaire rigides. Cela ne doit cependant pas nous empêcher d'avoir à l'esprit que différents mécanismes peuvent être à l'œuvre. Vous pourrez les envisager à titre d'exercice.

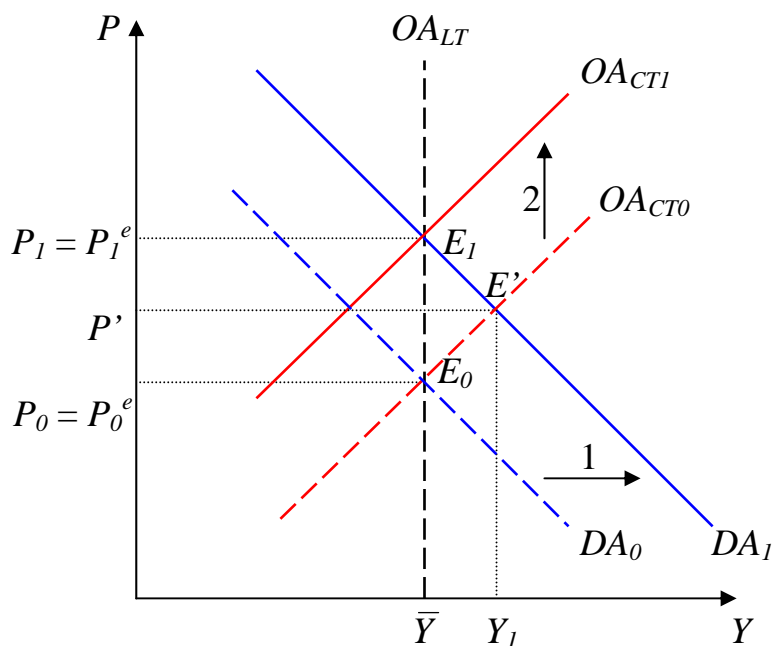
II- La politique monétaire

La politique monétaire est une politique de gestion de la demande agrégée au même titre que la politique budgétaire. Ses effets dans le modèle DA-OA sont donc très proches de ceux de la politique budgétaire.

⁹⁸ Cette conclusion ne concerne que la capacité de la politique budgétaire à servir d'instrument de politique conjoncturelle. Elle est due au fait que nous ne considérons les dépenses publiques que comme une composante de la demande. Evidemment, les dépenses publiques peuvent également constituer un investissement, par exemple dans des infrastructures. Dans ce cas, elles peuvent permettre d'augmenter les capacités de production de l'économie donc le revenu structurel.

Comme une politique budgétaire expansionniste, une politique monétaire expansionniste déplace la demande agrégée vers la droite. Ce résultat nous permet d'étudier graphiquement ses effets à court et long terme.

Graphique 11 : une politique monétaire expansionniste



Ainsi, DA_0 est remplacée par DA_1 . L'effet instantané en est de modifier l'équilibre de l'économie qui passe du point E_0 au point E' . Le niveau des prix et le revenu augmentent en même temps.

Cependant, au point E' , les anticipations sont erronées ($P' > P_0^e$). Le point E' ne peut donc pas représenter un équilibre de long terme. Les agents vont donc réviser leurs anticipations à la hausse, ce qui va déplacer vers le haut la courbe d'offre agrégée de court terme. Les prix vont alors augmenter et le revenu se rapprocher de sa valeur structurelle.

In fine, l'économie aura convergé vers un nouvel équilibre caractérisé par un niveau des prix plus élevé que son niveau initial et un revenu égal à sa valeur de long terme. Cet équilibre correspond au point E_1 .

Si la représentation graphique et les effets globaux de la politique monétaire sont comparables à ceux de la politique budgétaire dans le modèle DA-OA, les mécanismes en œuvre sont légèrement différents.

La situation de départ est la même. L'économie est en situation d'équilibre de court terme et long terme. La demande correspond à la production structurelle. Les anticipations sont validées et le marché du travail est en équilibre.

A court terme, l'augmentation de l'offre de monnaie provoque une offre excédentaire d'encaisses réelles qui est résorbée grâce à une diminution du taux d'intérêt. La diminution du taux d'intérêt stimule l'investissement, ce qui augmente la demande agrégée grâce à un mécanisme multiplicateur. La production doit donc augmenter pour satisfaire la demande.

Quel que soit le modèle de l'offre agrégée qu'on utilise, l'augmentation de la demande sera compensée par une augmentation de la production et des prix. Par exemple, dans le modèle à salaire rigide, l'augmentation de la demande induit une augmentation des prix, qui réduit le salaire réel, ce qui incite les entreprises à embaucher pour augmenter leur production. Les travailleurs peuvent alors revendiquer des salaires plus élevés, ce qui augmente les prix.

On passe ainsi du point E_0 au point E' . On peut donc conclure que **la politique monétaire est efficace à court terme mais qu'elle produit une augmentation des prix.**

Cet équilibre n'est soutenable qu'à court terme puisqu'il suppose que les anticipations sont erronées. Les agents doivent donc réviser leurs anticipations à la hausse. Cela va provoquer un déplacement vers le haut de la courbe d'offre agrégée de court terme tant que l'équilibre de long terme ne sera pas atteint.

Ainsi, dans le modèle à salaire rigide, les travailleurs révisent à la hausse le salaire nominal négocié. Le salaire réel augmente, ce qui incite les entreprises à débaucher et réduire leur production.

Dans tous les cas, le niveau des prix augmente. Cela réduit les encaisses réelles donc la demande de biens. L'économie se déplace alors le long de la courbe de demande agrégée.

Lorsque l'équilibre de long terme est atteint, la demande et la production ont retrouvé leur niveau initial. Seul le niveau des prix a augmenté. On peut donc conclure que **la politique monétaire est inefficace à long terme.**

Cela signifie-t-il que la politique monétaire a exactement les mêmes effets que la politique budgétaire à long terme ? En fait, au contraire de la politique budgétaire, la politique monétaire n'a non seulement aucun effet sur quantité produite mais n'influence pas non plus la composition de la demande globale.

En effet, les dépenses publiques sont restées constantes. La consommation n'a pas non plus varié puisque le revenu est retourné à son niveau initial. Puisque ni la demande globale, ni la consommation, ni les dépenses publiques n'ont été modifiées, on peut conclure que la dernière composante de la demande globale, l'investissement, n'a pas varié non plus.

L'ajustement a donc été réalisé entièrement par l'augmentation des prix. Ainsi, au fur et à mesure que les agents ont révisé leurs anticipations, les prix ont augmenté, ce qui a réduit l'offre d'encaisses réelles. Le taux d'intérêt a donc augmenté, ce qui a réduit l'investissement et la demande globale. Ce mécanisme s'est donc poursuivi jusqu'à ce que l'offre d'encaisses réelles ait retrouvé son niveau initial. Le taux d'intérêt a alors retrouvé son niveau initial, de même que l'investissement et la demande agrégée.

On peut donc conclure qu'à long terme le seul effet d'une augmentation de la masse monétaire est une augmentation proportionnelle du niveau des prix. A long terme, l'évolution de la masse monétaire n'affecte que le niveau des prix. **Il existe donc à long terme une dichotomie entre les variables réelles et nominales.**

III- Que faire face à un choc de demande ?

Si les politiques de demande sont inefficaces à long terme, cela signifie-t-il qu'elles sont inutiles ? La réponse est négative. Si elles ne peuvent influencer le revenu de façon systématique, elles ont un rôle conjoncturel à jouer.

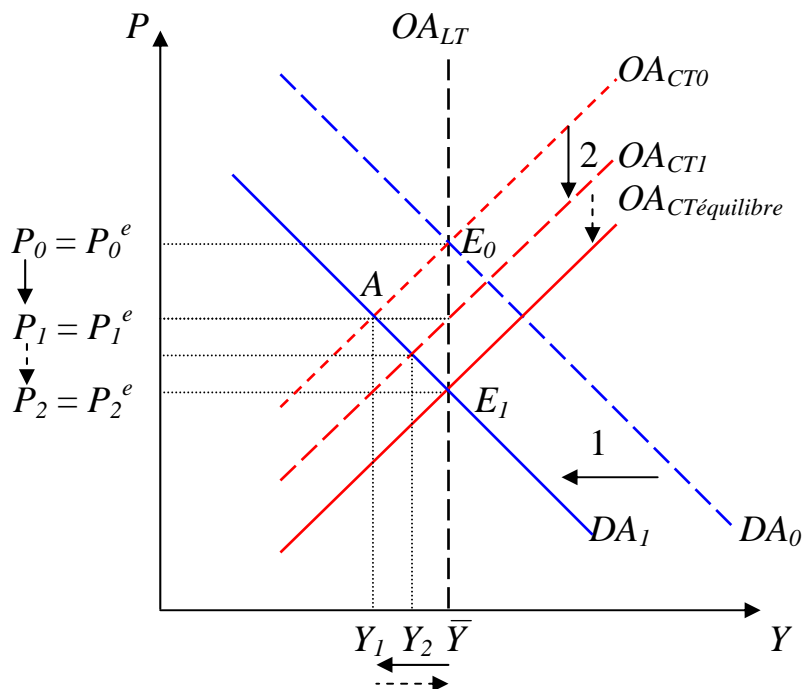
Pour s'en convaincre, supposons que l'économie soit touchée par une crise de confiance des entreprises qui réduisent leurs investissements. Il s'agit là d'*un choc de demande* négatif.

Confrontées à un tel choc, les autorités peuvent choisir entre deux réactions. Elles peuvent d'une part laisser l'économie s'ajuster spontanément. Elles peuvent d'autre part décider de compenser le choc en ajustant leur politique économique. Pour saisir les conséquences de ces deux scénarios, décrivons ce qui se passerait en l'absence de réaction des autorités et ce qui se passerait avec une politique économique judicieuse, en supposant que l'économie était initialement à l'équilibre.

Le choc sur l'investissement se traduit par une diminution de la demande de biens, qui réduit le revenu en étant amplifié par un mécanisme multiplicateur. Il déplace donc la courbe de demande agrégée vers la gauche.

A court terme, le graphique 12 montre que le choc se traduit par une diminution à la fois des prix et du revenu. L'économie passe du point E_0 au point A . En d'autres termes, l'économie entre en récession ($Y_1 < \bar{Y}$).

Graphique 12 : un choc de demande négatif



Comme les prix ont baissé par surprise, le niveau anticipé des prix surestime leur niveau réalisé. Les agents révisent alors à la baisse leurs anticipations, mais le processus est lent. Puisque le niveau anticipé des prix diminue, la courbe d'offre agrégée de court terme se déplace vers le bas. Par conséquent, l'écart entre les niveaux anticipé et réalisé des prix diminue et le revenu d'équilibre se rapproche de son niveau structurel.

A long terme, le revenu retrouve sa valeur initiale, qui correspond à son niveau structurel. On peut donc considérer que l'économie absorbe spontanément le choc de demande, grâce à une baisse des prix.

En l'absence de réaction des autorités, nous pouvons conclure qu'un choc de demande va provoquer à court terme une baisse des prix et une contraction du revenu. A long terme, le même choc ne provoque qu'une baisse des prix.

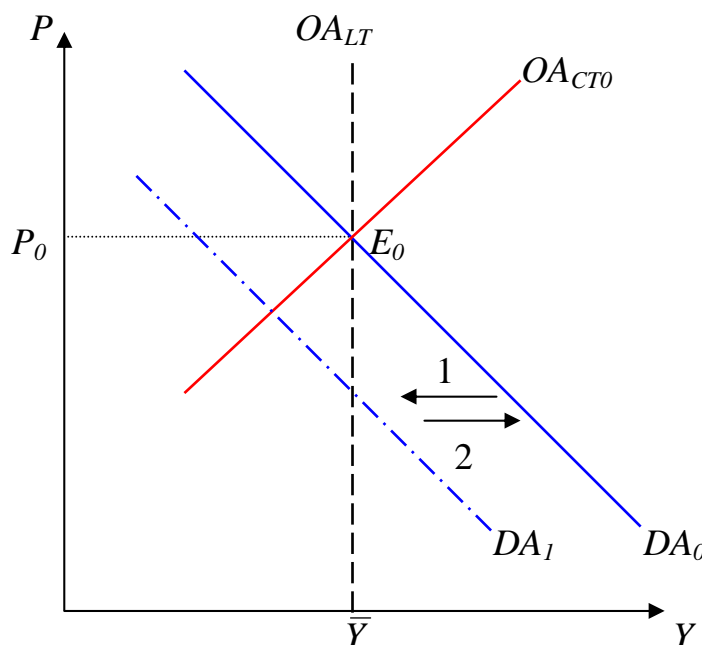
Supposons à présent que les autorités décident de réagir au choc. Puisque le choc a réduit la demande, elles ont intérêt à le compenser en pratiquant une politique expansionniste. Cette politique peut prendre la forme d'une politique budgétaire expansionniste, d'une politique monétaire expansionniste, ou encore d'une combinaison des deux.

Dans le premier cas, l'augmentation des dépenses publiques doit compenser la baisse de l'investissement. Dans le deuxième, la politique monétaire doit faire suffisamment baisser les taux d'intérêt pour que l'investissement retrouve son niveau initial. Dans le troisième, il

faut que la somme de l'augmentation des dépenses publiques et de l'investissement corresponde à la diminution initiale de la demande. Dans tous les cas, la courbe de demande va retrouver sa place initiale.

Par conséquent, si les autorités réagissent de façon suffisamment rapide, elles peuvent absorber entièrement le choc de demande. Les prix et la production restent ainsi à leur niveau initial.

Graphique 13 : un choc de demande négatif absorbé par la relance de la demande



Comme dans le cas précédent, la production retourne à son niveau initial. La grande différence est que ce retour est immédiat.⁹⁹ La politique de relance épargne ainsi à l'économie une douloureuse période de récession. Par conséquent, **lorsque la politique économique l'absorbe, un choc négatif de demande n'a d'effet ni sur le revenu ni sur les prix.**

On peut donc conclure que si les politiques de gestion de la demande ne permettent pas d'augmenter durablement le revenu, elles permettent d'en amortir les fluctuations. On peut ainsi justifier les politiques de *réglage fin* (ou « fine tuning ») qui consistent à ajuster en permanence les dépenses publiques et la politique monétaire afin de stabiliser le revenu.

⁹⁹ Il s'agit ici d'un résultat extrême qui suppose que la réaction des autorités au choc est immédiate et parfaitement bien calibrée. Pour être plus réaliste, il faudrait dire que l'économie retourne à l'équilibre plus rapidement qu'en l'absence de réaction des autorités.

Dans la mesure où les politiques macroéconomiques peuvent être lourdes à manipuler, on doit se demander si les autorités peuvent réagir à temps. La question qui se pose alors est de déterminer à la fois la vitesse de réaction des autorités, le temps qu'il faut aux politiques pour produire leurs effets, et, surtout, le temps qu'il faudrait à l'économie pour absorber spontanément le choc.

En d'autres termes, il s'agit de déterminer à quoi correspondent le long et le court terme. Malheureusement, cette question reste l'une des plus controversées de la macroéconomie.

Par ailleurs, la vitesse d'ajustement de l'économie dépendra dans une très large mesure de la façon dont les agents ajustent leurs anticipations. On voit pourquoi il est fondamental de bien comprendre comment ils les forment.

IV- Que faire face à un choc d'offre ?

La demande agrégée n'est pas le seul élément du modèle DA-OA à pouvoir être affecté par des chocs. Il existe également des chocs d'offre. On peut définir ces chocs comme l'ensemble des phénomènes qui affectent la position de la courbe d'offre agrégée mais cette définition reste quelque peu tautologique.

En revanche, elle permet de réfléchir à la nature des chocs d'offre. Comme la courbe d'offre agrégée est construite à partir de l'équilibre sur le marché du travail, tout événement qui affecte cet équilibre peut être à l'origine d'un choc d'offre. Ainsi toute modification dans l'organisation des négociations salariales, comme une réglementation des syndicats ou du temps de travail, ou tout phénomène susceptible d'affecter la productivité du travail, donc la demande de travail des entreprises, sera considéré comme un choc d'offre.

On peut remarquer que certains chocs d'offre sont transitoires alors que d'autres sont durables. Ainsi, une modification de la réglementation sur le temps de travail produira ses effets tant qu'elle n'aura pas été abolie. En revanche, les effets d'une mauvaise récolte ne dureront qu'une année, voire moins si plusieurs récoltes sont possibles au cours de la même année.

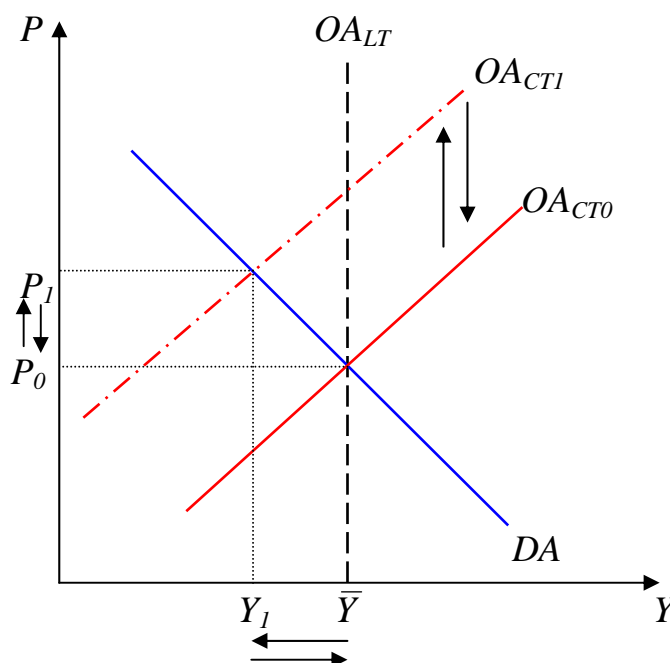
Cette distinction, qui n'était pas nécessaire dans le cas des chocs de demande, est essentielle pour les chocs d'offre. En effet, les chocs temporaires n'affecteront que la courbe d'offre agrégée de court terme alors que les chocs permanents affecteront également la production structurelle. La réaction de l'économie à ces deux types de chocs sera donc très différente.

Pour gagner du temps, nous n'allons nous intéresser ici qu'à un choc d'offre temporaire et considérer les effets d'une augmentation du prix du pétrole provoquée par une tension politique dans les pays producteurs.¹⁰⁰ Nous adoptons ici le point de vue des pays importateurs pour qui ce choc est négatif alors qu'il est bien entendu favorable pour les pays producteurs.

Pour analyser les effets d'un tel choc, nous devons commencer par nous demander comment ils peuvent être représentés dans notre modèle. Le pétrole est avant tout une matière première, c'est-à-dire une consommation intermédiaire. De façon équivalente, on peut donc affirmer qu'ils affectent la productivité marginale du travail ou le coût marginal de production des biens.

Dans le modèle à salaire rigide, qui reste notre modèle de référence, cela peut être intégré en supposant que la productivité marginale du travail diminue. On observe donc une diminution de l'emploi, ce qui implique une diminution de la production quel que soit le niveau des prix. On peut donc conclure que l'augmentation du prix du pétrole va déplacer la courbe d'offre vers le haut et la gauche.

Graphique 14 : Un choc d'offre non traité



¹⁰⁰ La première guerre du golfe en 1991 est un exemple de ce type de choc. Après une augmentation dans les mois précédant et suivant le conflit, le prix du pétrole a retrouvé son niveau initial. Les deux chocs pétroliers des années 70 ont en revanche eu des effets prolongés et peuvent être considérés comme des chocs durables. Vous trouverez une analyse de leurs effets dans le chapitre 8 du manuel d'O. Blanchard et D. Cohen.

Les notes du cours ne sont transmises qu'à titre indicatif.

Comme l'augmentation du prix du pétrole n'est que temporaire, le déplacement de la courbe d'offre agrégée l'est aussi. Cette dernière retrouvera donc sa position initiale lorsque le pétrole aura retrouvé sa valeur initiale.

On constate grâce au graphique 14 que le choc sur le prix du pétrole provoque une récession temporaire, puisque le revenu passe sous sa valeur structurelle ($Y_1 < \bar{Y}$). Cet effet d'un choc d'offre est comparable à celui d'un choc de demande négatif.

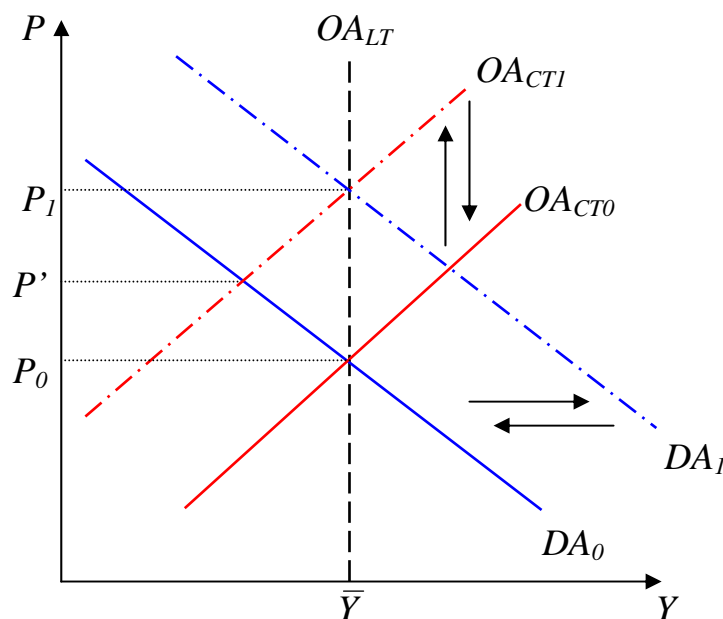
Une différence importante apparaît cependant puisque le choc d'offre est associé à une augmentation des prix, alors qu'un choc de demande est accompagné d'une diminution des prix. **En l'absence de réaction des autorités, nous pouvons donc conclure qu'un choc d'offre négatif va provoquer à court terme une contraction du revenu et une augmentation des prix.**

L'effet du choc d'offre est cependant temporaire. Lorsque le prix du pétrole retrouve sa valeur initiale, la courbe d'offre agrégée de court terme retourne à sa position initiale. Si les agents n'ont pas eu le temps d'adapter leurs anticipations, l'économie retrouve alors sa position initiale.

La récession n'est donc que temporaire. La question est alors de savoir si on pourrait l'éviter en ajustant à temps les politiques macroéconomiques. La réponse est positive. Il faut pour cela que les autorités pratiquent une politique budgétaire et/ou monétaire expansionniste. Dans tous les cas, le policy-mix doit être tel que la nouvelle courbe de demande agrégée croise la nouvelle courbe d'offre agrégée de court terme sur la courbe d'offre agrégée de long terme. Si la politique est judicieusement conçue, elle permet d'éviter au revenu de passer sous sa valeur structurelle, comme le montre le graphique 15.

Lorsque le choc est passé, c'est-à-dire quand le prix du pétrole a retrouvé sa valeur initiale, la courbe d'offre agrégée retourne à sa position initiale. Si les autorités reprennent au même moment leur politique initiale, le revenu aura constamment conservé sa valeur structurelle. On peut donc affirmer qu'**un policy-mix bien conçu permet d'éviter les effets sur le revenu d'un choc d'offre négatif.**

Graphique 15 : Un choc d'offre compensé par la relance de la demande



Cette absorption du choc d'offre n'est cependant pas gratuite. Elle se fait au prix d'une augmentation des tensions sur les prix. Le graphique 15 montre en effet que le niveau des prix si les autorités absorbent le choc est supérieur à ce qu'il serait en l'absence d'ajustement du policy-mix ($P_1 > P'$). On peut donc conclure que **l'absorption des effets réels d'un choc d'offre amplifie ses effets nominaux**.

Par conséquent, les autorités sont confrontées à un dilemme lorsque leur économie est frappée par un choc d'offre négatif. Elles doivent en effet choisir entre deux maux : l'augmentation des prix ou la diminution du revenu. Il apparaît donc un arbitrage entre croissance et inflation.

Il n'existe pas de stratégie universellement optimale face à un tel arbitrage. La politique retenue dépendra en fait des préférences des autorités. Plus elles seront soucieuses de maintenir l'activité, plus elles accepteront de voir l'inflation augmenter. A l'inverse, si elles présentent une forte aversion pour l'inflation, elles devront accepter une diminution de la croissance.

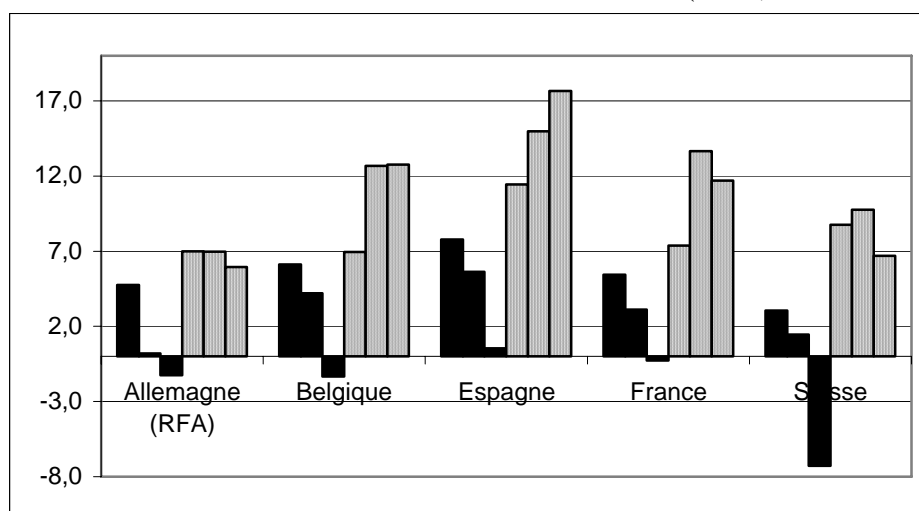
L'exemple du choc pétrolier de 1973 fournit une expérience en grandeur nature qui permet d'illustrer ces conclusions. Bien que ce choc ait été durable, la réaction des autorités

dans les années qui l'ont immédiatement suivi peut s'analyser à partir de notre raisonnement puisqu'il s'agit d'une réaction à court terme.

Du point de vue des pays européens, le choc pétrolier est un cas typique de choc symétrique, c'est-à-dire qui frappe toutes les économies en même temps et avec la même ampleur. Or, face à un même choc certains pays ont choisi de privilégier la croissance et d'autres la stabilité des prix.

Graphique 16

Evolution du taux de croissance et du taux d'inflation (1973, 1974 et 1975)



Croissance en noir et inflation en gris. Source : Banque Mondiale.

Le graphique 16 montre que les taux de croissance ont diminué dans tous les pays entre 1973 et 1975. Cependant cette baisse a été plus marquée dans certains pays que dans d'autres. La Suisse et l'Allemagne (mais aussi la Belgique) ont été particulièrement touchées, surtout si on considère 1975.

L'évolution de l'inflation s'est de même avérée très différente d'un pays à l'autre. Ainsi, certains pays, comme la France, l'Espagne et la Belgique ont laissé filer l'inflation alors que l'Allemagne et la Suisse l'ont maintenue sous contrôle.

Il est frappant de constater que les deux pays qui ont le plus privilégié la stabilité des prix sont aussi ceux qui ont connu les ralentissements les plus marqués de leur taux de croissance. A l'inverse, le seul pays qui n'a pas connu de croissance négative en 1975 est l'Espagne. C'est aussi celui qui a le plus laissé augmenter son inflation. Ces résultats correspondent précisément aux prévisions du modèle DA-OA en présence d'un choc négatif d'offre.

Conclusion

A l'issue de ce chapitre, nous disposons d'une explication cohérente des fluctuations conjoncturelles des prix et de la production. Cette explication repose sur l'idée que l'économie est soumise à différentes perturbations qui vont affecter soit l'offre soit la demande de biens. Il peut s'agir de changements de politique économique, de changements dans les habitudes des ménages ou le moral des entrepreneurs, de modifications des conditions de négociations sur le marché du travail etc. Dans tous les cas, ces chocs peuvent temporairement éloigner l'économie de son équilibre de long terme.

L'économie a tendance à revenir spontanément vers sa situation d'équilibre structurel, mais cela peut prendre du temps. En revanche, les politiques de gestion de la demande permettent d'accélérer cette tendance, à condition d'être bien conçues et suffisamment réactives. C'est particulièrement vrai pour les chocs de demande, qui peuvent être entièrement neutralisés grâce aux politiques budgétaire et monétaire. La gestion des chocs d'offre fait quant à elle apparaître un arbitrage entre stabilisation du revenu et des prix.

Une distinction fondamentale sur laquelle reposent tous ces résultats est celle qui oppose le court et le moyen/long terme. Cette distinction est cruciale, puisque les écarts par rapport à l'équilibre structurel ne peuvent être que transitoires.

Cependant, la définition opérationnelle du long terme reste controversée. On peut ramener cette question à celle de l'efficacité des mécanismes de marché. En effet, les différents modèles de l'offre agrégée que nous avons passés en revue reposent tous sur des imperfections de marché (concurrence imparfaite, information imparfaite, rigidité des salaires et des prix). Plus ces imperfections seront courantes et durables, plus le long terme sera « long ». Au contraire, si ces imperfections ne jouent qu'un rôle mineur, l'économie retournera plus rapidement vers l'équilibre.

Par conséquent, la définition du court et du long terme oppose les macroéconomistes qui font confiance aux mécanismes du marché à ceux qui accordent plus d'attention à ses défauts. Les premiers auront tendance à considérer que les politiques économiques sont trop lentes pour être utiles. Les seconds penseront au contraire qu'elles peuvent faciliter l'ajustement de l'économie. Ils considèrent comme Keynes que « à long terme nous serons tous morts ».

Au cours de la suite de vos études, vous reprendrez ces théories et les raffinerez. Vous améliorerez en particulier la prise en compte des anticipations et décrierez plus finement le

fonctionnement des différents marchés. Ces améliorations permettront de décrire avec toujours plus de réalisme et de détail les mécanismes fondamentaux de l'économie.

Cependant, il est aussi possible d'améliorer le réalisme de nos théories de façon tout aussi radicale sans pour autant devoir redéfinir nos principaux outils. Il s'agit d'abandonner l'hypothèse restrictive d'économie fermée que nous avons respectée jusqu'à présent. C'est précisément l'objet de la dernière partie du cours.

Références

Sur le cours

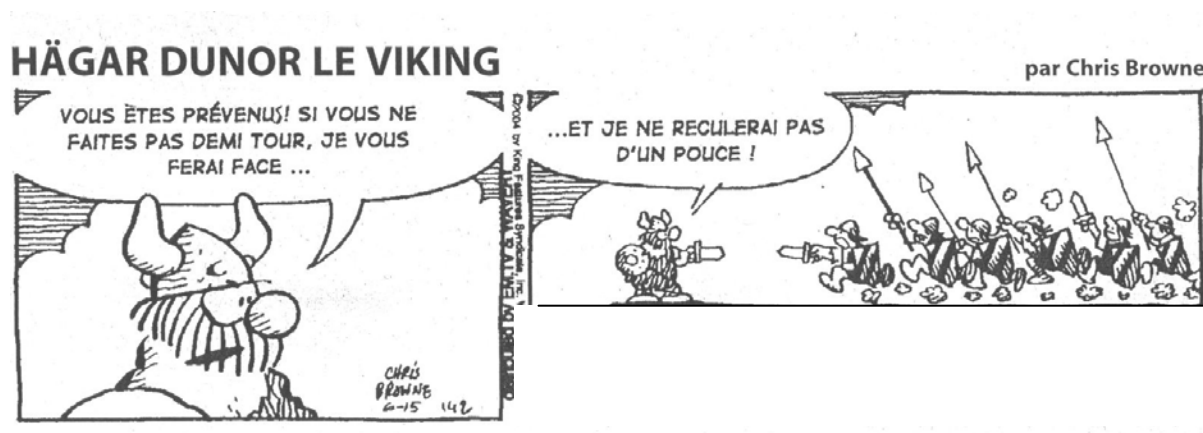
Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2004 : chapitre 8.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 10 et chapitre 11 (section 11.2 uniquement) et 13.

Annexe

Hägar est-il crédible ? (réponse dans l'annexe du chapitre 11)

Lorsque la banque centrale trouve un bénéfice à créer de l'inflation et qu'elle annonce son intention d'appliquer une politique monétaire restrictive, sa situation est comparable à celle d'Hägar. Personne ne la croira, à moins qu'elle ne trouve une astuce institutionnelle.



Les notes de cours ne sont transmises qu'à titre indicatif.

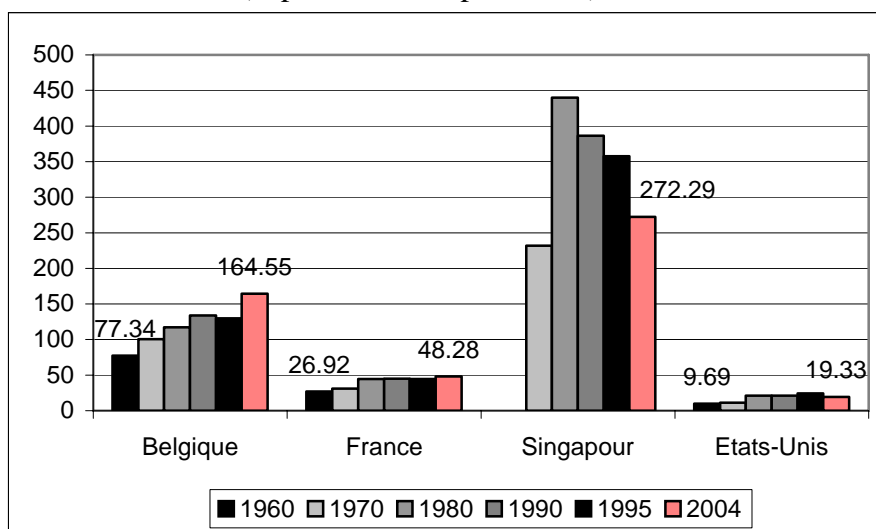
Troisième partie : L'économie ouverte

Introduction : L'ouverture des économies

Jusqu'à présent, nous avons construit toutes les théories que nous avons passées en revue en négligeant toutes les relations de l'économie avec l'extérieur. Cette hypothèse est bien sûr une simplification, aucun pays n'ayant jamais été totalement coupé de l'extérieur. En réalité, la seule économie réellement fermée sur elle-même est l'économie mondiale.

Si on a d'abord développé les théories macroéconomiques dans le cadre d'une économie fermée, c'est peut-être par commodité, mais c'est aussi parce que cette hypothèse n'était pas forcément très restrictive dans les années 30. En effet, le commerce extérieur ne jouait pas le même rôle il y a septante ans que de nos jours. De plus, certaines économies sont plus ouvertes que d'autres.

Graphique 1 : L'évolution de l'ouverture de quelques économies
(exportations+importations)/PIB



Note : Les chiffres pour 2004 représentent le rapport entre la somme des exportations et des importations et le PIB en PPA. Pour la Belgique, le chiffre le plus récent est relatif à l'année 2003.
Source : Banque Mondiale et CIA The World Factbook, 2004 et 2005.

Le graphique confirme ces impressions. Il semble que la part du commerce extérieur ait progressé dans tous les pays depuis les années 1960. C'est aux Etats-Unis que la progression est la plus importante. Même si les échanges représentent une part moins

Les notes du cours ne sont transmises qu'à titre indicatif.

importante du PIB dans cette économie que dans les autres, ceux-ci ne peuvent plus être négligés.

Ce qui frappe également, ce sont les différences qui apparaissent dans les degrés d'ouverture des différentes économies. Si les Etats-Unis n'échangeaient que l'équivalent d'un quart de leur PIB dans les années 1990, la Belgique en échangeaient 130%.¹⁰¹ Le cas de Singapour est encore plus frappant, puisque les échanges de ce pays équivalent à 3,5 fois son PIB.¹⁰²

Les chiffres suggèrent qu'un pays participera d'autant plus aux échanges internationaux que sa taille sera faible. Ce constat est l'un des faits stylisés bien établis du commerce international.

Nous pouvons donc conclure que la prise en compte des échanges avec le reste du monde est une étape nécessaire de notre raisonnement. L'hypothèse d'économie fermée n'est plus acceptable pour les grandes économies et ne l'a probablement jamais été pour les économies plus petites.

Dans ce chapitre, nous allons donc adapter nos théories du court terme afin de tenir compte des échanges avec le reste du monde.¹⁰³ C'est ce à quoi les chapitres 10, 11 et 12 seront consacrés. Auparavant, il faut apprendre à mesurer ces échanges. C'est ce à quoi sert la balance des paiements qui fait l'objet du prochain chapitre.

¹⁰¹ La même année, le ratio s'élevait à 16,96% pour l'Union Européenne à quinze membres. L'évolution de l'ouverture de cette zone est donnée en annexe.

¹⁰² Le fait que les échanges valent plus que le PIB n'est pas anormal. Un pays peut importer puis réexporter des biens qu'il n'a pas produits. C'est en raison de sa fonction d'entrepôt que Singapour montre un ratio commerce sur PIB si élevé.

¹⁰³ Nous ne discuterons pas de la relation entre ouverture et croissance économique à long terme bien qu'il s'agisse d'une des questions les plus anciennes et importantes de la science économique. La raison en est que l'impact de l'ouverture aux échanges internationaux sur la croissance est très difficile à appréhender et reste très controversé. Vous aurez l'occasion d'en discuter dans la suite de vos études.

Chapitre 9 : La comptabilité des transactions avec le reste du monde

Introduction : Un document comptable

On commente abondamment l'évolution des exportations et des importations de biens et de services, ainsi que les mouvements de capitaux. Sans le savoir, ces commentaires utilisent des données qui proviennent d'un même document comptable : la *balance des paiements*.

C'est ce document qui enregistre le bilan des transactions d'un pays avec l'étranger. Il enregistre à la fois les transactions qui portent sur les biens et services mais aussi celles qui portent sur les mouvements de capitaux.

Comme tout bilan, la balance des paiements n'est qu'un tableau qui présente des comptes. De plus, comme elle respecte le principe de la *comptabilité en partie double*, elle est toujours équilibrée. Cela signifie que son solde est toujours nul. Le montant des opérations dans un sens (par exemple en crédit) est toujours le même que celui des opérations dans l'autre sens (en débit). Le solde de la balance des paiements n'apporte par conséquent aucune information.

Si on veut pouvoir interpréter la balance des paiements, il faut y distinguer des *soldes intermédiaires* (section 2), c'est-à-dire regrouper les comptes qui la composent en sous-ensembles cohérents. Auparavant, il faut avoir compris comment on enregistre les transactions (section 1).

Section 1 : Construction de la balance des paiements

Pour construire la balance des paiements, il faut d'abord la définir (I), puis déterminer comment les opérations y sont enregistrées (II).

I- Définition de la balance des paiements

La définition de la balance des paiements est intuitive.

Balance des paiements : document qui recense l'ensemble des opérations économiques qui rassemblent des résidents et des non-résidents, au cours d'une période donnée.

Cette définition appelle plusieurs commentaires. Le premier est que la balance des paiements est définie sur une période donnée. Elle enregistre donc des flux et non des stocks.

Elle ne permet donc pas de déterminer par exemple le stock d'investissements étrangers dans un pays ou le stock des avoirs d'un pays à l'étranger.

Par ailleurs, la définition de la balance des paiements repose sur une distinction essentielle : celle qui oppose les résidents et les non-résidents.

Sont considérés comme *résidents* :

- les personnes physiques qui vivent sur le territoire du pays, quelle que soit leur nationalité (on écarte cependant les fonctionnaires étrangers).
- les entreprises qui exercent leur activité sur le territoire.

Sont considérés comme *non-résidents* :

- les personnes physiques qui habitent en dehors du territoire (exception faite des fonctionnaires nationaux en poste à l'étranger).
- les entreprises qui exercent leur activité en dehors du territoire.

II- L'enregistrement des transactions avec le reste du monde

On l'a vu, la balance des paiements respecte la règle de la comptabilité en partie double. Toute transaction entre une unité résidente et une unité non résidente est donc enregistrée deux fois, dans deux sens différents.

ex : la vente d'un bien d'une valeur de 1000€ payé comptant par un résident à un non-résident sera compté à la fois comme une exportation (+1000€) et comme un afflux de devises (-1000€).

La balance des paiements sera donc toujours équilibrée : son solde est par construction nul.

La grande question est de déterminer quelles opérations doivent être portées au crédit (+) et au débit (-) de la balance des paiements.

Il s'agit purement d'une convention, mais cette convention doit être cohérente. On peut expliquer la convention qui a été retenue en partant de l'observation du patrimoine des résidents :

- toute opération qui diminue le patrimoine des résidents est inscrite au crédit (+) de la balance des paiements.

- toute opération qui augmente le patrimoine des résidents est inscrite au débit (-) de la balance des paiements.

ex : une exportation réduit le patrimoine réel des résidents. Elle est donc portée au crédit de la balance des paiements (+).

Si elle est réglée en liquide, elle augmente les avoirs de devises des résidents. Le mouvement de devises est donc porté au débit de la balance des paiements (-).

Dans tous les cas, si une *opération autonome* (ex : une exportation) est portée au crédit, l'*opération induite* (ex : paiement de l'exportation) sera portée au débit, et vice versa.

Illustrons ces principes simples par un exemple. Une entreprise installée en Belgique exporte pour 1000€ d'acier spécial en Chine. Cet acier est réglé au comptant pour 250€. Le reste fait l'objet d'un crédit commercial de 750€

L'exportation réduit le patrimoine réel de l'entreprise. En effet, son stock d'acier a diminué. L'opération autonome est donc portée au crédit de la balance des paiements belge.

Le paiement au comptant augmente les avoirs de devises de l'entreprise belge. Le crédit commercial augmente le patrimoine financier de l'entreprise, puisqu'elle détient une créance sur sa cliente chinoise. Ces deux opérations sont donc portées au débit de la balance des paiements.

L'opération sera donc enregistrée par les trois écritures suivantes :

Exportations : +1000

Crédit commercial : -750

Avoirs de réserves : -250

On constate que l'opération autonome a bien été enregistrée en sens inverse des deux opérations induites. Le solde de la balance des paiements est donc resté nul.

Section 2 : Interprétation de la balance des paiements

On ne peut pas se contenter d'inscrire toutes les opérations une à une de façon anarchique, sinon la balance des paiements serait ininterprétable. On les regroupe donc en comptes cohérents dont les soldes peuvent être interprétés.

Il s'agit là encore de conventions. Cependant, ces conventions ont pour objet de regrouper les opérations qui sont comparables afin de faciliter l'interprétation de la balance des paiements.

Ces conventions sont parfois révisées. La dernière révision a été instaurée dans les années 90, suite aux directives du FMI. Il s'agissait principalement de proposer de nouveaux regroupements des comptes.

Il faut donc être vigilant lorsqu'on interprète des données concernant l'évolution des balances des paiements sur cette période et s'assurer que les données utilisées ont bien été construites à partir des mêmes conventions.

Cette remarque vaut aussi pour les cours et les manuels. Les manuels trop anciens présentent les anciennes conventions, ce qui pourrait vous déconcerter.

On distingue aujourd'hui *quatre comptes principaux* dans la balance des paiements dont l'ordre traduit le caractère plus ou moins définitif des opérations qui y sont enregistrées. Par convention, les échanges de biens et services sont comptabilisés en haut de la balance des paiements. Les mouvements monétaires tout en bas.

Le solde des quatre principaux comptes n'a aucune raison d'être nul. Il peut donc être interprété. Chaque compte peut de plus être subdivisé en comptes plus spécifiques, dont le solde peut également être interprété. Nous allons passer les comptes principaux en revue et préciser leur interprétation.

1. Compte des transactions courantes
 - 1.1. Biens
 - 1.2. Services
 - 1.3. Revenus (versés par et au reste du monde)
 - 1.4. Transferts courants
2. Compte de capital (brevets, transferts unilatéraux)
3. Compte financier
 - 3.1. Investissements directs
 - 3.2. Investissements de portefeuille
 - 3.3. Autres investissements
 - 3.4. Produits financiers dérivés
 - 3.5. Avoirs de réserve
4. Erreurs et omissions nettes

Le premier compte dont le solde est important est celui de la *balance des opérations courantes* (compte 1).

On y enregistre les importations (–) et les exportations (+) dans le compte 1.1. Le solde de ce compte est aussi appelé la *balance commerciale*.

Le compte suivant enregistre les échanges de services (compte 1.2). Il s'agit par exemple des services de transport, d'assurance, de tourisme.

Le compte 1.3. enregistre les revenus du travail et du capital. Un résident belge peut par exemple détenir des actions brésiliennes dont les dividendes seront comptabilisés au crédit de ce compte (+) dans la balance des paiements belge.

Enfin, le quatrième compte est consacré aux transferts unilatéraux privés (1.4.). C'est dans ce compte que sont par exemple comptabilisés les fonds qu'un travailleur immigré en Belgique pourrait renvoyer dans son pays d'origine (+).¹⁰⁴

Les trois comptes 1.2. à 1.4. constituent ce qu'on appelle la *balance des invisibles*. Les opérations qu'ils enregistrent sont en effet immatérielles, donc invisibles lors de leur passage à la douane.

¹⁰⁴ Cette opération est portée au crédit du compte, ce qui peut être étonnant. Il ne faut pas confondre le don en lui-même et le mouvement de monnaie qu'il induit. Dans notre exemple, le travailleur transfère des devises vers l'étranger. Son patrimoine de devises a donc diminué. On va donc porter l'opération au débit du compte « avoirs de réserves ». Le mouvement de devises est l'opération induite. L'opération autonome que constitue le don doit donc être portée au crédit du compte de transferts. On pourrait dire que son « patrimoine de dons » a augmenté.

Le deuxième compte est baptisé le *compte de capital*. Ce compte a été instauré lors de l'adoption de la nouvelle convention. Il était auparavant inclus dans le compte des opérations courantes.

Il enregistre les transferts entre états et les transactions portant sur les brevets. C'est dans ce compte que sont enregistrés les transferts de et vers le budget de l'Union européenne.

Il ne jouera qu'un rôle secondaire dans nos théories de l'économie ouverte.

Le troisième compte est le compte financier (3.). Il enregistre toutes les opérations financières, ce qui assez vaste. Une typologie utile consiste à isoler le cinquième compte (3.5. avoirs de réserves) et de regrouper les quatre premiers.

Les quatre premiers postes correspondent donc aux flux financiers hors avoirs de réserves. Il s'agit :

- des investissements directs (3.1.) : prises de contrôle d'entreprises à l'étranger, investissement dans la création ou le développement d'une filiale.
- des investissements de portefeuille (3.2.) : actions et obligations.
- des « autres investissements » (3.3.) : prêts, crédits commerciaux.
- des produits financiers dérivés (3.4.) : primes sur options, intérêts sur swaps.

Le dernier compte du compte financier (3.5.) enregistre les opérations sur les avoirs de réserve. Il s'agit là des opérations qui portent sur les devises, l'or monétaire et tous les actifs que les banques centrales peuvent utiliser comme réserve.

Ce compte est essentiel parce qu'il nous renseigne sur l'évolution d'une des composantes de la masse monétaire, comme nous le verrons dans le chapitre suivant.

Le quatrième et dernier compte est probablement le plus étonnant. Il est intitulé « erreurs et omissions ». Il n'aurait pas lieu d'exister si la comptabilité était parfaite, mais beaucoup d'opérations ne sont pas comptabilisées dans la balance des paiements. Il existe en effet deux grandes sources d'erreurs et d'omissions :

- la collecte des données. Certains documents peuvent être égarés, les informations transmises par les entreprises incorrectes etc. Comme il s'agit d'enregistrer toutes les opérations de tout un pays, cela fait beaucoup d'occasions de se tromper.
- la fraude. Les différents trafics (drogue, armes, contrebande etc.) ne sont pas directement enregistrés. En revanche, les mouvements de devises finissent par réapparaître dans le circuit économique légal et être enregistrés.

Pour respecter l'équilibre de la balance des paiements, on a créé le compte « erreurs et omissions » dont la seule fonction est de compenser le déséquilibre des autres comptes.

Il est loin d'être négligeable quantitativement. Ainsi, dans la balance des paiements de la zone euro en 2006, il était supérieur en valeur absolue au solde du compte des opérations courantes (-114,3 milliards d'euros pour les erreurs et omissions et -12,9 milliards d'euros pour le compte des opérations courantes).

Conclusion

La balance des paiements permet de présenter le bilan des flux entre un pays et le reste du monde sur une période donnée. Comme elle est par construction toujours équilibrée, il est nécessaire de distinguer des soldes intermédiaires si on souhaite utiliser les informations qu'elle contient.

En revanche, ces soldes intermédiaires ne permettent que des constats comptables. Ils ne permettent en particulier pas d'expliquer l'évolution de l'économie. Il faut pour cela disposer d'une théorie. C'est à cette théorie que les trois prochains chapitres sont consacrés.

Références

Sur le cours

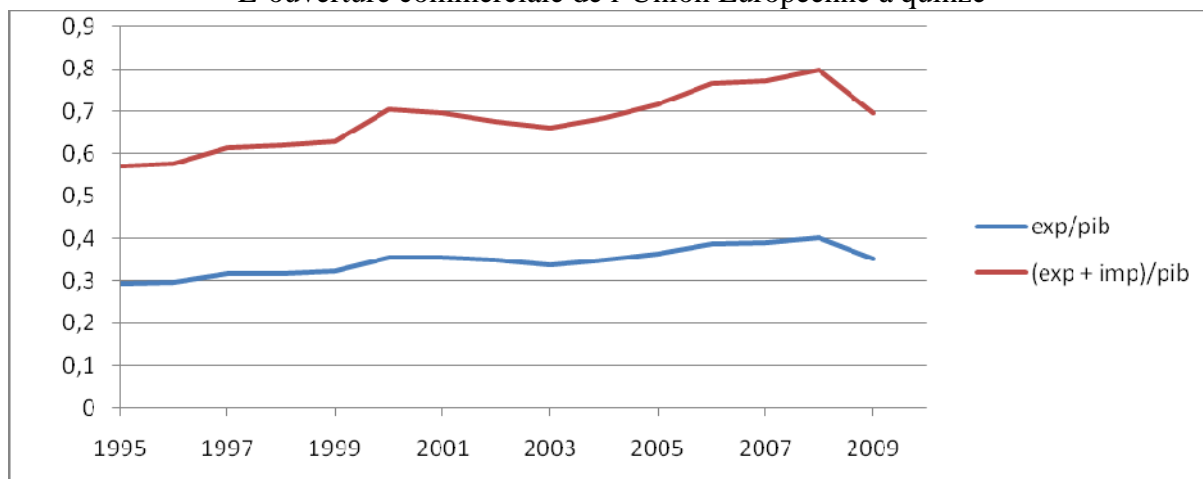
Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 15.

Pour les passionnés

Banque Nationale de Belgique : *L'établissement de la balance des paiements en Belgique*, notice méthodologique de deux pages téléchargeable gratuitement à l'adresse : <http://www.nbb.be/DQ/F/METHOD/M225F.pdf>

Annexe

Graphique A1
L'ouverture commerciale de l'Union Européenne à quinze



Source : Eurostat.

Balance des Paiements de la Belgique en septembre 2009
(en milliards d'euros)

1. Compte des transactions courantes	6,3
1.1. Marchandises	0,41
1.2. Services	3,5
1.3. Revenus	7,1
1.4. Transferts courants	-4,7
2. Compte de capital	0,228
3. Compte financier	-5,0
3.1. Investissements directs	9
3.2. Investissements de portefeuille	8,3
3.3. Produits financiers dérivés	-0,2
3.4. Autres opérations financières	-21,5
3.5. Avoirs de réserve de la BNB	-0,6
4. Erreurs et omissions	-1,5

Source : Belgostat, Banque Nationale de Belgique.

Les notes du cours ne sont transmises qu'à titre indicatif.

Chapitre 10 : Le marché des biens en économie ouverte

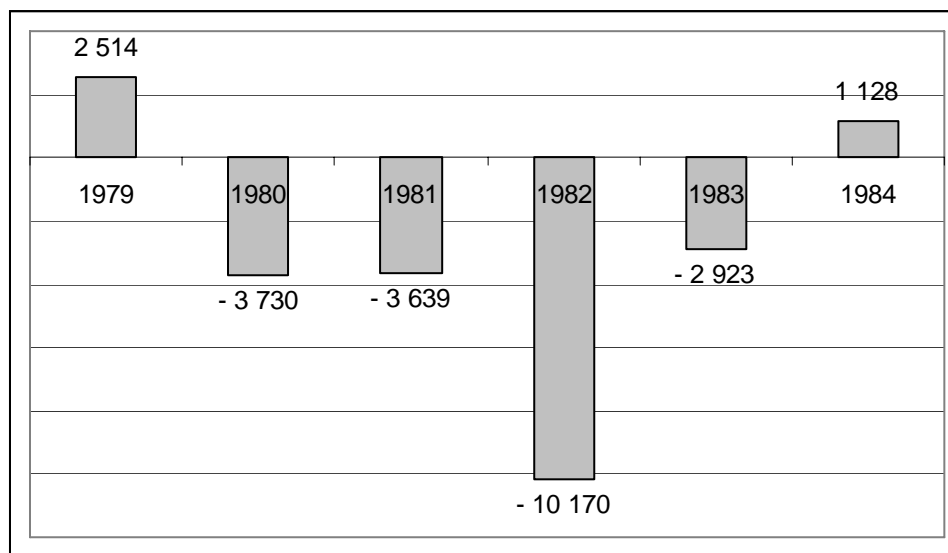
Introduction : La contrainte extérieure

L'expression de *contrainte extérieure* revient régulièrement dans les débats de politique économique. Il faut pourtant reconnaître qu'elle est assez vague. Appliquée au marché des biens, elle traduit la crainte qu'une relance économique se traduise par une dégradation du solde de la balance commerciale.

L'exemple typique de politique économique qui a buté sur la contrainte extérieure est la politique appliquée par le gouvernement français entre 1981 et 1982. Cette politique, censée relancer l'économie grâce à une augmentation de la dépense publique et de la consommation, s'est principalement traduite par une augmentation des importations françaises et un déficit commercial, comme le montre le graphique 1.

Graphique 1

Solde de la balance des échanges de biens et services en France (1979-1984)
(millions d'euros)



Source : Banque de France.

Confronté à une dégradation importante de sa balance commerciale, le gouvernement français a alors décidé de changer de politique. La politique dite de « rigueur » a ainsi succédé à la relance initiale.

Pour comprendre comment une relance budgétaire peut affecter la balance commerciale, il est nécessaire d'adapter la façon dont on représente le fonctionnement du marché des biens. Cela signifie qu'il faut, bien sûr, tenir compte de l'effet de la relance sur les importations et les exportations, mais aussi intégrer l'effet inverse, des échanges internationaux sur l'équilibre du marché des biens.

A cette fin, on doit d'abord décrire les déterminants des exportations nettes, définies comme la différence entre les exportations et les importations (section 1), avant de reprendre la modélisation de l'équilibre sur le marché des biens (section 2).

Section 1 : Les déterminants des exportations nettes

Schématiquement, les exportations nettes sont déterminées par les revenus du pays national et de ses partenaires (I), et par le coût relatif des productions des uns et des autres, qui est déterminé par le taux de change de la monnaie nationale (II).

I- Les revenus

Les importations sont des biens que les agents résidant sur le territoire national se procurent à l'étranger. Il s'agit pour une grande part de biens importés par les consommateurs. Ces derniers ne se rendent d'ailleurs même pas forcément compte qu'ils achètent des biens importés. Plus leur revenu disponible sera élevé, plus ils consommeront de biens importés.

Par ailleurs, si le revenu national augmente, il est probable que les entreprises souhaitent augmenter leurs capacités de production. Comme certains outils de production sont achetés à l'étranger, cela fera également augmenter les importations.

Il est donc raisonnable de considérer que le volume des importations est une fonction croissante du revenu intérieur, Y . En notant M les importations on peut donc écrire :

$$M = M(Y) \quad (1)$$

A l'inverse des importations, les exportations sont des biens produits sur le territoire national et achetés par des agents non résidents. On peut tenir le même raisonnement que pour les importations et considérer que le volume des exportations dépendra du revenu des pays avec lesquels le pays national entretient des relations commerciales.

Nous appellerons ces pays l'étranger et noterons toutes les variables et paramètres qui le concernent par un astérisque. Y^* mesure donc le revenu de l'étranger. On peut alors écrire que les exportations du pays national seront une fonction croissante du revenu de l'étranger :

$$X = X(Y^*) \quad (2)$$

II- Le taux de change

Les biens importés et les biens produits sur le territoire national ne sont pas des biens totalement différents. Ce sont au contraire bien souvent des biens substituables, au moins de façon imparfaite.

Cela signifie que les consommateurs vont tenir compte des prix relatifs de ces biens pour faire leur choix. Si le prix relatif des biens importés diminue, les consommateurs résidant sur le territoire national seront incités à substituer des biens étrangers aux biens nationaux.

ex : Pour un consommateur belge, un verre de bière blonde française ou allemande est un substitut imparfait aux bières blondes brassées en Belgique.

Bien qu'il préfère les bières de son pays aux bières importées, il se peut que le consommateur belge se mette à consommer des bières étrangères si leur prix est suffisamment inférieur à celui des bières belges.

Le volume des importations sera donc non seulement une fonction du revenu des consommateurs nationaux mais aussi du prix relatif des biens nationaux et étrangers.

Afin de pouvoir comparer le prix des biens nationaux et celui des importations, il faut les exprimer dans une unité commune. Il faut donc convertir en monnaie nationale le prix des importations qui est au départ donné en devise.

On utilise pour cela le *taux de change* des deux monnaies, c'est-à-dire leur taux de conversion. Par convention, nous allons mesurer le taux de change de la monnaie étrangère en monnaie nationale en notant e le prix en monnaie nationale d'une unité de monnaie étrangère.

Selon cette convention, une augmentation de e correspond à une *dépréciation*, c'est-à-dire une perte de valeur, de la monnaie nationale. A l'inverse, une diminution de e décrit une *appréciation* de la monnaie nationale, c'est-à-dire une augmentation de sa valeur.¹⁰⁵

¹⁰⁵ Cette façon de mesurer le taux de change est appelée « cotation à l'incertain ». C'est celle qui est utilisée en macroéconomie et sur les marchés financiers parce qu'elle permet d'exprimer tous les taux de change dans une seule unité, l'unité monétaire nationale. Cette convention a aussi l'avantage de bien montrer qu'un taux de change est un prix, le prix d'une monnaie étrangère.

On pourrait aussi utiliser la « cotation au certain » qui consiste à mesurer le prix d'une unité monétaire nationale en unités de devises. Cette convention est plus intuitive puisqu'une appréciation de la monnaie nationale

D'après cette convention, le prix moyen des importations exprimé en monnaie nationale sera donné par le niveau des prix étranger, P^* , multiplié par le taux de change, e : $e.P^*$.

On peut alors déduire le prix relatif des biens étrangers et nationaux en divisant le prix des biens étrangers par celui des biens nationaux. On obtient ainsi le *taux de change réel* de la monnaie nationale.

$$E = \frac{e \cdot P^*}{P} \quad (3)$$

Le taux de change réel est une des variables clefs dans une économie ouverte. En effet, c'est lui qui mesure la compétitivité des productions nationales par rapport aux productions étrangères.

On peut raisonnablement considérer que les importations dépendront du taux de change réel. Plus le prix relatif des biens nationaux sera élevé, plus les consommateurs nationaux auront tendance à préférer leurs substituts importés. On peut donc en déduire que le volume des importations sera non seulement une fonction croissante du revenu national mais aussi une fonction décroissante du taux de change réel :

$$M = M \left(Y, \frac{e \cdot P^*}{P} \right) \quad (4)$$

En effet, si le taux de change se déprécie, que le niveau des prix étrangers diminue ou que les prix nationaux augmentent, le prix relatif des biens étrangers diminue. Comme les produits nationaux sont moins compétitifs, les consommateurs auront tendance à consommer plus de biens étrangers.

Par ailleurs, on peut dire que les exportations du pays national sont les importations du reste du monde. Les exportations du pays national seront donc également une fonction du taux de change réel. Elles seront donc une fonction croissante du taux de change réel :

$$X = X \left(Y^*, \frac{e \cdot P^*}{P} \right) \quad (5)$$

correspond à une augmentation du taux de change et une dépréciation à une diminution du taux de change. Elle est en revanche peu pratique parce que chaque taux de change est exprimé dans une unité différente, l'unité monétaire du pays étranger concerné.

Au total, on peut déduire les déterminants des *exportations nettes*, ou *balance commerciale*. Pour cela, il faut exprimer les importations dans les mêmes unités que les exportations. Si on veut être cohérent, il faut exprimer toutes les quantités en unités de biens nationaux. On doit donc convertir les quantités importées de biens étrangers en leur valeur en biens nationaux. Il suffit pour cela de multiplier le volume des importations par le taux de change réel :¹⁰⁶

$$\begin{aligned}
 NX &= X(Y^*) - E.M(Y) \\
 &\quad \quad \quad - \quad + \quad ? \quad + \\
 NX &= NX\left(Y, Y^*, \frac{e \cdot P^*}{P}\right)
 \end{aligned}
 \tag{6}$$

En résumé, les exportations nettes seront donc :

- une fonction décroissante du revenu intérieur et du niveau des prix national.
- une fonction croissante du revenu étranger et du niveau des prix étranger.

L'effet du taux de change nominal est a priori ambigu. Pour s'en convaincre, supposons qu'une dépréciation se produise, c'est-à-dire que e augmente. Cette dépréciation a deux effets opposés :

- elle augmente la compétitivité des biens nationaux par rapport aux importations, puisqu'elle affecte le taux de change réel. Le volume des importations diminue donc alors que celui des exportations augmente.
- elle augmente la valeur des importations.

Pour que la dépréciation se traduise par une amélioration des exportations nettes, c'est-à-dire que la balance commerciale s'améliore, il faut que l'effet de la dévaluation sur les quantités exportées et échangées l'emporte sur son effet sur la valeur des importations. Il faut pour cela que le volume des exportations et celui des importations soient suffisamment élastiques aux variations du taux de change réel.

Cette condition a été définie et précisée par Alfred Marshall et Abba Lerner. C'est pourquoi on parle de la *condition de Marshall-Lerner*.¹⁰⁷ On utilise également l'expression d'*élasticités critiques*.

¹⁰⁶ C'est parce qu'on raisonne en volume qu'on utilise le taux de change réel. Si on raisonnait en valeur, nous utiliserions le taux de change nominal e .

¹⁰⁷ On exprime fréquemment cette condition en disant que la somme des élasticités-prix des importations et des exportations doit être supérieure à un. Vous aurez l'occasion de démontrer cette condition dans la suite de vos

Les travaux empiriques montrent que cette condition est en général remplie à moyen terme dans la plupart des économies. A l'horizon de quelques mois, une dépréciation permet ainsi d'améliorer le solde de la balance commerciale.

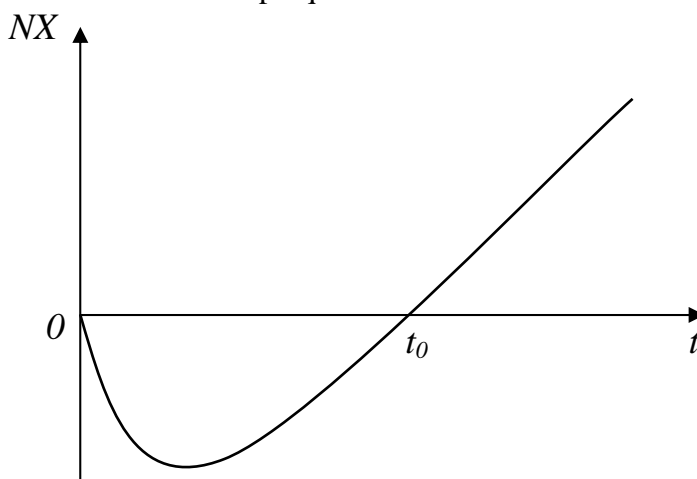
A plus court terme, il est fort probable que la condition ne soit pas respectée. En effet, ajuster les quantités importées et exportées prend du temps. Il faut que les agents réalisent que les prix relatifs ont changé, qu'ils renégocient leurs contrats ou changent de fournisseurs etc. Les quantités importées et exportées seront donc peu élastiques à court terme. La condition de Marshall-Lerner ne sera alors pas vérifiée.

Il est ainsi probable que l'effet de la dépréciation sur la valeur des importations l'emporte à court terme. Paradoxalement, la dépréciation de la monnaie nationale a donc toutes les chances de s'accompagner d'une détérioration de la balance commerciale.

Au fur et à mesure que le temps passe, les quantités vont peu à peu s'ajuster. L'élasticité-prix des quantités importées est plus élevée à long qu'à court terme. La condition de Marshall-Lerner finira alors par être respectée et la balance commerciale par s'améliorer.

Le graphique 2 représente graphiquement l'effet d'une dépréciation de la monnaie nationale sur le solde de la balance commerciale en fonction du temps, en supposant qu'il était nul au départ. Après une phase au cours de laquelle elle se détériore, et son solde devient négatif, la balance commerciale s'améliore et son solde devient positif.

Graphique 2 : La courbe en J



La forme de la courbe évoque un « J » majuscule. C'est pourquoi on parle de *courbe en J*.

études. Une approche analytique simple est en outre proposée dans l'annexe du chapitre 16 du manuel de Blanchard et Cohen (2004, p.357).

Dans la suite de ce chapitre et du cours, nous supposerons que les conditions de Marshall-Lerner sont vérifiées. Cela signifie que nos conclusions ne seront applicables que si nous raisonnons sur un horizon suffisamment long, c'est-à-dire plus long que t_0 sur le graphique.

Dans ces conditions, nous pouvons préciser l'expression 6 :

$$NX = NX \left(Y, Y^*, \frac{e \cdot P^*}{P} \right) \quad (6)$$

Les exportations nettes seront donc une fonction croissante du taux de change tel que nous l'avons défini. Cela signifie qu'une dépréciation de la monnaie nationale aura pour effet d'améliorer la balance commerciale.

Section 2 : L'équilibre du marché des biens en économie ouverte

A présent que nous avons défini les déterminants de la composante de la demande globale que nous avons négligés jusque là, nous pouvons reprendre l'équilibre du marché des biens (I) et en déduire les effets d'une augmentation exogène de la demande (II) et d'une dépréciation de la monnaie nationale (III).

I- Le revenu d'équilibre

La condition d'équilibre du marché des biens est la même, que ce soit en économie fermée ou ouverte. Il faut que la production soit égale à la demande globale. La différence est qu'il faut à présent ajouter les exportations (X) à la demande intérieure de biens nationaux ($C + I + G$) et en soustraire les importations (EM). La condition s'écrit alors :

$$Y = D^G$$

$$\begin{aligned}
 Y &= C(Y) + I(i) + G + X\left(Y^*, \frac{e \cdot P^*}{P}\right) - E M\left(Y, \frac{e \cdot P^*}{P}\right) \\
 Y &= C(Y) + I(i) + G + NX\left(Y, Y^*, \frac{e \cdot P^*}{P}\right)
 \end{aligned} \tag{7}$$

Pour simplifier la suite du raisonnement, nous allons préciser les fonctions de comportement en supposant qu'elles sont linéaires, comme dans le cas de l'économie fermée. Nous supposons donc que les fonctions d'importations et d'exportations s'écrivent respectivement de la façon suivante :

$$X = X_0 + xY^* \tag{8a}$$

$$M = M_0 + mY \tag{8b}$$

$$NX = NX_0 - EmY \quad \text{avec } NX_0 = X_0 + xY^* - EM_0 \tag{8c}$$

Le paramètre m mesure la propension marginale à importer du pays national. Le paramètre x mesure la propension marginale de l'étranger à importer des biens nationaux.

Par commodité, nous n'avons pas intégré explicitement l'effet du taux de change réel sur le volume des importations et des exportations. On peut cependant supposer que NX_0 est une fonction croissante de E .

Par ailleurs, il est raisonnable de supposer que NX_0 est positif. En effet, lorsque le revenu intérieur est nul, les importations résiduelles doivent être très faibles, alors que les exportations restent positives, puisqu'elles ne dépendent que du revenu étranger.

Les autres composantes de la demande globale sont décrites par les mêmes fonctions de comportement que dans le chapitre consacré au marché des biens en économie fermée : $I = I(r_0)$, $G = G_0$ et $C = c \cdot Y + C_0$.

La fonction de demande globale s'écrira donc :

$$\begin{aligned}
 D^G &= c \cdot Y + C_0 + I_0 + G_0 + NX_0 - EmY \\
 D^G &= (c - Em)Y + C_0 + I_0 + G_0 + NX_0
 \end{aligned} \tag{9}$$

Une dernière remarque est nécessaire. Nous traitons Y^* et P^* comme des paramètres constants. Cette hypothèse est réaliste si l'économie nationale correspond à ce qu'on appelle *une petite économie ouverte*. Sur le marché des biens, cette hypothèse se définit comme l'hypothèse que l'évolution de la demande nationale de biens est trop faible pour influencer les prix mondiaux ou le revenu étranger. On considère en particulier que la politique

économique d'une petite économie ouverte aura des effets négligeables sur ses partenaires. Toutes les variables qui décrivent la situation du reste du monde sont donc constantes.¹⁰⁸

Compte tenu de ces hypothèses, il n'y a alors plus qu'à écrire l'expression de l'équilibre du marché des biens. On obtient alors une équation à une inconnue qu'il suffit de résoudre pour obtenir l'expression du revenu d'équilibre :

$$Y = D^G$$

$$Y = (c - Em)Y + C_0 + I_0 + G_0 + NX_0$$

$$Y = \frac{C_0 + I_0 + G_0 + NX_0}{1 - c + Em} \quad (10)$$

Comme en économie fermée, le revenu d'équilibre est ici une fonction croissante de la consommation incompressible, de l'investissement et des dépenses publiques. C'est à présent aussi une fonction croissante des exportations nettes incompressibles et du revenu étranger.

Cette expression montre donc que la situation d'une économie dépend de celle de ses partenaires commerciaux. Elle explique pourquoi il est nécessaire de tenir compte de la croissance mondiale lorsqu'on souhaite prévoir la croissance belge.

L'expression (9) permet par ailleurs de déduire la valeur du multiplicateur des dépenses. Il suffit pour cela de dériver l'expression du revenu d'équilibre par rapport à l'une des composantes de la dépense, par exemple les dépenses publiques G_0 :

$$multi = \frac{1}{1 - c + Em} \quad (10)$$

On peut alors comparer la valeur du multiplicateur en économie ouverte avec celle qu'il avait en économie fermée. Pour mémoire, le multiplicateur était donné par $1 / (1 - c)$.

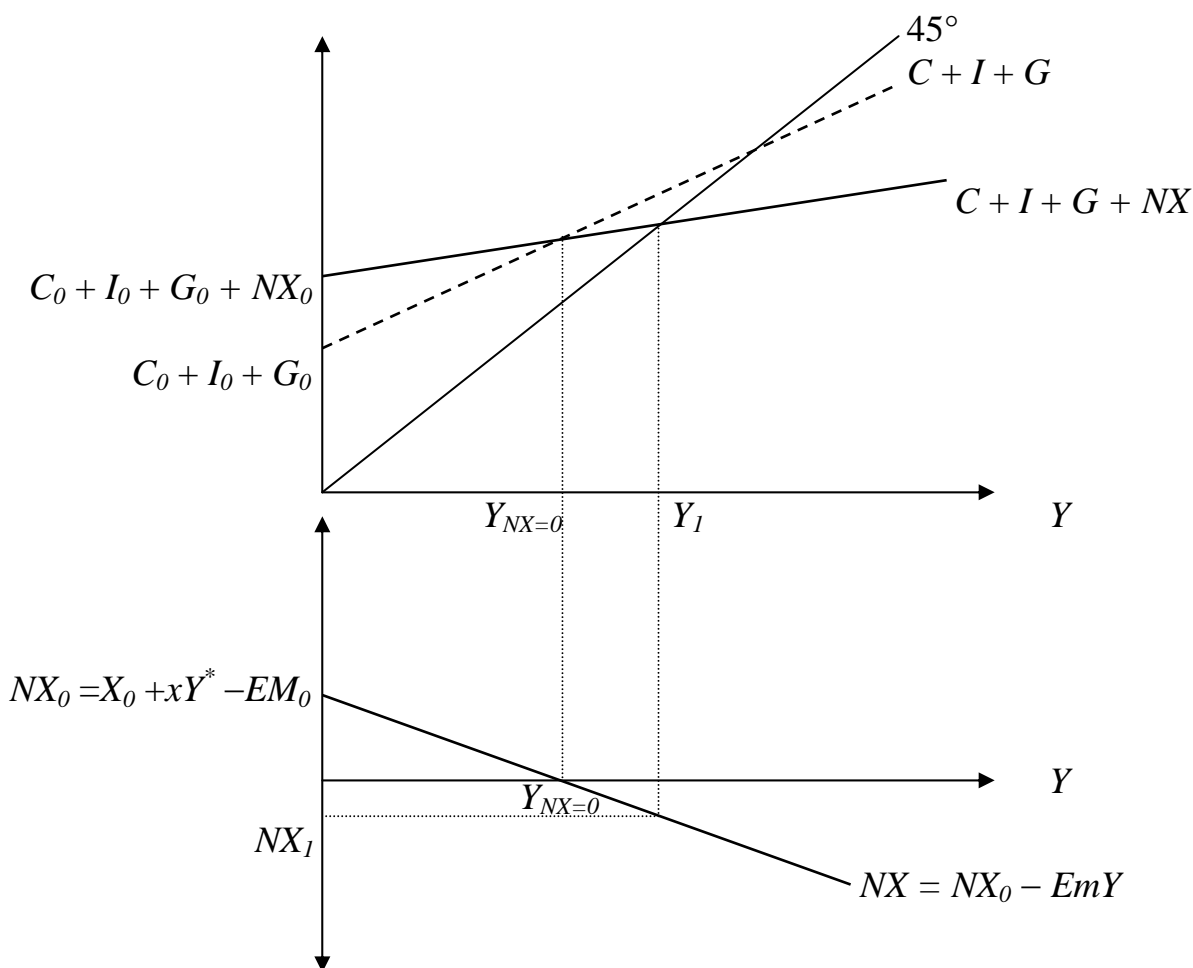
Comme on ajoute le terme Em qui est positif, le dénominateur est plus élevé en économie ouverte qu'en économie fermée. Par conséquent, **la valeur du multiplicateur est inférieure en économie ouverte** à ce qu'elle serait si la même économie était fermée.

¹⁰⁸ Pour prendre deux exemples extrêmes, on peut dire qu'à l'échelle de l'Europe le Luxembourg est une petite économie mais pas l'Allemagne. Les variations de la croissance allemande affectent l'économie de ses partenaires. C'est beaucoup moins vrai pour le Luxembourg. Certaines économies peuvent être petites sur certains marchés mais pas sur d'autres. L'Arabie Saoudite est vraisemblablement un petit pays sur tous les marchés de biens à l'exception de celui du pétrole. Elle pourrait doubler sa demande d'automobiles sans affecter significativement leur prix sur le marché mondial. En revanche, lorsqu'elle modifie sa production de pétrole, le prix mondial du pétrole en est affecté.

Intuitivement, ce résultat vient du fait que lorsque la demande totale de biens augmente, une partie de cette augmentation ne concerne pas les biens nationaux mais les importations. En économie ouverte, il existe donc deux « fuites » qui limitent l’augmentation de la demande globale de biens nationaux. La première existe aussi en économie fermée, c’est l’épargne. La seconde est constituée des importations. Elle est forcément typique de l’économie ouverte.

Pour visualiser graphiquement la demande globale et l’équilibre du marché des biens, on peut réutiliser le diagramme à 45°, à condition de l’adapter à une économie ouverte.

Graphique 3 : Le diagramme à 45° en économie ouverte



Le graphique 3 représente à la fois la demande globale de biens nationaux (en trait plein) et la demande intérieure (en pointillés), qui correspond en fait à la demande globale que nous prenions en compte lorsque nous considérons le cas d’une économie fermée.

Les notes du cours ne sont transmises qu’à titre indicatif.

Ce graphique permet d'observer la différence entre les deux :

- d'abord, la courbe de demande de biens nationaux est moins raide qu'en économie fermée. Cela provient du fait que lorsque le revenu augmente la consommation augmente d'un montant déterminé par la propension marginale à consommer, comme en économie fermée, mais il faut à présent tenir compte du fait que les importations augmentent également.

- ensuite, l'ordonnée à l'origine de la courbe de demande de biens nationaux est supérieure à ce qu'elle était en économie fermée. En effet, il faut ajouter les exportations nettes correspondant à un revenu nul à la consommation intérieure ($C_0 + I_0 + G_0$). Comme les importations seraient très faibles si le revenu était nul, les exportations nettes seraient donc positives.

Le quadrant inférieur du graphique 3 représente les exportations nettes. Il apparaît clairement qu'il s'agit d'une fonction décroissante du revenu. Plus le revenu augmente, plus les importations augmentent, ce qui dégrade le solde de la balance commerciale.

On constate que la valeur du revenu pour laquelle la balance commerciale est équilibrée correspond à celle pour laquelle la demande agrégée est égale à la demande agrégée intérieure. En effet, puisque les exportations nettes sont nulles, la quantité demandée totale de biens nationaux ne peut être différente de celle qui est demandée par les nationaux.¹⁰⁹

Comme dans le diagramme à 45° en économie fermée, le revenu d'équilibre est déterminé par l'intersection entre la courbe de demande agrégée et la première bissectrice, qui représente l'offre intérieure de biens.

On peut remarquer que le revenu d'équilibre sur le marché des biens ne correspond pas forcément au revenu qui équilibre la balance commerciale. Il n'y a en effet aucune raison à cela. Dans le graphique 3, l'équilibre sur le marché des biens est associé à des exportations nettes négatives. Cela signifie que le solde de la balance commerciale est négatif. L'économie importe donc plus qu'elle n'exporte.

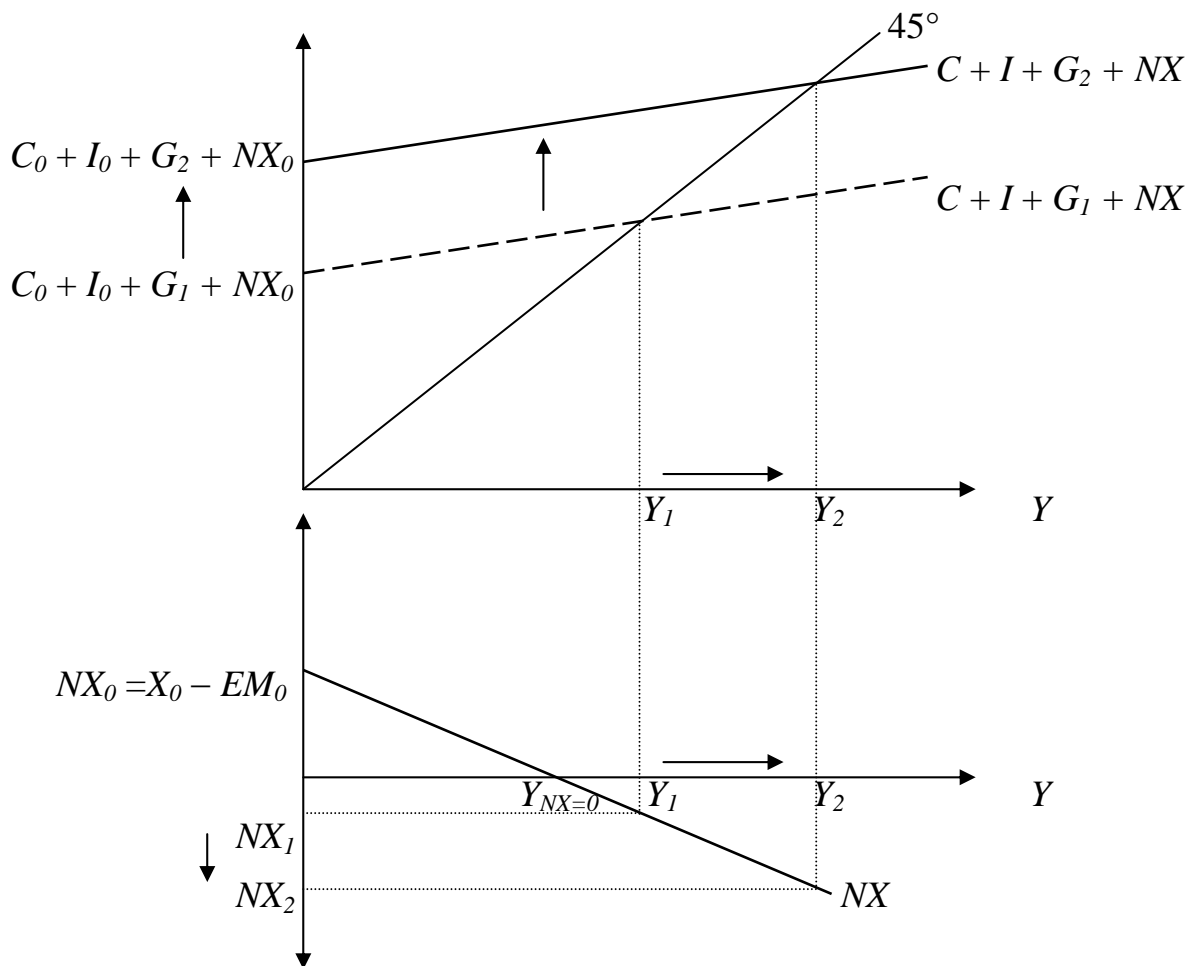
II- L'effet d'une augmentation exogène de la demande globale

L'équilibre du diagramme à 45° nous apprend que l'équilibre sur le marché des biens peut être incompatible avec l'équilibre de la balance commerciale. Il ne nous apprend en revanche pas grand' chose sur l'effet des politiques économiques. Il faut pour cela utiliser le

¹⁰⁹ Cela ne signifie pas que l'économie n'échange pas de biens avec l'extérieur. Cela signifie au contraire qu'elle exporte autant qu'elle importe.

diagramme pour étudier comment une modification des dépenses va affecter le revenu d'équilibre et la balance commerciale.

Graphique 4 : Une augmentation des dépenses

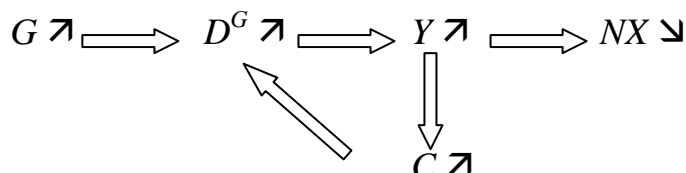


C'est ce que décrit le graphique 4, en simulant une augmentation des dépenses publiques. Une augmentation de l'investissement ou de la consommation produiraient exactement les mêmes effets. Nous parlons ici d'une augmentation des dépenses publiques par commodité, mais n'importe quelle augmentation des dépenses intérieures (consommation et investissement) serait représentée de la même façon.

On constate que l'augmentation des dépenses provoque une augmentation de la production. Le mécanisme à l'œuvre est le même qu'en économie fermée. L'augmentation des dépenses publiques augmente la demande globale, ce qui provoque une augmentation de la production. Comme la production augmente, le revenu augmente, ce qui fait augmenter la

consommation qui augmente à son tour la demande agrégée. Il apparaît donc un cercle vertueux qui aboutit à une augmentation du revenu grâce à un effet multiplicateur.

Il faut cependant garder à l'esprit que ce multiplicateur sera d'autant plus faible que la propension marginale à importer sera importante. En effet, l'augmentation du revenu fera certes augmenter la consommation mais aussi les importations. Or l'augmentation des importations ne contribue pas à faire augmenter la production intérieure.



On peut alors en déduire un deuxième effet de la relance budgétaire en économie ouverte. Celle-ci dégrade la balance commerciale. En effet, puisque le revenu augmente, les importations augmentent aussi. Comme les exportations sont constantes, il s'ensuit une diminution des exportations nettes, soit une dégradation du solde de la balance commerciale.

En définitive, **une augmentation de la demande intérieure relance la demande globale mais dégrade la balance commerciale.**

III- Une dépréciation

Comme le montre l'évolution du taux de change de l'euro par rapport au dollar, le cours d'une monnaie peut fluctuer de façon importante. Cette évolution du cours de la monnaie nationale va affecter la compétitivité de l'économie, ce qui va se traduire sur le marché des biens, via les exportations nettes.

Supposons que la monnaie nationale se déprécie, c'est-à-dire qu'elle perd de la valeur. Comme nous supposons que la condition de Marshall-Lerner est vérifiée, nous pouvons conclure que les exportations nettes augmenteront, quel que soit le revenu initial. Cela signifie que le solde de la balance commerciale sera toujours supérieur à sa valeur antérieure, à revenu constant.

Graphiquement, cela se traduit par un déplacement vers le haut de la courbe NX.

En toute rigueur, un deuxième effet plus subtil va accompagner l'effet précédent. Pour le visualiser, reprenons l'expression des exportations nettes :

$$NX = NX_0 - EmY \quad \text{avec } NX_0 = X_0 + xY^* - EM_0 \quad (8c)$$

Comme la dépréciation se traduit par une augmentation du taux de change réel E , elle va donc modifier la fonction d'exportations nettes de deux façons :

- elle va en faire augmenter l'ordonnée à l'origine (NX_0), comme nous l'avons vu ;
- elle va en modifier la pente (Em). Plus précisément, la valeur absolue de la pente de la courbe NX augmente. La courbe devient plus raide.

Puisque la dépréciation de la monnaie affecte les exportations nettes, elle affecte également la demande globale. Puisque les exportations nettes augmentent, la demande globale de biens nationaux augmente également. La dépréciation de la monnaie nationale va donc avoir un effet de relance sur l'économie.

Lorsqu'on représente graphiquement la fonction de demande globale, on doit donc la déplacer vers le haut. Par ailleurs, sa pente diminue, puisque celle de la courbe d'exportations nettes augmente en valeur absolue, comme le montre l'expression (8c).

Pour tracer un graphique cohérent, il est utile de représenter la demande intérieure et de se rappeler que son intersection avec la courbe de demande globale détermine l'équilibre de la balance commerciale. Suite à la dépréciation, la courbe de demande globale se déplace vers le haut et sa pente se tasse légèrement.¹¹⁰

Cela détermine une nouvelle intersection de la courbe de demande globale avec la courbe de demande intérieure qui indique la valeur du revenu pour laquelle le solde de la balance commerciale sera nul ($Y_{NX=0}$). On peut alors tracer la nouvelle courbe NX , qui coupe l'axe des abscisses au point d'abscisse $Y_{NX=0}$ et en se rappelant que sa pente s'accroît légèrement.

L'effet d'une dépréciation est représenté dans le graphique 5, en supposant que l'économie se trouve au départ en situation d'équilibre de sa balance commerciale.

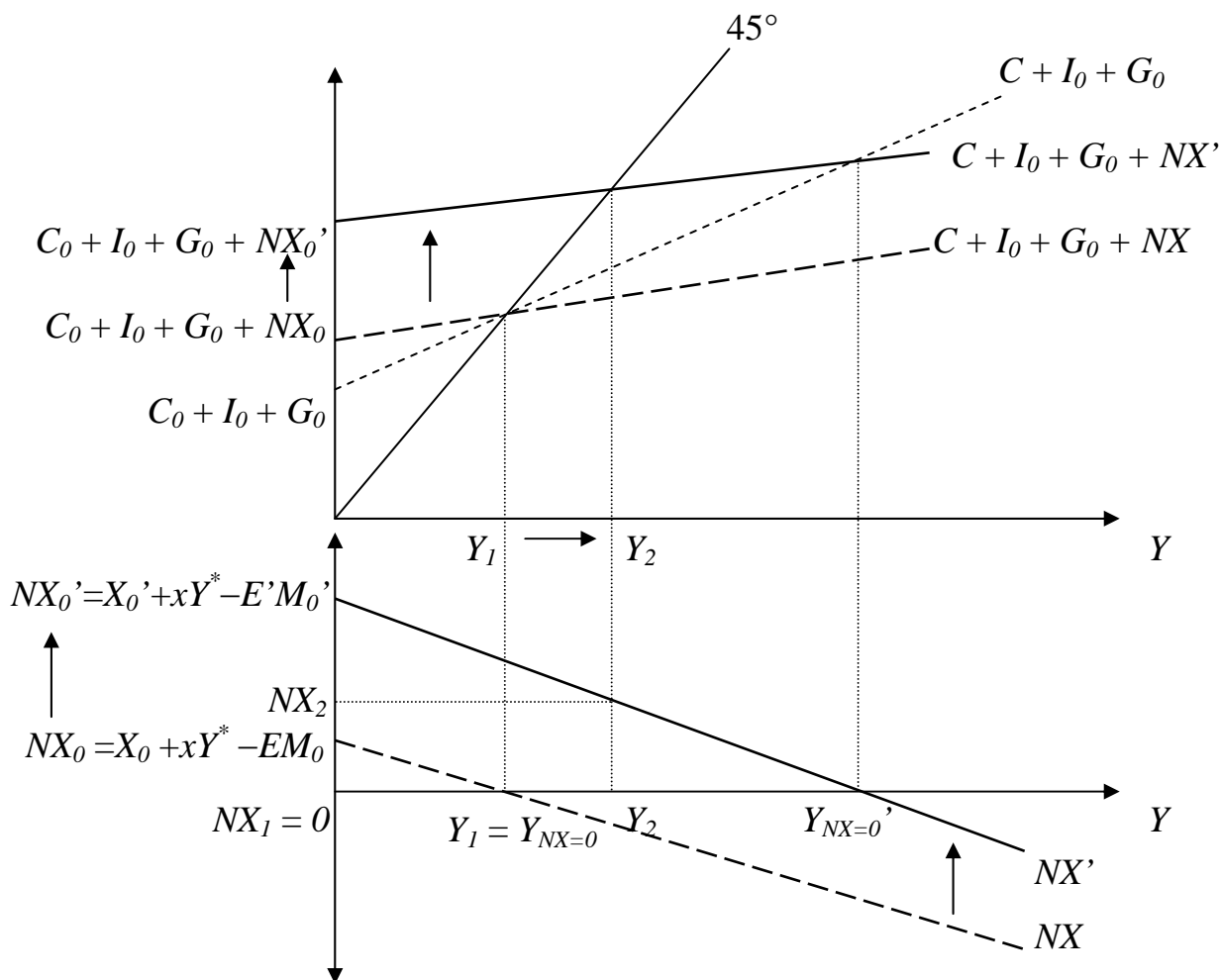
Une fois que les nouvelles courbes de demande agrégée et d'exportations nettes sont tracées, il n'y a plus qu'à lire le graphique.

L'intersection entre la nouvelle courbe de demande globale et la droite à 45° détermine le nouveau revenu d'équilibre. On constate alors que le revenu a augmenté ($Y_2 > Y_1$). La dépréciation de la monnaie nationale a donc permis de relancer la demande et l'économie.

¹¹⁰ L'effet de la dépréciation sur la pente de la courbe est secondaire. S'il vous pose un problème, vous pouvez le négliger en première analyse.

Comme la relance passe par la relance des exportations nettes, il est intéressant de déterminer son effet sur la balance commerciale. Il suffit pour cela de projeter la valeur du nouveau revenu sur la courbe NX et de comparer les nouvelles et les anciennes exportations nettes.

Graphique 5 : Une dépréciation de la monnaie nationale



On constate alors que la balance commerciale s'est améliorée ($NX_2 > NX_1$), mais que cette amélioration est moindre que celle qu'on aurait pu attendre en ne considérant que le déplacement de la courbe NX.

En effet, l'impact direct de la dépréciation sur la balance commerciale est en partie compensé par l'augmentation du revenu, qui fait augmenter les importations. Malgré tout, on peut retenir qu'une **dépréciation de la monnaie nationale relance la demande globale**. Ce résultat nous sera utile par la suite.

On peut remarquer qu'une augmentation du revenu étranger Y^* aurait un effet très proche de celui d'une dépréciation. Il se traduirait graphiquement de la même façon, à la différence près que la pente des courbes ne serait pas affectée.

Conclusion

Lorsqu'on s'intéresse à une économie ouverte, il est nécessaire de prendre en compte l'évolution de ses exportations nettes si on veut décrire correctement le fonctionnement du marché des biens.

La demande globale devient alors une fonction non seulement de la consommation, de l'investissement et des dépenses publiques, mais aussi du revenu des partenaires commerciaux de l'économie et du taux de change réel.

On constate alors que la relance budgétaire reste efficace mais que le multiplicateur est moins grand qu'en économie fermée. De plus, elle dégrade la balance commerciale. Une dépréciation de la monnaie nationale et une augmentation du revenu des partenaires relanceront, elles aussi, la demande globale. En revanche, elles ne dégraderont pas mais amélioreront la balance commerciale.

Ces résultats s'avéreront utiles pour étudier l'interaction du marché des biens avec celui de la monnaie.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 16.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 5 (sections 5.1 et 5.3 uniquement) et chapitre 12.

Chapitre 11 : Le marché des capitaux et de la monnaie dans une économie ouverte

Introduction : Presque autant de monnaies que de pays

La plupart des pays émettent leur propre monnaie. Il s'agit là d'un symbole de souveraineté nationale et c'est pourquoi la construction ou l'indépendance d'un état est presque toujours accompagnée de l'émission d'une nouvelle monnaie.

Il existe certes des exceptions. En effet, certains états ont adopté une monnaie étrangère, comme l'ont fait l'Equateur ou Panama avec le dollar américain. Il est aussi possible que plusieurs pays se rassemblent pour gérer une monnaie unique comme c'est le cas dans la zone euro. Il existe, à l'inverse, des cas où plusieurs monnaies circulent au sein d'un même état. A l'heure actuelle, Hong-Kong fait partie de la République Populaire de Chine mais le dollar hongkongais est une monnaie distincte du yuan.

Le résultat est qu'il existe dans le monde à peu près deux cents monnaies différentes en circulation.¹¹¹ Cela signifie qu'il est très probable qu'une transaction internationale nécessite de convertir deux monnaies entre elles et de déterminer leur taux de change. Cette nécessité de convertir les monnaies entre elles implique que la description de la gestion de la monnaie nationale que nous avons donnée dans le chapitre 5 doit être complétée pour s'adapter au cas d'une économie ouverte (section 2).

Par ailleurs, le marché des capitaux est lui aussi ouvert aux échanges internationaux. La circulation internationale des capitaux va se traduire par une relation entre le taux d'intérêt et le taux de change, qu'on peut déterminer indépendamment de la façon dont la monnaie nationale est gérée (section 1).

¹¹¹ Chacune est désignée par un code de trois lettres. En général, les deux premières lettres du code désignent le pays émetteur et la troisième l'initiale de la monnaie en question. Le code du dollar américain est ainsi USD. Le code de l'euro est EUR. Cette convention est définie par la norme ISO 4217 de l'organisation internationale des standards. Certaines monnaies sont également désignées par un symbole, comme \$ pour le dollar américain, € pour l'euro ou ¥ pour le yen japonais. Ces symboles ne font l'objet d'aucune convention internationale.

Section 1 : Les mouvements internationaux de capitaux

Dans le chapitre consacré au modèle IS-LM, nous avons vu que les agents devaient choisir la façon dont ils détiennent leurs encaisses. Dans une économie fermée l'alternative consistait à choisir entre la monnaie et les titres.

Dans une économie ouverte, le choix est plus large. Chaque agent doit choisir entre la monnaie nationale, les titres nationaux, les monnaies étrangères, ou devises, et les titres étrangers. Même dans le monde simplifié d'un modèle où n'existe que deux pays, le pays national et l'étranger, ce choix reste compliqué.

A part les voyageurs et les trafiquants en tout genre, qui se doivent de détenir des devises pour leurs transactions, la plupart des agents n'y ont aucun intérêt. Quitte à détenir des encaisses à l'étranger, ils n'ont aucune raison de préférer la monnaie qui ne rapporte rien aux titres qui rapportent un intérêt.

En gros, le choix se résume donc à un choix entre monnaie nationale, titres nationaux et titres étrangers. On peut même décomposer la décision en deux étapes. Les agents doivent d'abord déterminer la part de leurs encaisses qu'ils souhaitent conserver sous forme liquide et celle qu'ils désirent placer. Cette deuxième partie de leurs encaisses va alors être répartie entre titres nationaux et étrangers. Comme nous avons déjà décrit les déterminants de la demande de monnaie, la seule question qu'il nous reste à traiter est celle du choix entre titres nationaux et étrangers.

On peut facilement se faire une idée de ce qui va déterminer ce choix, en rappelant que la seule incitation à détenir des titres est qu'ils rapportent un intérêt. Par conséquent, les agents fonderont leur décision de détenir des titres nationaux et étrangers en comparant leurs rendements.

Cependant la comparaison des rendements est compliquée par le fait que les placements sont réalisés dans des monnaies différentes. Les agents doivent non seulement comparer les intérêts versés par les deux sortes de titres, mais tenir compte de l'évolution du taux de change.

Pour déterminer la relation entre les taux d'intérêt national et étranger et le taux de change, prenons l'exemple d'un habitant de la zone euro qui souhaite placer un euro et qui se demande s'il doit le placer dans la zone euro ou au Japon.

Le rendement du placement européen est noté i et celui du placement japonais i^* . Le taux de change de l'euro en yen est noté E . Comme nous nous plaçons du point de vue

européen, le taux de change mesure le prix d'un yen exprimé en euro. Le placement démarre à l'instant t et se termine un an plus tard, à l'instant $t + 1$.

Comme l'habitant de la zone euro réalise toutes ses transactions en euro, il sera intéressé par le rendement des deux placements exprimé en euro.

S'il place son euro en t dans la zone euro, l'agent possédera en $t + 1$:

$$\text{somme}_{\text{Euro}} = I + i \text{ €} \quad (1)$$

Si l'agent place son euro au Japon, il devra d'abord acheter des yens (¥). En t , il obtiendra ainsi (I / e_t) ¥. Il pourra alors placer ses yens au rendement i^* . A la fin de l'année, il possédera donc $(I / e_t) \cdot (I + i^*)$ ¥. Comme notre agent s'intéresse à la valeur de son placement exprimée en euro, il doit encore convertir cette somme au taux de change en vigueur en $t + 1$. Le rendement en euro du placement réalisé au Japon sera donc donné par l'expression suivante :

$$e_{t+1} \left(\frac{I}{e_t} \right) \cdot (I + i^*) = \frac{e_{t+1}}{e_t} (I + i^*)$$

On voit que la valeur en euro du placement au Japon dépend de l'évolution du taux de change entre l'euro et le yen. Or, cette évolution n'est connue avec certitude qu'en $t + 1$, c'est-à-dire à l'échéance du placement.

Au moment où l'agent fait son choix d'investir en Europe ou au Japon, en t , il doit donc anticiper le cours du yen en $t + 1$. Notons e_{t+1}^e le taux de change anticipé pour $t + 1$. La somme en euro que l'agent peut anticiper posséder à l'échéance du placement au Japon s'écrira alors :

$$\text{somme}_{\text{Japon}} = \frac{e_{t+1}^e}{e_t} (I + i^*) \text{ €} \quad (2)$$

L'agent choisira alors de placer son encaisse soit au Japon soit dans la zone euro en fonction de ce qui lui permettra d'obtenir le rendement le plus élevé. Il comparera donc les expressions (1) et (2) et prendra sa décision. On dit qu'*il arbitre* entre les deux placements.

Comme tous les agents dans l'économie font de même, le moindre écart de rendement entre les deux zones provoquera des mouvements immédiats de capitaux vers celle qui offre le rendement le plus élevé. Cela se traduira par une augmentation du prix de ses titres et/ou une appréciation de son taux de change courant, ce qui réduira l'écart entre les rendements des placements des deux zones. Le même phénomène se poursuivra de telle façon que l'écart de rendement disparaisse complètement.

L'arbitrage assurera donc l'égalité des rendements des titres de la zone euro et du Japon. A partir des expressions (1) et (2), on peut donc écrire :

$$\begin{aligned} \text{somme}_{\text{Japon}} &= \text{somme}_{\text{Euro}} \\ 1+i &= \frac{e_{t+1}^e}{e_t} (1+i^*) \end{aligned} \quad (3)$$

L'expression ci-dessus peut être manipulée pour l'exprimer de façon plus parlante :

$$\begin{aligned} 1+i &= \left(\frac{e_{t+1}^e}{e_t} - 1 + 1 \right) (1+i^*) \\ 1+i &= \left(\frac{e_{t+1}^e - e_t}{e_t} + 1 \right) (1+i^*) \end{aligned} \quad (3')$$

Le ratio qui apparaît dans la parenthèse ne mesure rien d'autre que la dépréciation anticipée de la monnaie nationale. L'expression (3') indique donc que le taux d'intérêt national dépend du taux d'intérêt étranger et de la dépréciation anticipée. Cette expression est peu commode et on a l'habitude d'utiliser une approximation, dont la démonstration est fournie en annexe :

$$\boxed{i \approx i^* + \left(\frac{e_{t+1}^e - e_t}{e_t} \right)} \quad (4)$$

Cette expression est appelée *condition de parité des taux d'intérêt non couverte* ou *PTI non couverte*. Elle implique que le taux d'intérêt sur les titres nationaux est approximativement égal au taux d'intérêt sur les titres étrangers auquel s'ajoute une prime égale à la variation anticipée du taux de change de la monnaie nationale.

Intuitivement, elle traduit le fait que si les investisseurs anticipent que les titres étrangers vont prendre de la valeur uniquement parce qu'ils sont exprimés dans une monnaie qui s'apprécie,¹¹² ils vont exiger que le rendement des titres nationaux soit plus élevé. Il s'agit pour eux de compenser la perte de valeur de la monnaie nationale.

¹¹² Rappelez-vous que si la monnaie nationale se déprécie par rapport à la devise étrangère, le taux de change à l'incertain, e , augmentera.

Encadré 1 : Démonstration de la formule approchée de la parité des taux d'intérêt non couverte.

Le point de départ est l'expression (3') qu'il suffit de développer :

$$1 + i = \left(\frac{e_{t+1}^e - e_t}{e_t} + 1 \right) (1 + i^*) \quad (3')$$

$$1 + i = \frac{e_{t+1}^e - e_t}{e_t} + \frac{e_{t+1}^e - e_t}{e_t} i^* + 1 + i^* \quad (E1)$$

$$i = \frac{e_{t+1}^e - e_t}{e_t} + \frac{e_{t+1}^e - e_t}{e_t} i^* + i^* \quad (E2)$$

Le deuxième terme du membre de droite de l'expression ci-dessus est négligeable. En effet, il s'agit du produit de deux termes qui sont eux-mêmes petits. Le taux d'intérêt ne dépasse en général pas quelques pourcents, de même que le taux de dépréciation. Leur produit est donc encore plus petit et le négliger ne représente pas une grosse erreur. On peut donc écrire la valeur approchée du taux d'intérêt intérieur de la façon suivante :

$$i \approx i^* + \left(\frac{e_{t+1}^e - e_t}{e_t} \right) \quad (4)$$

C'est telle qu'elle apparaît dans l'expression (4) que la PTI est en général utilisée.¹¹³ C'est cette formule que nous allons aussi utiliser dans la suite du cours.

Il faut cependant garder à l'esprit que cette forme de la PTI n'est qu'une approximation. Non seulement on néglige un terme, mais surtout on ne tient pas compte d'un grand nombre de facteurs qui peuvent affecter les écarts de taux d'intérêt.

Le coût des transactions internationales peut par exemple parfois être élevé, surtout pour des économies de petite taille dont les marchés sont peu liquides. La PTI fonctionnera donc moins bien pour ces économies.

De même, certains emprunteurs, y compris des états, peuvent être risqués. Cela signifie que les prêteurs ne peuvent pas être sûrs qu'ils seront remboursés. Ils exigeront *une prime de risque* en plus de la dépréciation anticipée.

¹¹³ Lorsqu'on étudie l'évolution des taux d'intérêt, on ne connaît en général pas la dépréciation anticipée. On observe en revanche l'évolution des taux d'intérêt nationaux et étrangers. C'est pourquoi la PTI non couverte est utilisée pour estimer la dépréciation anticipée.

Section 2 : la gestion de la monnaie dans une économie ouverte

La gestion de la monnaie dans une économie ouverte consiste non seulement à gérer la masse monétaire, comme dans une économie fermée, mais aussi à déterminer la valeur extérieure de la monnaie nationale, c'est-à-dire son taux de change. Pour comprendre comment la quantité de monnaie en circulation est déterminée, il faut comprendre quelles en sont les contreparties (I). La façon dont le taux de change est déterminé, dépendra du régime de change en vigueur (II).

I- Les contreparties de la masse monétaire en économie ouverte

Dans le chapitre consacré au marché de la monnaie, nous avons décrit l'activité de la banque centrale comme consistant à acheter et à vendre des titres financiers. Elle acquérait ainsi des créances sur le secteur privé. Cela lui permettait de contrôler l'offre de monnaie. Cette activité ne disparaît pas dans une économie ouverte. La banque centrale peut toujours acheter et vendre des titres.

Il se rajoute cependant une deuxième activité importante pour la banque centrale. Celle-ci intervient sur le marché des changes. Cela signifie qu'elle vend et achète des devises. En d'autres termes, elle échange de la monnaie étrangère contre de la monnaie nationale lorsqu'elle vend des devises, et elle échange de la monnaie nationale contre de la monnaie étrangère lorsqu'elle achète des devises.

Les raisons qui motivent l'intervention des banques centrales sur le marché des changes sont de deux ordres.

D'abord, les agents nationaux ont besoin de convertir de la monnaie nationale en devise à chaque fois qu'ils veulent réaliser une transaction à l'étranger ou en devise.

Si vous souhaitez acheter un CD par correspondance sur un site internet situé en dehors de la zone euro, vous devez d'abord vous procurer la monnaie dans laquelle le CD vous est facturé. Vous ne vous en rendez pas forcément compte si vous utilisez une carte de paiement car c'est votre banque qui se charge d'acquérir les devises dans ce cas. Plus généralement, toute importation donne lieu à un moment ou à un autre à un achat de devises, qui correspond à une vente de monnaie nationale.

A l'inverse, toute exportation donne lieu à une vente de devises, ce qui est équivalent à un achat de monnaie nationale.

Cependant, les échanges de biens ne représentent qu'une faible partie des opérations qui nécessitent des échanges de devises. Les transactions financières représentent aujourd'hui la majorité des transactions pour lesquelles on utilise le marché des changes.¹¹⁴

Ensuite, les banques centrales interviennent sur le marché des changes pour tenter d'influencer le cours de leur monnaie nationale. Ainsi, si une banque centrale souhaite que sa monnaie s'apprécie, elle en achètera sur le marché des changes. Elle vendra donc des devises puisées dans ses stocks de devises appelés *réserves de change*.

Si une banque centrale souhaite voir sa monnaie se déprécier, elle en vendra sur le marché des changes. Elle achètera donc des devises qui viendront ainsi s'ajouter à ses réserves de change.

On peut remarquer que toute intervention de la banque centrale sur le marché des changes se traduit par la mise en circulation ou le retrait de monnaie nationale.

ex : lorsqu'une entreprise européenne qui vient de vendre pour mille dollars de marchandises à un client américain, et qu'elle les convertit en euros, la quantité d'euros en circulation dans la zone euro augmente mécaniquement de la somme équivalente en euros.¹¹⁵

Lorsqu'une banque européenne achète pour mille euros de bons du trésor américain, la quantité d'euros en circulation diminue en revanche de mille euros.

Par conséquent, le fonctionnement du marché des changes est indissociable de l'évolution de la masse monétaire. Tout achat de devises par la banque centrale se traduit par une augmentation de la masse monétaire. A l'inverse, toute vente de devises se traduit par une diminution de la masse monétaire. Il est donc nécessaire d'en tenir compte pour comprendre ce qui détermine l'offre de monnaie dans une économie ouverte.

Pour résumer, on peut dire qu'il existe deux *contreparties de la base monétaire* dans une économie ouverte : les créances internes sur le secteur privé et les réserves de devises. Cela signifie que pour que la banque centrale émette une unité de monnaie nationale elle doit acquérir en contrepartie soit des créances internes, soit des devises. Au total la base monétaire (M_0) correspondra donc exactement à la somme des créances internes (D) et des réserves de devises détenues par la banque centrale (R). Les premières constituaient la seule contrepartie

¹¹⁴ Selon le dernier rapport triennal de la Banque des Règlements Internationaux, paru en mars 2005, la valeur quotidienne des transactions sur le marché des changes dépassait 1900 milliards de dollars en avril 2004. Cette somme dépasse largement le volume des échanges de biens et services. En 2005, la valeur des biens et services commerciaux exportés était de 12690 milliards de dollars et celles des biens et services importés de 12610 milliards de dollars.

¹¹⁵ La somme en euros dépendra du cours du dollar par rapport à l'euro. Au moment d'écrire ce cours, un euro valait 1,1842 dollars (02/01/2006). Les mille dollars correspondaient donc approximativement à 844,45 euros.

de la masse monétaire en économie fermée. Les secondes n'apparaissent qu'en économie ouverte.

$$M_0 = D + R \quad (5)$$

Cette relation est importante parce qu'elle détermine l'évolution de toute la masse monétaire. Elle montre que la politique monétaire et la politique de change sont intimement liées.

On peut par ailleurs remarquer qu'il existe une très forte ressemblance entre les opérations de la banque centrale sur le marché monétaire et sur le marché des changes. Sur le marché monétaire la banque centrale échange de la monnaie nationale contre des titres de créance. Sur le marché des changes, elles échangent de la monnaie nationale contre des devises. Dans les deux cas la base monétaire est affectée de la même façon.

La ressemblance entre les deux opérations apparaît également dans le bilan de la banque centrale. L'achat d'un même montant de titres ou de devises augmente le passif de la banque centrale du même montant. Ce qui change entre les deux opérations est le compte dans lequel elles sont enregistrées à l'actif (titres de créances ou réserves de change).

Ces résultats sont de nature comptable. Nous allons voir à présent que leurs implications dépendent du régime de change choisi par l'économie nationale.

II- Les régimes de change

Tant pour étudier le fonctionnement du marché des biens que du marché des capitaux, nous avons eu besoin de faire référence au taux de change. Nous n'avons en revanche pas encore décrit comment il était déterminé. La réponse dépend en fait d'un choix du pays, le choix de son *régime de change*.

Le régime de change est le système par lequel la valeur de la monnaie nationale par rapport aux autres devises est déterminée. Il existe schématiquement deux types de régimes de change :

- les *régimes de changes fixes* : les autorités choisissent le taux de change de la monnaie nationale et s'engagent à prendre les mesures nécessaires pour le défendre.
- les *régimes de changes flexibles* : les autorités laissent le cours de la monnaie fluctuer librement en fonction de l'évolution du marché des changes.

Le franc luxembourgeois et le franc belge ont été associés par un régime de changes fixes jusqu'à l'adoption de l'euro. Jusqu'à la crise de 2002, le peso argentin était attaché de façon très stricte au dollar américain. Le système d'étalon-or, selon lequel toutes les monnaies devaient être convertibles en or à un cours donné était un régime de changes fixes à l'échelle du monde entier qui est resté en vigueur de la fin du XIX^{ème} siècle au début de la première Guerre Mondiale. Après la deuxième Guerre Mondiale et jusqu'au début des années 70, les monnaies des économies occidentales participaient à un système de parités fixes connu sous le nom de système de Bretton Woods.¹¹⁶

A l'heure actuelle, le régime de change entre le dollar et la plupart des autres grandes monnaies du monde est un régime de changes flexibles. Le dollar, le yen, mais aussi la livre Sterling et la couronne suédoise sont censés flotter librement par rapport à la devise européenne.

Entre les deux cas d'école que sont les régimes de changes fixes et flexibles, il existe toute une variété de régimes de change. La monnaie nationale peut ainsi être rattachée à une seule devise ou à un panier de devises. La parité annoncée peut être assortie de marges de fluctuation qui donnent au taux de change une flexibilité limitée (comme dans le Système monétaire européen)¹¹⁷ etc.

Par ailleurs, il est très rare qu'un état laisse fluctuer sa monnaie totalement librement sans essayer d'en influencer le cours. On parle alors de flottement contrôlé ou de flottement impur. On sait par exemple que le cours des trois principales devises internationales que sont le dollar, le yen et l'euro est très surveillé par les banques centrales et qu'il donne de plus lieu à des négociations entre ces pays, notamment lors des réunions du G7.

¹¹⁶ Bretton Woods est une station de sports d'hiver du New Hampshire où les délégations de quarante-quatre pays s'étaient réunies en 1944 pour conclure les négociations sur l'organisation du système monétaire international d'après-guerre.

¹¹⁷ Le SME qui était né en 1979 n'est pas mort avec le lancement de l'euro. Il survit sous le nom de SME bis. A l'heure actuelle seule la couronne danoise participe à ce système de changes fixes avec l'euro.

Tableau 1

Le régime de change des pays membres de l'Union européenne par ordre de flexibilité croissante (2008)

Union monétaire	Allemagne, Autriche, Belgique, Chypre, Espagne, Finlande, France, Grèce, Irlande, Italie, Luxembourg, Malte, Pays-Bas, Portugal, Slovaquie, Slovénie
Caisse d'émission et taux fixe par rapport à l'euro	Estonie, Lituanie
Cours pivot par rapport à l'euro	Danemark, Hongrie, Pologne, République tchèque
Cours pivot par rapport à un panier de monnaies	Bulgarie, Lettonie (DTS)
Flottement	Roumanie, Royaume Uni, Suède

Le tableau 1 présente les régimes de change des pays membres de l'Union européenne. Il montre que même au sein de l'Union on peut trouver tous les degrés de flexibilité du taux de change.¹¹⁸

Par conséquent, la distinction entre changes fixes et flexibles est une distinction très théorique. Il faut donc considérer les régimes de change que nous allons étudier comme des cas d'école. Cependant, ils nous permettront d'obtenir quelques intuitions importantes sur les effets de la politique économique en fonction de la façon dont la valeur extérieure de la monnaie nationale est déterminée.

En termes de gestion de la masse monétaire, les deux régimes de change impliquent des contraintes tout à fait opposées.

En effet, en changes fixes, une fois que la parité de la monnaie nationale a été choisie, les autorités sont contraintes de satisfaire toutes les demandes de devise ou de monnaie nationale au prix annoncé. Elles s'engagent donc à émettre ou à retirer de la circulation n'importe quelle quantité de monnaie nationale, au gré des demandes. Par conséquent, la masse monétaire devient une fonction de la demande de monnaie nationale, que la banque centrale s'engage à satisfaire de façon automatique. **La masse monétaire est donc endogène en changes fixes.**

¹¹⁸ Le tableau 1, donné en annexe, présente l'évolution des régimes de change officiellement choisis par les pays du monde entier depuis 1974. Le tableau 2 qui montre les régimes effectivement adoptés suggère que les pratiques peuvent différer des annonces officielles.

En changes flexibles en revanche, les autorités s'engagent seulement à laisser flotter leur monnaie. La banque centrale peut donc choisir d'émettre la quantité de monnaie qu'elle souhaite. Le taux de change s'adaptera alors en fonction de la demande de monnaie nationale. Il s'appréciera si la demande augmente et se dépréciera sinon. **En changes flexibles, c'est le taux de change qui est endogène.**

Cette différence nous permet d'interpréter la parité des taux d'intérêt de façon différente selon le régime de change. En changes flexibles, on peut considérer que les taux d'intérêt national et étranger sont choisis avec plus ou moins de latitude par les banques centrales du pays national et de son partenaire. C'est donc le taux de change qui doit s'adapter pour rétablir l'égalité entre les rendements anticipés.

En changes fixes en revanche, la parité est donnée. Par conséquent, les autorités nationales sont contraintes d'ajuster le taux d'intérêt en fonction du taux d'intérêt étranger. L'adoption d'un régime de change fixe contraint donc la politique monétaire du pays national.

Cette différence entre les deux types de régime de change sera cruciale pour comprendre les effets des politiques économiques.

Conclusion

Dans une économie ouverte, les capitaux peuvent être échangés avec le reste du monde au même titre que les biens. Dans la mesure où ils sont placés en fonction de leur rendement, les arbitrages entre les placements internationaux aboutissent à la relation de parité des taux d'intérêts. Celle-ci est indépendante du régime de change en vigueur. La façon dont l'économie s'ajuste pour la faire respecter dépend en revanche du régime de change que le pays a adopté.

Il existe une grande variété des régimes de change, mais on peut tous les classer entre deux cas d'école : les changes fixes et les changes flexibles.

Ces deux régimes de change impliquent des contraintes très différentes sur le taux d'intérêt intérieur et donc sur l'équilibre macroéconomique. Ils vont donc affecter l'efficacité des politiques économiques. C'est ce que nous allons voir dans le prochain chapitre.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 15 (section 15.2).

Pour les passionnés

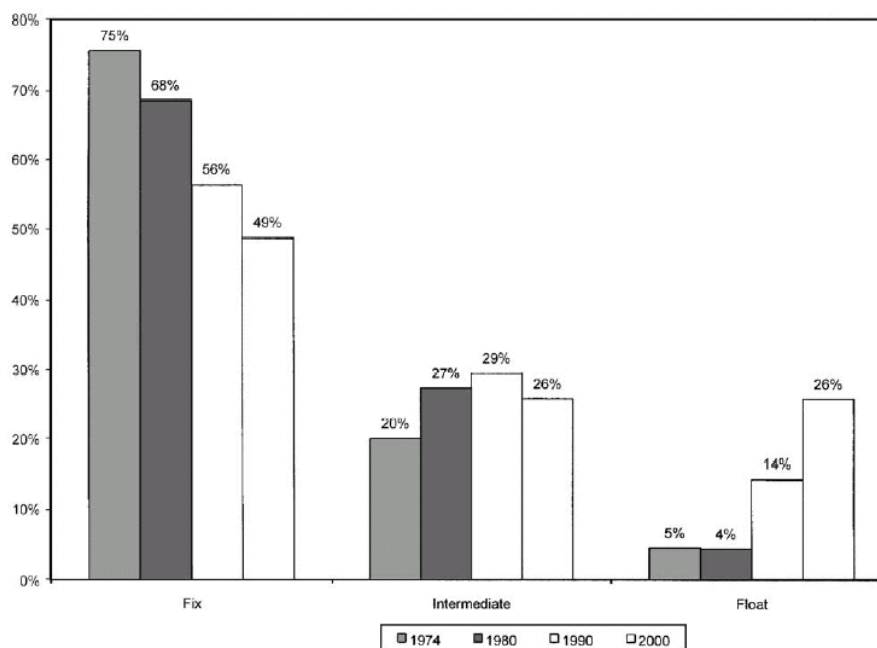
Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 12 (section 12.5) : pour une évocation rapide des avantages et des inconvénients des changes fixes et flexibles.

Mishkin F. : *Monnaie, banques et marchés financiers*, Pearson Education France, Paris, 2004 (chapitres 19 et 20) pour une évocation plus complète des avantages et des inconvénients des changes fixes et flexibles mais aussi des stratégies monétaires qui y sont associées.

Annexe

Graphique 1

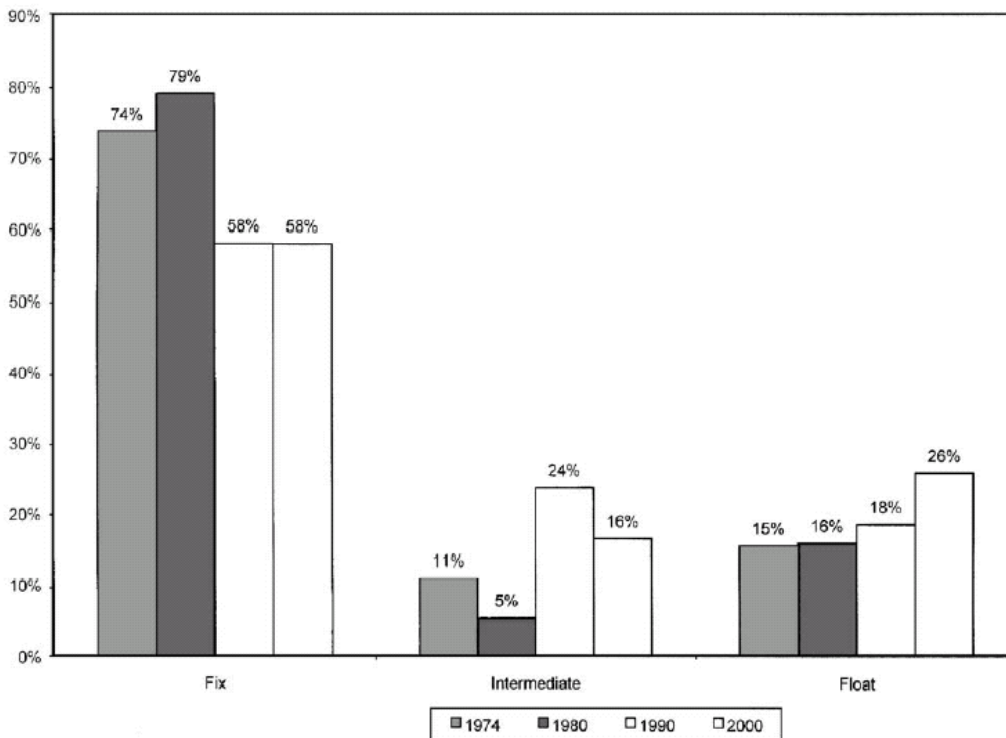
L'évolution des régimes de change officiellement déclarés au Fonds Monétaire International (1974-2000)



Source : Levy-Yeyati, E., et F. Sturzenegger "Classifying exchange rate regimes: Deeds vs. words", *European Economic Review*, vol. 49 n°6, p. 1603-1635, août 2005.

Les notes du cours ne sont transmises qu'à titre indicatif.

Graphique 2
L'évolution des régimes de change officiellement effectivement respectés (1974-2000)



Source : Levy-Yeyati, E., et F. Sturzenegger "Classifying exchange rate regimes: Deeds vs. words", *European Economic Review*, vol. 49 n°6, p. 1603-1635, août 2005.

Pourquoi Hägar est crédible :



Hägar est physiquement incapable de reculer. Son annonce est donc crédible. Certains pays ont choisi d'utiliser des stratagèmes similaires pour rendre crédible leur politique monétaire. Ils l'ont ainsi confiée à une banque centrale indépendante à qui il est interdit de créer de l'inflation. C'est notamment sur ce principe qu'est conçu le Système européen de banques centrales, qui gère la politique monétaire en Europe.

Les notes du cours ne sont transmises qu'à titre indicatif.

Chapitre 12 : La politique macroéconomique dans une économie ouverte : le modèle de Mundell-Fleming

Introduction : Une innovation théorique aux implications pratiques immédiates

Nous avons jusqu'à présent étudié la politique économique en nous appuyant sur le modèle IS-LM. Ce modèle permet d'obtenir des intuitions importantes sur les mécanismes qui sont à l'œuvre lorsque des politiques de gestion de la demande agrégée sont appliquées. Malheureusement, ce modèle repose sur l'hypothèse d'une économie fermée.

Au début des années soixante, les macroéconomistes n'étaient pas plus avancés que nous. Ils disposaient de théories macroéconomiques qui intégraient les échanges de biens. Ils disposaient également de théories des mouvements internationaux de capitaux. En revanche, ils ne disposaient pas d'une théorie de l'équilibre macroéconomique et prenant en compte les mouvements de biens et de capitaux.

C'est au travail de Marcus J. Fleming (1962) et de Robert A. Mundell (1963) que l'on doit la prise en compte des mouvements de capitaux dans la description des effets des politiques macroéconomiques.¹¹⁹ C'est pourquoi on parle aujourd'hui du modèle de Mundell-Fleming.¹²⁰

Au moment d'élaborer leurs théories, ces deux auteurs travaillaient tous les deux pour le FMI, même si Mundell a par la suite poursuivi une carrière plus universitaire qui lui a permis d'obtenir le prix Nobel en 1999. Leurs travaux, bien que théoriques, avaient donc une portée non seulement intellectuelle mais aussi pratique. Ils montrèrent contre toute attente que le régime de change en vigueur affectait considérablement le résultat des politiques économiques à partir du moment où on prenait en compte les mouvements internationaux de capitaux.

Sur le plan analytique, le modèle de Mundell-Fleming est une extension du modèle IS-LM au cas d'une économie ouverte. Sa nouveauté consiste à prendre en compte à la fois les échanges de biens et les mouvements de capitaux. Dans ce qui suit, nous allons d'abord

¹¹⁹ Les résultats de Fleming apparaissent dans l'article "Domestic financial policies under fixed and under floating exchange rates", *IMF Staff Papers*, vol. 9, pp.369-379, novembre 1962. Mundell a écrit au moins quatre articles complémentaires mais les résultats du modèle apparaissent le plus clairement dans "Capital mobility and stabilization policy under fixed and flexible exchange rates", *Canadian Journal of Economics and Political Science*, vol. 29, pp.475-485, novembre 1963.

¹²⁰ La première référence au modèle Mundell-Fleming remonterait à 1976. On parle parfois du modèle Fleming-Mundell, mais c'est plus rare. La raison pour laquelle l'ordre alphabétique n'est pas respecté reste d'ailleurs obscure.

montrer comment le modèle est construit à partir d'hypothèses très proches de celles du modèle IS-LM (section 1) avant de l'utiliser pour étudier les effets des politiques économiques (section 2).

Section 1 : Construction du modèle de Mundell-Fleming

Le modèle de Mundell-Fleming est une extension du modèle IS-LM. Pour adapter le modèle de base au cas d'une économie ouverte, il faut adapter la description du marché des biens par la courbe IS (I) et celle du marché de la monnaie par la courbe LM (II). Il faut enfin compléter le modèle en décrivant l'équilibre sur le marché des changes (III).

I- Le marché des biens : la courbe IS

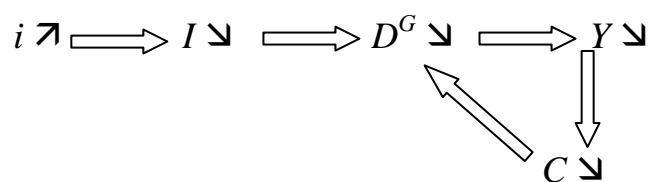
La définition de la courbe IS dans le modèle de Mundell-Fleming est la même que dans le modèle IS-LM en économie fermée. Sa construction repose donc également sur l'équilibre du marché des biens. On obtient ainsi l'équation de la courbe IS en écrivant l'égalité entre l'offre et la demande de biens :

$$Y = C(Y) + I(i) + G + NX\left(Y, Y^*, \frac{e \cdot P^*}{P}\right) \quad (1)$$

Cette équation est la même que dans le modèle IS-LM en économie fermée à ceci près que les exportations nettes sont à présent prises en compte. Cela ne change rien au principe de construction de la courbe. Il s'agit toujours d'exprimer l'effet sur le revenu d'équilibre sur le marché des biens d'une variation exogène du taux d'intérêt.

Graphiquement, la forme de la courbe IS se déduit par conséquent du diagramme à 45°, en partant du fait qu'une augmentation du taux d'intérêt provoque une diminution de l'investissement. Comme la méthode a été présentée dans le chapitre 4 du cours, et qu'il est pratiquement inchangé, nous n'allons pas le reprendre ici. Il suffit en fait de rappeler le raisonnement intuitif qui permet de conclure que la courbe IS est décroissante.

L'augmentation du taux d'intérêt réduit l'investissement, ce qui déprime la demande globale. La production s'ajuste à la demande, ce qui réduit le revenu puis la consommation. Il s'enclenche alors un mécanisme multiplicateur qui amplifie la réduction initiale de l'investissement. La diminution de la production va donc être égale à la variation de l'investissement multipliée par le multiplicateur. ($\Delta Y = multi \times \Delta I$).



La différence qu'il convient de souligner est que la valeur du multiplicateur est différente en économie ouverte et en économie fermée. En effet, puisqu'une partie du revenu supplémentaire est consacrée aux importations, la demande globale augmente moins à la suite de la baisse du taux d'intérêt.

Le raisonnement qui nous permet d'affirmer que la courbe IS est décroissante nous permet également d'isoler les déterminants de sa pente. Ainsi, la courbe IS sera d'autant moins raide qu'une diminution donnée du taux d'intérêt provoquera un accroissement important du revenu d'équilibre.

Par conséquent, la courbe IS sera d'autant plus plate que l'investissement sera sensible aux variations du taux d'intérêt et que le multiplicateur sera important.

Comme nous avons vu dans le chapitre 10 que le multiplicateur en économie ouverte est une fonction croissante de la propension marginale à consommer et décroissante de la propension marginale à importer, nous pouvons désormais affirmer que **la courbe IS sera d'autant plus plate que la propension marginale à consommer sera élevée et la propension marginale à importer sera faible.**

De même que ceux de sa pente, les déterminants de la position de la courbe IS seront les mêmes en économie ouverte et en économie fermée. En particulier, une augmentation des dépenses publiques, ou une augmentation exogène de la consommation ou de l'investissement, déplaceront la courbe vers la droite.

Un nouveau déterminant de la position de la courbe IS apparaît cependant en économie ouverte. C'est le taux de change. Comme le modèle de Mundell-Fleming est un modèle à prix fixes, les variations du taux de change nominal affectent directement le taux de change réel, donc la compétitivité des biens produits nationalement.

Une dépréciation de la monnaie nationale va donc augmenter les exportations nettes, ce qui augmente la demande globale de biens. Quel que soit le niveau du taux d'intérêt, la production va par conséquent augmenter. **Toute dépréciation de la monnaie nationale déplacera donc la courbe IS vers la droite.** A l'inverse, une appréciation de la monnaie déplacera la courbe IS vers la gauche.

II- Le marché de la monnaie et les marchés financiers : la courbe LM et la droite d'intégration financière

En économie fermée, on pouvait se contenter de déterminer l'équilibre sur le marché de la monnaie sans prêter attention au marché des titres. En raison de la loi de Walras, si le marché des biens et celui de la monnaie étaient équilibrés, le marché des titres l'était également. En économie ouverte, il existe deux marchés distincts des titres : le marché national et le marché étranger. Pour déterminer l'équilibre global de l'économie, il faut donc s'assurer que trois marchés sont en équilibre sur les quatre qui constituent le modèle.

En plus de l'équilibre du marché de la monnaie qui est représenté par la courbe LM (A), il faut s'assurer que le marché des changes est lui aussi équilibré. C'est ce que permet la droite d'intégration financière (B).

A- La courbe LM

Tout comme la courbe IS, la courbe LM ne subit guère de modifications entre le modèle IS-LM et le modèle Mundell-Fleming. Sa construction repose sur l'équation d'équilibre du marché de la monnaie. En reprenant les notations du chapitre 5, l'égalité entre l'offre et la demande d'encaisses réelles s'écrit :

$$M / P = L(i)Y \quad (2)$$

On peut cependant compléter l'équation de la courbe LM en remplaçant la masse monétaire par ses contreparties. Elle s'écrit alors :

$$(D + R) / P = L(i)Y \quad (3)$$

Cette modification ne modifie pas les déterminants de la pente de la courbe LM. Celle-ci représente la variation du taux d'intérêt provoquée par une augmentation exogène du revenu. Le raisonnement ci-dessous montre que cette variation est positive et que la courbe LM reste donc croissante :

$$Y \nearrow \implies M^D \nearrow \implies i \nearrow \implies M^D \searrow$$

L'augmentation du revenu augmente les besoins de monnaie pour les transactions. Il apparaît alors une demande excédentaire de liquidité qui provoque à son tour une augmentation du taux d'intérêt. Comme le taux d'intérêt augmente, la demande de monnaie de précaution et à des fins de spéculation diminue. On retourne ainsi à l'équilibre sur le marché de la monnaie.

La pente de la courbe LM va donc dépendre de la sensibilité de la demande de monnaie à des fins de transaction au revenu et, surtout, au taux d'intérêt. Ce résultat est le même que dans le modèle IS-LM en économie fermée.

On retrouve également le fait que la position de la courbe LM dépendra de l'offre d'encaisses réelles. Une augmentation de l'offre d'encaisses réelles provoquera une diminution du taux d'intérêt quel que soit le revenu d'équilibre initial. Elle déplacera donc la courbe LM vers le bas.

Il apparaît cependant une nouveauté par rapport à la situation d'une économie fermée. En effet, l'offre d'encaisses réelles peut varier soit parce que les créances au secteur privé (D) varient, soit parce que les réserves de devises (R) varient. Cette différence aura son importance au moment où nous étudierons l'équilibre de l'économie en changes fixes.

B- La droite d'intégration financière

Si nous avons déjà pris en compte les échanges de biens de l'économie nationale grâce à la fonction d'exportations nettes, nous n'avons pas encore intégré ses relations financières avec le reste du monde. Pour ce faire, nous devons faire deux hypothèses supplémentaires quant à la taille de l'économie et à la mobilité internationale des capitaux.

D'abord, comme nous supposons que l'économie nationale est une petite économie sur le marché des biens, nous allons supposer qu'elle l'est aussi sur les marchés financiers. Cela signifie qu'elle est trop petite pour influencer le taux d'intérêt mondial. Quelles que soient ses demandes de monnaie et de titres, le taux d'intérêt mondial i^* restera constant.

Deuxièmement, nous allons supposer une *parfaite mobilité internationale des capitaux*. Cette hypothèse signifie que les agents peuvent librement aller placer leurs capitaux dans le pays de leur choix. Ils choisiront donc d'investir dans le pays dans lequel le rendement espéré est le plus élevé.

Dans ces conditions, l'arbitrage implique que la relation entre le taux d'intérêt national et le taux d'intérêt mondial soit décrite par la relation de parité des taux d'intérêt définie dans le chapitre précédent :

$$i \approx i^* + \left(\frac{e_{t+1}^e - e_t}{e_t} \right) \quad (4)$$

Le taux d'intérêt national est donc égal au taux d'intérêt mondial auquel s'ajoute une prime de risque donnée par la dépréciation anticipée de la monnaie nationale. Pour simplifier le modèle, nous allons supposer que la dépréciation anticipée est constante et nulle. Le taux d'intérêt national est donc tout simplement égal au taux d'intérêt mondial :

$$i = i^* \quad (4)$$

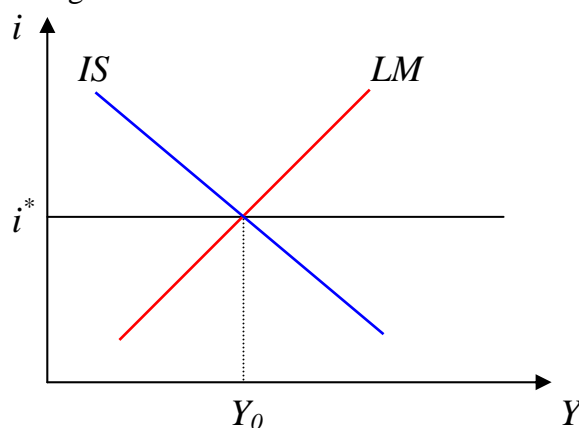
Cette égalité résume donc l'équilibre sur le marché des changes. Elle définit la *droite d'intégration financière*. Dans le repère (Y, i) , dans lequel le diagramme IS-LM est tracé, cette droite est une droite horizontale dont l'ordonnée correspond au taux d'intérêt mondial.

III- L'équilibre dans le modèle de Mundell-Fleming

L'équilibre global de l'économie est atteint quand tous les marchés sont équilibrés simultanément. Il faut donc que les marchés des biens, de la monnaie et des changes soient équilibrés. Si c'est le cas, la loi de Walras prévoit que le marché des titres sera aussi équilibré.

Comme la courbe IS représente les combinaisons du revenu et du taux d'intérêt national qui assurent l'équilibre sur le marché des biens, que la courbe LM décrit l'équilibre du marché national de la monnaie et que la droite d'intégration financière décrit la condition d'équilibre du marché des changes, l'équilibre global sera déterminé par l'intersection de ces courbes :

Graphique 1
L'équilibre global dans le modèle de Mundell-Fleming



Pour atteindre l'équilibre global de l'économie dans le modèle de Mundell-Fleming, il faut satisfaire trois conditions alors que deux suffisaient dans le modèle IS-LM de base. Toute la question est de déterminer comment ces trois conditions vont être satisfaites.

Graphiquement, cela revient à déterminer laquelle des courbes va se déplacer pour que les trois se coupent en un seul point.

La réponse à cette question dépend du régime de change en vigueur. Nous avons vu dans le chapitre précédent qu'en changes fixes, le taux de change est donné mais que la banque centrale doit ajuster en permanence l'offre de monnaie pour satisfaire les demandes de monnaie nationale ou de devises. En d'autres termes, le taux de change est exogène mais la masse monétaire endogène. C'est donc la courbe LM qui va s'ajuster en changes fixes pour assurer l'équilibre global.

En changes flexibles en revanche, les autorités laissent fluctuer le taux de change et ne sont pas contraintes d'ajuster l'offre de monnaie. Par conséquent, c'est le taux de change qui est endogène alors que la masse monétaire est exogène. Comme le taux de change détermine la compétitivité des produits nationaux, c'est la courbe IS qui va s'ajuster en changes flexibles.

Ces conclusions vont nous permettre d'analyser les effets des politiques économiques.

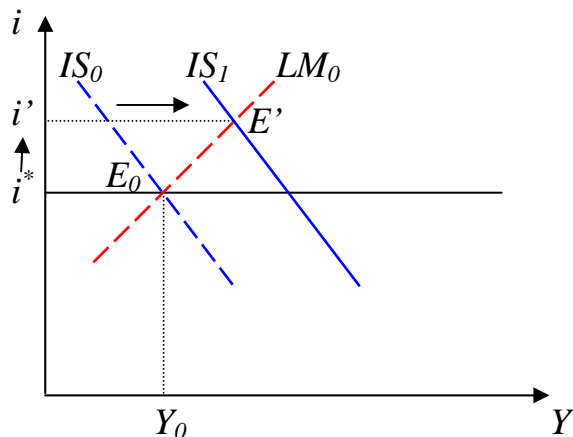
Section 2 : La politique économique dans le modèle de Mundell-Fleming

Comme le modèle IS-LM de base, le modèle de Mundell-Fleming permet d'analyser les effets des politiques budgétaires (I) et monétaires (II). La nouveauté est que leurs effets vont à présent dépendre du régime de change en vigueur. Lorsque les changes sont fixes, les autorités peuvent utiliser un troisième instrument de politique économique : la dévaluation (III).

I- La politique budgétaire

La politique budgétaire consiste à modifier le volume des dépenses publiques afin d'affecter la demande globale de biens. Elle se traduit donc par un déplacement vers la droite de la courbe IS et provoque donc un déséquilibre global puisque le taux d'intérêt domestique passe momentanément au-dessus du taux d'intérêt mondial, ce qui va provoquer un afflux de capitaux étrangers attirés par le rendement des titres nationaux.

Graphique 2
Effet immédiat d'une politique budgétaire expansionniste

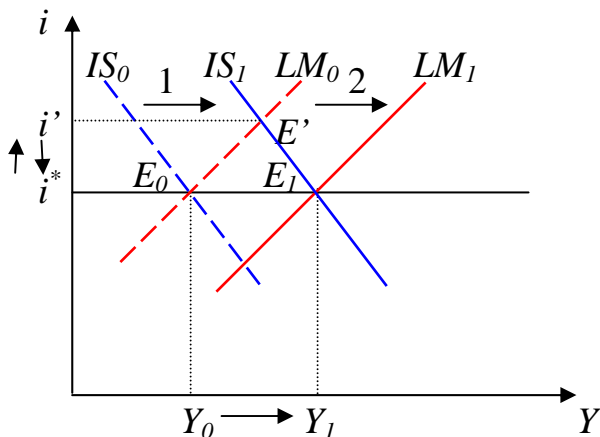


Cette situation n'est pas tenable, puisque le taux d'intérêt domestique est supérieur au taux d'intérêt mondial. Le retour à l'équilibre va s'opérer différemment selon le régime de change en vigueur.

A- Politique budgétaire en changes fixes

En changes fixes, l'afflux de capitaux ne peut par définition pas affecter le taux de change de la monnaie nationale. En revanche, il apparaît une demande excédentaire de monnaie nationale puisque les investisseurs étrangers ont besoin de convertir leurs devises avant de pouvoir acheter des titres nationaux. La banque centrale va donc devoir émettre de la monnaie nationale qu'elle échangera contre des devises. Elle va donc accumuler des réserves de change et faire augmenter la masse monétaire.

Graphique 3
Une politique budgétaire expansionniste en changes fixes

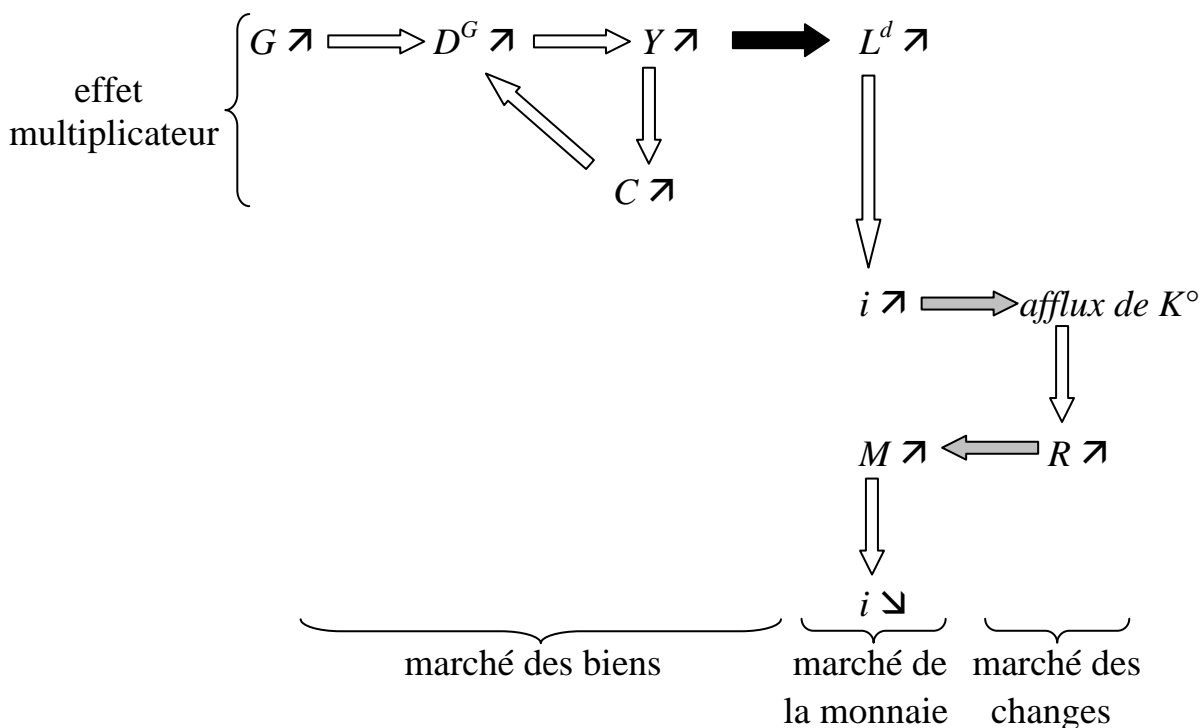


Les notes du cours ne sont transmises qu'à titre indicatif.

Comme le montre le graphique 3, l'augmentation de la masse monétaire nationale déplace la courbe LM vers le bas, ce qui permet au taux d'intérêt national de baisser. Le même mouvement se poursuit jusqu'à ce que le taux d'intérêt ait retrouvé son niveau initial.

Lorsque l'économie est retournée à l'équilibre, on constate que le revenu a augmenté. Comme le taux d'intérêt est resté constant, l'effet d'éviction a été neutralisé. On peut donc dire que l'augmentation du revenu correspond à l'augmentation des dépenses budgétaires multipliée par le multiplicateur.

Le schéma ci-dessous résume l'effet d'une politique budgétaire expansionniste dans le modèle de Mundell-Fleming en changes fixes. Il montre comment l'afflux de capitaux évite l'apparition d'un effet d'éviction.



On peut donc conclure que **la politique budgétaire est efficace en changes fixes** et qu'elle est accompagnée d'une augmentation des réserves.

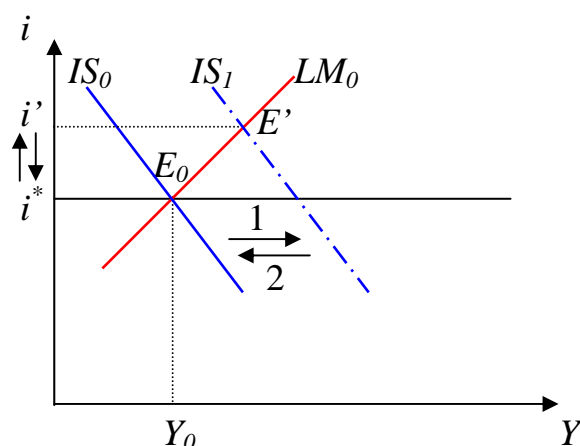
B- Politique budgétaire en changes flexibles

En changes flexibles, l'augmentation de la demande de monnaie nationale par les investisseurs étrangers se traduit par une appréciation du taux de change de la monnaie

nationale. Par conséquent, la compétitivité des produits nationaux diminue, ce qui réduit les exportations nettes et déplace la courbe IS vers la gauche. C'est cela qui permet au taux d'intérêt national de baisser. Le même mouvement se poursuit jusqu'à ce que la courbe IS ait retrouvé sa position initiale. Le taux d'intérêt est alors à nouveau égal au taux d'intérêt mondial. C'est ce que montre le graphique 4.

Graphique 4

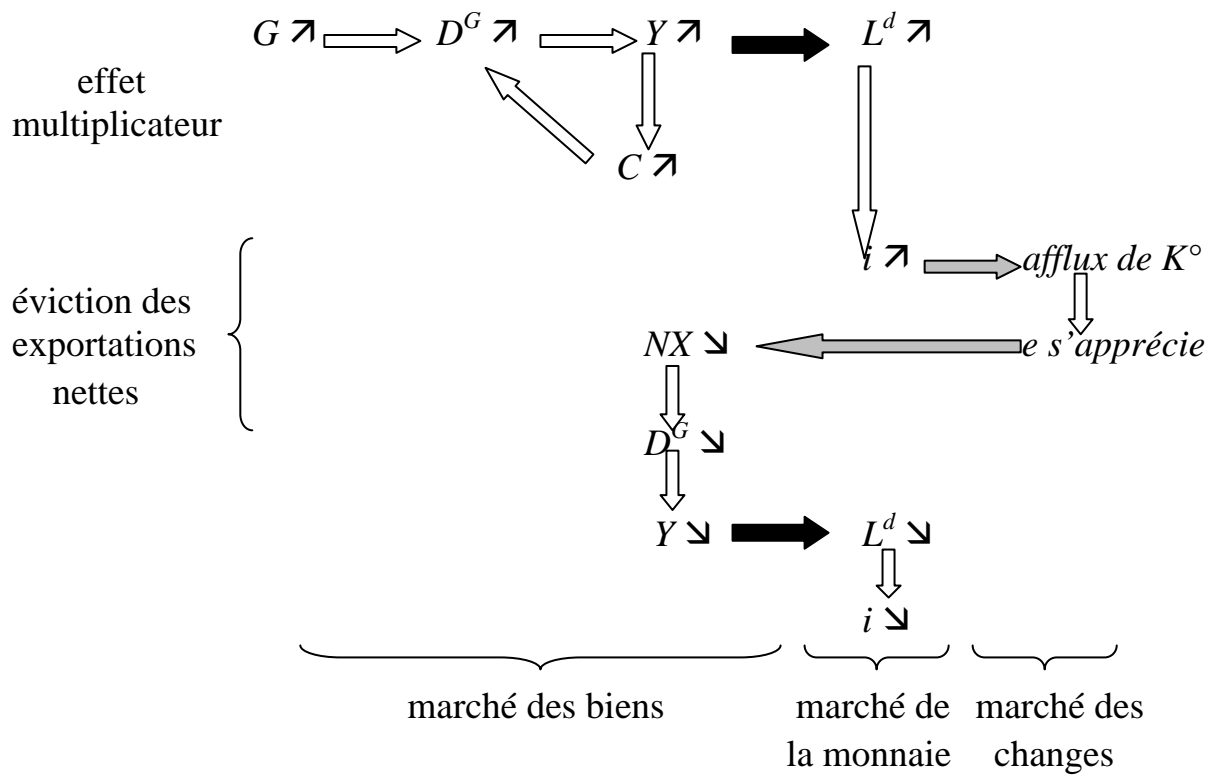
Une politique budgétaire expansionniste en changes flexibles



Lorsque l'économie est retournée à l'équilibre, on constate que le revenu a repris sa valeur initiale. Au lieu d'une éviction de l'investissement par l'augmentation du taux d'intérêt, on peut dire qu'il y a eu éviction d'une partie des exportations nettes par l'appréciation du taux de change.

Le schéma ci-dessous résume l'effet d'une politique budgétaire expansionniste dans le modèle de Mundell-Fleming en changes flexibles.

Il montre pourquoi on peut conclure que **la politique budgétaire est inefficace en changes flexibles** et qu'**elle est accompagnée d'une appréciation de la monnaie nationale**.

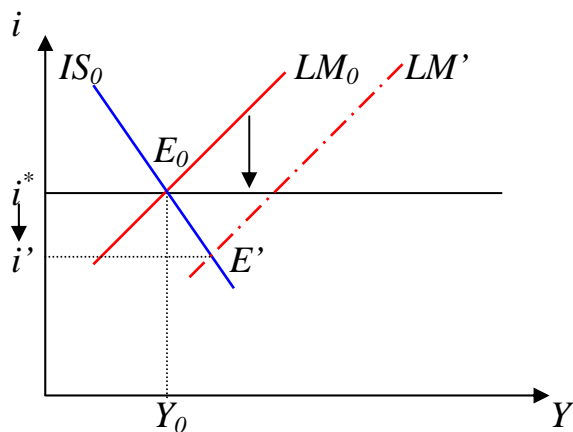


II- La politique monétaire

Dans une économie ouverte, comme dans une économie fermée, la politique monétaire consiste à manipuler la masse monétaire dans le but de modifier le taux d'intérêt. Dans une économie ouverte cependant, il faut tenir compte du fait que la masse monétaire a deux contreparties, les crédits internes et les réserves. Comme le stock de réserves est limité et difficile à contrôler, c'est le volume des crédits internes que la banque centrale va manipuler.

Si la politique monétaire est expansionniste, la banque centrale va augmenter le volume des crédits internes. La masse monétaire va donc augmenter. Comme l'offre d'encaisses réelles augmente, la courbe LM va se déplacer vers le bas et le taux d'intérêt momentanément diminuer. On assistera alors à une fuite de capitaux, attirés par le rendement des titres étrangers.

Graphique 5
Effet immédiat d'une politique monétaire expansionniste

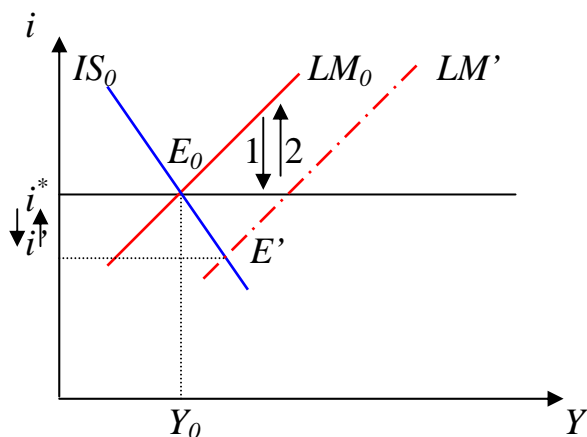


Le retour à l'équilibre va s'opérer différemment selon le régime de change en vigueur.

A- Politique monétaire en changes fixes

En changes fixes, la fuite des capitaux se traduit par une diminution du stock de réserves de la banque centrale. En effet, les investisseurs qui veulent acquérir des titres étrangers vendent leurs titres nationaux et se tournent vers la banque centrale pour échanger leurs avoirs en monnaie nationale contre des devises. La banque centrale qui s'est engagée à maintenir un taux de change constant doit satisfaire ces demandes. Elle doit donc puiser dans ses réserves de changes pour fournir les devises qu'on lui demande. Ce faisant, elle réduit l'offre de monnaie nationale.

Graphique 6
Une politique monétaire expansionniste en changes fixes

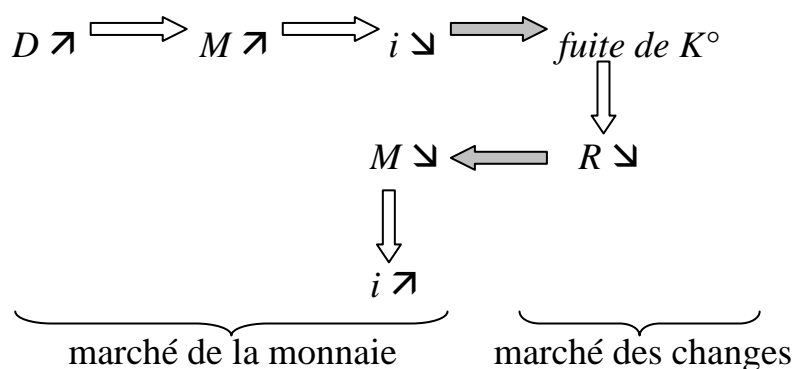


Les notes du cours ne sont transmises qu'à titre indicatif.

Comme l'offre d'encaisses réelles diminue, la courbe LM se déplace vers le haut et le taux d'intérêt augmente. Le même mécanisme se poursuit jusqu'à ce qu'il ait retrouvé son niveau initial. In fine, la courbe LM est revenue à sa position initiale. Cela est dû au fait que la masse monétaire a retrouvé son niveau initial. Le seul changement est à chercher dans ses contreparties. Les crédits internes ont certes augmenté, mais la diminution des réserves a complètement compensé cette augmentation. La somme des deux contreparties de la masse monétaire est donc restée constante.

Lorsque l'économie est retournée à l'équilibre, la courbe LM a retrouvé sa position initiale, comme le montre le graphique 6. Le revenu reste donc à sa valeur initiale.

L'enchaînement des effets d'une politique monétaire expansionniste est résumé par le schéma ci-dessous.

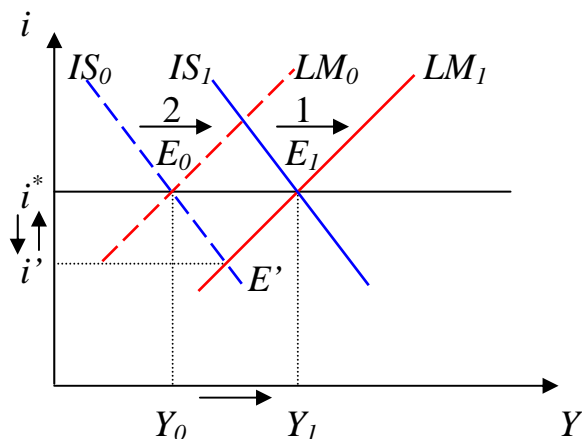


L'augmentation initiale de la masse monétaire est entièrement neutralisée par le fonctionnement du marché des changes. On peut donc conclure que **la politique monétaire est inefficace en changes fixes** et qu'**elle est accompagnée d'une diminution des réserves**.

B- Politique monétaire en changes flexibles

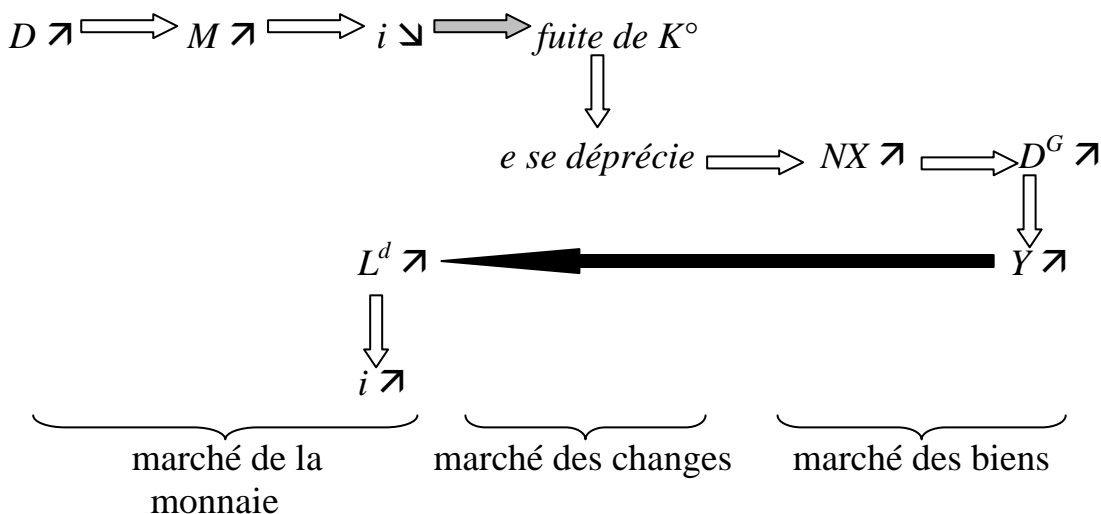
En changes flexibles, la fuite des capitaux nationaux se traduit par une dépréciation de la monnaie nationale. En effet, les investisseurs qui vendent leurs titres nationaux pour les remplacer par des titres étrangers mieux rémunérés doivent convertir leurs avoirs en monnaie nationale en monnaie étrangère. Il apparaît donc une demande excédentaire de devises et une offre excédentaire de monnaie nationale qui ne peut se résorber que par la dépréciation de la monnaie nationale.

Graphique 7
 Une politique monétaire expansionniste en changes flexibles



Comme la monnaie nationale se déprécie, la compétitivité des produits nationaux augmente. Les exportations progressent, ce qui augmente la demande de biens nationaux et déplace donc la courbe IS vers la droite. Comme la production augmente, la demande de monnaie nationale augmente aussi, ce qui pousse le taux d'intérêt national à la hausse. Ce mouvement se poursuit tant que le taux d'intérêt national n'a pas retrouvé sa valeur initiale, correspondant au taux d'intérêt mondial. C'est ce que représente le graphique 7 ci-dessus.

Le schéma ci-dessous résume l'effet d'une politique monétaire expansionniste dans le modèle de Mundell-Fleming en changes flexibles. Il montre comment la fuite des capitaux complète l'augmentation de l'offre de crédits internes.



Les notes du cours ne sont transmises qu'à titre indicatif.

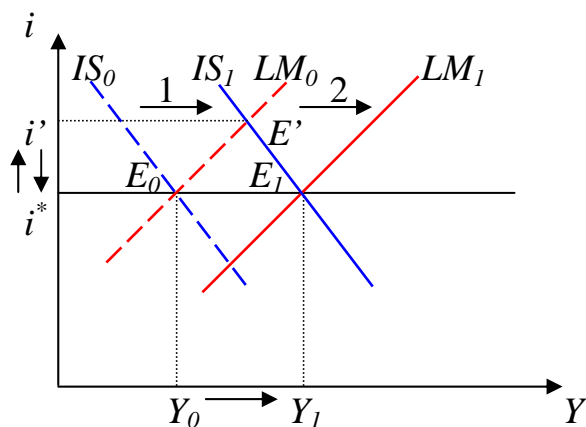
On voit que, même si le taux d'intérêt reste égal au taux d'intérêt mondial, l'augmentation de la masse monétaire se transmet au marché des biens via la dépréciation de la monnaie nationale. On peut donc conclure que **la politique monétaire est efficace en changes flexibles** et qu'elle est accompagnée d'une dépréciation de la monnaie nationale.

III- La dévaluation

La *dévaluation* est un instrument de politique économique qui ne peut être utilisé qu'en changes fixes. Elle consiste à réduire de façon délibérée la valeur de la monnaie nationale. Il s'agit donc d'une dépréciation volontaire de la monnaie.¹²¹ La logique de cette politique est un peu différente des deux précédentes. Elle consiste à partir du marché des changes pour obtenir des effets sur les autres marchés.

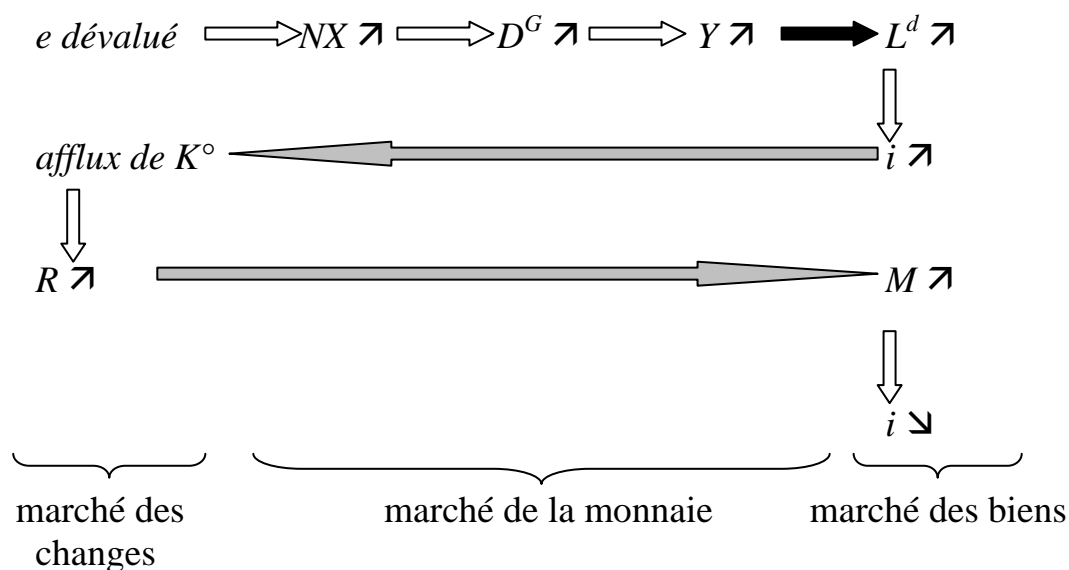
Le point de départ d'une dévaluation est l'annonce que la monnaie nationale est à présent échangée à un cours moins avantageux qu'auparavant. On provoque donc de façon délibérée une augmentation de la compétitivité des produits nationaux, c'est-à-dire une dépréciation réelle de la monnaie nationale.

Graphique 8
Une dévaluation



Comme le montre le graphique 8, cette augmentation de la compétitivité des produits nationaux se traduit par une augmentation des exportations nettes, ce qui déplace la courbe IS vers la droite.

¹²¹ La politique inverse, qui consiste à apprécier la monnaie nationale en changes fixes est appelée une *réévaluation*.



Le taux d'intérêt national devient alors supérieur au taux d'intérêt mondial, ce qui provoque un afflux de capitaux et une demande de monnaie nationale, que la banque centrale satisfait. Cette dernière voit donc ses réserves de changes augmenter, ce qui signifie que la masse monétaire augmente. La courbe LM se déplace alors vers le bas et le taux d'intérêt national diminue. Le mécanisme se poursuit jusqu'à ce que le taux d'intérêt national ait retrouvé sa valeur initiale.

On constate donc que la dévaluation provoque indirectement une augmentation de la masse monétaire. On peut donc conclure qu'**une dévaluation est efficace**.

Globalement, on peut remarquer que les effets d'une dévaluation sont les mêmes que ceux d'une politique monétaire expansionniste en changes flexibles. En effet, une politique monétaire expansionniste augmente le revenu via une dépréciation de la monnaie nationale. Lors d'une dévaluation, le revenu augmente également ainsi que la masse monétaire, qui est gonflée par l'afflux des capitaux étrangers.

Conclusion

Lorsque l'on adapte le modèle IS-LM au contexte d'une petite économie ouverte, on est amené à nuancer et compléter les résultats obtenus en économie fermée. D'abord, l'efficacité des politiques économiques dépend du régime de change en vigueur. Une même politique efficace dans un régime de change donné peut devenir totalement inefficace dans un autre.

La politique budgétaire est ainsi particulièrement efficace en changes fixes mais totalement inefficace en changes flexibles parce que ses effets sont neutralisés par l'appréciation de la monnaie nationale.

A l'inverse, la politique monétaire est inefficace en changes fixes parce que l'augmentation des crédits intérieurs est neutralisée par la diminution des réserves de changes. Elle est en revanche particulièrement efficace en changes flexibles parce qu'elle provoque une dépréciation de la monnaie nationale qui relance les exportations nettes.

La dévaluation est un troisième instrument de politique économique qui n'est disponible par définition qu'en changes fixes. Ses effets sont comparables à ceux d'une politique monétaire pratiquée en changes flexibles. Elle est donc efficace.

Toutes ces conclusions doivent être nuancées parce qu'elles ne sont valables que dans le cas d'une petite économie ouverte et en présence d'une mobilité parfaite des capitaux.

Si l'économie est suffisamment grande pour influencer le revenu et le taux d'intérêt du reste du monde, elle retrouvera une certaine efficacité de ses politiques économiques quel que soit le régime de change en vigueur. En revanche, le régime de change affectera l'efficacité relative des différentes politiques. La politique monétaire sera relativement plus efficace en changes flexibles qu'en changes fixes et ce sera l'inverse pour la politique budgétaire. Bien que nous n'ayons pas traité le cas d'une grande économie, on peut y réfléchir en considérant un cas intermédiaire entre ceux d'une petite économie ouverte et d'une économie fermée.

La mobilité imparfaite des capitaux peut également redonner de l'efficacité à des politiques qui n'en auraient pas autrement. Elle permet au taux d'intérêt national de différer du taux d'intérêt mondial. Par conséquent, elle redonne de l'autonomie en changes fixes et permet à la politique budgétaire de produire des effets en changes flexibles.

Enfin, le modèle de Mundell-Fleming ne prend en compte ni les contraintes d'offre de l'économie, ni la formation des anticipations. La suite de votre cursus vous permettra d'étudier ces raffinements.

Références

Sur le cours

Blanchard O. et D. Cohen : *Macroéconomie*, Pearson Education France, Paris, 2010 : chapitre 17.

Mankiw N.G. : *Macroéconomie*, De Boeck, Bruxelles, 2010 : chapitre 12.

Attention : la présentation du modèle Mundell-Fleming dans ces deux manuels est différente de celle du cours. La présentation la plus proche de celle du cours peut être trouvée dans le manuel suivant :

Burda M. et C. Wyplosz : *Macroéconomie : une perspective européenne*, De Boeck, Bruxelles, 2009 : chapitre 11.

Pour les passionnés

Boughton J.M. "On the Origins of the Fleming-Mundell Model", *IMF Staff Papers*, vol. 50 n°1, p.1-9, 2003.
