ECOLE CENTRALE PARIS

2ème ANNÉE d'ÉTUDES

INTRODUCTION A L'ANALYSE ECONOMIQUE

Gilles KOLÉDA

2007 - 2008

Introduction à l'Analyse Économique

(Seconde version)[1]

Gilles Koléda
Maître de Conférences à l'Institut Universitaire de Technologie de Tours
Membre de l'équipe ERASME (Ecole Centrale Paris)

[1] Merci à tous les élèves qui ont, par leurs remarques ou corrections, permis l'amélioration de la première version de ce polycopié d'*Introduction à l'Analyse Economique*. Vos remarques, suggestions d'ajouts ou corrections sont toujours les bienvenues : gkoleda@ecp.fr
J'ai bénéficié pour la rédaction de certaines annexes et d'encadrés de ce polycopié, de l'aide de **Pascal da Costa** et **Florent Pratlong**. Qu'ils soient ici remerciés pour cette aide précieuse.

Préambule

Pourquoi étudier l'économie ?

Dans un cursus tel que le vôtre, vous abordez de nombreuses matières et bénéficiez d'enseignements techniques et théoriques qui vous serviront directement dans votre vie professionnelle future. *Quid* de l'économie ? Pourquoi intégrer l'économie dans votre formation d'ingénieur ? Rares sont ceux d'entre vous qui seront économistes : relativement peu d'entre vous utiliseront directement dans le cadre de leur métier les développements théoriques et les bagages techniques acquis dans ce cours d'introduction.

Cependant, vous allez être confrontés durant toute votre vie (professionnelle ou non) à des questions économiques (lorsque vous allez voter, vous faîtes un choix économique ; lorsque vous devrez choisir votre futur emploi également ; lorsque vous investirez ou achèterez des biens, c'est encore des choix économiques[2]…). Par ailleurs la globalisation et les mutations sectorielles actuelles (essor des secteurs technologiques, apparition de nouveaux concurrents au niveau mondial…) rendent l'appréciation des phénomènes économiques de plus en plus complexe. Alors mieux vaut vous armer dès maintenant pour être à même d'aborder ces questions et ces situations avec le recul et les connaissances nécessaires pour ne pas être submergé ou perdu devant elles (comprendre pour dominer ces problèmes). Car si tout le monde peut avoir un avis sur l'économie ou la politique économique, il paraît essentiel que le vôtre ait été confronté, au moins une fois, aux raisonnements des véritables économistes. Libre ensuite à vous de retomber dans la facilité de la vulgate économique véhiculée par les médias, ou bien de continuer à approfondir ces questions économiques en appliquant les méthodes et raisonnements de ce cours aux différentes situations auxquelles vous serez confrontés.

Comment étudier l'économie ?

L'économie est une science. Le temps où la discipline s'appelait encore l'Economie Politique a vécu. Désormais, on parle de science économique. Il existe même un prix décerné par la Banque de Suède en l'honneur d'Alfred Nobel qui récompense chaque année un chercheur en économie pour ses apports à cette discipline. Il y aurait donc la possibilité de débuter l'apprentissage de cette discipline scientifique, d'aborder certains de ces développements fondamentaux, d'acquérir la maîtrise de ses protocoles. Nous ferons en partie cet apprentissage. Mais étudier l'économie de cette manière est long et fastidieux, notamment parce que vous n'avez acquis aucune des bases de cette discipline durant votre enseignement secondaire, contrairement à la physique, à la chimie ou aux mathématiques. Il nous faut donc commencer… par le début. Or ce cours est…court (neuf séances). Il doit donc être rapide et efficace. Son intérêt ne peut vous apparaître que s'il permet de répondre rapidement à certaines des questions économiques que vous vous posez.

Plus que le désir de maîtriser les protocoles les plus fondamentaux de la science économique, il me semble que vous devez brûler de pouvoir répondre à certaines interrogations économiques (Pourquoi existe-t-il du chômage ? Quels sont les moteurs de la croissance ? Quel est l'impact des politiques économiques sur la croissance et l'emploi ? Pourquoi des crises financières surviennent-elles ? La globalisation est-elle un phénomène nouveau et inéluctable ?…), que vous avez envie de pouvoir mieux comprendre les débats économiques, d'être à même d'évoluer sans être complètement perdus dans un environnement professionnel où l'économie a un certain poids et une certaine incidence. Ce cours a donc été bâti dans la perspective d'être directement utile au plus grand nombre d'entre vous, tout en permettant aux plus motivés d'acquérir les protocoles fondamentaux de la science économique afin de poursuivre leur apprentissage de l'économie s'ils le désirent.

Pour cela, nous n'aborderons la technique et les protocoles scientifiques de l'économie qu'à travers l'examen d'un certain nombre de thématiques, lorsqu'ils apparaîtront nécessaires pour formuler la réponse à une question précise ou éclairer une situation donnée.

---

[2] A la suite de Becker (1962), certains économistes considèrent même que faire des études est une décision économique (détermination de l'investissement optimal en capital humain sur la base d'un arbitrage coûts/bénéfices). Certains économistes vont même jusqu'à analyser économiquement le nombre d'enfants optimal pour un couple ! La méthode et les raisonnement économiques peuvent s'appliquer à de nombreux domaines au-delà même du champ économique considéré comme traditionnel.

Comment sera organisé l'enseignement de l'économie ?

L'approche pour ce cours sera donc thématique. Ce faisant nous dépassons les clivages habituels de l'enseignement d'économie, entre microéconomie et macroéconomie (voir le chapitre introductif pour une description des deux domaines). Nous aborderons toujours un thème, donnant lieu à un chapitre, de la même façon. En donnant en premier lieu l'éclairage empirique nécessaire (les faits, les chiffres), puis en dégageant une ou des problématiques, c'est-à-dire des questionnements qui traversent ce thème, enfin en exposant les représentations théoriques principales de la science économique sur cette question (c'est-à-dire les modèles et les théories économiques permettant de trancher le débat ou de donner des éléments de réponses aux questions ouvertes dans ce domaine).
Les modèles théoriques éclairant la problématique d'un chapitre sont présentés en annexe de celui-ci et seront abordés plus intensivement lors des petites classes, donnant lieu à des applications et des exercices.
Bien entendu, certains raisonnements et protocoles de la science économique sont utiles à l'examen de plusieurs des thèmes retenus dans ce polycopié et certaines techniques doivent être acquises sans qu'elles soient directement utiles à l'examen d'une problématique donnée (le BA-Ba de l'économie en somme…). Celles-ci figureront alors en annexe générale et pourront être abordées une fois la lecture du chapitre introductif finie. Celui-ci se propose de vous donner quelques clés de lecture indispensable pour cette « introduction à l'analyse économique »…

Comment utiliser au mieux ce polycopié ?

Pour que ce polycopié soit le plus utile et le plus complet possible tout en demeurant agréable à lire, les développements techniques ont été regroupés en *annexes* (annexes de fin de chapitre ou annexes générales). Lorsqu'une annexe doit être lue, cela est signalé dans le corps du texte (ou bien dans une note de bas de page) en **gras et encadré** comme dans l'exemple suivant :

2.1.3.1 Tarification au coût moyen dans des structures intégrées

L'annexe **la fixation d'un prix par un monopole public** décrit précisément les conditions qui aboutissent à la fourniture d'un bien ou d'un service par un monopole public. Les économies d'échelle dues aux conditions techniques existent dans de nombreux secteurs d'activité. On qualifie alors la situation de « monopole naturel » au sens où le monopole s'avère être la structure de marché la plus pertinente. L'exemple de tel monopole naturel est donné par les industries de réseau (télécommunications, transport, distribution de gaz ou d'électricité) où les coûts fixes sont extrêmement importants et les rendements d'échelles sont croissants (voir l'annexe sur **la fonction de coût et les économies d'échelle**). Il est alors plus efficace économiquement de ne pas gaspiller de ressources en dupliquant le réseau. La concurrence est alors difficile à mettre en place et la structure économique la plus efficace est un monopole.

Que ce soit dans le corps du texte ou bien dans des notes de bas de page, il y aura aussi parfois des renvois aux articles parus dans les revues scientifiques (comme dans l'exemple ci-dessous) ou bien à des livres. Ces *références* sont le plus souvent disponibles au Centre de Documentation et peuvent intéresser ceux qui souhaitent aller plus loin ou approfondir certaines questions.

[22] Aghion Ph., Harris C., Howitt P. et Vickers J. (2001), « Competition, Imitation and Growth with Step-by-Step Innovation », *Review of Economic Studies*. Une application empirique de ce modèle est fournie dans Aghion Ph., Bloom N., Blundell R., Griffith R. et Howwit P. (2003), « Competition and Innovation : An Inverted-U Relationship », *The Quaterly Journal of Economics*, vol.120, pp.701-728.
[23] Eto (2004), « Innovation by leaders », *Economic Journal*.

Enfin, à la fin du polycopié, figure une *bibliographie* avec entre parenthèses les chapitres du polycopié qui sont approfondis dans chacun des ces ouvrages.

Table des matières

PREAMBULE **3**

Pourquoi etudier l'economie ? 3

Comment etudier l'economie ? 3

Comment sera organise l'enseignement de l'economie ? 4

Comment utiliser au mieux ce polycopie ? 4

1 INTRODUCTION **11**

1.1 Vers une definition de l'analyse economique 12

1.2 Economie positive et normative 12

1.3 Macroeconomie et microeconomie 13

1.4 Analyses conjoncturelles et structurelles 14

1.5 L'economie est-elle une science ? 14

1.6 Tres (trop) breve histoire de la pensee economique 16

1.7 L'economie neoclassique, courant dominant 18

1.8 Annexes 19

1.8.1 Les lauréats du prix de la Banque de Suède en l'honneur d'Alfred Nobel 19

2 MARCHE ET REGULATION **21**

2.1 Les defaillances de marche 23

2.1.1 La non optimalité des mécanismes de marchés en présence d'externalités 24
2.1.2 La remise en cause de l'hypothèse d'atomicité et les différentes formes de marché 29
2.1.3 Monopoles publics et ouverture à la concurrence 32

2.2 « Asymetries d'information et nouvelle microeconomie » 34

2.3 Concurrence sur le marche des biens et concurrence pour l'innovation 34

2.4 Annexes 36

2.4.1 Fonctions de coût et économies d'échelle 36
2.4.2 Rendements de la fonction de production 36
2.4.3 La notion d'élasticité-prix de la demande 37
2.4.4 Le surplus du consommateur 38
2.4.5 La situation de concurrence pure et parfaite 39
2.4.6 La situation de monopole 40
2.4.7 La concurrence monopolistique 41
2.4.8 La fixation des prix par un monopole public 43
2.4.9 L'oligopole 44

3 ROLE DE LA MONNAIE ET FINANCEMENT DE L'ECONOMIE **51**

3.1 La monnaie 52

3.1.1 La fonction d'intermédiaire des échanges de la monnaie 52
3.1.2 La fonction d'unité de compte 53
3.1.3 La fonction de réserve de valeur 53

3.2 La demande d'encaisses monetaires 53

3.2.1 La monnaie est neutre pour les auteurs « classiques » 53
3.2.2 La monnaie est « active » pour Keynes 54
3.2.3 Friedman et la vision monétariste 55

3.3 Le processus de creation monetaire 55

3.3.1 Masse monétaire et agrégats monétaires 55
3.3.2 Evolutions des agrégats monétaires et inflation 56
3.3.3 Création monétaire dans un système bancaire hiérarchisé 56
3.3.4 Le rôle de la banque centrale 57
3.3.5 Les instruments de la Banque centrale 58

3.4 Le processus de financement de l'economie 58

3.4.1 Besoins et capacités de financement des agents 59
3.4.2 Les deux formes de financement et le rôle de l'intermédiation 60
3.4.3 Le passage de l'économie d'endettement à celle de marchés financiers 62
3.4.4 La globalisation financière et ses risques 64

3.5 Annexe 67

3.5.1 Les agrégats monétaires 67
3.5.2 Le principe du multiplicateur de crédit 67
3.5.3 L'Open market et le comportement des agents 68
3.5.4 Evolution inverse de la valeur d'une obligation et des taux d'intérêt 69
3.5.5 Valeur d'une action et condition de non arbitrage 70

4 CONJONCTURE ET POLITIQUES ECONOMIQUES 71

4.1 LA POLITIQUE BUDGETAIRE 74

4.2 L'ENDETTEMENT DE L'ETAT 75

4.2.1 Stabilisateurs économiques 76
4.2.2 Les politiques budgétaires volontaristes 77
4.2.3 Les limites de la politique budgétaire 77
4.3 LA POLITIQUE MONETAIRE 78

4.3.1 Les objectifs et les instruments de la politique monétaire 78
4.3.2 La politique de lutte contre l'inflation 81
4.3.3 La construction de l'UEM et les contraintes associées 82
4.3.4 La politique monétaire dans un environnement complexe 82
4.4 CONCLUSION 84

4.5 ANNEXES 86

4.5.1 Le modèle IS-LM 86
4.5.2 Le chômage keynésien 90
4.5.3 Le modèle offre globale/demande globale 92
4.5.4 La règle de Taylor 93
4.5.5 Le modèle IS-LM…sans LM 94
4.5.6 La courbe de Phillips 95
4.5.7 Soutenabilité de la dette publique 98
4.5.8 Le pacte de stabilité : pourquoi 3% ? 98

5 LE CHOMAGE 99

5.1 QUELQUES DEFINITIONS 100

5.2 LA REALITE DU MARCHE DU TRAVAIL 102

5.3 LES REPRESENTATIONS DU FONCTIONNEMENT DU MARCHE DU TRAVAIL 104

5.3.1 La représentation néo-classique du marché du travail 104
5.3.2 « Chômage classique » et « chômage keynésien » 105
5.3.3 Le chômage d'équilibre et les approches récentes 107
5.4 LES POLITIQUES DE L'EMPLOI 108

5.4.1 Un bref état de la situation 108
5.4.2 La baisse du temps de travail : bonne ou mauvaise solution ? 110
5.4.3 Vers des réformes plus structurelles. 111
5.5 CONCLUSION 113

5.6 ANNEXES 114

5.6.1 Le modèle néo-classique du marché du travail 114
5.6.2 Le modèle de salaire d'efficience 115
5.6.3 Le modèle de négociation salariale 116
5.6.4 Le modèle WS-PS 117
5.6.5 La courbe de Beveridge 118
5.6.6 Coût d'appariement et chômage d'équilibre 119

6 OUVERTURE INTERNATIONALE ET GLOBALISATION 121

6.1 QUELQUES FAITS 122

6.1.1 D'une mondialisation à l'autre 122
6.1.2 Les aspects essentiels des échanges mondiaux 123
6.2 LE COMMERCE INTERNATIONAL 125

6.2.1 Les arguments pour le libre échange 125
6.2.2 Les nouvelles théories du commerce international 126
6.2.3 Les arguments en faveur du protectionnisme 126
6.2.4 Le rôle des Firmes multinationales 127
6.3 LA REGULATION DU COMMERCE MONDIAL 128

6.3.1 Le GATT 128
6.3.2 La mise en place de l'OMC et son agenda 129
6.4 FINANCE INTERNATIONALE 130

6.4.1 Le taux de change 130
6.4.2 La Balance des paiements 131

6.4.3 Le système monétaire et financier international ... 133
6.4.4 Vie et mort du système de Bretton Woods ... 133
6.4.5 Le régime actuel de changes flottants ... 134
6.5 ANNEXES ... 137
6.5.1 Le principe des avantages comparatifs et les gains à l'échange ... 137
6.5.2 Effets de la mise en place d'un tarif douanier ... 138
6.5.3 Effets de la création d'une union douanière ... 140
6.5.4 Le modèle Mundell - Flemming : IS-LM-BP ... 141
6.5.5 Le théorème des élasticités critiques et la courbe en J ... 145
7 LES MOTEURS DE LA CROISSANCE : EDUCATION ET INNOVATION ... 147
7.1 QUELQUES FAITS STYLISES ... 148
7.1.1 □-convergence, □-convergence et rattrapage ... 148
7.1.2 Convergence et développement ... 148
7.1.3 Résidu de Solow et paradoxe de la productivité ... 149
7.1.4 « Nouvelle économie » et fin du paradoxe de Solow ? ... 150
7.1.5 Le retard européen en matière d'éducation et d'innovation ... 152
7.1.6 Croissance et environnement ... 153
7.2 LES EFFETS ECONOMIQUES DE L'EDUCATION ... 155
7.3 LES THEORIES DE LA CROISSANCE ... 156
7.3.1 Le modèle néoclassique : l'accumulation du capital ... 156
7.3.2 Les modèles de croissance endogène fondée sur l'innovation ... 157
7.4 CONCLUSION : QUELLE POLITIQUE DE CROISSANCE ? ... 159
7.5 ANNEXES ... 160
7.5.1 Le modèle de Solow ... 160
7.5.2 L'exercice de comptabilité de la croissance ... 161
7.5.3 Le modèle de croissance endogène fondée sur l'innovation : Romer (1990) ... 162
7.5.4 Le modèle schumpétérien de croissance : Aghion et Howitt (1992) ... 166
7.5.5 Education, innovation et distance à la frontière ... 167
8 ANNEXES GENERALES ... 169
8.1 LE CADRE DE COMPTABILITE NATIONALE ... 169
8.1.1 Le Produit Intérieur Brut ... 169
8.2 LE PROGRAMME DU CONSOMMATEUR ... 171
8.2.1 Le programme statique : détermination des demandes de biens ... 171
8.2.2 Effet Revenu / Effet Substitution ... 173
8.2.3 Le programme dynamique du consommateur : Modèle de Ramsey. ... 173
8.3 LE PROGRAMME DU PRODUCTEUR ... 175
8.3.1 Programme statique : détermination des demandes de facteurs ... 175
8.4 LA BOITE D'EDGEWORTH ... 176
8.5 BREVE INTRODUCTION A LA THEORIE DES JEUX ... 177
8.5.1 Représentation d'un jeu ... 177
Détermination des équilibres ... 178
8.5.2 Le « dilemme du prisonnier » ... 179
8.5.3 La « guerre des sexes » ... 180
8.6 LES ASYMETRIES D'INFORMATION ET LA THEORIE DES CONTRATS ... 181
8.6.1 La « sélection adverse » ... 182
8.6.2 Le « hasard moral » ... 182
BIBLIOGRAPHIE ... 185

Index des figures

Figure 1 : Les démarches scientifiques en économie 15
Figure 2 : Classement des biens en fonction du degré de rivalité et d'exclusivité de leur usage 25
Figure 3 : Parts de marché dans quelques secteurs 30
Figure 4 : Fonctions de coût 36
Figure 5 : Zones des rendements d'échelle 36
Figure 6 : Demande inélastique et infiniment élastique 37
Figure 7 : Comparaison de deux élasticités 37
Figure 8 : Surplus du consommateur 38
Figure 9 : Situations de concurrence et de monopole 40
Figure 10 : Analyse des surplus 40
Figure 11 : Equilibre de concurrence 41
Figure 12 : Equilibre de monopole 41
Figure 13 : Equilibe de concurrence monopolistique 42
Figure 14 : Tarification de « moindre mal » 43
Figure 15 : L'équilibre de Cournot 45
Figure 16 : Equilibre de Stackelberg *vs.* équilibre de Cournot 49
Figure 17 : Bulle sur les valeurs 52
Figure 18 : Indice des prix à la consommation et taux directeurs de la BCE 52
Figure 19 : Evolution de M3 (haut) et de l'inflation (indice IPCH, bas) 56
Figure 20 : Financement et refinancement dans un système bancaire hiérarchisé 57
Figure 21 : Besoin net et capacité de financement 60
Figure 22 : Rencontre entre besoins et capacités de financement 60
Figure 23 : Taux d'intermédiation en France source : Comptes nationaux et Bulletin de la Banque de France 61
Figure 24 : Une nouvelle forme d'intermédiation : les OPCVM 64
Figure 25 : Indices des valeurs américaines (S&P500) 65
Figure 26 : Evolution du cours des actions et des taux d'intérêts 66
Figure 27 : « Carrés magiques de Kaldor », (années 2005 et 2006) 71
Figure 28 : Taux de croissance du PIB (rythme annuel) 72
Figure 29 : Taux de croissance de l'indice des prix à la consommation (rythme annuel) 72
Figure 30 : Déficit public en% du PIB 75
Figure 31 : Dette publique en% du PIB 75
Figure 32 : Evolution de la croissance l'indice des prix à la consommation 81
Figure 33 : Le triangle des incompatibilités de Mundell 83
Figure 34 : Arbitrage croissance – stabilité : France et Etats-Unis 84
Figure 35 : Construction de la courbe IS 87
Figure 36 : Construction de la courbe LM 88
Figure 37 : L'équilibre IS-LM 88
Figure 38 : Effet d'une politique budgétaire restrictive 89
Figure 39 : Effet d'une politique monétaire expansionniste 90
Figure 40 : Contrainte de débouchés et chômage keynésien 91
Figure 41 : Le modèle Offre Globale - Demande Globale 93
Figure 42 : Passage de la relation de Phillips originelle à la courbe inflation-chômage 96
Figure 43 : Courbe de Phillips – Etats-Unis (1961-95) (données source OCDE) 96
Figure 44 : Courbe de Phillips – France (1961-2005) (données source OCDE) 96
Figure 45 : Dérive de la relation inflation – chômage 97
Figure 46 : Taux de chômage en France (déc.67-juin2006) en % 99
Figure 47 : Chômeur, actif occupé, inactif… 100
Figure 48 : Stocks et flux de chômage, d'actifs employés et d'inactifs 101
Figure 49 : Proportion du chômage de longue durée parmi les chômeurs selon l'âge et le sexe (France, 2005) 101
Figure 50 : Equilibre du marché du travail en concurrence parfaite 104
Figure 51 : Chômage volontaire 104
Figure 52 : Le chômage « keynésien » 105
Figure 53 : Le chômage « classique » 106
Figure 54 : Taux d'activité 110
Figure 55 : Le modèle WS-PS 117
Figure 56 : Courbe de Beveridge 118
Figure 57 : Courbe de Beveridge et courbe VS 119
Figure 58 : Evolution du $ euros (1€ = x$) 134
Figure 59 : Evolution du $ et du Yen en euros 135
Figure 60 : Equilibre d'un pays en autarcie 137
Figure 61 : Equilibre d'un pays en économie ouverte cas de l'avantage comparatif en bien 1 138
Figure 62 : La mise en place d'un tarif douanier 139
Figure 63 : Mise en place d'un tarif douanier (analyse des surplus) 140
Figure 64 : Effet de la mise en place d'une union douanière 141
Figure 65 : La droite BP 142
Figure 66 : Le rééquilibrage par entrée de capitaux 143
Figure 67 : Efficacité de la politique budgétaire en régime de change fixe (mobilité imparfaite des capitaux) 143
Figure 68 : Efficacité de la politique budgétaire en régime de change fixe (mobilité parfaite des capitaux) 144

Figure 69 : Inefficacité de la politique budgétaire en régime de changes flottants (mobilité parfaite des capitaux) 144
Figure 70 : Efficacité de la politique monétaire en régime de changes flottants (mobilité parfaite des capitaux). 145
Figure 71 : La courbe en J 146
Figure 72 : Rattrapage et stagnation 149
Figure 73 : Courbe de Kuznets « environnementale » 153
Figure 74 : Effets de subventions à la R&D quand l'offre de travail qualifié est inélastique 158
Figure 75 : Dynamique du modèle de Solow 161
Figure 76 : Courbe d'indifférence 171
Figure 77 : Programme du consommateur 172
Figure 78 : Effet revenu et effet substitution 173
Figure 79 : Diagramme d'Edgeworth 176
Figure 80 : Représentation d'un jeu sous forme d'un arbre de décision 178
Figure 81 : Représentation d'un jeu sous forme matricielle 178
Figure 82 : Dilemme du prisonnier 180
Figure 83 : La « guerre des sexes » 181

Index des tableaux

Tableau 1 : Taxonomie des structures de marché 29
Tableau 2 : Agrégats monétaires de la zone euro en 2005 (en Milliards d'euros) 67
Tableau 3 : Les critères de convergence 82
Tableau 4 : Mouvements de main d'œuvre par grands secteurs d'activité (2004) 103
Tableau 5 : Déformation de la structure des emplois en France (1982-2004) Progression des CSP qualifiées (tableau du haut) Progression de la part des ouvriers qualifiés parmi les emplois ouvriers (tableau du bas) 103
Tableau 6 : Quelques rapports récents sur l'emploi 109
Tableau 7 : Taux d'indemnisation du chômage 112
Tableau 8 : L'essor du commerce mondial et des investissements directs dans différents pays développés et en développement (source : Banque Mondiale (2003), CNUCED (2003)) 123
Tableau 9 : Evolutions comparées, du commerce international et des IDE dans le monde (taux de croissance) (source : World Investment Report (2003)) 123
Tableau 10 : Tarifs douaniers sur les biens manufacturés (évolution sur longue période) (source : Bairoch (1994), OMC (2002)) 124
Tableau 11 : Parts des biens manufacturés dans le commerce 124
Tableau 12 : Parts des exportations intra-régionales 124
Tableau 13 : Les négociations commerciales multilatérales 129
Tableau 14 : Les différents comptes et soldes de la balance des paiements 132
Tableau 15 : La balance des paiements pour la France (2001) 133
Tableau 16 : Comptabilité de la croissance (1913-1987), France – source : Maddison (1991) 150
Tableau 17 : Comptabilité de la croissance (1913-1987), Etats-Unis – source : Maddison (1991) 150
Tableau 18 : Effet des TIC sur la PGF, Etats-Unis – source : Oliner et Sichel (2000) 151
Tableau 19 : Source de la croissance de la productivité du travail dans l'UE15 et aux Etats-Unis – source : van Ark et Inklaar (2005) 151
Tableau 20 : Dépenses intérieures brutes de R&D (en % du PIB) – source : OCDE, Eurostat 152
Tableau 21 : Dépenses intérieures brutes de R&D (en % du PIB) des entreprises– source : OCDE, Eurostat .. 152
Tableau 22 : Nombre de chercheurs en entreprises (pour 1000 emplois) – source : OCDE, STI 152
Tableau 23 : Dépenses annuelles au titre des établissements d'enseignements par élève/étudiant (en équivalent $ convertis sur base PPA) (4 grands pays pour la zone euro) – source : OCDE 152

1 Introduction

L'économie mondiale a connu de profondes mutations depuis deux décennies :

- La chute du mur de Berlin a entériné le triomphe de l'économie de marché sur l'économie centralisée (ainsi, même la Russie et la Chine basent désormais leur essor sur le développement de leur économie de marché),
- On a assisté à des développements importants des transferts internationaux : la circulation des biens et des services, des capitaux et des personnes n'a cessé de croître rendant le monde plus globalisé, plus interconnecté,
- A partir du milieu de la décennie 90, la diffusion de nouvelles technologies basées sur l'essor du progrès scientifique et technique (nouvelles technologies de l'information et de la communication, biotechnologies) a plongé l'économie mondiale dans une nouvelle « révolution industrielle », créant de nouveaux marchés et ouvrant des perspectives de croissance importantes,
- La construction européenne s'est poursuivie. Le marché unique puis la monnaie unique ont été mis en place, facilitant les échanges intra-européens et augmentant la taille du marché domestique pour les entreprises européennes.

A l'énoncé de ces faits, il pourrait sembler qu'un ensemble de forces positives placent l'économie européenne, et donc également française, dans une situation idéale et qu'une période de prospérité et de création de richesses sans précédente est en marche. Ce n'est pas exactement le cas ou, tout au moins, ce n'est pas la perception qu'on a de l'évolution de l'économie européenne de prime abord (il suffit pour s'en convaincre de regarder les niveaux des taux de chômage dans certains grands pays européens, constater l'atonie de la croissance européenne, se remémorer les débats et le résultat du référendum sur le projet de Constitution européenne en 2005…).

Notre vision est pessimiste, à tort ou à raison, parce qu'il existe également un certain nombre d'écueils et de problèmes qui viennent contrecarrer les éléments positifs décrits ci-dessus, éléments pourtant susceptibles de soutenir le développement économique et d'occasionner un progrès social :

- Le vieillissement de la population occidental entraîne un problème de financement des retraites et à terme un besoin de migration de main d'œuvre auquel une partie de la population européenne a parfois du mal à se résoudre,
- La concurrence des pays émergents dans les secteurs les plus traditionnels occasionne des pertes de parts de marché et donc d'emplois, ou des délocalisations qui cristallisent les angoisses des travailleurs les moins qualifiés,
- Les deux dernières décennies ont été marquées par une instabilité financière mondiale toujours plus importante (crises dans les pays émergents (crise asiatique, crise russe, crise argentine…), déficits jumeaux américains comme une épée de Damoclès sur l'économie mondiale, crise du *subprime* …) qui ont également des répercussions sur l'économie réelle,
- L'évolution préoccupante de l'environnement (dérèglement climatique, érosion de la biodiversité, épuisement progressif des ressources…) a abouti à la mise en place de mesures qui peuvent parfois handicaper les entreprises et oblige à prendre en compte ce problème structurel majeur dans les décisions les plus courantes.

La politique économique a donc encore un rôle important à jouer pour orienter les économies et les placer sur une trajectoire vertueuse (développement durable, croissance soutenable). Cependant la politique économique est rendue plus difficile par la compétition qui s'est instaurée entre les territoires nationaux (problème des délocalisations qui aboutissent à la nécessité d'attirer des investissements étrangers, à l'obligation d'être « compétitif » au niveau réglementaire et fiscal et au désir de posséder des infrastructures de pointe). De plus, la globalisation qui s'accompagne d'entreprises de taille de plus en plus importante et de flux financiers gigantesques et volatiles semble parfois incontrôlable et gêne les adaptations structurelles et les politiques conjoncturelles nécessaires au niveau national.

Ces problèmes, ces sujets, vous les connaissez ou vous en avez entendu parler. En effet, l'attention de plus en plus grande portée aux phénomènes économiques a rendu les concepts de l'analyse économique omniprésents. Ainsi le chômage, la mondialisation, la croissance, l'actualité boursière font-ils désormais partie de l'information quotidienne. Cette approche journalistique contribue à vous familiariser avec le vocabulaire des économistes, à vous donner un aperçu des problèmes existants, mais ne suffit pas à permettre une vraie compréhension des raisonnements et des mécanismes économiques sous-jacents.

Pour cela, il est nécessaire de parcourir les chemins de la pensée et de l'analyse économiques. C'est ce que proposent, autant que faire se peut[3], ce cours et ce polycopié « d'introduction à l'analyse économique ».
En tout premier lieu, il convient de s'entendre sur une définition de l'économie avant d'aborder les différents champs de l'analyse économique.

1.1 Vers une définition de l'analyse économique

La science économique est une science humaine. Elle a pour objet d'étude l'être humain. C'est également une science sociale puisqu'elle étudie le comportement des individus au sein de la société.
Toute science se donne un objet précis qu'elle tente d'analyser. La science politique procède à l'étude du pouvoir. La physique étudie les propriétés générales de la matière, de l'espace et du temps. La science économique se concentre, elle, sur l'étude de la rareté.
Le constat de départ de l'analyse économique est que les hommes éprouvent des besoins illimités mais, hélas, les ressources dont ils disposent pour les satisfaire n'existent qu'en nombre limité (d'où le phénomène de rareté). En conséquence, ils doivent faire des choix dans l'utilisation de leurs ressources (temps, budget) et dans les besoins qu'ils souhaitent satisfaire.
La rareté impose des choix. La science économique se donne pour objectif de résoudre le problème de l'allocation des ressources, c'est-à-dire d'analyser les choix imposés par la rareté, et cela en répondant aux trois questions suivantes (P. Samuelson) : *Quels biens produire ? Comment les produire ? Pour qui les produire ? (Ces questions couvrent l'ensemble des problèmes économiques de l'allocation des ressources dans la consommation et la production, jusqu'au problème de distribution et d'équité…).*

L'analyse économique **se propose d'établir la façon dont la société décide ce qu'elle doit produire, comment et pour qui elle doit le faire.**

Cette définition est cependant bien générale et il convient de réaliser plusieurs distinctions tant sur le plan des méthodes de l'analyse économique que sur le champ de l'analyse.

1.2 Economie positive et normative

La première distinction que l'on peut réaliser est celle entre analyse positive et normative.
Une analyse **positive** explique pourquoi les choses et les comportements sont ce qu'ils sont. Elle vise à montrer (décrire) le monde tel qu'il est. Une analyse **normative** cherche à définir ce que doivent être les choses et les comportements, à expliquer comment doit être le monde.

L'économie positive a donc trait aux explications objectives ou scientifiques du fonctionnement de l'économie. Elle se fixe pour objet d'expliquer comment la société prend des décisions relatives à la consommation, à la production et à l'échange de biens. Cette investigation a deux finalités : satisfaire notre curiosité en expliquant pourquoi l'économie fonctionne comme elle le fait, offrir une base pour prévoir comment l'économie réagira à des changements de situation.

L'économie normative fournit pour sa part des prescriptions ou recommandations fondées sur des jugements de valeur personnels.

Avec l'analyse positive, l'économiste espère agir comme un scientifique dépassionné. Quelle que soit son orientation politique, quelles que soient ses idées sur l'avenir ou sur ce qui est « bien », il veut d'abord s'intéresser à la façon dont le monde fonctionne dans la réalité. Il n'accorde aucune place aux jugements de valeur personnels. Il s'agit de propositions telles que: si l'on change *ceci,* alors il arrivera *cela.* En ce sens, l'économie positive se compare aux sciences comme la physique, la géologie ou l'astronomie.
La démarche positive relève seule de la science. L'économie normative est pour sa part trop influencée par des valeurs que l'on cherche à respecter.
L'économiste n'a pas pour vocation première d'édicter des règles de bonne conduite. Il peut, du reste, laisser ce soin aux autres puisque les décisions en matière de politique économique sont en définitive

[3] Il convient de garder à l'esprit que ce cours n'est qu'une introduction. On ne pourra aborder et expliquer l'ensemble des problèmes et phénomènes économiques. Les différents chapitres et sujets traités ne sont qu'une sélection bien entendue non exhaustive. Les références en bas de page et la bibliographie en fin de polycopié peuvent vous permettre d'approfondir vos connaissances en analyse économique.

prises par les hommes politiques qui détiennent le pouvoir politique, les décisions en matière de production et d'investissement étant prises par les industriels et les entrepreneurs.
Les économistes émettent des avis. Les hommes politiques et les entrepreneurs (mais en fait tous les individus !) prennent les décisions économiques.
Les économistes conçoivent cependant souvent comme un prolongement naturel de leur travail le fait de donner des conseils en matière de politique économique. Il paraît en effet absurde de connaître en profondeur les mécanismes à l'origine d'un problème (le chômage, la pauvreté, les volumes trop importants de gaz à effet de serre dans l'atmosphère, l'absence d'innovation ou de croissance…) sans tenter de le résoudre... Il est donc parfois difficile de séparer totalement le positif du normatif. L'économie conserve forcément un lien avec l'action politique même si elle aspire à être une science.
Cela ne veut pas dire que la science économique ne peut jeter aucune lumière sur des questions normatives. Nous pouvons recourir à l'économie positive pour exposer les conséquences détaillées de l'un ou l'autre choix. L'économie positive peut servir à clarifier la liste des opinions parmi lesquelles la société devra finalement faire un choix normatif.

La seconde distinction a trait au niveau retenu pour l'analyse économique.

1.3 Macroéconomie et microéconomie

Les sciences sociales opposent presque toujours des conceptions **holistes** et des conceptions **individualistes**. L'économie n'échappe pas à cette opposition. L'**individualisme méthodologique** est une méthode d'analyse des faits économiques et sociaux qui part du principe que les phénomènes étudiés peuvent être expliqués à partir de l'analyse des comportements individuels. Le **holisme** est au contraire une méthode d'analyse qui considère que les comportements individuels s'inscrivent dans un contexte global prédéterminé. En conséquence, l'étude du contexte global est nécessaire pour comprendre les actes individuels. En science économique, ce débat renvoie à la séparation initiale existant entre microéconomie et macroéconomie.
La microéconomie relève de l'individualisme méthodologique et prend pour point de départ l'analyse économique à l'échelle de l'agent (l'individu représentatif que celui-ci soit un consommateur, un producteur ou les deux en même temps). Elle choisit un agent type puis propose d'agréger les décisions individuelles pour passer aux grandeurs globales. Quand l'analyse microéconomique néglige des effets induits indirects[4], on la qualifie *d'analyse partielle* (Raisonnement « toute chose égale par ailleurs »). Quand l'analyse microéconomique prend en compte l'ensemble des interactions sur tous les marchés, on débouche sur des analyses dites d'*équilibre général*.

A l'inverse la macroéconomie s'intéresse aux relations entre les grandeurs globales. La *macroéconomie* insiste sur les interactions dans l'ensemble du système économique. Elle simplifie délibérément les blocs constitutifs de l'analyse pour étudier, dans un cadre qui reste maniable, la totalité des interactions au sein de l'économie.
Il n'est pas toujours possible de donner des fondements microéconomiques à l'analyse macroéconomique, ce qui peut justifier une approche purement macroéconomique pour expliquer des phénomènes que ne peut expliquer la microéconomie. En effet, les conclusions de l'analyse microéconomique ne sont pas forcément généralisables au niveau macroéconomique (problème du *no bridge*). **Exemple :** *Une baisse des salaires à l'intérieur d'une entreprise peut faire augmenter les profits de celle-ci (si les salariés conservent la même productivité). Toutes les entreprises peuvent faire ce raisonnement. Cependant, les salaires des unes représentent la demande qui s'adresse aux autres. Une baisse généralisée des salaires peut faire diminuer les profits.* Ce qui était vrai au niveau individuel ne l'est plus au niveau global.
Cependant les recherches les plus récentes en économie tendent à faire disparaître cette dichotomie entre microéconomie et macroéconomie (intégration des comportements des agents dans les modèles macroéconomiques…).
A propos de cette opposition entre macroéconomie et microéconomie (qui donc tend à disparaître), une petite remarque. Il est de bon ton de penser que la microéconomie et son prolongement, l'économie industrielle, vous seraient seules nécessaires dans votre formation d'ingénieur. Il me semble au contraire que la seule connaissance de la modélisation des comportements individuels ne saurait vous suffire. Il vous faut à la fois comprendre les mécanismes de détermination des choix individuels, c'est entendu, mais également savoir comment ceux-ci répondent à l'environnement

[4] Dans certains cas, il se peut que les effets indirects ne soient pas très importants et les économistes pourront alors raisonnablement concentrer leurs efforts sur une analyse très détaillée d'activités ou de secteurs particuliers.

économique qui entoure les individus. La macroéconomie semble donc tout aussi importante dans votre formation, car elle va vous donner les bases pour une juste appréciation de l'environnement économique de l'entreprise. Par ailleurs, le parti pris de ce cours d'aborder des problématiques économiques majeures et non des techniques économiques essentielles évacue ce débat. Ce parti pris nous amènera souvent à aborder les questions d'un point de vue d'abord macroéconomique avant de s'intéresser aux comportements microéconomiques et à leur capacité explicative des phénomènes globaux.

1.4 Analyses conjoncturelles et structurelles

Une troisième distinction peut être ajoutée aux deux précédemment abordées (analyse normative ou positive, analyse macro ou microéconomique). C'est la distinction que l'on peut faire entre l'analyse de problèmes conjoncturels et l'analyse de problèmes structurels.
Cette distinction renvoie à l'horizon d'analyse nécessaire à l'éclairage du problème abordé. Ainsi les problèmes de développement, d'environnement, de financement des retraites s'inscrivent dans le long terme. Ce sont des problèmes structurels dont l'horizon dépasse largement ce qu'on peut entendre comme le court terme. En revanche, la prévision du taux de croissance au prochain trimestre, les raisons d'une possible remontée des taux d'intérêt de la Banque Centrale Européenne, les explications possibles à l'augmentation des prix des matières premières[5] sont des problèmes plus conjoncturels qui s'inscrivent dans le court ou le moyen terme. Les outils d'analyse ne sont pas les mêmes.
Les différents chapitres de ce polycopié utiliseront les analyses des deux types. Le chapitre sur « les déterminants de la croissance » est bien sûr un chapitre éminemment structurel, tandis que le chapitre « Conjoncture et politiques économiques » est, comme le titre le suggère, essentiellement fondé sur des éléments d'analyse conjoncturelle. Mais nous verrons que cela n'est pas aussi tranché qu'on pourrait le penser car, par exemple, le niveau de la dette publique (plutôt un élément structurel) influe sur le degré de liberté de la politique économique de court terme. A l'opposé, les politiques tournées vers la croissance (éducation, innovation) obligent à des investissements qui ont des répercussions sur l'allocation des ressources à court terme et donc des effets conjoncturels…

Ces distinctions faites, on peut s'intéresser rapidement à la méthodologie de l'analyse économique en tentant de répondre à la question suivante : l'économie peut-elle être considérée comme une science ?

1.5 L'économie est-elle une science ?

On assimile communément science et connaissance exacte, universelle et vérifiable (pouvant s'exprimer par des lois). Le terme de science vient de *scientia* qui signifie connaissance correcte. L'absence de laboratoire et la rareté de l'expérimentation (malgré le recours à l'économétrie, c'est-à-dire au traitement statistique des données économiques) paraissent devoir restreindre la portée de l'économie comme science.
De plus, l'analyse économique bénéficie de l'avancée progressive de la connaissance sans que l'on puisse parler de découvertes (une découverte est en effet une connaissance de portée générale s'appliquant au monde réel, une connaissance ayant un caractère irréversible qui permet de trouver ce qui était caché ou inconnu).
L'économie n'entre pas dans une telle problématique. Le discours en analyse économique est une accumulation de thèses et de connaissances qui coexistent pendant de longues périodes sans forcément se substituer. Il y a multiplication des écoles de pensée et des théories contradictoires sur un même sujet ce qui peut paraître étonnant aux représentants des sciences exactes. Il peut exister un courant de pensée dominant durant une certaine période (voir le bref historique de la sous-section suivante) mais sans pour autant qu'on puisse parler de paradigme scientifique s'imposant à tous les économistes.
L'économie peut malgré les objections ci-dessus être considérée comme une science de par la démarche méthodologique qu'elle adopte lorsque celle-ci est rigoureuse.
L'analyse scientifique commence par des hypothèses et doit se prolonger par des expérimentations. Pour Karl Popper [1963], une proposition est scientifique lorsqu'elle peut être réfutée (c'est à dire quand il est possible de la confronter aux faits). Une hypothèse scientifique est donc une proposition qui comporte en elle la possibilité de sa réfutation (Il est toujours possible de prouver une erreur, alors qu'on ne peut jamais établir une vérité…). Une analyse est donc scientifique lorsqu'elle peut être réfutée par l'observation des faits. L'expérimentation en science économique étant cependant difficile,

[5] Sauf si elle provient de l'épuisement des ressources auquel cas c'est plutôt un problème structurel…

le caractère scientifique de l'économie pourrait sembler s'éloigner. Mais l'expérimentation peut également se révéler impossible dans certaines sciences dures (clonage humain, essais nucléaires, climatologie…) ce qui remet en cause le critère de scientificité de Popper. Ce n'est pas parce qu'on ne peut pas faire d'expériences que la démarche en économie n'est pas scientifique (ouf !...).
L'expérimentation redevient possible si l'on construit des modèles qui permettent des simulations à l'aide d'outils mathématiques et statistiques. L'économiste peut s'appuyer sur l'histoire et sur les séries statistiques pour valider son modèle théorique et ses hypothèses. Les hypothèses en économie doivent s'entendre comme des probabilités (exemple : le consommateur est le plus souvent rationnel. Ce n'est pas parce qu'un individu ne l'est pas que l'hypothèse de rationalité doit être rejetée).

Nous pouvons préciser les démarches et méthodes adoptées par la science économique en définissant rapidement les termes d'hypothèse, de loi et de modèle.
Une **loi** est un raisonnement théorique destiné à donner une représentation des liens entre les variables économiques. Les lois sont fondées sur des hypothèses plus ou moins contraignantes. Une **hypothèse** est une simplification de la réalité qui rend possible la formulation de théories compréhensibles et utilisables. Les hypothèses et les lois permettent de construire des **modèles**. Un modèle est un ensemble d'hypothèses et de lois qui donne une représentation théorique du fonctionnement de l'économie. Les modèles peuvent être confrontés aux faits. La validité d'une théorie repose sur la capacité de ses conclusions à expliquer les faits. Le raisonnement en économie peut donc être soit inductif (les faits induisent la théorie) soit déductif (on déduit la théorie des hypothèses) (cf. figure (1)).

La scientificité de l'économie repose donc à la fois sur les démarches inductive et déductive (particulièrement cette dernière), appuyées par un processus de vérification empirique des hypothèses permise par le recours à l'économétrie ou les expériences.

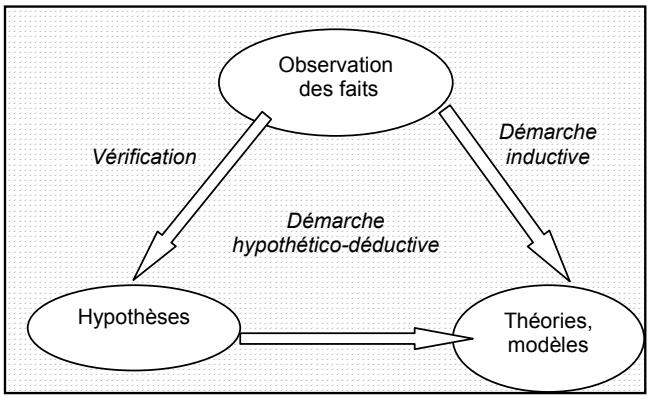


Figure 1 : Les démarches scientifiques en économie

Ce qui est souvent reproché à l'économie, et ce qui peut expliquer que certains nient à l'économie son statut de science, est qu'elle tend parfois à se focaliser, plus particulièrement le courant néo-classique dominant, sur une démarche hypothético-déductive sans toujours recourir à la vérification empirique des hypothèses. La cohérence interne des modèles depuis les hypothèses jusqu'aux résultats, appuyée sur des raisonnements mathématiques rigoureux, est seule invoquée pour justifier des résultats normatifs, sans qu'aucun lien avec la réalité ne vienne corroborer ceux-ci (démarche hypothetico-déductive isolée).

Le questionnement sur le caractère scientifique ou non de l'économie ne doit pas masquer une question tout aussi importante : l'économie est-elle utile ?
Les décisions de politiques économiques (ou bien les décisions d'une agence de régulation sectorielle, du conseil de la concurrence, de la banque centrale…) peuvent s'inspirer des théories économiques. Les gouvernants ou les responsables de ces institutions prennent la plupart de leurs décisions en référence à des courants théoriques (théorie Keynésienne conduisant aux politiques de relance ou de redistribution, monétarisme conduisant à la rigueur monétaire…). C'est d'ailleurs l'objectif de ce cours de pouvoir interpréter et comprendre ces décisions avec l'arrière-plan théorique nécessaire. Il ne faut cependant pas exagérer le poids de la théorie économique dans la conduite des politiques économiques. Les gouvernants ou les entrepreneurs prennent en considération de multiples contraintes qui découlent de la situation réelle et non théorique.

Il convient que vous ayez à l'esprit les grands courants de pensée qui ont jalonné la construction de l'analysé économique depuis Adam Smith, fondateur de l'école classique à la fin du XVIII^{ème} siècle, jusqu'aux plus récents. Pour ce faire, la section suivante dresse un panorama rapide (et bien sûr largement non exhaustif) de l'histoire de l'analyse économique.

1.6 Très (trop) brève histoire de la pensée économique

Une réflexion économique véritablement autonome n'apparaît qu'à la fin de XVIII^{ème} siècle avec les **économistes classiques** contemporains de la révolution industrielle (on peut citer les analyses des mercantilistes ou des physiocrates au XVI^{ème} et XVII^{ème} comme précurseurs). Les économistes classiques, Adam Smith puis David Ricardo, appuient leurs raisonnements sur une théorie de la valeur dite de la valeur travail : La valeur d'un bien est la quantité de travail incorporé dans celui-ci. Les richesses produites proviennent donc de la plus ou moins grande disponibilité de travail dans l'économie. Smith met ainsi en avant l'importance de la division du travail, source d'augmentation de la productivité. Les auteurs classiques (A. Smith, D. Ricardo, J.-S. Mill en Angleterre, J.-B. Say en France…) considèrent que le libéralisme économique est l'état idéal et naturellement favorable à l'augmentation des richesses produites. Pour eux l'Etat doit intervenir aussi peu que possible et aucun obstacle à la circulation des marchandises ne doit être érigé (les auteurs classiques sont donc également libre-échangistes). Dans la Recherche sur la nature et les causes de la richesse des nations (1776), Adam Smith énonce la parabole de la main invisible : Tout individu poursuivant son intérêt purement égoïste et individuel œuvre, malgré lui, pour l'intérêt collectif et la prospérité générale comme guidé par une main invisible (apologie du marché et du laisser-faire). Le français Jean-Baptiste Say énonce pour sa part la loi des débouchés (les produits s'échangent contre des produits, car la monnaie ne peut être désirée pour elle-même dans le schéma d'analyse classique, de sorte que les crises de surproduction sont impossibles et les interventions de l'État ne sont que perturbantes). Keynes reformulera quelques années plus tard la loi des débouchés sous la forme suivante ; « l'offre crée sa propre demande » pour la critiquer et la réfuter.

Au début des années 1870, l'école **néo-classique** marque une certaine rupture dans l'évolution de l'analyse économique (jusque-là science de l'accumulation des richesses, l'économie devient la science de la rareté et de l'allocation des ressources rares). Les auteurs néo-classiques s'appuient sur une nouvelle approche conceptuelle de la valeur. Ce qui fait la valeur d'un bien n'est pas la quantité de travail nécessaire à sa conception mais l'utilité qu'il procure à celui qui le consomme (passage de la « valeur travail » à la « valeur utilité »). Les auteurs néo-classiques (Stanley Jevons en Angleterre, Carl Menger en Autriche et Léon Walras, un français exilé en Suisse) introduisent également le raisonnement à la marge d'où le nom de marginalistes souvent retenu pour les désigner : l'utilité procurée par la prochaine unité de bien consommé (utilité marginale) doit être supérieure au prix de ce bien (désutilité marginale) pour que le consommateur réalise cet achat[6]. L'approche néo-classique utilise le raisonnement à partir d'un individu représentatif rationnel cherchant à maximiser sa satisfaction (approche microéconomique) et introduit l'utilisation des mathématiques en économie. Léon Walras, dans ses Eléments d'économie politique pure (1874) envisage une approche en termes d'équilibre général, c'est-à-dire un équilibre simultané sur tous les marchés de l'économie (les différents marchés de biens et services, le marché des capitaux, le marché du travail). La flexibilité des prix fait que l'équilibre obtenu est un équilibre stable. Les prix s'ajustent jusqu'à obtenir l'égalité entre offre et demande sur tous les marchés. Dans cette économie où le rôle de la monnaie est joué par un bien étalon, la loi de Walras indique que si n-1 marchés sont en équilibre alors le n^{ème} marché l'est également. Il s'agit d'une conséquence de la loi de Say (l'offre est globalement égale à la demande). Vilfredo Pareto, successeur de Léon Walras à l'Université de Lausanne, introduit pour sa part la notion d'équilibre socialement le plus profitable qu'on qualifiera par la suite d'optimum de Pareto : l'optimum de Pareto est une situation dans laquelle on ne peut améliorer la satisfaction d'un individu sans détériorer celle d'au moins un autre. Sous certaines conditions, notamment l'existence d'une situation de concurrence pure et parfaite, le premier théorème de l'économie du bien-être montre que tout équilibre de marché est une situation optimale au sens de Pareto. Comme les classiques, les néo-classiques sont donc dans l'ensemble des

[6] L'individu utilitariste cherche toujours à maximiser son utilité. Il est de plus rationnel, ce qui signifie qu'il prend toujours les bonnes décisions pour parvenir à cet objectif. C'est le fameux *homo economicus* largement critiqué par les contempteurs du modèle néo-classique. De la même façon, le raisonnement marginal s'applique à la décision de production. Le producteur, également rationnel, embauche un salarié supplémentaire uniquement si le supplément de production réalisé grâce à ce travailleur (production marginale) est supérieur au coût de ce salarié (le salaire versé).

défenseurs du libéralisme économique puisque les marchés concurrentiels permettent d'aboutir à la meilleure situation d'un point de vue social.

John Maynard **Keynes**, contemporain de la crise des années 1930, se refuse à accepter les préceptes classiques et néo-classiques d'ajustement automatique des marchés par les prix, sensés résoudre les problèmes du chômage et de la déflation. Il adopte dans La Théorie Générale de l'emploi, de l'intérêt et de la monnaie (1936) une approche macroéconomique afin de mettre en exergue les relations entre les grandes variables économiques. Keynes, en prenant pour point de départ la demande et non l'offre, montre qu'il peut exister un équilibre de sous-emploi et de sur-production durable. Il réfute donc la loi de Say en mettant en avant le fait que la monnaie peut être désirée pour elle-même et thésaurisée, et qu'elle n'est pas un simple intermédiaire des échanges comme chez les classiques. Pour soutenir la demande (consommation des ménages et investissement), Keynes pense qu'il peut être judicieux d'agir sur les anticipations des entrepreneurs pour rendre ceux-ci plus optimistes quant à leurs perspectives de débouchés. Keynes envisage ainsi d'utiliser une politique budgétaire expansionniste pour stimuler la demande, accompagnée par une politique monétaire accommodante pour relancer l'investissement. Keynes met ainsi en avant le rôle positif d'une intervention de l'Etat pour réguler les économies de marché en situation de crise.

Dès 1937, John Hicks[7] tente de montrer que l'analyse Keynésienne n'est qu'un cas particulier de l'approche néo-classique en élaborant le modèle IS/LM (voir annexe 1 du chapitre 4). A partir des années 1950, une « **synthèse néo-classique** » se met en place qui incorpore l'approche keynésienne au corpus néo-classique. Paul Samuelson[8], auteur de nombreux ouvrages et manuels est l'une des chevilles ouvrières de cet amalgame. Les idées keynésiennes vont inspirer les politiques économiques menées dans les pays développés durant les trois décennies qui suivent la seconde guerre mondiale, avec un certain succès (période des Trente Glorieuses en France selon l'expression de Jean Fourastié).

La théorie et les politiques keynésiennes vont cependant être critiquées à partir des années 1970, lorsque la *stagflation* (inflation et chômage simultanément) montre l'inefficacité voire la nocivité des recettes keynésiennes. Milton **Friedman**[9], chef de file des **monétaristes**, mettant l'accent sur la monnaie, conteste l'efficacité des politiques keynésiennes de soutien à la demande. L'inflation engendrée par ces relances fait que le supplément de revenu dont bénéficient les agents n'est que transitoire. Les ménages sachant que leur pouvoir d'achat sera érodé par la hausse des prix conséquente à la relance vont progressivement épargner plutôt que de consommer de sorte que l'effet de la relance sera de moins en moins important voire nul. Seule l'inflation demeure (l'inflation importante des années 70 semble lui donner raison et place la théorie Keynésienne sur la sellette). Friedman indique qu'il faut procéder à un contrôle strict de la masse monétaire en circulation car si la monnaie peut avoir un rôle à court terme, elle est neutre à long terme.

Les débats entre les tenants de l'intervention de l'Etat dans l'économie (héritiers de Keynes) et les partisans d'une économie plus libérale perdurent à travers les courants les plus contemporains : nouveaux classiques et nouveaux keynésiens.
La **nouvelle école classique**, parfois qualifiée d'école des anticipations rationnelles en raison de l'importance qu'elle accorde à cette hypothèse, continue la critique des politiques économiques interventionnistes. Former une anticipation rationnelle revient à connaître le vrai modèle de l'économie pour décider de son action future[10]. En partant des hypothèses que les agents économiques forment des anticipations rationnelles les principaux auteurs de ce courant (notamment Robert Lucas[11] et Robert Barro) se sont attachés à montrer l'inefficacité des politiques conjoncturelles. Lucas a ainsi approfondi la critique monétariste en montrant que la monnaie est neutre aussi bien à court terme qu'à long terme tandis que Barro a montré l'inefficacité des politiques budgétaires expansionnistes (lorsqu'il y a un déficit budgétaire les agents anticipent une augmentation future de leurs impôts et épargne plutôt que de consommer).

[7] « Prix Nobel » d'économie en 1972 (voir l'annexe en fin de ce chapitre sur **les lauréats des « prix Nobel » d'économie**). Le prix de la Banque de Suède en l'honneur d'Alfred Nobel est décerné chaque année à un ou plusieurs économistes vivants pour leur contribution significative à la discipline.
[8] « Prix Nobel » d'économie en 1970.
[9] « Prix Nobel » d'économie en 1976.
[10] L'anticipation rationnelle n'est pas nécessairement l'anticipation exacte de l'avenir puisqu'elle n'est rationnelle que par rapport à l'information disponible.
[11] « Prix Nobel » d'économie en 1995.

Les **nouveaux keynésiens** (Joseph Stiglitz[12] par exemple) s'attachent pour leur part à étudier les rigidités de prix et de salaires qui freinent les ajustements sur le marché des biens et services et sur le marché du travail. Du fait de l'existence de ces rigidités, le fonctionnement des marchés ne peut aboutir spontanément à l'équilibre de plein emploi ce qui justifie une intervention de l'État, censé permettre de suppléer aux défaillances de marché.

1.7 L'économie néoclassique, courant dominant

Un rapide survol des lauréats du prix de la Banque de Suède en l'honneur d'Alfred Nobel (« prix Nobel » d'économie) qui récompense chaque année, depuis 1969, un ou plusieurs économistes pour leur contribution significative à la discipline (cf. annexe de ce chapitre) permet de s'apercevoir de la place importante occupée par les économistes d'obédience plutôt classique (auteurs de la synthèse ; Hicks, Samuelson, Nouveaux classiques ; Lucas, Kydland, Prescott, théoricien des choix publics ; Buchanan, théoriciens de l'équilibre général : Debreu, Allais…) auxquels il convient d'ajouter les économistes qui utilisent pour leurs travaux l'arrière plan méthodologique offert par le corpus théorique néoclassique (Ohlin en économie internationale, Becker utilisant la microéconomie de manière extensive dans de nombreux domaines…).
Un autre survol des manuels universitaires les plus répandus finirait de vous convaincre que l'économie qui s'est imposée (que l'on s'en félicite ou que l'on regrette cet état de fait par manque de pluralisme méthodologique) est le résultat de la synthèse néoclassique.
Nous allons donc aborder l'économie principalement en nous basant sur les apports, les théories et les méthodes du courant néo-classique. Nul intérêt de vous imposer les discours d'une chapelle ou ses éléments d'analyse s'ils ne sont pas les plus usités et les plus répandus. D'autres types d'analyses existent (institutionnalisme, régulationisme, évolutionnisme…) mais aucun de ces courants ne couvre totalement, avec à sa disposition un ensemble cohérent d'outils et de méthodes, le champ d'analyse de l'économie néoclassique.

En résumé, l'économie est une discipline qui se construit comme une science des choix individuels et collectifs. C'est un ensemble de connaissances quantitatives et qualitatives, organisées de manière scientifique, sur les trois niveaux de la construction scientifique : (i) la mesure et la description, (ii) l'explication, et (iii) la prescription.
Plus on va de la connaissance empirique vers l'analyse théorique, puis de la théorie économique vers la politique économique, plus les débats apparaissent et deviennent vifs. L'économie est un débat vivant et passionnant et non une science abstraite ou un champ de connaissances désincarnées. C'est ce que nous allons voir dans les différents chapitres qui vont suivre.

[12] « Prix Nobel » d'économie en 2001.

1.8 Annexes

1.8.1 Les lauréats du prix de la Banque de Suède en l'honneur d'Alfred Nobel

1969	**Ragnar FRISCH** (1895-1973, Norvège) et **Jan TINBERGEN** (1903-1994, Pays-Bas)	Pour avoir développé et appliqué des modèles dynamiques pour l'analyse des processus économiques.
1970	**Paul A. SAMUELSON** (1915-, USA)	Pour les travaux scientifiques à travers lesquels il a développé une théorie statique et dynamique de l'économie et activement contribué à l'élévation du niveau d'analyse en économie.
1971	**Simon KUZNETS** (1901-1985, Russie)	Pour son interprétation, fondée sur l'analyse empirique de la croissance économique qui a permis de développer et d'approfondir la connaissance de la structure et du processus socio-économique du développement.
1972	**John R. HICKS** (1904-1989, GB) et **Kenneth J. ARROW** (1921-, USA)	Pour leur contributions pionnières à la théorie de l'équilibre général et la théorie du bien-être.
1973	**Wassily LEONTIEF** (1901-1985, Russie)	Pour l'élaboration de la méthode *input-output* et son application aux problèmes économiques importants.
1974	**Gunnar MYRDAL** (1898-1987, Suède) et **Friedrich August Von HAYEK** (1899-1992, Autriche)	Pour leur travaux pionnier sur la théorie monétaire et des fluctuations économiques et pour leur analyse pénétrante de l'interdépendance des phénomènes économiques, sociaux et institutionnels.
1975	**Leonid V. KANTOROVICH** (1912-1986, Russie) et **Tjalling C. KOOPMANS** (1910-1985, Pays-Bas)	Pour leur contribution à la théorie de l'allocation optimale des ressources.
1976	**Milton FRIEDMAN** (1912-2006, USA)	Pour ses réalisations dans le champ de l'analyse de la consommation, de l'histoire et de la théorie monétaire et pour sa démonstration de la complexité de la politique de stabilisation.
1977	**Bertil OHLIN** (1899-1979, Suède) et **James E. MEADE** (1907-, GB)	Pour leurs travaux ouvrant de nouvelles voies dans les domaines du commerce international et des mouvements internationaux de capitaux.
1978	**Herbert SIMON** (1916-, USA)	Pour sa recherche pionnière sur le processus de prise de décision dans les organisations économiques.
1979	**Theodore W. SCHULTZ** (1902-, USA) et **Athur LEWISE** (1915-1991, GB)	Pour leur recherche pionnière en économie du développement avec une considération particulière pour les problèmes des pays en développement.
1980	**Lawrence R. KLEIN** (1920-, USA)	Pour la création des modèles économétriques et leur application à l'analyse économique des fluctuations et des politiques économiques.
1981	**James TOBIN** (1918-2002, USA)	Pour son analyse des marchés financiers et leurs relations avec les décisions de dépenses, l'emploi, la production et les prix.
1982	**George J. STIGLER** (1911-1991, USA)	Pour ses études décisives sur la structure industrielle et le fonctionnement des marchés, ainsi que sur les causes et les effets de la réglementation publique.
1983	**Gérard DEBREU** (1921-, France/USA)	Pour l'introduction de nouvelles méthodes d'analyse dans la théorie économique et pour sa reformulation rigoureuse de la théorie de l'équilibre général.
1984	**Richard STONE** (1913-1991, GB)	Pour sa construction fondamentale à l'élaboration de systèmes de comptabilité nationale qui ont permis d'améliorer sensiblement les conditions d'analyse empirique de l'économie.
1985	**Franco MODIGLIANI** (1918-2003, Italie)	Pour son œuvre de pionnier en matière d'analyse de l'épargne et des marchés financiers.
1986	**James M. BUCHANAN Jr.** (1919-, USA)	Pour son développement des bases contractuelles et constitutionnelles de la théorie de la prise de décision politique et économique.
1987	**Robert M. SOLOW.** (1924-, USA)	Pour ses contributions à la théorie de la croissance économique.
1988	**Maurice ALLAIS** (1911-, France)	Pour ses contributions pionnières à la théorie des marchés et l'utilisation efficace des ressources.
1989	**Trygge. HAAVELMO** (1911-, Norvège)	Pour sa clarification de la théorie des probabilités, les fondations de l'économétrie et son analyse des structures économiques simultanées.

1990	**Harry M. MARKOVITZ** (1927-, USA), **Merton M. MILLER** (1923-2000, USA) et **William F. SHARPE** (1934-, USA)	Pour leur travaux pionniers en théorie de la finance.
1991	**Ronald H. COASE** (1910-, GB)	Pour sa découverte et clarification de l'importance des coûts de transaction et des droits de propriété pour la structure institutionnelle et le fonctionnement de l'économie.
1992	**Gary S. BECKER** (1930-, USA)	Pour avoir étendu le domaine de l'analyse microéconomique à un vaste champ de comportements et interactions humains, incluant les comportements hors-marché.
1993	**Robert F. FOGEL** (1926-, USA) et **Douglass C. NORTH** (1920-, USA)	Pour avoir renouvelé la recherche en histoire économique en appliquant la théorie économique et les méthodes quantitatives afin d'expliquer les changements institutionnels.
1994	**John C. HARSANYI** (1920-, Hongrie), **John F. NASH** (1928-, USA) et **Reinhard SELTEN** (1930-, Allemagne)	Pour leur analyse pionnière des équilibres dans la théorie des jeux non-coopératifs.
1995	**Robert LUCAS** (1937-, USA)	Pour avoir développé et appliqué l'hypothèse d'anticipations rationnelles, et par là même avoir transformé l'analyse macroéconomique et approfondi notre compréhension de la politique économique.
1996	**James A. MIRRLEES** (1936-, GB) et **William VICKREY** (1914-1996, Canada)	Pour leurs contributions fondamentales à la théorie économique des incitations en présence d'asymétrie d'information.
1997	**Robert C. MERTON** (1944-, USA) et **Myron S. SCHOLES** (1941-, USA)	Pour une nouvelle méthode de détermination de la valeur des produits dérivés.
1998	**Amartya SEN** (1933-, Inde)	Pour sa contribution à l'économie du bien-être.
1999	**Robert A. MUNDELL** (1932-, Canada)	Pour son analyse des politiques fiscale et monétaire en régimes de change différenciés et son analyse des zones monétaires optimales.
2000	**James J. HECKMAN** (1944-, USA) et **Daniel L. MacFADDEN** (1937-, USA)	Pour son développement de la théorie et des méthodes pour analyser des échantillons sélectifs. Pour son développement de la théorie et des méthodes pour analyser les choix discrets.
2001	**George A. AKERLOF** (1940-, USA), **A. Michael SPENCE** (1943-, USA) et **Joseph E. STIGLITZ** (1943-, USA)	Pour leur analyse des marchés avec asymétrie d'information.
2002	**Daniel KAHNEMAN** (1934-, Israël) et **Vernon L. SMITH** (1927-, USA)	Pour avoir intégrer des éléments de la recherche en psychologie en science économique, particulièrement en ce qui concerne le jugement humain et la prise de décision en incertitude. Pour avoir fait des expériences en laboratoire un outil dans l'analyse économique empirique, spécialement dans l'étude des mécanismes alternatifs au marché.
2003	**Robert F. ENGLE** (1942-, USA) et **Clive W.J. GRANGER** (1934-, GB)	Pour ses méthodes d'analyse des séries temporelles économiques avec une volatilité dépendant du temps (ARCH). Pour sa méthode d'analyse des séries temporelles économiques avec un trend commun (cointégration).
2004	**Finn E. KYDLAND** (1943-, Norvège) et **Edward C. PRESCOTT** (1940-, USA)	Pour leurs contributions à la macroéconomie dynamique : la consistance temporelle des politiques économiques et les forces conduisant les cycles d'affaires.
2005	**Robert J. AUMANN** (1930-, Israël) et **Thomas C. SCHELLING** (1921-, USA)	Pour avoir amélioré notre compréhension des conflits et coopérations à travers l'analyse de la théorie des jeux.
2006	**Edmund S. PHELPS** (1933-, USA)	Pour son analyse des arbitrages intertemporels en matière de politique macroéconomique.

2 Marché et régulation

Dans une économie de marché, les volumes de l'offre et de la demande s'ajustent grâce à des variations du niveau des prix. C'est une loi que l'on retrouve en premier lieu au niveau d'un secteur donné ou d'une production particulière (niveau microéconomique), mais qui transposée au niveau macroéconomique traduit un ajustement mutuel (équilibre général). L'offre et la demande sont tous deux des moteurs de l'économie à travers leur action dans cet ajustement mutuel. L'importance de leur rôle respectif dépend cependant des situations et des périodes.

Dans le long terme ou durant les périodes de forte innovation, on assiste à l'émergence de nouveaux acteurs produisant de nouveaux biens. L'accent est alors mis sur l'offre (rôle moteur de l'investissement, des profits). On retrouve la logique sous-jacente à la loi des débouchés de J.-B. Say ; « Tout offre crée sa propre demande », car les revenus distribués par la mise sur le marché d'un nouveau produit ouvre des débouchés pour les autres produits.

Dans le moyen ou le court terme, il peut cependant apparaître un décalage entre les capacités de production et la demande effective de sorte que la demande peut s'avérer inférieure à l'offre et contraindre celle-ci (notion d'équilibre de sous-emploi que nous reverrons dans le chapitre sur la conjoncture et les politiques économiques). Si le cadre d'analyse requis dans ce cas est le plus souvent considéré comme le cadre d'analyse keynésien, certains économistes (économistes de l'offre, A. Laffer par exemple) ont cherché des réponses théoriques privilégiant une stimulation de l'offre.

Dans les périodes d'application des innovations ou lorsqu'on tend vers une surproduction du fait de la consommation de masse (trente glorieuses), l'accent est donc mis sur la demande (consommation, revenu des ménages) qui constitue le moteur de l'économie.

Quel que soit le moteur considéré et quels que soient les primats théoriques retenus, l'ajustement correct de l'offre et de la demande sur les marchés des biens et des services requiert de préserver la concurrence, qui apparaît comme une condition première de l'efficacité économique.

Cependant, on va le voir, la concurrence ne doit pas être forcément considérée comme un état naturel de l'économie. Elle doit être organisée, soutenue, surveillée, par les autorités publiques (politique de la concurrence). C'est ce qu'on nomme la régulation de l'économie.

Se pose alors la question de la meilleure régulation possible de l'économie pour que celle-ci soit la plus efficace possible et utilise pleinement ses facteurs de production. La construction du grand marché européen a amené l'Europe à se focaliser sur la mise en place de marchés concurrentiels « partout où cela est possible », au détriment peut-être parfois de la promotion de l'innovation qui selon l'économiste Schumpeter apparaît le plus souvent dans les secteurs ou dans les entreprises disposant d'un certain degré de monopole. « *L'Europe est Smithienne alors que les Etats-Unis sont Schumpétériens* » relevait Daniel Cohen dans *Le monde* en 2002. Depuis près de vingt ans, l'Europe a fait le choix politique de bâtir un vaste marché de concurrence sur son territoire. Certains semblent même regretter qu'elle n'ait fait que ça, assujettissant le plus souvent sa politique économique à cette ambition et abandonnant sur l'autel de cette construction bon nombre d'autres objectifs. On peut ainsi considérer que la mise en place de la monnaie unique (1999 puis 2002 pour la circulation des euros) n'est apparue nécessaire que pour parachever l'Acte Unique signé en 1986 (aboutissant en 1992 à la constitution du grand marché, libre circulation des biens, des services et des capitaux).

Pourquoi l'Union Européenne a-t-elle consacrée autant d'énergie à l'établissement de ce grand marché concurrentiel ? Celui-ci a-t-il réellement profité aux européens ? Qu'attendre des prochaines ouvertures à la concurrence des marchés de l'énergie par exemple ? Pourquoi d'ailleurs certains secteurs étaient historiquement en position de monopole et appartenaient au secteur public ? C'est à l'ensemble de ces questions que nous allons répondre à travers ce chapitre consacré au marché des biens et services, et à la nécessité ou non de sa régulation. Pour répondre à ces questions, il convient de s'intéresser aux objectifs et aux modalités de la politique de la concurrence.

L'actualité des dernières années, des derniers mois, a été rythmée par les projets de concentrations, les batailles boursières entre entreprises au niveau européen ou mondial. Désormais, les gouvernements nationaux ne sont plus les seuls arbitres de ces « grandes affaires économiques et financières ». La Commission européenne, ou des autorités indépendantes des gouvernements (conseil de la concurrence en France par exemple) ont également leur mot à dire sur ces rapprochements et ces fusions, amicales ou non, au regard du respect de la concurrence.

A Bruxelles, des équipes d'experts sont chargés de préparer la position de la Commission et travaillent sur ces dossiers (analyse des parts de marché et des conséquences des fusions).

La stratégie des entreprises en matière de développement est rendue extrêmement difficile dans le contexte de concurrence mondiale. Elles doivent acquérir une taille suffisante pour pouvoir bénéficier d'économies d'échelle et donc abaisser leurs coûts tout en veillant à ne pas s'imposer comme un acteur trop omniprésent sur un secteur, au risque de se voir accuser de fausser le jeu concurrentiel.
La surveillance de la Commission s'est accrue en matière d'ententes illicites. Le montant des amendes infligées aussi (en 2001, 855 millions d'euros d'amendes pour huit entreprises s'étant entendues pour se partager le marché des vitamines ; en 2002, amende de 472 millions d'euros pour des sociétés de plaques de plâtre, amende de 497 millions d'euros pour Microsoft en 2004…).

La concurrence est donc le résultat fragile de la régulation publique indispensable. L'entreprise, à la recherche de profits est tentée d'agir au détriment du consommateur par des pratiques visant à élever son prix au dessus de son coût marginal[13], à accroître son pouvoir de marché.
Le grand résultat de l'économie classique, on va le voir, est que la concurrence parfaite est la structure de marché optimale pour la société car elle permet de maximiser le surplus global (surplus des consommateurs[14] et profits des entreprises). A long terme, lorsque les ajustements économiques ont éliminé les entreprises les moins efficaces et que la libre entrée sur un marché a permis l'installation des entreprises qui le souhaitaient, les profits sont nuls en situation de concurrence. C'est là un autre résultat de l'analyse microéconomique classique. L'entreprise à long terme est juste capable de rémunérer ses facteurs de production (capital, travail), c'est-à-dire de verser intérêts, dividendes et salaires, mais ne fait pas de surprofits. Bien entendu, on va le voir l'ensemble des hypothèses requises, pour que le cadre de concurrence parfaite s'impose, est si important que cette situation est rarement rencontrée dans la réalité. Cependant, on peut considérer qu'un certain nombre de marchés sont dans une situation proche de la concurrence parfaite décrite par la théorie économique.
Or les entreprises recherchent le profit. C'est leur motivation. C'est ce qui les incite à répartir de la manière la plus efficace leurs ressources. Elles seront toujours tentées d'exploiter les opportunités d'accroître leurs parts de marché mais également leur pouvoir de marché. Cela peut se faire grâce à la publicité, à l'innovation, à la croissance externe (rachat de concurrents), à la mécanisation ou à la rationalisation (afin de faire baisser les coûts de production en abaissant le recours au facteur travail) mais quelquefois, cela les amène à des pratiques anticoncurrentielles (entente collusive ou cartel, dumping, échange d'informations…). L'acquisition par une entreprise d'une position privilégiée sur un marché est donc parfois l'aboutissement d'une stratégie vertueuse (domination par les coûts ou stratégie de différenciation de produits efficace) et parfois le résultat d'un comportement délictueux. Les autorités de la concurrence doivent pouvoir faire la différence.

La politique de la concurrence vise dans ce contexte à favoriser l'efficience économique et à protéger les intérêts des consommateurs en leur permettant d'acquérir les biens et services aux meilleures conditions. Elle concerne donc les pratiques anti-concurrentielles qu'elle pourchasse (entente entre entreprises, pratiques concertées), les aides d'Etat incompatibles avec le marché intérieur européen (puisqu'elles faussent la concurrence entre entreprises de nations différentes) qu'elle condamne, les anciennes entreprises publiques en position de monopole dans certains secteurs qu'elle cherche à mettre en compétition avec d'autres acteurs.
En France, l'année 2005 a été marquante pour le Conseil de la concurrence pour deux raisons. D'une part, on a atteint un record en termes du montant d'amendes infligées (754,4 millions d'euros d'amende infligés[15] dont 534 millions pour Orange, SFR et Bouygues, condamnés pour entente et échanges d'informations). D'autre part parce que le principe de clémence pour les entreprises qui acceptent de dénoncer une entente a été appliqué pour la première fois[16].

---

[13] Le coût marginal est le coût de la dernière unité produite. Lorsque l'entreprise est en situation de concurrence parfaite, son coût marginal est égal au prix de marché.
[14] Voir l'annexe en fin de chapitre sur le **surplus des consommateurs**.
[15] En 2005, avec treize décisions et avis, le BTP arrivait en tête des affaires traitées. Mais les affaires étaient plutôt anciennes et B. Lasserre, le président du conseil de la concurrence, notait que le Conseil était moins saisi qu'il y a cinq ou six ans pour des affaires dans ce secteur. La distribution, les services et tous les secteurs oligopolistiques (chimie, pharmacie) sont les secteurs qui semblent susciter le plus d'affaires en ce moment. Le risque d'entente est fort car les acteurs sont peu nombreux. Pour le premier semestre 2006, le montant des sanctions s'élève déjà à 127 millions d'euros tous secteurs confondus.
[16] Des entreprises françaises avaient déjà eu recours à cette procédure dans des dossiers traités par les autorités de la concurrence américaine ou européenne. Aventis avait ainsi, dans l'affaire des vitamines en 2001, bénéficié d'une clémence de la part de Bruxelles pour avoir coopéré et fourni des preuves décisives pour l'enquête. En 2005, une entreprise a bénéficié pour la première fois de cette procédure dans une affaire française. Pour avoir dénoncé une entente sur les prix des portes et apporté au Conseil de la concurrence des éléments probants, la

L'ajustement de l'offre et de la demande sur les marchés, par les prix et les volumes, suppose que la concurrence fonctionne correctement ce qui passe par une vigilance permanente des pouvoirs publics en raison des nombreuses défaillances de marché.
En effet, la régulation des marchés de biens et services, qui ne s'identifie pas à la seule politique de la concurrence, ne s'arrête pas à la traque des entreprises se comportant de manière non concurrentielle. L'Etat a également, on va le voir un rôle à jouer lorsque pour diverses raisons la concurrence ne peut s'appliquer car le marché présente des failles : présence d'externalités[17], secteur en situation de monopole naturelle, absence de réciprocité de l'information…

A ce propos, un autre grand sujet d'actualité est ainsi l'ouverture à la concurrence de secteur anciennement en situation de monopole (aujourd'hui marché de l'énergie, hier marché du transport aérien, demain celui de l'acheminement du courrier ou du transport ferroviaire). Il conviendra d'expliquer d'une part pourquoi ces secteurs étaient anciennement en situation de monopole et pourquoi, il semble qu'aujourd'hui une mise en concurrence au niveau européen soit souhaitable. Les réticences des utilisateurs de ces services ou de ces biens sont grandes. Les peurs des salariés de ces entreprises aussi. Nous essaierons de comprendre les raisons de ces changements et d'apprécier leur pertinence.

2.1 Les défaillances de marché

La vision classique a défini l'économie de concurrence pure et parfaite comme l'économie idéale (parabole de la « main invisible » d'Adam Smith : l'égoïsme de chacun et la recherche de son propre intérêt débouche sur le bien-être collectif comme si chaque individu dans son action était guidé par une main invisible). Cette vision classique a été mise en équations dans un modèle où tous les marchés sont en équilibre, par Walras (modèle d'équilibre général de l'économie néo-classique). Pareto, autre auteur néo-classique, a montré que l'équilibre de concurrence pure et parfaite coïncidait avec l'optimum social ; c'est-à-dire à la situation préférable d'un point de vue social puisqu'on ne peut accroître le bien-être d'un individu sans détériorer celui d'au moins un autre (optimum de Pareto). Le grand résultat de la représentation théorique de référence (aboutissement théorique des écoles classique et néo-classique) est donc la coïncidence entre équilibre de concurrence parfaite et optimum social (premier théorème du bien-être). Il faut néanmoins ne pas interpréter ce résultat « à la lettre » et conclure que le laisser-faire entraîne automatiquement une situation efficace et que, par conséquent, toute intervention des pouvoirs publics est inutile voire nocive. Bien au contraire. La microéconomie néo-classique traditionnelle montre que les échanges marchands et le libre jeu du marché, n'aboutissent à une allocation efficace des ressources que dans un environnement très particulier de concurrence pure et parfaite, lorsque six conditions sont respectées.

Les hypothèses nécessaires à l'existence de la concurrence parfaite sur un marché sont les suivantes

- **Atomicité** des participants : le marché comprend un grand nombre de vendeurs et d'acheteurs, de sorte que le volume de transactions pour chacun d'eux est négligeable par rapport au volume global des échanges. L'action de chacun des participants n'a ainsi aucune incidence notable sur les quantités et le prix d'équilibre. Les agents sont dits « *price-takers* » car il considère le prix de marché comme une donnée exogène.
- Les biens doivent être **homogènes** sur ce marché, c'est-à-dire qu'ils doivent être rigoureusement identiques, de façon à ce que les acheteurs soient indifférents entre tous les biens. Le prix est alors le critère ultime de décision, ce qui oblige chacun des participants à accepter le prix de marché pour ne pas être exclu des échanges.
- Il doit y avoir **libre entrée** ou libre sortir sur le marché : les acteurs de l'échange ne doivent pas pouvoir adopter des comportements collusifs empêchant un concurrent d'intervenir ou d'entrer sur le marché,
- L'**information** doit être totalement **transparente** : tous les participants à l'échange sont parfaitement informés du prix et de la qualité des produits, de façon à pouvoir arbitrer entre les différents biens offerts et ainsi permettre la réalisation de l'équilibre.
- Les **biens** doivent être **exclusifs** ce qui signifie qu'une même unité de bien ne peut être consommé par plusieurs individus en même temps.

---

société France Portes a obtenu l'immunité et échappé à une amende de 1,25 millions d'euros (neuf autres entreprises on été condamnées à hauteur de 5 millions d'euros).

[17] Les externalités sont des phénomènes hors marché, c'est-à-dire des phénomènes dont les effets ne transitent pas par les prix. Voir la sous-section à ce sujet à la page suivante.

La validité du résultat « concurrence parfaite = optimum » nécessite donc le respect de conditions très strictes. Ainsi, penser que l'absence d'intervention des pouvoirs publics permet à elle seule l'instauration de la concurrence parfaite peut raisonnablement être considéré comme une ineptie (quelque chose d'absolument pas sérieux pour un économiste…). L'objectif européen d'assurer la « concurrence partout où cela est possible » ne saurait être réalisé que par des interventions accrues des pouvoirs publics et des autorités de régulation, et non par le renoncement de ceux-ci ou leur désengagement[18].
L'intervention des pouvoirs publics est rendue nécessaire en raison de l'existence des défaillances de marché qui vont perturber l'équilibre naturel du marché et rendre l'équilibre concurrentiel non optimal dans certaines situations.

Nous allons étudier trois cas particuliers de défaillances de marché, à travers les problèmes et les conséquences économiques soulevés : (2.1.1) lorsqu'il existe des **externalités** positives ou négatives, les allocations de ressources et l'équilibre qui ressort de l'ajustement par les prix ne sont alors pas optimaux, (2.1.2) lorsque **l'hypothèse d'atomicité n'est plus respectée**, c'est-à-dire lorsque les acteurs, offreurs ou demandeurs, ne peuvent plus être considérés comme des entités à même de subir les prix, alors, la fixation des prix ne correspond plus à la situation de concurrence pure et parfaite, et d'autres situations de marché émergent : monopole, oligopole, monopsone…, (2.1.3) lorsque les **rendements d'échelle croissants** ne permettent pas à la situation de concurrence d'être rentable et donc qu'émerge naturellement une structure monopolistique. Dans ce cas, il revient aux pouvoirs publics de contraindre celle-ci à œuvrer pour le bien-être des consommateurs afin qu'ils ne pâtissent pas de cet état de fait (notamment si le service rendu par le **monopole naturel** est un « service public »).
Nous verrons ensuite que lorsque l'information sur les biens échangés n'est pas identique entre les protagonistes de l'échange, c'est-à-dire lorsqu'il existe une **asymétrie d'information**, émerge une rente informationnelle pour celui qui détient l'information, ce qui fausse le libre jeu de la concurrence. La « nouvelle microéconomie », apparue dans les années quatre-vingt propose des outils pour prendre en compte ces situations.
Enfin la seule appréciation d'efficacité statique d'un marché ne saurait suffire pour conclure quant à l'efficacité d'une économie. L'objectif de l'efficience statique ne doit pas occulter la recherche de l'efficacité dynamique de l'économie. La dynamique intrinsèque de l'économie repose en partie sur la capacité d'innovation des entreprises. La question qui divise depuis de nombreuses années les économistes est alors de savoir quelle est la structure de marché la plus incitative pour l'innovation : la concurrence ou le monopole ?

2.1.1 La non optimalité des mécanismes de marchés en présence d'externalités

Une externalité, ou effet externe, est un effet subi par un agent économique A au cours d'une action de production ou de consommation effectuée par un autre agent économique B, qui n'est pas compensé dans le cadre d'un échange ou d'un marché. L'effet peut être négatif ou positif sur l'utilité de l'agent qui reçoit l'externalité.

La notion d'externalité remonte à Alfred Marshall, économiste anglais néo-classique du début du XX^{ème} siècle, qui mettait plutôt l'accent sur les externalités positives pour les entreprises. Le développement général des connaissances scientifiques et techniques est par exemple un effet externe positif pour les entreprises, puisque cela est susceptible d'améliorer leur productivité ou de les aider dans le cadre de leur propre programme de recherche et développement.
Une nuisance ou une pollution due à un acte de production ou de consommation est en revanche une externalité négative, lorsqu'elle porte atteinte à la satisfaction d'un agent tiers non concerné par l'échange ou la production. L'économie de l'environnement s'attache donc beaucoup à l'étude des externalités et à la possibilité d'internaliser leurs effets par une politique appropriée.
Avant de développer ces deux exemples génériques d'externalités dans des activités habituelles de l'entreprise (externalité positive dans l'activité d'innovation et externalité négative générée par certaine production), nous devons examiner le classement des biens selon leurs degrés de rivalité et d'exclusivité (voir la figure en haut de la page suivante). Les biens et services peuvent en effet être classés selon ces deux caractéristiques de rivalité et d'exclusivité.

[18] L'idée qu'on aboutit à la concurrence par le seul « laisser-faire » est bien souvent un fantasme.

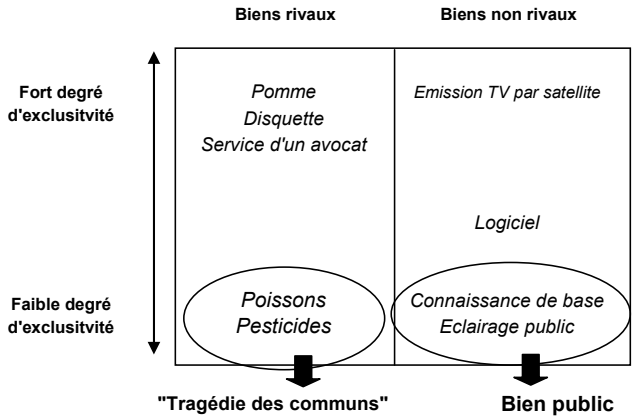


Figure 2 : Classement des biens en fonction du degré de rivalité et d'exclusivité de leur usage

(d'après Jones (2000) et Romer (1993))

Le **principe de rivalité** signifie que deux agents ne peuvent bénéficier en même temps du même bien. Le **principe d'exclusion par le prix** (exclusivité) exprime le fait qu'un consommateur ne peut disposer d'un bien économique que si il en paie le prix. L'offreur de ce bien est capable ou non d'exclure des agents qui ne voudraient pas payer de sa consommation. Les biens et services de consommation privée traditionnels sont des biens à la fois exclusifs et rivaux (pomme, disquette, service d'un avocat…). Si je mange une pomme, ou que j'utilise un baladeur MP-3 c'est que je les ai achetés (exclusivité). Vous ne pouvez, à moins de m'assommer préalablement, manger cette même pomme ou utiliser mon baladeur MP-3 (principe de rivalité dans la consommation[19]). Les biens de consommation collective (biens collectifs) correspondent au contraire aux biens économiques dont plusieurs consommateurs peuvent jouir en même temps à un prix faible ou nul (quelle entreprise privée accepterait de produire ces biens dont le prix tend vers zéro ?). On les identifie généralement aux biens publics car ils sont souvent produits par les pouvoirs publics. Ce sont des biens non rivaux et non exclusifs, et souvent indivisibles.

On dit qu'un bien est indivisible, lorsque les avantages que procure un bien sont non séparables, c'est à dire que l'utilisation de ce bien par un agent n'affecte en rien l'utilisation de ce même bien par d'autres agents (défense nationale, éclairage public…). Il peut néanmoins y avoir des phénomènes d'encombrement (une route gratuite peut être embouteillée).

Des biens peuvent également être non rivaux dans leur utilisation mais pouvoir faire l'objet d'une exclusion par les prix. C'est le cas d'une émission de télévision par satellite. Tout le monde peut regarder cette émission sans que cela n'entrave l'utilité de chacun des individus. Il faut cependant disposer d'une parabole ou payer un abonnement (voir aussi l'exemple des chaînes payante de type Canal +, il faut disposer d'un décodeur et payer l'abonnement).

Enfin, une quatrième série de biens apparaît. Les biens qui sont à la fois non exclusifs mais rivaux dans leur consommation. L'exemple est celui d'une zone poissonneuse dans des eaux non territoriales dans lesquelles les pêcheurs de toute nationalité peuvent venir jeter leurs filets. On est donc ici en présence d'une ressource (les poissons) qui ne fait pas l'objet d'un marché (on ne paye pas pour venir pêcher, biens non exclusifs). En revanche, c'est une ressource épuisable. Si les pêcheurs utilisent de manière trop intensive cette ressource, elle risque de disparaître. Or, ils ne sont pas incités à économiser cette ressource car elle est sans prix et ils pensent que leurs concurrents pêcheront de toute façon, même si eux adoptent un comportement plus raisonnable. C'est ce qu'on appelle la « tragédie des biens communs » mise en lumière par Hardin[20].

A côté de cette classification des biens qui détermine la façon dont les biens seront offerts sur le marché en fonction de leurs caractéristiques, il convient d'étudier les phénomènes d'externalités qui échappent aux mécanismes de marché.

---

[19] Il existe cependant des lecteurs MP3 avec des sorties pour deux casques ou bien des enceintes sur lesquelles je peux brancher mon lecteur. Cela permet d'abaisser légèrement le caractère exclusif de ce bien. Vous seriez néanmoins forcés d'écouter la même musique que moi (un tel abaissement du degré d'exclusivité vous permet donc de bénéficier d'une externalité, puisque vous ne payeriez pas pour écouter la musique de mon baladeur, mais cette externalité peut être positive ou négative selon que vous partagez ou non mes goûts musicaux…).

[20] Hardin G. (1968), « The tragedy of the commons », *Science*, Vol.162, pp.1243-48. Hardin prend l'exemple d'un pré, propriété collective d'un village de paysans, qui y font paître leur bétail. Les animaux dégradent les communs, en arrachant l'herbe et en laissant des portions boueuses derrière eux, qui ne recouvreront que lentement leur couverture végétale. En l'absence de politique consentie (et appliquée, fût-ce par la force) pour allouer des droits et des limites à chacun, l'intérêt de chacune des parties est d'y mener un maximum de têtes le plus vite possible, pour en extraire toute la valeur avant que les communs ne soient réduits à une mer de boue.

2.1.1.1 Un exemple d'externalité négative : Les émissions de polluants

L'environnement se compose de nombreux biens utiles mais non attribués par le marché. L'utilisation de ces biens (forêt, air, océan…) est gratuite, soit comme réceptacle de déchets, soit comme biens utiles directement dans la production (bois des forêts), soit comme support de services environnementaux de type loisirs (au contraire d'une piscine, vous ne payez pas pour vous baigner dans l'Atlantique ou la Méditerranée…à la différence d'un parc d'attraction, une promenade en forêt ou dans la campagne est également gratuite… enfin, bien entendu, vous ne devez payer à personne l'air que vous respirez…). Toute action de production ou de consommation peut donner lieu à des externalités environnementales, négatives lorsqu'elle porte atteinte aux services environnementaux, ou positives si elle les renforce.

Bien entendu, en matière d'environnement, nous sommes le plus souvent en présence d'externalités négatives puisque les agents économiques dans leur choix de consommation ou de production ignorent, ou feignent d'ignorer, les effets de leurs actions. Une usine qui déverse des polluants dans une rivière évite ainsi un coûteux processus de retraitement mais l'ensemble des agents qui utilisait les services environnementaux de la rivière (pêcheurs, baigneurs,…) voient leur utilité diminuée lorsque la rivière est polluée. Pourtant, ils ne sont pas indemnisés pour cette perte d'utilité car la rivière ne leur appartient pas, ils ne sont que des utilisateurs de services liés à cette rivière.

Les biens dont la production engendre une dispersion des coûts sont souvent produits en quantité excessive par le marché (cas des dégâts causés à l'environnement par notre production et notre consommation). L'intervention publique s'avère alors nécessaire afin de compenser l'insuffisante définition des droits de propriété. Deux solutions se dégagent : la définition de droits de propriété ou bien la mise en place de taxe permettant d'internaliser les effets externes négatifs sur l'environnement.

R. Coase (prix Nobel d'économie en 1991) a en effet mis en exergue le fait que les externalités pouvaient provenir d'une défaillance de droits de propriété. Cette intuition est à la base de l'idée de mettre sur place des marchés de permis d'émission négociables (« droits à polluer ») afin de trouver une solution au problème du réchauffement climatique.

C. Pigou (1920), auteur néo-classique, a montré que les externalités se présentent comme une divergence entre le produit marginal social et le produit marginal privé. Il a préconisé de corriger ces externalités par une intervention de l'Etat soit par des réglementations, soit par la mise en place de taxation des activités conduisant à des externalités négatives (« taxes pigouviennes ») ou par la subvention des activités conduisant à des externalités positives. Le principe dit d'internalisation des effets externes, est d'augmenter le produit marginal privé d'un pollueur en le taxant jusqu'à ce que son produit marginal privé égalise le produit marginal social. L'ancien pollueur produit alors la quantité « optimale » du bien qui engendre la pollution.

Les marchés de permis d'émission négociable et le Protocole de Kyoto

L'économie du bien-être montre, qu'en présence d'externalités négatives, le coût marginal social de la production est plus élevé que le coût de production marginal privé. Cet écart dans les coûts ne permet plus d'assurer l'optimalité de l'équilibre économique. Coase (1960)[21] prolonge cette proposition et établit que le recours à la délimitation juridique de droits de propriété est une condition nécessaire pour que les agents économiques puissent remédier aux effets externes de la pollution. Cette conception tranche ainsi avec l'idée que l'air, comme "bien libre", n'est pas appropriable. Pour Coase, l'internalisation va résulter de la négociation bilatérale consécutive à la mise en place de droits de propriété sur les attributs environnementaux. Cependant, cette analyse n'est valable qu'en l'absence de coûts de transaction.

Dales (1968)[22] élabore le principe de marché de droits à polluer. Dales attribue l'existence des externalités à « l'absence ou la mauvaise définition de droits de propriété sur les biens » et envisage alors d'établir un marché de droits à polluer, consistant « à mettre en place un système complet de marché concurrentiel flexible qui marchandise les effets externes de la pollution ». Conformément à cette conception il propose d'établir sur des biens dont l'usage était auparavant illimité, un faisceau de droits de propriété exclusifs et échangeables sur un marché concurrentiel. Au prix d'équilibre, l'externalité est réintégrée dans le calcul économique des agents et son caractère hors marché disparaît. La démonstration de la condition d'efficacité économique de tels marchés de

[21] Coase R. (1960), « The problem of social cost », *Journal of Law and Economics*, Vol.3, pp.1-44.

[22] Dales J. H. (1968), Pollution, Property and Prices, University of Toronto Press.

droits à polluer a été formulée par l'apport théorique de Montgomery (1972)[23]. Le fonctionnement concurrentiel du marché des droits à polluer permet d'atteindre à moindre coût les objectifs de réduction des émissions et favorise une répartition optimale des efforts de dépollution entre les sources polluantes en raison de l'égalisation de leurs coûts marginaux de dépollution. Dans ce paradigme concurrentiel, l'allocation initiale des permis est neutre car les effets allocatifs et les effets distributifs sont disjoints. La répartition initiale des droits de propriété ne conditionne donc pas l'équilibre économique des marchés. Depuis ces travaux initiaux, les marchés de droits à polluer ont gagné une importance croissante dans la hiérarchie des instruments économiques de la réglementation environnementale. Ils sont passés du stade de la représentation théorique à celui d'instrument de marché utilisé à grande échelle pour réduire les émissions polluantes.

Un marché de permis découle de la fixation d'un objectif quantifié de réduction des émissions, correspondant au montant total de permis initialement distribués aux différentes sources polluantes. La possibilité que ces dernières satisfassent collectivement l'objectif total d'émission est assurée par les échanges de permis, résultant d'une hétérogénéité des coûts de dépollution. Si le marché fonctionne parfaitement, les coûts marginaux de dépollution entre les agents s'égalisent au prix des permis. L'échange de permis s'apparente alors à une compensation, dont la valeur est comprise entre le coût marginal de réduction évité par l'acheteur de permis et le coût marginal de réduction encouru par le vendeur de permis. Le prix des permis devient un indicateur utilisé pour évaluer le degré d'internalisation des émissions car il fournit une valeur *in situ* à la contrainte de réduction des émissions. Ce système conduit ainsi à une répartition optimale des efforts de dépollution entre les agents. Les échanges de permis permettent une réduction des émissions là où les coûts de réduction sont les plus faibles et abaissent les coûts de réduction de l'ensemble des agents. Le transfert des permis encourage, par conséquent, l'adoption de politiques de réduction plus ambitieuses. L'avantage d'un marché de permis réside essentiellement dans le fait que, contrairement à une taxe environnementale, la connaissance des dommages environnementaux n'est pas nécessaire. Le prix des permis s'ajuste automatiquement et ce mécanisme permet d'atteindre avec certitude l'objectif d'internalisation des émissions. Les permis négociables apparaissent aussi préférables aux régulations administratives directes qui imposent des standards d'émission rigides parce que les échanges de permis permettent aux firmes de choisir leur niveau de pollution, en accordant avec leurs demandes spécifiques, et leurs propres coûts de production, qui sont typiquement de l'information privée. Le marché de permis vérifie de cette façon les conditions d'efficience environnementale (la définition de la meilleure allocation possible permettant la réalisation de l'objectif de réduction des émissions) et d'efficacité économique (la répartition optimale des efforts de réduction des émissions au coût minimum).

Deux formes de marchés de permis d'émission négociables peuvent être distinguées. Le premier système correspond à un marché de permis de type limite et échange ("***cap and trade***"). Les autorités publiques repèrent les sources d'émission et fixent un plafond total sur les émissions de polluants, correspondant à un nombre déterminé de permis répartis ensuite entre ces sources réglementées. Les agents doivent alors, pour émettre une certaine quantité de polluant, détenir un montant de permis équivalent. Les échanges de permis sont autorisés entre les participants. Ce mécanisme s'inscrit dans une perspective dynamique car le plafond total sur les émissions diminue de période en période, incitant l'ensemble des sources à réduire progressivement leurs émissions polluantes. Ce principe permet aussi d'assurer le résultat environnemental total de réduction des émissions, car le plafond des émissions est fixé au début du marché et la certification des crédits de réduction n'est pas nécessaire. Ce type de marché de permis a été utilisé notamment aux Etats-Unis dans le cadre des programmes *Acid Rain* sur les émissions de SO2 et dans le cadre du programme RECLAIM en Californie sur les émissions de NOx ainsi qu'en Europe dans le cas de la directive 2003/87/CE sur les émissions de CO2. Le second système s'apparente à un marché de permis de type niveau de base et crédits ("***based and credits***"). Les crédits de réduction des émissions sont créés et accordés aux agents par rapport à une base d'émission. Ce niveau de base, assimilable à une norme d'émission, est fixé de manière réglementaire ou résulte de négociations entre les entreprises. Les sources qui émettent moins que leur niveau de base peuvent obtenir des crédits de réduction d'émission pouvant être alors vendus à d'autres sources d'émission pour satisfaire leur niveau de base.

Le protocole de Kyoto établit un objectif de réduction des émissions pour six gaz à effet de serre via l'instauration d'un marché international de permis négociables. Collectivement les pays dits de l'Annexe-1 (représentant 38 pays de l'OCDE et pays en transition) se sont engagés à réduire en moyenne de - 5.2% leurs émissions de gaz à effet de serre par rapport à leur niveau de 1990 d'ici la période 2008-2012.

La répartition des objectifs ou contraintes de réduction des émissions de gaz à effet de serre dans quelques pays industrialisés et en transition par rapport à 1990, résultant du protocole de Kyoto sont les suivants pour la période 2008-2012: Union Européenne (-8%), Etats-Unis (-7%), Canada (-6%), Japon (-6%), Russie (0%), Ukraine (0%), Nouvelle-Zélande (+1%), Australie (+8%).

---

[23] Montgomery W.D. (1972),: « Markets in licences and efficient pollution control programs », *Journal of Economic Theory*, Vol.3, pp.395-418.

2.1.1.2 Un exemple d'externalité positive : La diffusion des connaissances

Un bien faiblement exclusif, on l'a vu, ne permet pas à son propriétaire de tirer tous les bénéfices possibles de sa détention ou de sa production. Lorsque les biens ne sont pas exclusifs, les bénéfices se dispersent et ne sont pas intégralement récupérés par leurs propriétaires (la dispersion des bénéfices est qualifiée d'externalité positive). Les biens dont les bénéfices se dispersent sont souvent produits en quantité insuffisante par le marché. C'est le cas de la connaissance, fruit des recherches privées menées par les entreprises.

Sans avoir à investir dans de coûteux programme de recherche et développement, des entreprises peuvent capter les connaissances, utiles pour le développement de certains biens ou services, mises à jour par leurs concurrents. Cette possibilité est offerte par la nature même de la connaissance, un bien non exclusif et non rival. L'entreprise qui met à jour une connaissance n'est donc pas la seule à bénéficier de celle-ci car la connaissance se diffuse (processus d'externalité positive). La connaissance ne peut donc être pleinement appropriée.

Ce phénomène de diffusion de la connaissance est un problème économique majeur car si aucune entreprise ne peut s'approprier pleinement les résultats de ses recherches du fait de la nature diffuse de la connaissance, alors les entreprises seront réticentes à se lancer dans de coûteux programmes de recherche. Même si certaines osent, il n'en demeure pas moins que la quantité de recherches privées menées va être plus faible que ce qui serait optimal pour la société. Si ce problème n'est pas résolu, cela va nuire au développement scientifique et technique de la nation et obérer son développement économique puisque le progrès technique est un déterminant essentiel de la croissance (voir le chapitre 7)

Le système de brevets permet heureusement de s'extraire de ce problème économique. L'entreprise qui dépose un brevet obtient le droit exclusif d'utiliser sa nouvelle connaissance technique, son nouveau produit ou son nouveau procédé, pour une période donnée (la durée de vie d'un brevet est de vingt ans). Elle dispose ainsi d'un pouvoir de marché (elle peut produire en position de monopole) et va pouvoir amortir le coût fixe initial engendré par la recherche et développement. L'entreprise qui dépose un brevet doit par ailleurs rédiger une demande dans laquelle elle décrit avec précision son invention de manière à ce qu'un « homme de l'art » (un ingénieur dans ce domaine technique) soit à même de la reproduire. Le système de brevets est ainsi un instrument économique extrêmement intéressant. Il permet de sécuriser les bénéfices de l'inventeur empêchant la production par des tiers (l'entreprise innovante dispose d'une preuve ; le brevet, pour porter l'affaire ne justice en cas de contrefaçon) tout en aidant à la diffusion de l'information scientifique et technique. Les futurs inventeurs peuvent utiliser la gigantesque base de données que constitue les brevets déjà déposés pour s'aider dans leurs futures recherches (puisque la connaissance est cumulative). Le brevet est donc un instrument économique qui permet de définir un droit de propriété là où il pouvait en manquer. Il offre la possibilité d'effectuer un arbitrage entre efficacité statique et efficacité dynamique (le brevet améliore l'efficacité dynamique de l'économie (stimule le progrès technique en laissant entrevoir la possibilité de tirer profit de ses inventions) au prix d'une diminution provisoire de son efficacité statique (le bien innovant est produit en condition de monopole, et non de concurrence, durant toute la durée de vie statutaire du brevet).

Le système de brevets s'avère pourtant une solution de moins en moins satisfaisante du point de vue de la politique de la concurrence dans certains secteurs ou domaines techniques en plein essor. Un premier point problématique est qu'il était initialement impossible de déposer des brevets pour les logiciels en Europe (cela est possible au Etats-Unis) à moins que les programmes « aient un effet technique ». Cette interdiction semble bien disparaître dans les faits puisqu'on a constaté de nombreux dépôts de brevets à l'Office Européen des Brevets, sans que la question de la brevetabilité des brevets soit complètement tranchée d'un point de vue politique. Il convient d'avoir cependant à l'esprit, qu'aux Etats-Unis où cette possibilité était offerte depuis longtemps, la brevetabilité des logiciels a été concomitante d'une diminution des dépenses de R&D dans ce domaine[24]. On examinera un peu plus loin le lien entre concurrence et innovation qui permet d'expliquer cet apparent paradoxe. D'autre part, dans le domaine de la brevetabilité des gènes ou bien dans le secteur des technologies de l'information et de la communication, un autre problème surgit. Les biens de plus en plus complexes nécessitent le recours à de multiples brevets ce qui peut aboutir à un blocage si l'un des détenteurs d'un brevet incontournable refuse d'accorder des licences. La stratégie qui consiste à

[24] Voir l'article de Bessen et Maskin (2000), « Sequential innovation, patents and imitation *», MIT working paper* sur ce sujet. Ces auteurs concluent que dans un monde dynamique comme celui de l'industrie du logiciel, les entreprises peuvent avoir beaucoup de rendement sans brevet. Les brevets peuvent, bien au contraire, étrangler l'innovation lorsque celle-ci présent un caractère complémentaire (des innovations sont nécessaires pour d'autres innovations). La protection par simple copyright des programmes logiciels peut ainsi atteindre un meilleur équilibre que la protection par brevet.

définir des « pools de brevets » (c'est-à-dire un ensemble de brevets nécessaires à la production d'une technologie donnée) regroupant les brevets déposés par plusieurs organismes ou entreprises peut apparaître problématique car elle peut constituer une forme de « barrière à l'entrée » dans le secteur et s'accompagner d'une régression de la concurrence dans le secteur, ce qui a des effets ambigus sur l'incitation à l'innovation et des effets négatifs pour les consommateurs.
Nous venons de voir l'émergence de monopole légal, permis par l'existence de brevets, qui réintroduit une certaine incitation à l'innovation pour les entreprises, malgré le caractère diffus de la connaissance. Nous allons à présent étudier une autre défaillance de marché : lorsqu'une structure monopolistique émerge au détriment des consommateurs.

2.1.2 La remise en cause de l'hypothèse d'atomicité et les différentes formes de marché

L'hypothèse d'atomicité indique que les intervenants sur un marché ont une taille si modeste et sont présents en si grand nombre que l'influence de chacun sur les prix et les quantités est négligeable. En concurrence pure et parfaite, les entreprises sont ainsi dites « preneuses de prix ». Elles considèrent le prix comme une donnée et n'ont pas conscience qu'elles participent à sa fixation par la quantité de biens qu'elles mettent sur le marché. Au contraire, lorsqu'on commence à examiner des structures de marché où le nombre de participants est faible, on ne peut plus faire l'hypothèse que les entreprises n'ont pas d'incidence sur les prix. Elles connaissent alors la fonction de demande inverse (p(Q)) et donc la façon dont le prix va réagir aux quantités offertes.

2.1.2.1 Différentes structures de marché

La situation concurrentielle peut être différente selon les marchés ou les secteurs considérés et selon le degré de développement du secteur ou la période considérée. Deux critères permettent de définir une taxonomie des régimes de concurrence que l'on peut rencontrer. Le premier critère est celui du **nombre d'acteurs**. Le deuxième critère concerne la **nature du produit** (hypothèse d'homogénéité remise en cause ou non). Lorsque les produits échangés sont homogènes, le nombre d'offreurs ou d'entreprises concurrentes détermine le type de structure de marché (voir le tableau ci-dessous).

Nombre de firmes	Situation
une infinité	Concurrence parfaite
quelques unes	Oligopole
2	Duopole
1	Monopole

Tableau 1 : Taxonomie des structures de marché

L'annexe **la situation de concurrence** en fin de chapitre présente le programme de l'entreprise en concurrence parfaite. L'annexe **la situation de monopole**, en fin de chapitre, présente le comportement d'une entreprise en monopole. L'annexe **L'oligopole** décrit pour sa part les différents duopoles (**duopole de Cournot** lorsque les entreprises se font une concurrence en quantité, **duopole à la Bertrand** lorsque la concurrence porte sur les prix, **équilibre de Stackelberg** lorsque l'une des firmes possède un avantage sur l'autre et prend ses décisions en connaissant la réaction de sa concurrente) ainsi que les possibilités d'**ententes collusives** et leurs conséquences.
Lorsque les produits présentent un certain degré d'hétérogénéité, alors le cas considéré est celui de la concurrence monopolistique (voir annexe sur **la concurrence monopolistique** en fin de chapitre). Puisque les consommateurs perçoivent les produits comme différenciés, chaque producteur a un certain pouvoir de marché sur son segment de produit.
La structure des grands secteurs de production est rarement totalement concurrentielle. Les concentrations, qui limitent la concurrence, ont le bénéfice de pouvoir s'accompagner de réduction de coûts grâce à des économies d'échelle importantes (c'est d'ailleurs souvent la raison de ces concentrations, voir par exemple Air France et KLM qui ne cessent de revoir à la hausse leurs synergies, ce qui s'accompagne d'une efficacité économique accrue et de résultats financiers au rendez-vous).
Cependant, de nombreux éléments corroborent la pertinence de la vision théorique de la primauté de la concurrence. D'une part, le gain pour les entreprises si il est obtenu au détriment des consommateurs n'est pas un gain pour la société. D'autre part, à l'échelle internationale, l'avantage des entreprises est souvent lié au degré de concurrence intérieur du secteur.

Quoi qu'il en soit la dynamique de l'économie s'accompagne d'évolution dans la structure de marché. Lorsqu'un nouveau secteur apparaît, suite à une innovation par exemple, une entreprise unique (monopole) ou quelques entreprises (oligopole) se font concurrence. Puis la libre entrée va permettre l'arrivée de plus en plus de concurrents ce qui aura tendance à augmenter l'offre et donc à peser sur les prix. Lorsque les profits sont faibles ou nuls, les entrées sur le marché cessent. Il peut cependant encore y avoir des mouvements occasionnés par la sortie d'entreprises non rentables.
La figure ci-dessous présente les parts de marché dans certains secteurs en France (marché de l'accès à internet par l'ADSL, marché de la téléphonie mobile, secteur automobile).

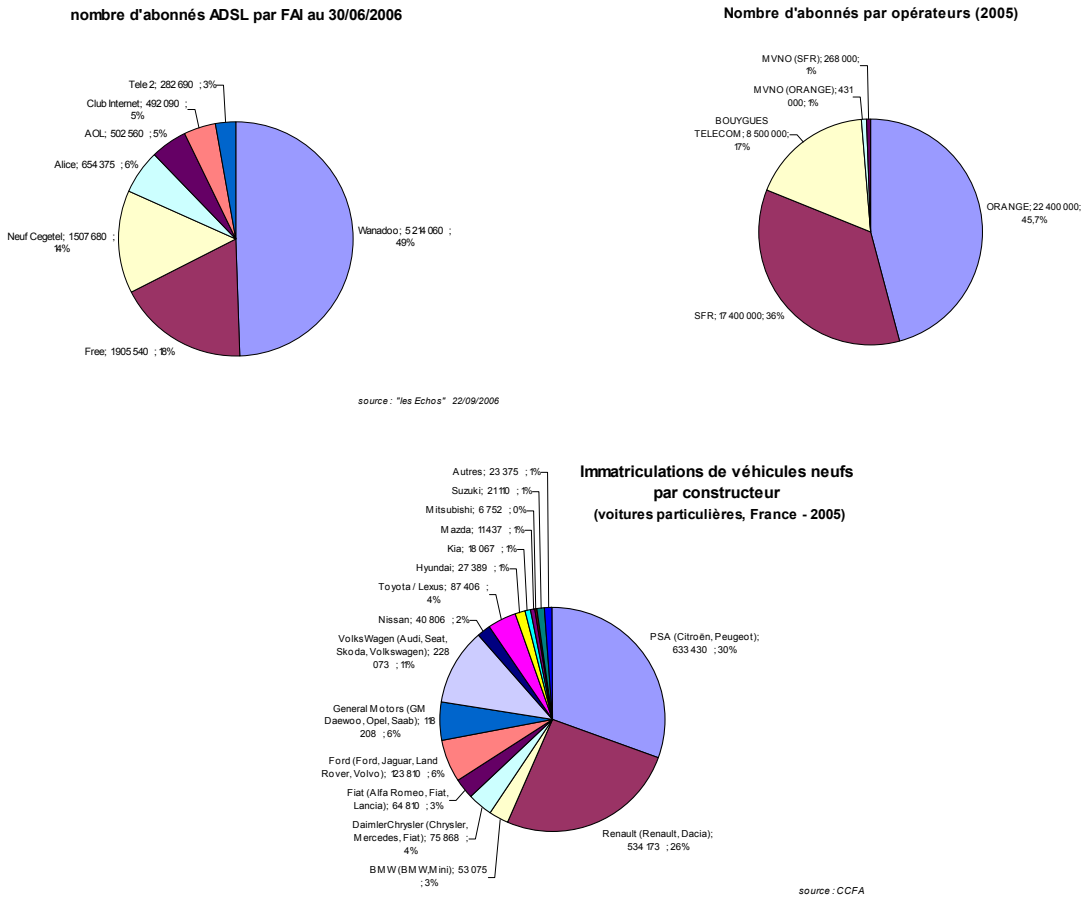


Figure 3 : Parts de marché dans quelques secteurs[25]

Dans ces exemples, on constate qu'un seul marché pourrait raisonnablement être considéré en situation de concurrence : celui de l'automobile (on constate d'ailleurs qu'actuellement les profits y sont faibles voire que certains constructeurs font des pertes). Un marché est en situation oligopolistique : la téléphonie mobile. Ce marché à même fait l'objet de pratiques collusives entre acteurs (mise en place de ce que la théorie économique appelle un cartel : plusieurs entreprises se comportent comme un monopole en place grâce à une entente. Une condamnation pour entente à l'automne 2005 par le Conseil de la concurrence semble avoir mis fin à cet état de fait). Le marché de la fourniture d'accès internet par ADLS est également un marché oligopolistique, avec un leader et une frange compétitive. Le leader (Wanadoo) n'est autre que l'ancien monopole (France Telecom). Mais l'arrivée de concurrents a rendu ce secteur concurrentiel. On a constaté que la guerre des prix avait abouti une baisse importante des tarifs d'abonnements et ainsi permis un équipement rapide des ménages en connexions haut débit. On constate cependant que ces configurations de marché ne sont pas figées. L'apparition des MVNO (opérateurs virtuels qui utilisent l'un des trois réseaux) a introduit une nouvelle concurrence dans le secteur de la téléphonie mobile. Dans l'ADSL, il semble qu'au contraire on rentre dans une phase de concentration puisque les baisses de prix obligent désormais à chercher des économies d'échelle pour accroître la rentabilité[26].

[25] Concernant le nombre d'abonnés ADSL, les chiffres au printemps 2007 pour quelques FAI sont les suivants. Orange: 6,575 millions d'abonnés, 9 Cegetel qui a absorbé AOL et Club-Internet : 3,03 millions d'abonnés, Alice : 847 000 abonnés, Télé 2 : 337 000 abonnés. Le secteur, on le voit est en phase de consolidation. Les structures de marché ne sont pas figées, particulièrement dans un secteur récent basé sur des technologies récentes.
[26] Mais déjà, la technologie « fibre optique » devrait venir bousculer ce marché en pleine croissance

2.1.2.2 Les interactions stratégiques

Dans la situation de concurrence parfaite, l'hypothèse de l'atomicité de l'offre permet de ne pas avoir à prendre en considération les interactions stratégiques. Dans la « concurrence parfaite » tout ce passe, comme si, du fait de sa trop faible taille et de son poids négligeable sur le marché, chaque firme n'avait pas conscience de l'effet sur le prix de marché de son propre comportement. Lorsque le nombre de firmes présentes sur un marché est faible (oligopole, monopole), bien entendu cette hypothèse n'est plus tenable et il faut prendre en considération ces interactions.

La prise en compte des interactions stratégiques est actuellement l'une des pistes les plus fécondes en matière de recherche en économie. L'apport de la théorie des jeux (voir l'annexe **Brève introduction à la théorie des jeux** en fin d'ouvrage) permet de déterminer les résultats et la stabilité des équilibres qui résultent de ses interactions[27]. Le résultat le plus important est que, dans le cadre d'une concurrence en quantité à la Cournot, l'entente collusive n'est pas stable car chacun des participants est incité à tricher (voir le « dilemme du prisonnier »).

2.1.2.3 La discrimination

Un monopole produit une quantité trop faible de biens parce qu'il souhaite limiter sa production à un niveau où les consommateurs sont pourtant prêts à payer pour une quantité supplémentaire une somme supérieure au coût de production. Le monopole ne souhaite pas produire ces quantités supplémentaires car cela ferait alors baisser le prix pour l'ensemble de ses ventes (on fait bien sûr toujours l'hypothèse que tous les achats se font au même instant). Si le monopoleur arrive à vendre différentes unités de biens à des prix différents, il peut être amener à augmenter la quantité de biens qu'il met sur le marché. On parle de discrimination par les prix. Il existe en fait trois types de discrimination par les prix :

- **La discrimination au premier degré** ou discrimination parfaite. Chaque unité de bien est vendue au consommateur qui lui attribue la valeur la plus élevée et au prix maximum que cet agent est prêt à payer pour cette unité. Cela revient pour le monopole à accaparer l'ensemble du surplus des consommateurs puisqu'il est à même de vendre à des prix différents (égaux aux dispositions à payer) à chacun des consommateurs (situation éminemment théorique et donc rare dans la réalité…).
- **La discrimination au second degré**. Les prix diffèrent selon les quantités achetées (non plus selon les acheteurs). La tarification est alors non linéaire (prix unitaire non constant). Cela correspond au système de rabais pour achats de gros, ou à des tarifs comportant deux parties : l'abonnement et la partie variable (abonnement Fréquence de la SNCF par exemple).
- **La discrimination au troisième degré**. Le monopole est alors capable de pratiquer des prix différents en fonction de l'appartenance des consommateurs à une catégorie. Chaque unité de bien est vendue au même prix à l'intérieur d'un groupe (ou catégorie) et le prix est d'autant plus élevé que l'élasticité-prix de la demande est faible (voir l'annexe sur l'**élasticité-prix de la demande** en fin de chapitre).

Le profit du monopole est bien entendu supérieur dans le cas de discrimination mais, paradoxalement, la situation est également profitable à certains des consommateurs. En l'absence de discrimination du troisième degré, par exemple, c'est-à-dire lorsque le monopole fixe un tarif unique relativement élevé, une partie de la population des consommateurs peut être exclue de la consommation du bien ou du service vendu par le monopole (ceux qui ont l'élasticité prix la plus forte, c'est-à-dire ceux qui ont la plus grande sensibilité au prix peuvent être sortis de ce marché). Lorsque le monopole discrimine au troisième degré, cela permet d'atteindre ces populations, quitte à faire payer plus cher la population plus « captive » (celle dont l'élasticité-prix est faible). Ainsi, les cartes « 12-25 », les cartes « vermeil » permettent au secteur des transports de proposer des prix plus bas à ces populations qui sinon prendraient moins facilement le train au tarif « normal ».

[27] La nature des interactions stratégiques est également importante. Le duopole de Cournot est un jeu dans lequel chaque entreprise choisit les quantités qu'elle va mettre sur le marché en fonction de celles de l'autre entreprise. Les quantités choisies par l'une et l'autre firme sont des substituts stratégiques (si l'une des firmes « inonde » le marché, l'autre devra restreindre son offre pour éviter que le prix ne s'écroule). Au contraire la compétition à la Bertrand se réalise au niveau des prix que fixent les deux entreprises. Les prix choisis par l'une et l'autre sont des compléments stratégiques (si l'une des entreprises élève son prix, l'autre peut faire de même sans risquer de perdre des clients).

2.1.3 Monopoles publics et ouverture à la concurrence

En France, il y a souvent confusion sur la notion de « service public ». Certains considèrent qu'il s'agit d'un service rendu par une entreprise ou une administration publique, alors que la notion clairement définie en droit français est la suivante : « *Activité d'intérêt général exercée par une personne publique ou sous le contrôle d'une personne publique, suivant un régime juridique dérogatoire au droit commun* ». Une entreprise privée peut donc tout à fait assurer une mission de service public, dans le cadre d'une délégation de service public et sous le contrôle d'une personne publique (c'est par exemple le cas pour le service de distribution d'eau ou celui du traitement des déchets ménagers…). Pourtant, un certain nombre de service, entrant plus ou moins dans le cadre de ce qu'on appelle le service public était encore récemment ou sont encore assurés par des entreprises publiques (fourniture d'énergie, acheminement du courrier,…). Sous la pression de la Communauté européenne, cela devient de plus en plus rare[28]. Il convient alors de s'interroger sur deux points. Quelles étaient les raisons économiques pour lesquelles certains services étaient fournis par des entreprises publiques en situation de monopole, puisque la théorie traditionnelle suggère que le monopole est inefficace d'un point de vue économique ? Pourquoi est-on, ces dernières années, passé de ce système de monopoles publics à un système de fourniture des mêmes services en situation de concurrence ? Comment cela a-t-il été rendu possible ?

2.1.3.1 Tarification au coût moyen dans des structures intégrées

L'annexe **la fixation d'un prix par un monopole public** décrit précisément les conditions qui aboutissent à la fourniture d'un bien ou d'un service par un monopole public. Les économies d'échelle dues aux conditions techniques existent dans de nombreux secteurs d'activité. On qualifie alors la situation de « monopole naturel » au sens où le monopole s'avère être la structure de marché la plus pertinente. L'exemple de tel monopole naturel est donné par les industries de réseau (télécommunications, transport, distribution de gaz ou d'électricité) où les coûts fixes sont extrêmement importants et les rendements d'échelles sont croissants (voir l'annexe sur **la fonction de coût et les économies d'échelle**). Il est alors plus efficace économiquement de ne pas gaspiller de ressources en dupliquant le réseau. La concurrence est alors difficile à mettre en place et la structure économique la plus efficace est un monopole.

Cependant, en l'absence de régulation, la situation de monopole autorise des rentes qui peuvent être dissipées de plusieurs manières (« surprofits » aux actionnaires sous forme de dividende, « sursalaires » aux salariés ou avantages sociaux), et cela au détriment de l'intérêt général et du bien-être des consommateurs. Or certains de ces secteurs font partie de ce qu'on peut qualifier de « service essentiel » ou de « service public » (tout le monde en France doit pouvoir recevoir une lettre et disposer d'électricité…).

La solution choisie en France avait consister, au sortir de la seconde guerre mondiale, à mettre sur pied des monopoles institutionnels dont l'objectif était de fournir au plus grand nombre les services primordiaux à des prix abordables, tout en étant rentables. Cela a abouti à fixer un prix des services rendus par ces entreprises publiques égal non plus au coût marginal comme en concurrence parfaite mais au coût moyen, ce qui permet d'amortir les importants coûts fixes tout en offrant le service au plus grand nombre.

L'approche traditionnelle française (tarification au coût moyen dite tarification à la Ramsey-Boiteux) partait donc de l'idée d'un monopole intégré. La même entreprise (EDF, SNCF) avait la charge de la gestion du réseau (lignes électriques, lignes de chemin de fer… à la base du coût fixe important) et la fourniture des services (électricité, transport de voyageurs…)[29]. L'Etat, un peu schizophrène, était à la fois l'actionnaire unique (ou principal) du monopole et le régulateur garant de l'intérêt général. La résolution de deux objectifs antagonistes (être rentable, logique de l'actionnaire, fournir le service à tous les concitoyens, rôle du régulateur et de l'administration de tutelle) trouvait sa solution dans la détermination d'un tarif au coût moyen déjà évoqué. Les pratiques discriminantes appropriées (voir plus haut) et le fait que les monopoles étaient souvent multi-produits permettaient de compléter au mieux l'objectif de services publics (la SNCF exploitent des TGV rentables mais aussi des lignes entre villes de provinces peu fréquentées et non rentables. Les profits des lignes TGV permettent de combler les pertes sur les lignes non rentables et de ne pas fermer des gares dans des zones moins denses en matière de population, c'est ce qu'on nomme les péréquations tarifaires géographiques du service universel).

[28] Souvenez-vous que l'Union européenne souhaite introduire la concurrence « partout où cela est possible ».

[29] Le monopole pouvait même dans certains cas être une administration (les PTT avant la scission ayant abouti à la création de La Poste et de France Telecom).

2.1.3.2 Mise en concurrence et démantèlement partielle ou totale du monopole

Le principe d'ouverture à la concurrence consiste à limiter le monopole légal historique dans le seul maillon en monopole naturel ; la gestion de l'infrastructure, et à instaurer une concurrence pour la fourniture des services utilisant ce réseau. Cela peut déboucher :

- soit sur le démantèlement du monopole historique en plusieurs entreprises concurrentes (cas de ATT au début de la décennie quatre-vingt aux Etats-Unis),
- soit sur la conservation du monopole historique et la tentative de faire émerger de nouveaux opérateurs concurrents.
- Soit sur la seule levée des obstacles juridiques interdisant l'arrivée de nouveaux concurrents (concurrence uniquement potentielle sensée discipliner le monopole en place dans la droite ligne de la théorie des marchés contestables de Baumol[30])

Les Etats-Unis et le Royaume-Uni ont souvent opté pour le démantèlement lors des processus d'ouverture à la concurrence (découpage par zones géographiques ou par métiers). Ces démantèlements se sont souvent accompagnés de la privatisation des opérateurs. Il a par ailleurs été mis sur pieds des agences de régulation spécialisées indépendantes par secteur, pour promouvoir la concurrence et défendre les consommateurs.

L'Union européenne, et donc la France, se sont plus orientées vers une conservation des monopoles historiques et la tentative de faire émerger des concurrents. L'Union européenne a progressivement mis en place un cadre réglementaire incitant à ouvrir à la concurrence les principaux services en réseaux (télécoms, transport aérien, poste, énergie, transport ferroviaire…). Les autorités européennes préconisent une gestion des réseaux fondées sur la mise en place d'autorités sectorielles de régulation indépendantes des opérateurs et des autorités de tutelle (ART, Autorité de Régulation des Télécoms par exemple en France), ainsi qu'une transparence des relations financières entre l'Etat et les entreprises en charge du service universel. La Communauté est en revanche neutre à l'égard du régime de propriétés des entreprises. Elle ne s'oppose pas à la présence de l'Etat comme actionnaire (minoritaire ou majoritaire). La privatisation des monopoles, en l'absence d'ouverture à la concurrence, n'est jugée en aucun cas comme la solution souhaitable.

Il est important de bien comprendre que dans le processus d'ouverture à la concurrence, l'acte primordial est la séparation de l'industrie entre l'amont (gestion de l'infrastructure) et l'aval (la fourniture du service). Ainsi en France, a-t-on assisté soit à la séparation complète entre fourniture de service et activités de réseaux (exemple SNCF et RFF, Réseau Ferré de France, chargé de l'entretien des lignes ferroviaires), soit à l'obligation de fourniture d'accès au réseau par l'opérateur historique lorsque celui-ci a conservé pour des raisons d'efficacité économique l'entretien de celui-ci (« boucle locale »[31], pour la téléphonie filaire).

2.1.3.3 Les bénéfices à attendre de l'ouverture à la concurrence

Les activités de réseaux (eau, gaz, électricité, poste, télécoms, transports) représentent en France 8% du PIB et sont déterminantes pour l'attractivité du territoire national.

La concurrence peut, si elle est correctement mise en place, stimuler les entreprises d'un secteur et les obliger à améliorer la qualité de leurs services ou à baisser leurs coûts de production. Il peut donc ressortir de cette ouverture à la concurrence dans les industries de réseaux en Europe une amélioration de l'efficacité économique et un accroissement du surplus des consommateurs.

La concurrence peut cependant également déstabiliser un secteur. Un pays comme la France dont la diversité géographique est importante doit veiller à ce que les péréquations tarifaires puissent demeurer, de façon à ce que chaque concitoyen puisse continuer à disposer des services universels.

Il existe un risque que sur un marché concurrentiel non régulé, les nouveaux opérateurs concurrents se placent uniquement sur les segments les plus rentables, abandonnant la desserte des usagers à faible pouvoir d'achat ou situés dans des zones à faible densité de population (il est évident que les futurs concurrents de la SNCF vont venir la concurrencer sur ses lignes TGV rentables et non sur la ligne « Tours-Caen » dont la survie ne tient qu'à des interventions « politiques ». La SNCF si elle est forcée de maintenir les lignes non rentables pour des raisons « politiques » sans recevoir les compensations nécessaires de la part des régions ou de l'Etat sera alors en position délicate).

Un autre risque est de voir apparaître des sous-capacités de production dans des secteurs fortement capitalistiques. Les entreprises en place peuvent être tentées de limiter les capacités de production

[30] L'économiste Baumol a montré qu'une entreprise seule sur un marché pouvait se comporter comme une entreprise en situation de concurrence, pour peu que le marché soit « contestable », c'est-à-dire qu'il n'y ait aucun obstacle à la libre entrée de sorte que la menace de l'entrée de concurrents en cas de prix trop élevé pousse l'entreprise à se discipliner.

[31] « La boucle locale » est la paire de fils de cuivre qui sépare la prise téléphonique d'un abonné du répartiteur, à partir duquel les communications sont dirigées vers le réseau régional ou national.

existantes pour éviter un écroulement des prix avec l'arrivée de nouveaux concurrents. Cela peut donc amener à retarder des investissements pourtant essentiels (notamment dans le domaine de l'énergie). L'Etat doit donc accompagner l'ouverture à la concurrence du développement des capacités de production par des programmes pluri-annuels d'investissements.
Il convient enfin d'avoir à l'esprit qu'un véritable marché concurrentiel dans les industries de réseau ne peut apparaître qu'au niveau européen. La taille des marchés nationaux est parfois trop faible pour permettre l'émergence d'une véritable concurrence. Il convient donc de réaliser au plus vite les interconnexions des marchés nationaux afin d'unifier les marchés européens (par exemple dans le domaine de l'électricité).

2.2 « Asymétries d'information et nouvelle microéconomie »

L'une des hypothèses nécessaire à la réalisation de la concurrence parfaite est que l'information soit parfaite. Or, dans la réalité, en de nombreuses situations, il existe des asymétries d'information (assurance, crédit, marché de l'occasion, relation de travail…). Depuis G. Akerlof (1970), qui a le premier modélisé les échanges en présence d'asymétrie d'information (en prenant pour exemple le marché des voitures d'occasion), les économistes ont élaboré des modèles à même de décrire les effets de ces asymétries d'information sur le niveau et le prix des échanges. L'agent qui détient la rente informationnelle peut utiliser celle-ci pour améliorer son sort lors de l'échange. De l'autre côté, l'agent qui se doit de connaître l'information que détient son partenaire dans l'échange peut amener celui-ci à la révéler en mettant en place un mécanisme incitatif. Une annexe générale en fin d'ouvrage décrit la problématique des asymétries d'information (**sélection adverse** et **aléa moral**) et présente brièvement le modèle principal - agent qui décrit la relation entre un individu (le principal) qui mandate, contre paiement, un agent pour effectuer une tâche stipulée dans un contrat. Le principal est confronté au problème de risque moral (situation où un côté du marché ne peut pas observer les comportements de l'autre côté.) lorsqu'il observe imparfaitement l'action entreprise par l'agent. Le problème du principal est donc de déterminer la forme d'un contrat qui incite l'agent à agir dans l'intérêt du principal (**théorie des contrats**).

2.3 Concurrence sur le marché des biens et concurrence pour l'innovation

En matière de relation entre concurrence et innovation, un premier argument remonte à Schumpeter (1942)[32]. L'**argument schumpétérien** est le suivant : qu'on s'en félicite ou qu'on le regrette (Schumpeter le regrettait puisqu'il voyait dans la généralisation de ce phénomène la fin du capitalisme et le probable passage au socialisme), la structure de marché la plus appropriée pour l'innovation est le monopole. Seule l'entreprise de grande taille ou le monopole qui dégage des profits est à même de disposer de la surface financière suffisante pour mener des programmes de recherche et développement en continu et amortir par un portefeuille de recherches diversifiées le risque inhérent à l'activité d'innovation.
Cet argument schumpétérien « justifie » ainsi un pouvoir de marché *ex ante* de manière à fournir les conditions d'apparition des innovations. L'idée le monopole est la structure de marché la plus adéquate pour l'innovation a cependant été remise en cause par Arrow (1962) qui a avancé l'argument de **l'effet de remplacement**. Le monopole est moins incité qu'une firme en situation de concurrence à innover. En réalisant une innovation de procédé (qui abaisse son coût marginal) le monopole s'assure un différentiel de gain plus faible que celui d'une entreprise en concurrence parfaite. En effet, le monopole se remplace lui-même en passant d'un profit de monopole avant l'innovation à un profit de monopole plus élevé après l'innovation (uniquement un gain marginal), alors qu'une entreprise en concurrence (ou bien un nouvel entrant) obtient un différentiel de profit plus élevé (passe d'un profit nul avant l'innovation à un profit de monopole après l'innovation).
Gilbert et Newbery (1982) mettent en avant un *effet d'efficacité* qui, sous certaines conditions, peut cependant expliquer la persistance du monopole et redonnent ainsi du crédit à l'argument schumpétérien. Lorsqu'un monopole est contraint par la menace d'entrée d'un concurrent potentiel, deux cas peuvent apparaître. Soit le monopole est le plus incité à innover et il conservera son monopole. Soit c'est le concurrent qui a la plus forte incitation à l'innovation. A la suite de celle-ci, il fait son entrée sur le marché qui devient un duopole. Pour le monopole en place, l'incitation à l'innovation est la différence entre le profit si elle réalise l'innovation et le profit de duopole. Si cette différence est supérieure au profit de duopole qu'obtient le concurrent en effectuant l'innovation,

[32] Schumpeter (1942), Capitalisme, socialisme et démocratie.

condition exprimée sous le nom d'**effet d'efficacité**, le monopole est le plus incité à réaliser l'innovation (il se maintien en tant que monopole). En cela, la disponibilité à investir d'un monopole est la valeur qui annule le profit actualisé du concurrent potentiel (stratégie de prolifération de brevets : Intel, Microsoft, Xerox).
Les théories économiques qui voyaient dans les firmes en place et le monopole les structures les plus aptes à faire naître l'innovation ont cependant paru de plus en plus obsolètes à mesure que de nouveaux secteurs d'activité basés sur le développement de nombreuses innovations sont apparus à la fin de la décennie quatre-vingt-dix et au début du XXI^{ième} siècle. Les nombreuses « start-up » dans le domaine des biotechnologies ou des technologies de l'information et de la communication s'avèrent des vecteurs importants de l'innovation, dans un environnement relativement concurrentiel (nouveaux entrants venant concurrencer des entreprises installées qui n'ont pas la réactivité en matière d'innovation permise par la flexibilité et la petite taille des « start-up »).
Les recherches les plus récentes (Boone (2000) , Aghion, Harris, Howitt et Vickers (2001)[33]) sur le lien entre concurrence et innovation tendent cependant à réconcilier la théorie avec la réalité des faits empiriques. L'idée est alors que la concurrence est un bon aiguillon de l'innovation mais uniquement jusqu'à un certain **seuil**.
Boone (2000) a montré que la relation entre l'intensité de la concurrence et les incitations à innover est de nature non monotone. Cela provient du fait que lorsqu'on passe d'un régime de concurrence à un autre, l'identité de l'innovateur se modifie ainsi que la valeur de l'innovation et son type. Lorsque l'intensité de l'innovation est forte, c'est la firme qui possède les coûts de production les plus faibles qui est la plus incité à mettre en œuvre l'innovation. Dans les secteurs où l'intensité de l'innovation est faible, c'est au contraire, les firmes qui ont les coûts de production les plus élevés qui innovent et tentent de rattraper les leaders. Lorsque la firme possède un avantage important en matière de coût, l'incitation à réaliser une innovation de produit est d'autant plus forte que la concurrence est élevée sur le marché des produits. Lorsque le coût de la firme est proche de celui de ses concurrents, l'incitation à réaliser une innovation de procédé est d'autant plus forte que la concurrence est élevée sur le marché des produits. L'augmentation de la pression concurrentielle sur le marché des produits ne peut accroître en même temps les innovations de produits et de procédés dans un secteur donné.
Aghion, Howitt, Harris, Vickers (2001) établissent également que le lien entre le degré de concurrence et l'intensité de l'innovation serait plutôt en forme de U-inversé. Lorsque le degré de concurrence est faible dans un secteur, le degré d'innovation l'est aussi. L'augmentation de la concurrence exerce une incitation positive sur l'activité d'innovation jusqu'à un certain seuil au-delà duquel, la concurrence « lamine » les profits et limite la possibilité de recherche et d'innovation (zone correspondant au schéma schumpétérien).
Cependant, Eto (2004)[34] a réintroduit les arguments en faveur de la plus grande incitation à l'innovation des firmes en place à la façon de Gilbert et Newberry, conditionnant l'efficacité à l'existence d'une libre entrée. Les entreprises leaders, du fait de leur capacité d'innovation, sont fortement incitées à investir en recherche et développement, lorsque la libre entrée sur le marché est permise (pas de barrières institutionnelles ou d'entraves à la concurrence) car le leader ne peut alors conserver sa position dominante que si il gagne la course à l'innovation. Lorsque l'entrée sur le marché est au contraire verrouillée, l'entreprise dominante a tendance à rechercher à bénéficier le plus possible de sa position et retarde ses projets d'innovation. La concurrence entendue comme la libre entrée dans un secteur demeure donc un facteur nécessaire à l'efficacité du processus d'innovation même si l'innovation provient d'une firme installée en position dominante.

Dans une économie centrée sur l'innovation, les dimensions traditionnelles de la concurrence en prix ou en quantités cèdent le pas à de nouvelles formes de concurrence portant sur de nouveaux produits, de nouvelles opportunités technologiques, des accords de coopération, des transferts de technologies, la conquête de nouveaux marchés. La concurrence sur le marché est remplacée par la concurrence pour le marché. La définition même du marché pertinent permettant de mesurer le pouvoir de monopole peut s'avérer difficile car la concurrence la plus probable n'est pas encore matérialisée. Pour les autorités de la concurrence, il convient alors d'examiner avec recul des positions qui peuvent paraître dominantes tout en préservant les possibilités d'une concurrence provenant d'un nouvel entrant, c'est-à-dire lever tous les obstacles ou barrières à l'entrée qui pourraient empêcher l'innovation d'une entreprise extérieure.

---

[33] Aghion Ph., Harris C., Howitt P. et Vickers J. (2001), « Competition, Imitation and Growth with Step-by-Step Innovation », *Review of Economic Studies*. Une application empirique de ce modèle est fournie dans Aghion Ph., Bloom N., Blundell R., Griffith R. et Howwit P. (2003), « Competition and Innovation : An Inverted-U Relationship », *The Quaterly Journal of Economics*, vol.120, pp.701-728.
[34] Eto (2004), « Innovation by leaders », *Economic Journal*.

2.4 Annexes

2.4.1 Fonctions de coût et économies d'échelle

Soit une entreprise dans un secteur en situation de concurrence pure et parfaite. Sa fonction de coût total est : $CT(Q) = CF + CV(Q)$ où CF est le coût fixe et CV(Q) le coût variable en fonction du niveau de production. On peut définir deux fonctions : (i) la fonction de **coût moyen** : CM(Q)=CT(Q)/Q ; (ii) la fonction de **coût marginal** : Cm(Q)=dCT(Q)/dQ. Cette dernière indique le coût supplémentaire de produire une unité en plus.

Le coût fixe moyen est décroissant avec les quantités produites (on amortit le coût fixe en produisant en grandes séries). En revanche, le coût variable moyen est plutôt croissant avec les quantités produites car les facteurs de production donnés (capital, travail) doivent être utilisés plus intensivement (recours au heures supplémentaires, usure plus importante des machines…). Il en résulte que la fonction de coût moyen a un profil d'abord décroissant puis croissant. Le coût marginal (qui ne prend pas en compte les coûts fixes) est pour sa part croissant. La propriété suivante peut être notée : **la fonction de coût marginal coupe la fonction de coût moyen en son minimum** (figure 1).

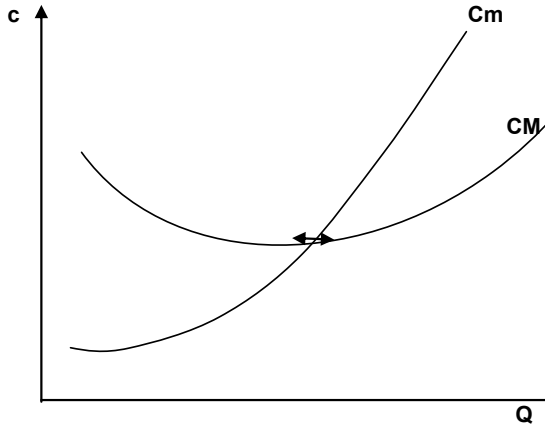


Figure 4 : Fonctions de coût

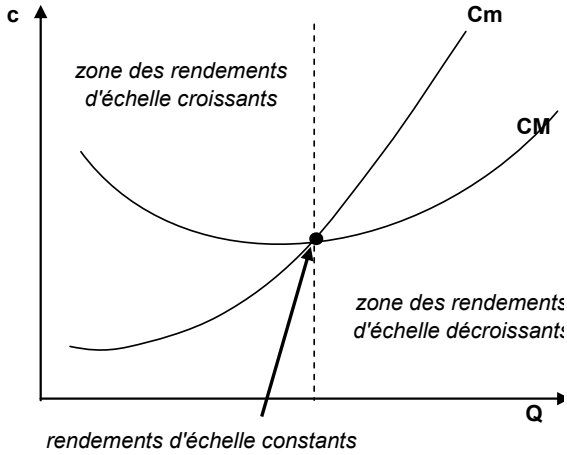


Figure 5 : Zones des rendements d'échelle

On peut situer sur ce schéma les zones d'économies d'échelle ou de déséconomies d'échelle. Lorsque la courbe de coût marginal se situe en dessous de la courbe de coût moyen, l'entreprise a intérêt à accroître son volume de production car le coût de la prochaine unité produite (Cm) est inférieur au coût moyen de l'ensemble des unités déjà produites (CM). La production de cette unité supplémentaire aura pour conséquence de faire baisser le coût moyen. On dit qu'on est en présence d'économies d'échelle ou que les rendements d'échelle sont croissants (le coût moyen est décroissant). Au contraire lorsque le coût marginal se situe au dessus du coût moyen, la production d'une unité supplémentaire fait croître le coût moyen de l'ensemble des unités produites. On est alors en présence de déséconomies d'échelle. On dit qu'on est dans la zone des rendements d'échelle décroissants. Pour l'entreprise, il est intéressant de faire diminuer son volume de production. Lorsque le coût moyen égalise le coût marginal, on est en présence de rendements d'échelle constants.

On peut ainsi déterminer la taille optimale de l'entreprise qui correspond au minimum du coût moyen de l'entreprise. Plus la zone des rendements d'échelle croissants est importante et plus le fait d'avoir une grande capacité de production permet de réduire le coût unitaire. L'entreprise a dans ce cas une taille importante.

2.4.2 Rendements de la fonction de production

Soit une entreprise dont la fonction de production est Q=F(K,L). On dit que la fonction de production F(.,.) est à **rendements constants** si lorsque les quantités utilisées de chacun des facteurs de production sont multipliées par un facteur λ>1, la production est également multipliée par λ : $\lambda Q=F(\lambda K, \lambda L)$.

La fonction est à rendements croissants si $\lambda Q<F(\lambda K, \lambda L)$, et à rendements décroissants si $\lambda Q>F(\lambda K, \lambda L)$.

La fonction de production Cobb-Douglas souvent retenue ($Q=K^{\alpha}L^{1-\alpha}$) est une fonction à rendements constants, c'est-à-dire qu'une augmentation du capital et du travail dans une même proportion induit également une hausse proportionnelle de la production.

2.4.3 La notion d'élasticité-prix de la demande

On définit l'élasticité prix de la demande de la manière suivante : $\varepsilon = \frac{dQ}{dp}\frac{p}{Q}$. Cette élasticité ε mesure la sensibilité de la demande Q(p) à une variation de prix. Comme la pente de la fonction de demande est toujours négative, cette élasticité a un signe négatif. Cependant on omet très souvent son signe (puisqu'il va de soi ; lorsque le prix augmente, la demande diminue) et on exprime donc l'élasticité prix de la demande en valeur absolue.

Lorsque l'élasticité-prix de la demande est supérieure à 1 en valeur absolue ($\varepsilon \in [-\infty,-1]$), on dit que la demande est élastique puisqu'elle réagit fortement à une évolution de prix. Lorsque l'élasticité-prix de la demande est inférieure à 1 en valeur absolue ($\varepsilon \in [-1,0[$), on dit que la demande est faiblement élastique car elle réagit relativement peu aux évolutions de prix. L'élasticité-prix de la demande est un concept important car il permet d'apprécier le pouvoir d'un marché d'un monopole. Plus la demande en face d'une entreprise en situation de monopole est peu sensible au prix et plus le monopole peut augmenter son prix sans que ses ventes ne chutent drastiquement. Lorsque l'élasticité-prix de la demande est nulle, on dit que la demande est inélastique.

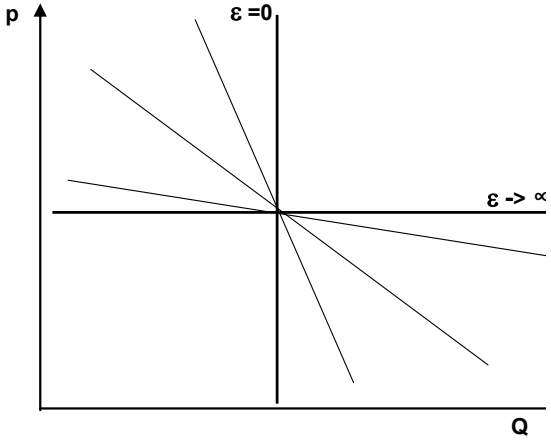


Figure 6 : Demande inélastique et infiniment élastique

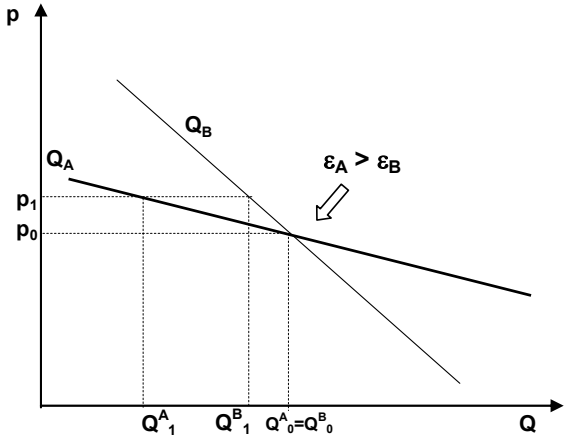


Figure 7 : Comparaison de deux élasticités

La figure ci-dessus permet d'apprécier la situation des droites de demande dont les élasticités tendent vers les limites 0 et -∞ (ou plus exactement la situation de droites de demande inverse car p(Q) est appelée fonction de demande inverse et Q(p) la fonction de demande). La fonction de demande inverse est verticale dans le plan (Q,p) lorsque la demande est totalement inélastique (cas ε=0). Les variations de prix n'ont aucun effet sur la quantité demandée qui est fixée. Au contraire lorsque la demande devient infiniment élastique (cas ε->∞), une infime variation de prix occasionne une réponse en quantité infinie. Entre ces deux cas limites se situent l'ensemble des fonctions de demande réalistes dont la pente est plus ou moins forte. La pente de la fonction de demande dépend de l'élasticité même si elles ne sont pas exactement égales (puisque $\frac{dQ}{dp} = \varepsilon\frac{Q}{p}$). De la même manière la pente de la fonction inverse ne se confond pas avec l'inverse de l'élasticité-prix de la demande (puisque $\frac{dp}{dQ} = \frac{1}{\varepsilon}\frac{p}{Q}$). Ainsi plus on remonte le long d'une fonction de demande inverse et plus l'élasticité devient importante (car p/Q tend vers l'infini).

On ne comparera donc la valeur des élasticités de deux fonctions de demande uniquement à leur point d'intersection. Sur la figure deux sont tracées deux fonctions de demande inverse (celles de deux individus A et B par exemple). L'élasticité-prix de la demande de la fonction de demande de A est plus importante que celle de la fonction demande de B au point d'intersection (situation 0). Ainsi la réaction à une hausse de prix (de p_0 à p_1) est plus forte pour l'individu A que pour l'individu B. La fonction de demande de A est plus élastique que la fonction de demande de B ($\varepsilon_A > \varepsilon_B$).

Voici quelques exemples d'élasticité-prix de la demande de quelques fruits : pommes (-0,21), cerises (-0,77), raisins (-0,16), pêches (-0,24), poires (-1,64), calculées dans une étude aux Etats-Unis au début des années 80. Une étude parue dans Economie et Statistiques (n°324-325, 1999) a calculé l'élasticité-prix des différents types de boissons consommées par les ménages français : vins de

consommation courante (-0,963), vin d'appellation (-0,516), alcools forts (-0,497), alcools doux (-0,675), anis (-0,435), bière (-0,746), eaux (-0,939).
On notera par ailleurs que les élastictés-prix de la demande peuvent évoluer dans le temps (exemple de la hausse du prix du tabac). L'impact de la hausse initiale du prix du tabac est relativement fort. Les achats baissent immédiatement (si hausse de 1%, baisse de 0,5%). Cela est souvent dû au contrecoup des achats massifs réalisés juste avant l'augmentation… Puis les achats repartent et au bout de 6 mois, la baisse de la consommation n'est plus que de 0,2% inférieure à son niveau d'origine. A long terme, la consommation rebaisse légèrement (et se stabilise autour d'un niveau inférieur de 0,3% au niveau d'origine). Cette baisse est souvent le fait du découragement de fumeurs potentiels (les jeunes par exemple).

Il existe d'autres élasticités. L'**élasticté-prix croisés** de la demande permet de saisir la façon dont la demande d'un bien réagit à l'évolution de prix d'un autre bien ($\varepsilon_{ij} = \frac{dQ_i}{dp_j}\frac{p_j}{Q_i}$). Lorsque l'élasticité-prix croisés est positive, on dit que les biens sont **substituables** (exemple : le beurre et la margarine), lorsqu'elle est négative on dit qu'ils sont **complémentaires** (exemple : le picon et la bière pour peu que tout le monde mette du picon dans la bière…). L'**élasticté-revenu** de la demande permet d'appréhender la façon dont la consommation d'un bien répond aux évolutions de revenu ($\varepsilon_R = \frac{dQ}{dR}\frac{R}{Q}$). Si l'élasticité-revenu est supérieure à l'unité, le bien est dit « **de luxe** » (exemple : toiles de maîtres, bijoux, grands crus…), si elle est négative, le bien est dit « **inférieur** » (exemples : pommes de terre, vins de table…). Lorsque le revenu augmente ; dans le premier cas, la demande augmente plus que proportionnellement ($\varepsilon_R > 1$), dans le second cas, la demande augmente moins que proportionnellement ($\varepsilon_R < 0$)

2.4.4 Le surplus du consommateur

Le concept théorique d'utilité introduit pour mesurer la satisfaction ressentie par un agent lors de la consommation d'un bien ou de plusieurs bien est assez peu opératoire. Il est en effet difficile voire impossible de déterminer empiriquement les fonctions d'utilité. Pour mesurer la satisfaction ou le bien-être associée à la consommation d'une certaine quantité de biens à un prix donné, les économistes disposent cependant d'un outil : le surplus du consommateur.
Le **surplus du consommateur** est la différence entre le prix auquel le consommateur était prêt à payer le bien (sa disponibilité à payer) et le prix qu'il paye effectivement (le prix de marché). Il est facilement calculable pour peu qu'on connaisse précisément la fonction de demande. La formule du surplus est :

$S = \int_0^{Q^*} \left(p(Q) - p^*\right) dQ$ où p* et Q* sont les prix et quantité de cet échange.

Graphiquement, c'est la zone hachurée sur le schéma suivant :

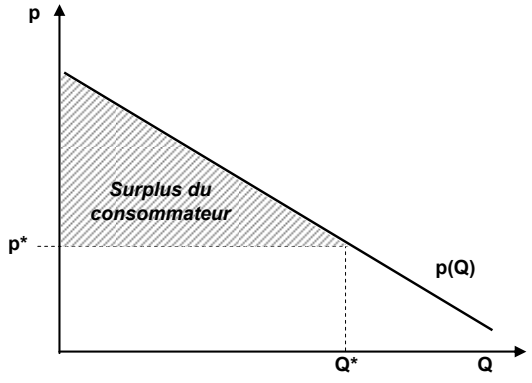


Figure 8 : Surplus du consommateur

On peut considérer que la fonction de demande inverse représente soit l'agrégation des différents consommateurs pour un bien donné soit la quantité de biens demandée par un consommateur donné en fonction du prix de marché. La fonction de demande inverse, dans le premier cas, peut être vue comme un ordonnancement des consommateurs en fonction de leurs disponibilités à payer décroissantes. Le point le plus haut de la droite p(Q) représente un consommateur prêt à mettre un prix très important pour acquérir ce bien. Puis viennent les consommateurs prêts à payer un peu

moins. Le dernier consommateur à acheter le bien est celui dont la disponibilité à payer est égal au prix de marché. Bien entendu tous les consommateurs payent leurs unités au prix de marché. Il y a donc une satisfaction à acquérir pour un prix inférieur un bien pour lequel on était prêt à mettre un somme plus importante. Cette différence est le surplus individuel. Lorsqu'on agrège sur l'ensemble des consommateurs, on obtient le surplus du consommateur. Un raisonnement analogue pourrait être mené en considérant un consommateur unique achetant des unités de biens tant que sa disponibilité à payer est supérieure au prix de marché.
On constate que toute augmentation de prix ou de réduction de la quantité consommée réduit l'aire hachurée et abaisse donc le surplus du consommateur (ce sera par exemple le cas lorsqu'un marché en situation de concurrence devient monopolistique).
Il existe également un surplus du producteur qui se confond avec les profits lorsque les coûts fixes sont nuls.

2.4.5 La situation de concurrence pure et parfaite

L'objectif d'une entreprise est de maximiser son profit. Le profit réalisé par une entreprise est la différence entre ses recettes et son coût de production.
Le profit d'une entreprise s'écrit $\pi(Q) = p.Q - CT(Q))$ où p.Q sont les recettes réalisées par cette entreprise (Q unités de biens vendues au prix p) et CT(Q) le coût de production associé.
Lorsque l'entreprise est en situation de concurrence pure et parfaite, on considère que son poids est trop faible pour que ses décisions aient une quelconque influence sur le prix de marché p (hypothèse d'atomicité. L'entreprise en concurrence considère donc le prix p constaté sur le marché comme une donnée (on dit qu'elle est *price taker*) et doit déterminer la quantité optimale à produire de manière à maximiser son profit.
La condition du premier ordre lorsqu'on dérive le profit par rapport à la variable de choix (Q) est donc :

$$\frac{d\pi(Q)}{dQ} = 0 \Leftrightarrow p = Cm(Q)$$

L'entreprise va produire jusqu'à ce que son coût marginal égalise le prix de marché (rappelez-vous que le coût marginal est croissant en Q).
Quel sera le profit réalisé par cette entreprise ? Avant de répondre à cette question, on peut exprimer le profit moyen : $\frac{\pi(Q)}{Q} = p - CM(Q)$. Le profit par unité produite est la différence entre le prix de marché et le coût moyen de l'entreprise.
Aussi longtemps que le prix de marché est supérieur au coût moyen de l'entreprise, celle-ci réalise des profits. Cependant, comme nous sommes en situation de concurrence pure et parfaite, les profits réalisés dans ce secteur vont attirer d'autres entreprises. C'est ce qu'on appelle la **libre-entrée**. On considère qu'il existe une multitude de firmes identiques en termes de technologie (même fonction de coût) qui sont prêtes à se créer et à entrer dans ce secteur si les conditions de marché le permettent. L'entrée de nouvelles firmes dans ce secteur aura pour conséquence d'augmenter la quantité totale de biens offerte par les entreprises et donc de faire baisser le prix (à demande constante). Les profits réalisés par les entreprises de ce secteur vont donc s'amoindrir à mesure que de nouvelles firmes s'installent et font baisser le prix. A long terme, il y aura suffisamment d'entreprises installées pour que le prix de marché soit égal au coût moyen de chacune de ces entreprises (celles qui auraient un coût moyen supérieur aux autres auraient disparu…).
On peut donc formuler les résultats suivants pour la situation de concurrence :

- Les profits ne peuvent exister qu'à court terme (durant le processus de libre entrée).
- **A long terme, les profits sont nuls** en situation de concurrence et **p=Cm=CM**.

Mais alors, à quoi sert de produire pour faire des profits nuls ? Attention, il ne s'agit en fait que des surprofits ! Dans la fonction de coût sont intégrées les rémunérations des facteurs de production (capital=>distribution des dividendes, remboursement des intérêts ; travail=>rémunération des salariés), c'est-à-dire que même si les profits sont nuls, les salariés sont payés et les détenteurs du capital rémunérés… Ouf !

2.4.6 La situation de monopole

En situation de monopole, c'est-à-dire lorsqu'une firme unique fait face à la demande, il n'est bien entendu plus possible de considérer que le comportement de l'entreprise dominante n'a pas d'influence sur le prix de marché. L'entreprise a pleinement conscience du fait que ses décisions en matière de quantités servies influencent le prix auquel elle va pouvoir vendre ses unités. On va donc considérer que l'entreprise en situation de monopole connaît la fonction de demande et intègre cette connaissance dans son programme de maximisation du profit.

Celui-ci s'écrit alors : $\underset{Q}{Max}\ \pi(Q) = p(Q).Q - CT(Q)$

La condition du premier ordre est donc : $\frac{d\pi(Q)}{dQ} = 0 \Leftrightarrow \underbrace{\frac{dp(Q))}{dQ}Q + p(Q)}_{Rm(Q)} = Cm(Q)$.

L'entreprise en situation de monopole n'égalise plus son coût marginal au prix de marché mais à la recette marginale. La condition du premier ordre du programme de maximisation du profit d'une entreprise en situation de monopole est donc : **Recette marginale = Coût marginal**.

On peut exprimer la recette marginale de la façon suivante :

$Rm(Q) = \frac{dp(Q)}{dQ}Q + p(Q) = \left(\frac{1}{\varepsilon} + 1\right)p(Q)$ où ε<-1 est l'élasticté-prix de la demande.

On a donc : $p(Q) = \frac{Cm(Q)}{\left(\frac{1}{\varepsilon} + 1\right)} > Cm(Q)$. En monopole, le prix est strictement supérieur au coût marginal (et donc au prix de concurrence). Le rapport entre le prix et le coût marginal est appelé le taux de marge. On constate que moins la demande à laquelle fait face le monopole est élastique (ε tend vers -1) et plus le taux de marge de celui-ci est élevé.

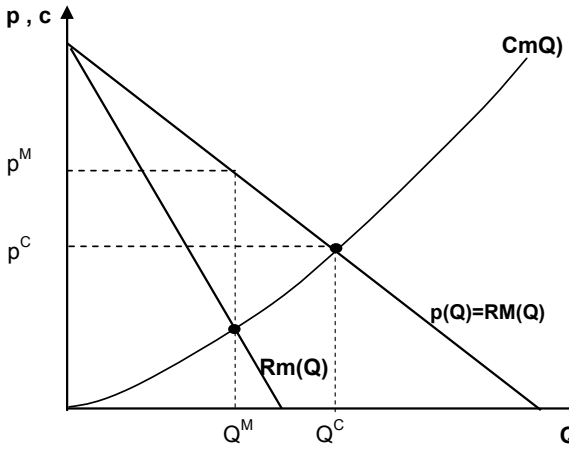


Figure 9 : Situations de concurrence et de monopole

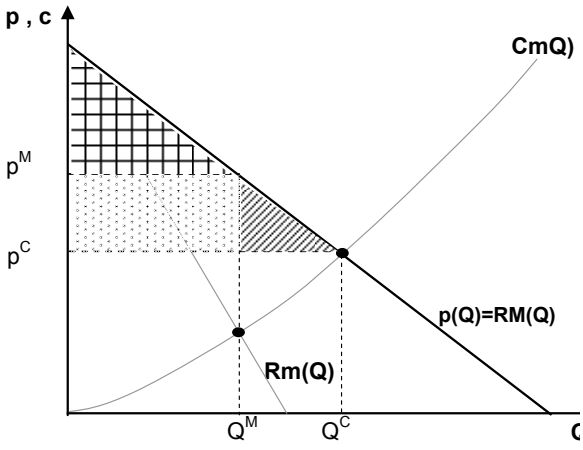


Figure 10 : Analyse des surplus

Sur la figure ci-dessus (à gauche), on visualise la façon dont sont déterminés les équilibres de concurrence et de monopole. En considérant que le coût moyen égalise le prix de concurrence (c'est-à-dire en postulant que l'équilibre de long terme était atteint), on peut procéder à une analyse en terme de bien-être. Le surplus du consommateur en concurrence pure et parfaite est le large triangle rectangle au dessus de p^C. Lorsqu'on passe à une situation de monopole, la surface du surplus du consommateur en concurrence est la somme de trois aires : (i) le surplus du consommateur dans cette situation de monopole (triangle rectangle quadrillé), (ii) le profit du monopole (le rectangle pointillé) et (iii) la **charge morte du monopole** (le triangle rectangle hachuré)

Cette surface de l'ancien surplus du consommateur est désormais perdue pour la société. C'est le fait qu'une partie du surplus ne soit plus accaparé par le producteur ou les consommateurs qui rend la situation de monopole moins efficace au point de vue du bien-être que la situation de concurrence. En termes de bien-être, peu importe que le surplus aille au consommateur ou au producteur (vous voulez bien payer vos achats plus chers pour peu que vous touchiez des dividendes plus importants…). En revanche, qu'une partie du surplus disparaisse permet d'affirmer que la situation de concurrence est préférable à celle du monopole.

Le monopole réduit donc les quantités de biens offertes sur le marché de manière à accroître le prix de vente. La quantité produite en monopole diminue par rapport à la concurrence tandis que le prix de marché augmente ($Q^M < Q^C \Leftrightarrow p^M > p^C$).

2.4.7 La concurrence monopolistique

Nous pouvons examiner plus précisément le lien entre économies d'échelle et structure de marché. Prenons l'exemple d'une firme dont la fonction de coût possède des coûts fixes. Le coût moyen est alors toujours supérieur au coût marginal. Des firmes en concurrence, déterminant leurs quantités par l'égalisation de leur coût marginal au prix de marché, feront toujours des pertes. La concurrence est donc impossible. Prenons une fonction de la forme $CT(Q_i) = CF + c.Q_i$ où c est le coût marginal, il est évident que $CM(Q_i) = \frac{CF}{Q_i} + c$. Seule une firme en position de monopole peut produire en réalisant des profits (cf. figures ci-dessous).

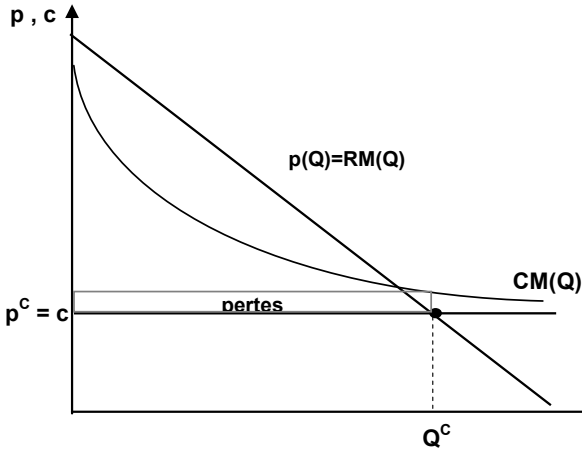


Figure 11 : Equilibre de concurrence

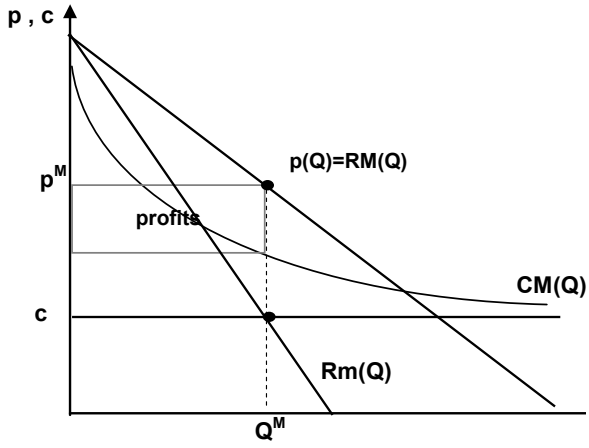


Figure 12 : Equilibre de monopole

Un modèle se situe entre ces deux cas polaires et permet de décrire une situation de concurrence avec rendements d'échelles croissants : c'est le modèle de **concurrence monopolistique** introduit par l'économiste Chamberlin.

Les hypothèses du modèle sont les suivantes ; il existe n firmes identiques dans un secteur donné de l'économie. Soit $\overline{p}$ le prix moyen pratiqué dans ce secteur et p, le prix de la firme représentative. Soit Q les ventes totales du secteur et Q_i les ventes de la firme représentative. La fonction de demande adressée à la firme représentative prend la forme suivante : $Q_i = Q\left(\frac{1}{n} - \beta(p - \overline{p})\right)$. En concurrence imparfaite, une firme vend d'autant plus que la demande totale est élevée et que ses concurrents pratiquent des prix élevés. Elle vendra d'autant moins qu'elle pratique elle-même un prix élevé et que le nombre de concurrents est élevé. Si toutes les entreprises adoptent le même prix, elles ont chacune une part de marché identique 1/n. Si l'entreprise représentative pratique un prix supérieur au prix moyen du secteur alors cette entreprise détient une part de marché inférieure (le paramètre β permet de calibrer la sensibilité par rapport à la déviation moyenne du prix). On fait l'hypothèse que les ventes totales du secteur sont constantes et que les entreprises n'ont pas conscience de leur potentielle influence sur le prix de marché (en cela on conserve l'hypothèse d'atomicité de la concurrence parfaite).

Le programme de maximisation de la firme représentative s'écrit :

$\underset{Q_i}{Max} \quad \pi(Q_i) = p.Q_i - c.Q_i$ où $p = \frac{1}{\beta n} - \frac{1}{\beta}\frac{Q_i}{Q} + \overline{p}$ (fonction de demande inverse).

La condition du premier ordre, $\frac{d\pi(Q_i)}{dQ_i} = 0$, peut être exprimée sous la forme suivante : $p = \frac{Q_i}{\beta Q} + c$. Or toutes les firmes sont symétriques ($Q_i = \frac{Q}{n}$) et donc, on a le prix d'équilibre

suivant : $p = \frac{1}{\beta n} + c$. Le prix fixé est donc supérieur au coût marginal. Cependant, le prix fixé et donc la marge réalisée par les firmes est une fonction décroissante du nombre de concurrents.
Il existe des rendements d'échelle croissants dans ce secteur puisque le coût moyen de chaque entreprise est croissant avec le nombre d'entreprises du secteur : $CM(Q_i) = \frac{n.CF}{Q} + c$

Ces deux relations d'équilibre (prix et coût moyen) sont satisfaites au point d'équilibre. Ce point se traduit par un nombre d'entreprises n* pour lequel les profits du secteur sont nuls. En dessous les entreprises font des profits ce qui suscite des entrées et donc une baisse du prix et une hausse du coût moyen. Inversement, si le nombre d'entreprises présentes dans ce secteur est supérieur à n*, les entreprises font des pertes et certaines sortent jusqu'à ce que le secteur redevienne rentable.

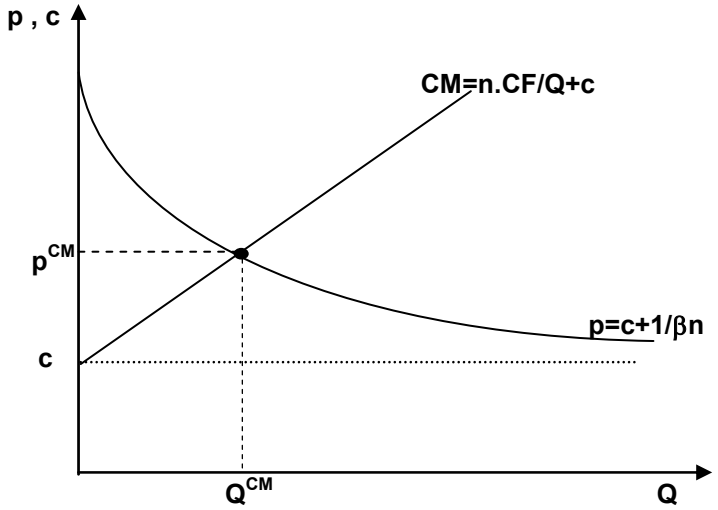


Figure 13 : Equilibe de concurrence monopolistique

A l'équilibre p=CM=c. Le nombre de firmes présentes sur ce secteur est : $n^* = \sqrt{\frac{Q}{\beta.CF}}$. Le prix d'équilibre est : $p = \sqrt{\frac{CF}{\beta.Q}}$, tandis que les ventes par firme sont : $Q_i = \sqrt{\beta.Q.CF}$. En raison des rendements croissants, une augmentation de la taille du marché entraîne une augmentation du nombre de firmes moins que proportionnelle.

2.4.8 La fixation des prix par un monopole public

Un certain nombre de secteurs de l'économie sont caractérisés par l'importance des coûts fixes à la base de la production du bien ou du service. C'est particulièrement le cas des industries qui nécessitent préalablement l'établissement de réseaux (télécommunications, distribution d'énergie, transports…). Comme nous l'avons vu précédemment, la présence de coûts fixes importants correspond à une situation de rendements d'échelle croissants pour laquelle la production en concurrence est impossible. La production sous forme de monopole s'impose donc d'un point de vue de l'efficacité économique. Cependant, nombre de ces services ou biens sont essentiels pour le bon fonctionnement de l'économie et nécessaires à la vie quotidienne des citoyens. L'Etat s'est donc porté garant de la fourniture de ces biens et services au plus grand nombre à un prix acceptable, en établissant l'obligation de service public. On voit bien là deux objectifs antagonistes. D'un côté le monopole souhaite restreindre les quantités servies pour élever le prix au-delà du coût moyen et ainsi réaliser des profits. De l'autre l'Etat souhaite que cette entreprise, producteur unique du fait de l'existence de rendements croissants, servent le plus grand nombre au prix le plus bas possible. La théorie économique permet d'établir le moyen de fixer le prix optimal pour un monopole public de ce type.

Pour compléter l'objectif de service public (servir le plus grand nombre) tout en respectant les contraintes économiques (ne pas faire de pertes), le monopole public doit **fixer son prix au niveau de son coût moyen**. C'est ce qu'on appelle la tarification de « moindre mal » (tarification à la Ramsey-Boiteux).

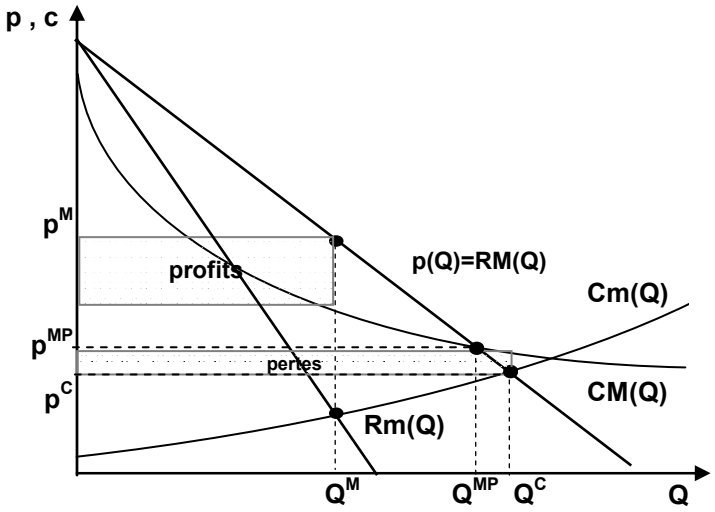


Figure 14 : Tarification de « moindre mal »

Sur le schéma ci-dessus, on voit que cette situation (MP) se situe entre la tarification monopolistique (M) permettant les profits mais écartant bon nombre de consommateurs de l'accès à ce bien ou services, et la tarification concurrentielle (C) qui permet de servir un maximum de consommateurs mais aboutit à des pertes pour l'entreprise. Etant données les fonctions de coûts et de recettes adoptées pour cette illustration, la solution de tarification au coût moyen n'est ici pas très éloignée de la tarification concurrentielle.

Par ailleurs, le schéma ci-dessus permet d'illustrer le comportement optimal d'un monopole public mono-produit. Or, dans la réalité, le monopole public fait face à une clientèle différenciée et offre plusieurs biens et services (la SNCF vend différents trajets de rentabilités différentes, EDF fournit les particuliers et les entreprises, France Telecom fait de la téléphone fixe, du mobile et de l'ADSL…). Ainsi la tarification au coût moyen s'entend sur l'ensemble des produits et services de l'entreprise publique. Celle-ci peut ainsi financer des services non-rentables entrant dans l'obligation de service public par des services plus rentables et commerciaux (la tarification sur les lignes TGV permet de maintenir certaines lignes régionales non rentables par exemple…).

Ce bel ordonnancement est cependant remis en question pat l'obligation d'ouverture à la concurrence imposée par l'Union Européenne.

2.4.9 L'oligopole

La concurrence oligopolistique est caractérisée par des interactions stratégiques entre un petit nombre limité d'entreprises sur un marché. Ce nombre limité d'entreprises peut s'expliquer par des coûts d'entrée importants. Les exemples les plus connus d'industrie oligopolistique sont l'industrie automobile, la production d'acier, d'aluminium et des produits chimiques, les groupes d'équipements électroniques (ordinateurs, téléphonie), etc. Deux cas de figures sont possibles. Soit les entreprises sont interchangeables et ont un rôle symétrique dans un marché (le même poids), soit elles ont un rôle asymétrique car l'une d'elles peut avoir un rôle dominant. Le premier cas renvoie à une représentation simultanée de l'équilibre économique caractérisé par l'oligopole de Cournot ou de Bertrand alors que le dernier renvoie à une illustration séquentielle de l'équilibre économique au sens de l'oligopole de Stackelberg. Les comportements stratégiques des firmes peuvent être de deux sortes : non-coopératif (chaque entreprise maximise son profit individuel en tenant compte du comportement de ses concurrents) ou coopératif (les entreprises décident d'entreprendre des décisions qui facilitent la coordination de leurs actions dans une industrie et qui limitent en conséquence la concurrence).

2.4.9.1 Le duopole de Cournot

Hypothèses

Considérons deux entreprises $i = 1,2$ qui produisent un bien homogène en quantité q_1 et q_2 ayant respectivement des coûts de production $C_1(q_1) = c_1 q_1$ et $C_2(q_2) = c_2 q_2$. La quantité totale de bien disponible sur le marché est $Q = q_1 + q_2$. Les entreprises connaissent la courbe de demande inverse de bien adressé au marché $p(Q) = p(q_1 + q_2) = A - (q_1 + q_2)$. Par conséquent, la distribution des prix que sont prêts à payer les consommateurs est connue pour chaque quantité totale offerte sur le marché. Les variables stratégiques de chacune des firmes sur le marché sont leur niveau de production. Le choix de production d'une entreprise se réalise en prenant en compte la décision du concurrent. Le bien produit dans la branche est parfaitement homogène (c'est-à-dire parfaitement substituable). Chaque firme a pour objectif la maximisation de son profit en s'adaptant aux conditions de marché. Les choix des firmes vont dépendre des conjectures sur le comportement stratégique des concurrents.

L'équilibre de Cournot

Chaque entreprise a pour objectif de maximiser son profit en s'adaptant aux conditions du marché et sous la contrainte du niveau de production anticipée de l'entreprise concurrente.
Le programme de maximisation du profit de l'entreprise 1 est :

$$\max_{q_1} \prod_1 = p(Q)q_1 - C_1(q_1) = p(q_1 + q_2)q_1 - C_1(q_1)$$

Le programme de maximisation du profit de l'entreprise 2 est :

$$\max_{q_2} \prod_2 = p(Q)q_2 - C_2(q_2) = p(q_1 + q_2)q_2 - C_2(q_2)$$

Le profit de la firme 1 dépend de la quantité produite par la firme 2 au travers de la recette totale (qui dépend de la demande totale et de la quantité offerte individuellement par chaque entreprise). Chaque entreprise maximise donc son profit pour un niveau de production donné du concurrent.
Les conditions de maximisation indépendante du profit conduisent au résultat d'égalisation de la recette marginale au coût marginal. Les conditions du premier ordre s'écrivent :
Pour la firme 1 :

$$\frac{\partial \prod_1}{\partial q_1} = p(q_1 + q_2) + p'(q_1 + q_2)q_1 - C_1'(q_1) = 0$$

$$= A - (q_1 + q_2) - q_1 - c_1 = 0$$

$$\Leftrightarrow q_1 = R_1(q_2) = \frac{A - c_1}{2} - \frac{1}{2} q_2$$

Pour la firme 2 :

$$\frac{\partial \prod_2}{\partial q_2} = p(q_1+q_2) + p'(q_1+q_2)q_2 - C_2'(q_2) = 0$$
$$= A-(q_1+q_2)-q_2-c_2 = 0$$
$$\Leftrightarrow q_2 = R_2(q_1) = \frac{A-c_2}{2} - \frac{1}{2}q_1$$

Ces conditions du premier ordre expriment pour chaque entreprise sa production optimale en fonction de ses anticipations sur le choix de production de son concurrent. Ces deux équations donnent la fonction de réaction de la firme 1 $q_1 = R_1(q_2)$ compte tenu de ses différentes conjectures sur le choix possible de la firme, et identiquement pour la firme 2 $q_2 = R_2(q_1)$. Ces fonctions de réaction sont considérées comme des fonctions implicites de meilleures réponses à la quantité offerte par l'entreprise concurrente. Graphiquement, les fonctions de réaction sont décroissantes dans le plan (q_1, q_2).

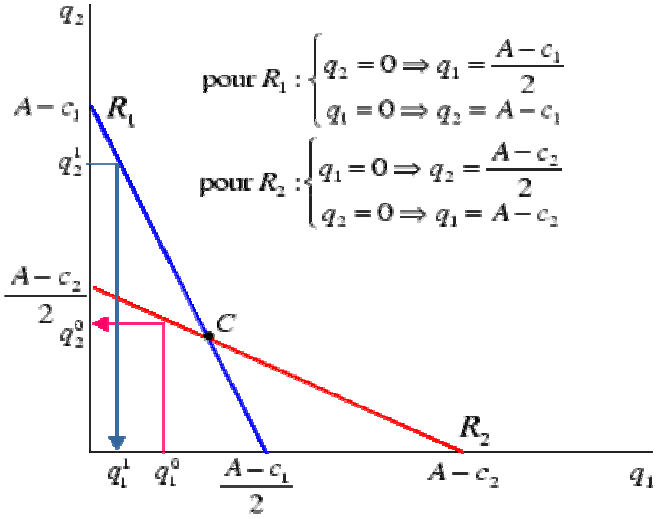


Figure 15 : L'équilibre de Cournot

L'équilibre est un cas particulier de ces fonctions de réaction. Ainsi la première firme, en fonction de ses coûts, va déterminer la quantité à produire qui maximisera son profit. La deuxième firme, connaissant la demande, déterminera avec sa fonction de réaction la quantité à offrir en fonction de la quantité offerte par la première. Mais celle-ci aura également un impact sur la fonction de réaction de la première firme, qui devra elle aussi ajuster alors sa quantité offerte, etc. L'équilibre de marché doit être une situation telle qu'une fois atteinte, aucune firme ne doit avoir envie de s'éloigner de cet état, aucune firme ne doit pouvoir améliorer son profit en produisant une quantité autre que sa quantité d'équilibre.

Cette situation (q_1^C, q_2^C) est un **équilibre de Cournot**. La quantité d'équilibre de chaque firme est sa meilleure réaction à la quantité d'équilibre de son concurrent et la firme ne peut plus améliorer son profit en modifiant ses quantités. Pour trouver l'équilibre, il faut donc résoudre le système posé par les fonctions de réaction. En substituant $q_2 = R_2(q_1)$ dans la fonction de réaction de la firme 1, $q_1 = R_1(R_2(q_1))$, la production d'équilibre de la firme 1 est déterminée par q_1^C. En remplaçant dans la fonction de réaction de la firme 2, la production de la firme 2 est déterminée par q_2^C.

La production de la firme est déterminée par :

$$q_1 = R_1(q_2) = \frac{A-c_1}{2} - \frac{1}{2}\left(\frac{A-c_2}{2} - \frac{1}{2}q_1\right)$$
$$\Leftrightarrow q_1^C\left(1-\frac{1}{4}\right) = \frac{A-2c_1+c_2}{4}$$
$$\Leftrightarrow q_1^C = \frac{A-2c_1+c_2}{3}$$

La production de la firme 2 est obtenue par :

$$q_2^C = R_2(q_1^C) = \frac{A - 2c_2 + c_1}{3}$$

L'offre et le prix d'équilibre :

$$Q^C = q_1^C + q_2^C = \frac{2A - c_1 - c_2}{3}$$

$$p^C = \frac{A + c_1 + c_2}{3}$$

Les profits sont donnés par :

$$\prod_1^c = p^C q_1^C - c_1 q_1^C = \left(\frac{A - 2c_1 + c_2}{3} \right)^2$$

$$\prod_2^c = p^C q_2^C - c_2 q_2^C = \left(\frac{A - 2c_2 + c_1}{3} \right)^2$$

Cet équilibre de marché apparaît donc dans une situation où les firmes prennent leur décision de production de manière isolée, sans communication entre elles.

2.4.9.2 Le duopole de Bertrand

Dans le modèle de Bertrand, la variable stratégique est le prix, le comportement des firmes est symétrique et les firmes ont une capacité de production suffisante pour couvrir la totalité du marché. Cette situation renvoie à des situations où les firmes vendent un produit homogène similaire (par exemple, Avis et Hertz pour la location de voiture).

Hypothèses

La demande est contingente, c'est-à-dire qu'elle est dépendante du niveau de prix décidé par l'autre firme. Ainsi, si une firme fait payer un prix supérieur au prix fixé par son concurrent la demande pour son produit sera nulle ; si elle fait payait un prix inférieur au prix fixé par son concurrent, elle s'emparera de toute la demande du marché. Si les deux entreprises fixent le même prix pour le produit, elles se répartiront équitablement la demande sur le marché. Chaque entreprise fixe un prix pour ses produits p_j avec *j=1* ou *2*. Notons $D(p)$ la fonction de demande qui s'adresse aux deux entreprises lorsqu'elles choisissent le même prix *p*. Dans ce cas, la demande se répartit à l'identique entre les entreprises, chacune percevant une demande égale à $D(p)/2$. Lorsque les prix p_1 et p_2 diffèrent, la demande s'adresse en priorité à l'entreprise qui a fixé le prix le plus bas (car les firmes produisent des biens homogènes). Notons D_1 et D_2 les demandes adressées respectivement à la première et à la deuxième entreprise.
La répartition de la demande est donc définie par :

$$\begin{cases} D_1 = D(p_1) \quad et \quad D_2 = 0 & si \quad p_1 < p_2 \\ D_1 = D_2 = D(p_1)/2 & si \quad p_1 = p_2 \\ D_1 = 0 \quad et \quad D_2 = D(p_{12}) & si \quad p_1 > p_2 \end{cases}$$

Toutes les firmes sont supposées avoir assez de capacité de production pour fournir la totalité du marché. La variable stratégique de chacune des firmes sur le marché est le prix. Le bien produit dans la branche est parfaitement homogène (=parfaitement substituable). Chaque firme va chercher à maximiser le profit contingent qu'elle pourrait réaliser dans les circonstances créées par le duopoleur.

Equilibre de Bertrand

Considérons un cadre de duopole symétrique où deux entreprises produisent un bien homogène et ont des fonctions de coûts identiques. Supposons que le coût moyen est constant pour les deux entreprises tel que $CT_1(q_1) = cq_1$ et. Le paramètre *c* représente le coût moyen (égal au coût marginal), commun aux deux firmes. Les entreprises suivent des stratégies non-coopératives et décident du prix auquel elles vendent leurs produits. Par ailleurs chaque firme considère que la décision de son concurrent est donnée. Chaque entreprise anticipe donc que son concurrent maintiendra le prix qu'il a fixé, indépendamment du prix qu'elle détermine elle-même pour ses propres produits.

Chaque entreprise satisfait la demande qui s'adresse à elle et fixe son prix de manière à maximiser son profit. Celui-ci s'écrit $\prod_1 = (p_1 - c)q_1$ pour l'entreprise 1 et $\prod_2 = (p_2 - c)q_2$ pour l'entreprise 2.

Un équilibre de Bertrand est alors caractérisé par des prix p_1^* et p_2^* tel que pour chaque entreprise fixer son prix au niveau de l'équilibre est une décision optimale si celle-ci anticipe que son concurrent va agir de même. En d'autres termes, (p_1^*, p_2^*) constitue un équilibre de Bertrand si, en maximisant son profit, l'entreprise 1 choisit $p_1 = p_1^*$ en considérant le prix de l'entreprise 2 comme fixé à p_2^*, et symétriquement. L'équilibre de Bertrand est défini de manière unique par la condition $p_1^* = p_2^* = c$. Ainsi, à l'équilibre de Bertrand, les deux firmes fixent un même prix égal au coût marginal, c'est-à-dire le prix qui prévaut en concurrence pure et parfaite.

Démonstration

- Si $p_1 > p_2 > c$, la firme 1 ne vendra pas de bien et fera donc un profit nul. Or en fixant un prix $p_1 = p_2 - \varepsilon$, la firme 1 peut s'emparer de la totalité du marché et faire un profit positif. C'est la même chose pour la firme 2 qui répondra en fixant un prix encore plus bas. Par conséquent, une telle situation ne peut pas être un équilibre car les deux firmes continueront simplement à pratiquer des réductions de prix jusqu'à ce que le prix atteigne le coût marginal.
- Si $p_1 = p_2 > c$, les deux firmes vont se partager le marché, cependant un tel arrangement n'est pas stable car si l'une des deux réduit son prix, elle s'emparera de la totalité du marché et donc des profits encore plus élevés. Ce n'est donc pas non plus un équilibre.
- Si $p_1 > p_2 = c$, la firme 2 ne gagnera aucun profit (car prix=Cm) et la firme 1 ne réalisera aucun profit car son prix est trop élevé. Cependant, la firme 2 a intérêt à augmenter son prix en le conservant toujours inférieur à p_1 pour s'emparer de la totalité du marché. Ce n'est pas un équilibre.
- $p_1 = p_2 = c$, c'est la seule possibilité d'obtenir un équilibre (équilibre de Bertrand). Dans cette situation, les deux firmes ne feront pas de profit mais seront indifférentes entre rester sur le marché ou en sortir.

A l'équilibre de Bertrand, le duopole conduit donc au même niveau de prix qu'un marché de concurrence parfaite, les firmes obtiennent un profit nul. Ceci tient au fait qu'une entreprise peut toujours obtenir et satisfaire la totalité de la demande en fixant un prix légèrement inférieur à celui de son concurrent.

2.4.9.3 **Le modèle Stackelberg** (rôle asymétrique des entreprises)

Le modèle de Stackelberg est une extension du modèle de Cournot mais qui tient compte d'un comportement asymétrique de la part des deux firmes sur un marché duopolistique. Comme à l'équilibre de Cournot, chaque entreprise est supposée décider de la quantité qu'elle désire produire mais l'une des deux entreprises choisit son niveau de production en supposant que son concurrent va prendre sa décision selon sa fonction de réaction. En effet, la dynamique industrielle crée souvent des firmes dominantes, soit parce qu'elles ont un poids quantitatif important (part de marché élevée – Microsoft dans le secteur des systèmes d'exploitation pour les compatibles PC), soit parce qu'elles ont un comportement agressif et innovateur (Dell contre IBM dans le secteur des ordinateurs compatibles PC).
Le modèle de Stackelberg suppose donc qu'une firme joue un rôle actif sur le marché (le leader ou meneur) et l'autre firme joue un rôle passif (le follower ou suiveur). La firme 1 est supposée être l'entreprise dominante sur le marché. La firme 2 follower réagit comme précédemment en considérant la production de la firme 1 comme donnée, elle se contente de suivre le comportement du leader Une interprétation possible de cette situation est que la firme 1 choisit son volume de production avant la firme 2 et que cette décision est irréversible : la firme 2 sait qu'elle ne pourra conduire la firme 1 à modifier sa production et elle la considère donc comme une donnée. Ce qui confère statut de firme leader, c'est donc la capacité de s'engager sur une décision non susceptible d'être remise en cause.

La résolution de l'équilibre de Stackelberg s'apparente à la résolution d'un jeu en deux étapes.

Le problème de l'entreprise follower

La firme suiveuse détermine son niveau de production à partir de la maximisation de son profit, en tenant compte des choix effectué par la firme leader $\overline{q}_1$:

$$\max_{q_2} \prod_2 = p(\overline{q}_1 + q_2)q_2 - C_2(q_2)$$

La fonction de réaction de la firme follower est déterminée à partir de la condition du premier ordre :

$$\frac{\partial \prod_2}{\partial q_2} = p(\overline{q}_1 + q_2) + p'(\overline{q}_1 + q_2)q_2 - C_2^{'}(q_2) = 0$$

$$\Leftrightarrow q_2 = R_2(\overline{q}_1) = \frac{A - c_2}{2} - \frac{1}{2}\overline{q}_1$$

Le problème de l'entreprise leader

La firme dominante 1 tient compte du fait que ses décisions influences le choix de production de la firme suiveuse 2. En d'autres termes, l'entreprise 1 connaît la fonction de réaction de la firme follower $q_2 = R_2(\overline{q}_1)$. La firme 1 essaie donc d'atteindre le niveau le plus élevé de profit en respectant la fonction de réaction de son concurrent. En fait,la firme 1 tient compte du fait que le suiveur n'acceptera jamais de produire une quantité qui ne maximise pas son profit. La firme 1 essaie donc de se déplacer sur sa courbe d'iso-profit correspondant au profit le plus élevé possible, qui a au moins un point d'intersection avec la courbe de réaction de l'entreprise 2 suiveuse.Le programme de maximisation du profit du leader est par conséquent : $\max_{q_1} \prod_1 = p\left(q_1 + R_2(q_1)\right)q_1 - C_1(q_1)$

La condition de premier ordre de la firme dominante est :

$$\frac{\partial \prod_1 \left(q_1, R_2(q_1)\right)}{\partial q_1} = \frac{\partial \prod_1 \left(q_1, R_2(q_1)\right)}{\partial q_1} + \frac{\partial \prod_1 \left(q_1, R_2(q_1)\right)}{\partial q_2}\frac{dq_2}{dq_1} = 0$$

$$\Leftrightarrow q_1^S = \frac{A - 2c_1 + c_2}{2} > q_1^C$$

D'où la production de la firme suiveuse :

$$q_2^S = R_2(q_1^S) = \frac{A - 3c_2 + 2c_1}{4}$$

Le prix et les profits à l'équilibre de Stackelberg sont déterminés par :

$$p^S = \frac{A + 2c_1 + c_2}{4}$$

$$\prod_1^S = \frac{\left(A - 2c_1 + c_2\right)^2}{8} > \prod_1^c$$

$$\prod_2^c = \left(\frac{A - 3c_2 + 2c_1}{4}\right)^2$$

Le leader prend donc en compte le fait que lorsqu'il choisit son niveau de production q_1, la production de l'entreprise suiveuse est modifiée selon la fonction de réaction de cette dernière.

Application numérique : $A = 100, \quad c_1 = 5, \quad c_2 = 10$

$$q_1^C = 33.3 < q_1^S = 50$$

$$q_2^C = 28.3 > q_2^S = 20$$

$$\prod_1^c = 1111 < \prod_1^S = 1250$$

$$\prod_2^c = 802.77 > \prod_2^S = 400$$

La position du meneur améliore donc la situation de la firme 1 par rapport à l'équilibre de Cournot.

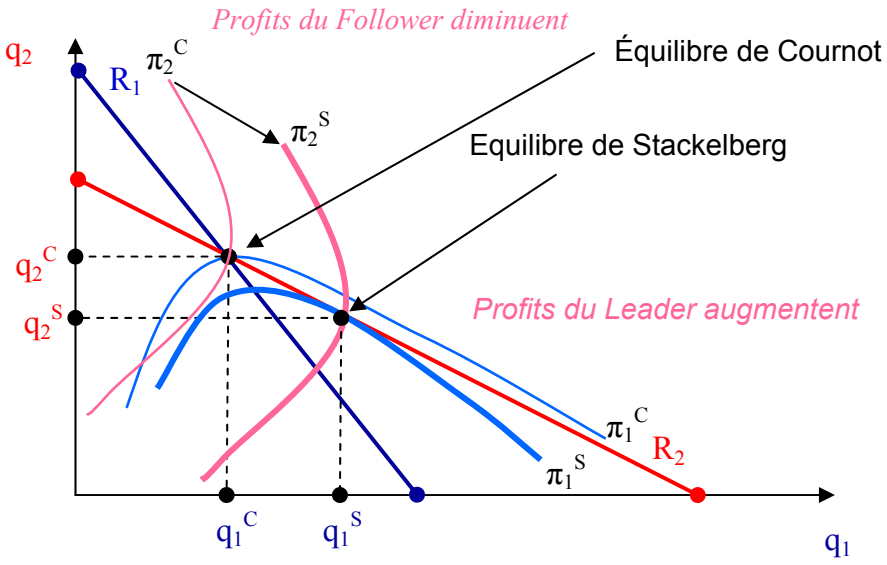


Figure 16 : Equilibre de Stackelberg *vs.* équilibre de Cournot

2.4.9.4 L'entente collusive ou cartel

Lorsque les entreprises entreprennent des **stratégies coopératives** sur le marché, elles forment une entente visant à choisir les niveaux de production pour lequel les profits totaux sont maximums. Dans cette situation, les entreprises forment un cartel de façon à se comporter comme un monopole unique sur le marché et à maximiser la somme de leur profit.

Le problème de maximisation du profit pour les deux entreprises consiste dès lors à choisir leurs productions q_1 et q_2 afin de maximiser les profits totaux du secteur :

$$\max_{q_1,q_2} \prod = \prod_1 + \prod_2 = p(q_1+q_2)(q_1+q_2) - C_2(q_1) - C_2(q_2)$$

Ceci détermine les conditions d'optimalité suivantes :

$$\frac{\partial \prod}{\partial q_1} = p(q_1+q_2) + p'(q_1+q_2)(q_1+q_2) - C_1'(q_1) = 0$$

$$\frac{\partial \prod}{\partial q_2} = p(q_1+q_2) + p'(q_1+q_2)(q_1+q_2) - C_2'(q_2) = 0$$

L'interprétation de ces conditions est intéressante. Si la firme 1 envisage d'accroître sa production, elle est confrontée à deux effets : le profit supplémentaire découlant de la vente d'un output supplémentaire et la diminution du profit consécutive à la réduction du prix. Mais dans ce second effet, elle tient désormais compte de l'impact de la réduction du prix à la fois sur son propre niveau de production mais aussi sur celui de son concurrent. Ce résultat découle de l'intérêt de la firme 1 pour la maximisation du profit total et non plus simplement de son profit. A l'équilibre du cartel, il y a égalité des coûts marginaux de production. Si une entreprise a un avantage en terme de coût, elle produira nécessairement davantage à l'équilibre dans une solution de cartel.

En supposant $c_1 = c_2 = c,\ P = A - (q_1+q_2)$ le problème du cartel devient :

$$\max_{q_1,q_2} \prod = \prod_1 + \prod_2 = (A - q_1 + q_2)(q_1+q_2) - c(q_1+q_2)$$

$$\frac{\partial \prod}{\partial q_1} = (A - q_1 + q_2) - (q_1+q_2) = c$$

$$\frac{\partial \prod}{\partial q_2} = (A - q_1 + q_2) - (q_1+q_2) = c$$

En posant $Q = q_1 + q_2$, ces conditions d'optimalité qui sont redondantes donnent quand même une condition d'optimalité pour l'output total : $A - 2Q^* = c$. Ce qui donne l'équilibre de collusion :

$$Q^* = \frac{A-c}{2} \quad p^* = \frac{A+c}{2} \quad \Pi^* = \left(\frac{A-c}{2}\right)^2$$

La solution de l'équilibre de collusion correspond à celle du monopole. Donc, en s'associant, les deux firmes sont capables d'atteindre ensemble le profit de monopole. La répartition du profit total entre les firmes dépend du pouvoir de négociation de chacune d'elles. S'il y a des possibilités de transfert de revenus entre les firmes, le pouvoir de négociation de chaque firme va encore jouer dans la détermination des profits individuels.

La cartellisation est en général interdite par la réglementation de la concurrence. C'est le cas en Europe et aux États-Unis. Les cartels ont un autre problème : la **stabilité**. Nous pouvons voir cela en reprenant les conditions d'optimalité et en imaginant que la firme envisage d'augmenter sa quantité à partir de la solution de cartel. Dans ce cas, la condition d'optimalité du cartel implique :

$$p\left(q_1 + q_2^*\right) + (q_1 + q_2^*)\frac{\partial p}{\partial q_1} - C_1'\left(q_1\right) \quad \text{ou} \quad p\left(q_1 + q_2^*\right) + q_1\frac{\partial p}{\partial q_1} = -q_2^*\frac{\partial p}{\partial q_1} > 0$$

Le membre de gauche de cette condition est le profit marginal de la firme et ce profit marginal est donc positif à l'optimum du cartel. Ce qui veut dire que la firme sera incitée à augmenter sa production si elle pense que son partenaire ne va pas modifier la sienne. Par conséquent, si les firmes ne peuvent pas observer les quantités individuelles, cela va déboucher dans un duopole de Cournot où chacun va obtenir, en fin de compte, des profits plus faibles que dans le cartel. C'est un cas typique de dilemme du prisonnier.

La théorie des jeux (voir en annexes générales, la présentation sur la théorie des jeux) permet de prendre en compte ces situations d'interactions stratégiques et l'incitation à ne pas respecter l'accord collusif (voir le problème dit du « dilemme du prisonnier »)

3 Rôle de la monnaie et financement de l'économie

La monnaie occupe une place prépondérante dans nos économies développées. Elle est en effet au cœur de tous les mécanismes économiques, dans la mesure où les opérations d'achat et de vente s'effectuent presque exclusivement en monnaie (le troc a pratiquement disparu dans nos économies…). La monnaie est donc un sujet important pour l'analyse économique. Plusieurs questions émergent immédiatement. A quoi sert la monnaie ? Qui est garant de la valeur de la monnaie ? La quantité de monnaie en circulation a-t-elle des effets sur l'économie réelle ? Si oui, ces effets sont-ils durables ? Nous allons voir que ces questions font l'objet d'un débat déjà ancien entre économistes. Le rôle de la monnaie dans l'économie est bien loin de faire l'objet d'un consensus entre économistes. C'est au contraire une ligne de fracture importante entre économistes keynésiens et classiques.

Nous allons donc, dans ce chapitre nous interroger sur l'identité des agents qui créent de la monnaie et la manière dont le processus de création monétaire fonctionne. Nous verrons ainsi le rôle du système bancaire dans ce processus puis examinerons le processus de financement de l'économie. Les deux phénomènes sont en effet intimement liés. Lorsque les banques commerciales accordent des crédits à des particuliers ou des entreprises, elles augmentent la quantité de monnaie en circulation puisqu'elles accordent des liquidités immédiates contre une promesse de remboursements futurs. Le système de financement par le crédit peut donc s'avérer inflationniste si il n'est pas contrôlé par une autorité. Le système bancaire est ainsi hiérarchisé. Au dessus des banques commerciales dont l'un des objectifs est d'accorder des crédits et donc financer l'activité, se trouve la banque centrale qui va être le garant de la valeur de la monnaie et lutter contre le phénomène d'inflation.

Ainsi, nous verrons les modalités d'intervention de la banque centrale et la façon dont elle peut réguler la quantité de monnaie en circulation. L'octroi de crédit n'est cependant pas l'unique moyen de financer l'économie. Les entreprises peuvent aller directement à la rencontre des offreurs de capitaux sur les marchés financiers sans passer par l'intermédiaire des banques. Il conviendra donc d'étudier le fonctionnement des marchés financiers et d'appréhender certaines pratiques sur ceux-ci qui peuvent s'avérer dangereuses pour le système financier global et donc également pour l'économie réelle.

Plusieurs phénomènes d'actualité seront abordés dans ce chapitre. Premièrement, comment et pourquoi une bulle financière peut-elle se former ? A titre d'exemple, la figure de gauche ci-dessous montre l'évolution de la valorisation des sociétés « technologiques » entre 1999 et 2006. On voit bien qu'entre janvier 1999 et mars 2000, un épisode de bulle financière[35] a eu lieu (large décrochage entre l'EURO STOXX Technology, indice des valeurs « technologiques » et l'indice EURO STOXX large, prenant en compte les sociétés de l'ensemble des secteurs). Deuxièmement, pourquoi les entreprises, les journalistes, les économistes, les politiques sont-ils régulièrement dans l'attente d'une décision de la banque centrale en matière de taux d'intérêt directeurs[36] ? Quels sont les éléments qui décident les gouverneurs qui siègent au conseil de la BCE (Banque Centrale Européenne) à un

[35] On pense également immédiatement à la désormais fameuse crise du *subprime* qui a agité les bourses mondiales cet été. A la base, c'est une crise qui concerne les ménages américains les moins solvables qui s'étaient endettés à taux variable pour acheter des biens immobiliers, souvent en utilisant des crédits hypothécaires (crédits gagés sur la valeur du logement de l'emprunteur). La baisse des prix de l'immobilier américain ainsi que la remontée des taux d'intérêt depuis plus d'un an ont rendu la situation de ces ménages intenable. Mais pourquoi cette simple crise de surendettement a provoqué une panique sur l'ensemble des bourses mondiales ? Parce que ces crédits avaient ensuite été titrisés, c'est-à-dire convertis en titres regroupant de nombreux emprunts afin d'en diminuer le risque. Ces titres avaient attirés de nombreux investisseurs (notamment des *hedge funds*) à la recherche de rentabilités élevées. Lorsque les ménages endettés ont commencé à ne pas pouvoir rembourser, cela a abouti à des faillites des organismes prêteurs mais également à une panique boursière puisque les véhicules de cette titrisation (les fonds en partie investis dans ces *subprimes*) se révélaient remplis de créances douteuses. Les conditions du crédit se sont alors durcies entraînant une crise de liquidité sur le marché monétaire. Les banques centrales (Fed et BCE notamment) ont ainsi du intervenir massivement pour fournir des liquidités aux banques (processus de refinancement) afin d'éviter que cette crise ne se propage à l'économie réelle en durcissant les modalités d'accès au crédit pour l'ensemble des acteurs de l'économie.

[36] Voir l'effet salvateur de la baisse des taux de la Fed (la Banque centrale américaine) intervenue mi-août 2007 et qui a permis (momentanément ?) de stabiliser les marchés financiers. Voir également les appels de la Ministre de l'économie et des finances française, Madame Lagarde, pour une baisse des taux de la part de la BCE en septembre.

relèvement de taux ? La figure de droite ci-dessous met ainsi en regard la croissance d'un indice des prix à la consommation (HICP) et les différents taux d'intérêt directeurs de la zone euro. Nous allons tenter d'expliquer en partie les interactions entre inflation et taux d'intérêt en décrivant précisément les processus de création monétaire et de financement de l'économie.

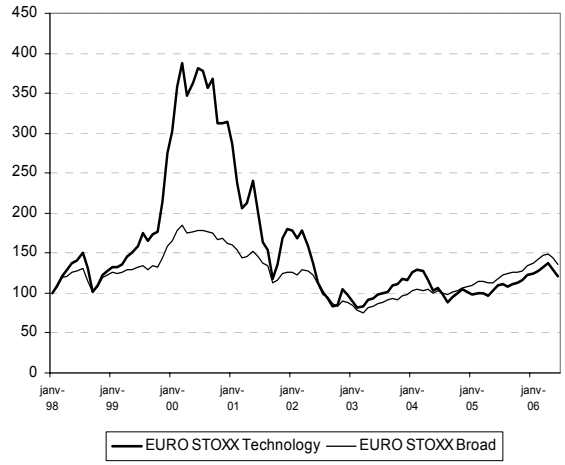


Figure 17 : Bulle sur les valeurs technologiques en 2000

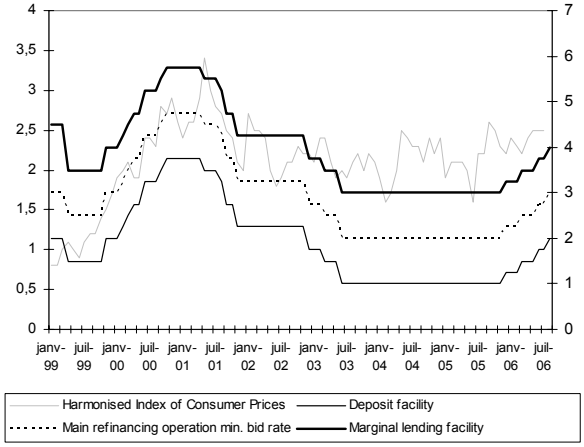


Figure 18 : Indice des prix à la consommation et taux directeurs de la BCE

De la monnaie au niveau des prix, en passant par l'efficacité de la politique monétaire, nous allons tenter de percevoir les enjeux économiques derrière les débats théoriques, afin que vous soyez à même de percevoir les enjeux économiques derrière les évolutions du cours des actifs boursiers et des taux d'intérêt directeurs de la Banque centrale.
Nous devrons pour cela nous intéresser aux modalités du financement de l'économie. Comment se finance l'activité économique ? Quelles ont été les mutations en matière de financement de l'économie depuis une vingtaine d'année ? Est-ce que l'accent mis sur la lutte contre l'inflation par la politique monétaire a eu des répercussions sur les modalités de financement de l'économie ? Quelles sont les incidences de ces changements d'une part sur le fonctionnement des entreprises et d'autre part sur la conduite de la politique économique? Le système financier actuel est-il compatible avec le vieillissement de la population d'un certain nombre de pays développés et est-il capable d'être un adjuvant au financement de la retraite par répartition ?

3.1 La monnaie

Les définitions de la monnaie sont nombreuses. Elle peut être aussi considérée comme « l'ensemble des moyens de paiement dont disposent les agents économiques pour régler leurs transactions » ou comme « le bien d'échange généralement accepté par une communauté de paiement ». La première définition a le défaut de limiter l'utilisation de la monnaie à son rôle dans les transactions, la seconde fait apparaître la notion de communauté de paiement (espace géographique ou zone monétaire comme l'UEM) mais précise peu le rôle spécifique de la monnaie. En fait, la monnaie se définit au regard des trois grandes fonctions qu'elle remplit dans l'économie : (i) la fonction d'intermédiaires des échanges, (ii) la fonction d'unité de compte et (iii) la fonction de réserve de valeur.

3.1.1 La fonction d'intermédiaire des échanges de la monnaie

Les transactions se sont multipliées dans l'économie à mesure que la division du travail s'est développée. Les pratiques de troc se sont alors montrées insuffisantes et l'utilité d'un instrument qui pouvait servir d'intermédiaire entre les échanges est apparu. L'économie de troc requiert en effet que les besoins des agents se correspondent exactement, c'est-à-dire qu'il existe une double coïncidence des besoins, ce qui est rare ou nécessite des coûts de recherche (et donc une perte de temps) importants. Par ailleurs pour qu'un troc entre deux marchandises se réalise, il convient que les besoins des deux protagonistes se correspondent quantitativement, et cela en intégrant les valeurs respectives. L'indivisibilité de certains biens peut donc aboutir à ce que ne se réalisent pas certains échanges.
Pour toutes ces raisons, le développement de l'économie a nécessité la mise en place de monnaies. Nous verrons plus loin, les différentes formes qu'a revêtu la monnaie au cours des époques et en fonction du degré de développement de l'économie.

3.1.2 La fonction d'unité de compte

En l'absence de monnaie dans une économie, la valeur d'un bien quelconque échangé dans cette économie doit s'exprimer en fonction de la valeur de tous les autres biens susceptibles de faire l'objet d'un échange avec lui. La valeur d'un bien est donc relative dans une économie sans monnaie.
L'introduction de la monnaie permet d'exprimer la valeur absolue de chaque bien dans une seule unité et il en résulte une grande simplification du système de prix (dans une économie à *n* biens sans monnaie, il existe *n(n-1)/2* prix relatifs alors que le nombre de prix relatifs dans une économie à *n* biens comprenant la monnaie est *n-1*).

3.1.3 La fonction de réserve de valeur

La monnaie est un actif particulier du patrimoine des agents. La monnaie détenue par un agent représente en effet le pouvoir d'achat dont il dispose. La monnaie a donc l'avantage de permettre le stockage dans le temps du pouvoir d'achat. Elle constitue un lien entre le présent et le futur puisqu'elle conserve au pouvoir d'achat toute sa valeur, si on fait, en première approximation, l'hypothèse qu'il n'existe pas d'inflation.
Si les deux premières fonctions de la monnaie (intermédiaires des échanges et unité de compte) ont fait l'objet d'une reconnaissance de la part de l'ensemble des économistes, les économistes classiques ont souvent négligé le rôle de réserve de valeur de la monnaie.

Une fois admis le rôle essentiel de la monnaie dans l'économie et déterminé ses principales fonctions, nous allons pouvoir nous interroger sur plusieurs points concernant la monnaie. En fonction de leurs besoins économiques, les agents vont demander plus ou moins de monnaie pour réaliser leurs transactions. Nous allons donc examiner les ressorts de la **demande de monnaie** et les conséquences de la quantité de monnaie en circulation sur l'économie réelle (c'est-à-dire le monde de la production et de la consommation de biens et services). Les économistes ne sont pas unanimes sur ce point et les conclusions généralement admises sur celui-ci ont évolué au cours des périodes. Nous allons ensuite examiner les conséquences d'une **variation de la masse monétaire** sur l'économie réelle. En fonction de la quantité de monnaie en circulation, les agents détiennent plus ou moins d'encaisses monétaires et leur pouvoir d'achat évolue. L'offre de monnaie peut-elle jouer un rôle (important ?) de stimulant de l'économie ? Nous allons voir si cette possibilité existe et si elle peut donner lieu à la définition d'une politique monétaire active et discrétionnaire[37].

3.2 La demande d'encaisses monétaires

Les auteurs classiques et les auteurs néo-classiques croient en la neutralité de la monnaie, et donc postulent l'existence d'une déconnection entre les phénomènes réels et monétaires. L'une des grandes nouveautés de l'analyse keynésienne fût de remettre en cause cette dichotomie entre sphère réelle et sphère monétaire en décrivant plus précisément les motifs de demande d'encaisses monétaires de la part des agents et en introduisant des arguments expliquant l'instabilité de la demande de monnaie.
Friedman, chef de file des monétaristes, poursuivant la logique des classiques aboutit à des conclusions en opposition avec celles de Keynes. Nous le verrons, dans le prochain chapitre (Conjoncture et politiques économiques) que ces débats sur le rôle de la monnaie et son incidence sur l'économie réelle sont importants. L'aspect dichotomique ou non des sphères monétaires et réelles conditionne en effet l'efficacité et donc l'utilisation de la politique monétaire à des fins de relance ou de réglage de la conjoncture.

3.2.1 La monnaie est neutre pour les auteurs « classiques »

Pour les économistes classiques la monnaie ne remplit que les fonctions d'intermédiaires des échanges et d'unité de compte (la monnaie n'est qu'un « voile » permettant d'exprimer des prix absolus puisque les biens s'échangent contre des biens sur la base de leur valeur travail, voir le bref historique des courants de pensée en introduction). La fonction de réserve de valeur n'est pas reconnue. Cette conclusion est à l'origine de la « loi des débouchés de J.-B. Say ». « *L'offre crée sa propre demande* » uniquement si la monnaie n'est pas désirée pour elle-même et ne peut être conservée (la thésaurisation monétaire ne doit pas exister). Dans ce cas, l'ensemble des revenus issus de la production d'un bien quelconque est distribué sous forme de salaires ou de dividendes et sert soit à la consommation d'autres biens soit à l'épargne (qui donne lieu à des investissements et à

[37] On dit d'une politique qu'elle est discrétionnaire si une certaine latitude est laissée aux autorités dans l'intensité et la direction de celle-ci.

l'achat de biens d'équipement). Sans fuite dans ce système, tout revenu distribué est automatiquement consommé ou investi. Il en résulte que toute production trouve toujours en face d'elle une demande équivalente. Il n'y a jamais de crise de surproduction.
La neutralité de la monnaie chez les auteurs classiques trouve son aboutissement dans l'expression de l'équation quantitative de la monnaie :

$$M \times v = p.T$$

Avec *M* la masse monétaire en circulation, *v* la vitesse de circulation de la monnaie (combien de fois une même unité monétaire est utilisée dans l'intervalle de temps considéré), *p* le niveau général des prix et *T* le volume de transactions.
D'après cette relation, une augmentation de la quantité de monnaie en circulation aboutit à une hausse du niveau général des prix équivalente, pour peu qu'on considère le volume de transactions inchangé et la vitesse de circulation de la monnaie constante (ce qui est le cas à long terme (travaux de M. Friedman) mais pas à court terme (voir la sous-section sur l'évolution des agrégats monétaires et inflation)).

3.2.2 La monnaie est « active » pour Keynes

Rompant avec l'analyse classique, qui ne considérait que deux des trois fonctions de la monnaie (intermédiaire des échanges, unité de compte, mais pas la fonction réserve de valeur), Keynes réintroduit un motif de demande de monnaie qui traduit la préférence pour la liquidité des agents. Le niveau élevé des taux d'intérêts peut expliquer le renoncement des agents à la détention de monnaie. Le raisonnement de Keynes s'appuie sur l'arbitrage que les agents effectuent entre détention de liquidités et détention d'actifs de type obligataire (titres à revenus fixe).

Pour Keynes, les agents ont une préférence pour la liquidité. Ils peuvent désirer la monnaie pour elle-même, et pas simplement pour la possibilité qu'elle offre de réaliser des transactions. Keynes va insister sur les conséquences de la fonction de réserve de valeur de la monnaie. Pour Keynes, il existe trois motifs de détention d'encaisses monétaires : (i) le **motif de transaction**, (ii) le **motif de précaution** et (iii) le **motif de spéculation**. Les agents désirent des encaisses monétaires pour réaliser leurs achats. La demande de monnaie pour motif de transaction sera donc croissante avec le niveau de production de l'économie. Les agents peuvent également vouloir détenir des encaisses pour faire face à des risques. Cette demande est plus ou moins forte selon le revenu et la possibilité de recours au crédit des agents. Si le motif de transaction est proche du rôle traditionnel de la monnaie comme intermédiaire des échanges, le motif de précaution s'éloigne déjà un peu de la perception traditionnelle du rôle de la monnaie par les classiques. Mais le véritable apport de Keynes est l'introduction d'une demande de monnaie pour motif de spéculation. Pour Keynes, les agents qui ont une préférence pour la liquidité peuvent accepter de renoncer à la détention de monnaie pour réaliser des placements dans des titres à revenus fixes (titres sans risque de type placement obligataire). Le volume de la demande d'encaisses pour motif de spéculation va donc dépendre directement du taux d'intérêt qui rémunère ces titres (le taux d'intérêt est le prix du renoncement à la liquidité).
La valeur des titres à revenus fixes de type obligataire évolue inversement avec la valeur du taux d'intérêt (voir l'annexe **Evolution inverse de la valeur d'une obligation et des taux d'intérêt** en fin de chapitre pour une explication sur ce point).
Les agents achètent donc des titres lorsque les taux d'intérêt sont élevés (la valeur des titres est faible) avec pour but de les céder lorsque les taux baisseront (hausse de la valeur et donc gain en plus-value). La demande d'encaisses monétaires dépend donc en partie du niveau des taux d'intérêt. La demande de monnaie pour motif de spéculation tend à s'accroître quand les taux d'intérêt sont à la baisse et à faiblir quand les taux d'intérêt sont à la hausse.
Les conséquences de la mise en exergue d'une demande de monnaie pour motif de spéculation par Keynes sont importantes. La notion de monnaie oisive (thésaurisation) apparaît. Lorsque les taux d'intérêt sont élevés, les agents détiennent de la monnaie dont ils ne font aucune utilité, car ils attendent la baisse des taux d'intérêt pour investir dans des titres à revenus fixes. La monnaie remplit donc bien une fonction de réserve de valeur et la loi des débouchés n'est plus valide. L'analyse keynésienne remet donc en cause la dichotomie sphère réelle / sphère monétaire au cœur de l'analyse classique. Comme par ailleurs, la demande de monnaie des agents dépend en partie du niveau des taux d'intérêts (qui sont relativement volatiles), la demande de monnaie est instable. En face, l'offre de monnaie de la part de la Banque centrale se doit de prendre en compte cet état de fait et s'ajuster au mieux. Cette conclusion a donc des conséquences importantes en matière de préconisation de politique monétaire puisqu'elle justifie les politiques monétaires actives et discrétionnaires.

3.2.3 Friedman et la vision monétariste

Milton Friedman est le fondateur du courant de pensée monétariste dont les conclusions ont commencé à être prises en considération dans les années soixante-dix lorsque la pensée économique keynésienne a semblé marquer le pas. Friedman a, entre autres apports à la science économique, introduit le concept de revenu permanent[38] qui s'oppose à l'idée de revenu courant si importante chez Keynes. Pour M. Friedman, la demande de monnaie est grandement déterminée par le revenu permanent (ainsi que par le taux d'intérêt et le taux d'inflation). Puisque le revenu permanent (beaucoup moins sensible aux fluctuations de court terme et aux aléas de la conjoncture que le revenu courant) est un déterminant essentiel de la demande de monnaie, celle-ci est considérée comme stable.

Avec Friedman, la théorie quantitative de la monnaie devient une théorie de la demande de monnaie. Lorsque la masse monétaire augmente, les agents modifient la structure de leur portefeuille. Pour rééquilibrer celui-ci, les agents utilisent l'excès de monnaie pour acheter des titres et biens ce qui fait augmenter les prix. L'accroissement de la masse monétaire entraîne donc également une hausse des prix, sans aucun impact sur les phénomènes réels à long terme.

Friedman rejoint donc les auteurs classiques quant aux préconisations concernant l'offre de monnaie : celle-ci doit évoluer de manière stable car son accroissement n'a pas d'effet sur l'économie réelle à long terme et ne provoque que de l'inflation.

3.3 Le processus de création monétaire

3.3.1 Masse monétaire et agrégats monétaires

Les formes de monnaie ont grandement évolué au cours du temps (bétail, puis apparition des monnaies métalliques, puis apparition de la **monnaie papier**[39], apparition de la **monnaie scripturale**[40] enfin). On a assisté à une dématérialisation croissante de la monnaie. Aujourd'hui plus de 80% de la monnaie en circulation est de la monnaie scripturale. La valeur du support de la monnaie a complètement disparu (à la différence d'une pièce d'or, le chèque en tant que tel est un bout de papier qui ne vaut rien. C'est son acceptation et la confiance de la société dans cette forme de monnaie qui lui confère la possibilité de porter une certaine valeur nominale). La tenue des comptes et l'existence de la monnaie scripturale ne sont possibles que par l'existence du système bancaire.

La **masse monétaire** se définit comme l'ensemble des moyens de paiement en circulation dans une économie et les placements facilement transformables en liquidité.

La **liquidité** d'un actif se définit comme son aptitude à se transformer en moyen de paiement aisément, rapidement et sans coût. La monnaie est l'actif est le plus liquide, mais un certain nombre d'actifs possède, à des degrés divers la caractéristique de liquidité. On parle alors de **quasi-monnaie** puisqu'une transformation, qui s'effectue aisément et sans perte de valeur, est néanmoins nécessaire.

La masse monétaire se définit comme l'ensemble des moyens de paiement en circulation dans une économie plus les placements facilement transformables en liquidité. La masse monétaire intègre donc les différentes formes de monnaie (fiduciaire, scripturale) et de quasi-monnaie. La masse monétaire se décompose en différents agrégats (voir l'annexe sur **les agrégats monétaires**).

---

[38] Le **revenu permanent** (Y_P) est le revenu normal attendu du patrimoine humain (profession, formation, aptitudes…) et du patrimoine matériel (actifs financiers, immobiliers ou monétaire…) sur toute sa durée de vie (notion de revenu proche de celle d'un « patrimoine »). Il existe une composante aléatoire de ce revenu (liée au risque), que Friedman qualifie de revenu transitoire (Y_T). Le revenu effectif (Y) diffère du revenu permanent : $Y= Y_P + Y_T$. Pour Friedman, dans la mesure où la consommation s'ajuste sur le revenu permanent et non sur le revenu effectif de la période, cela permet de stabiliser l'économie et donc la demande d'encaisses monétaires.

[39] On appelle la monnaie papier (les billets donc…) la **monnaie fiduciaire** car elle repose sur la confiance (*fiducia*) qu'ont les agents dans le système bancaire puisque la valeur faciale (10, 20 ou 500€) est largement supérieure à la valeur du support (le coût de production d'un billet est de quelques centimes d'euro).

[40] La monnaie scripturale est une forme de monnaie qui s'appuie sur un jeu d'écriture dans les livres d'une banque. Lorsqu'un client ouvre un compte dans une banque et réalise un dépôt, la banque reconnaît la dette à son égard. Ce compte devient de la monnaie lorsque l'agent, par l'intermédiaire de différents instruments (chèque, virement, RIB, TIP), l'utilise pour réaliser des transactions

3.3.2 Evolutions des agrégats monétaires et inflation

On observe des changements importants dans la demande de monnaie. Pourquoi ? Car, comme on vient de le voir, un certain nombre d'actifs financiers sont très proches de la monnaie. Ils ne peuvent pas être utilisés pour les transactions mais ils peuvent être échangés contre de la monnaie à moindre coût. Plus ils sont liquides et plus ils sont des substituts intéressant de la monnaie. Les échanges massifs de monnaie contre ces actifs sont le principal facteur des fluctuations de la demande de monnaie (par exemple des entreprises ou des particuliers réduisent leurs encaisses monétaires sur leur compte courant et les déplacent vers des SICAV monétaires lorsque ces dernières ont un rendement attractif).
L'existence de déplacement entre la monnaie et d'autres actifs liquides a obligé les banques centrales à développer des indicateurs qui ne contiennent pas simplement la monnaie mais également les autres actifs liquides (les fameux agrégats monétaires que nous avons défini).

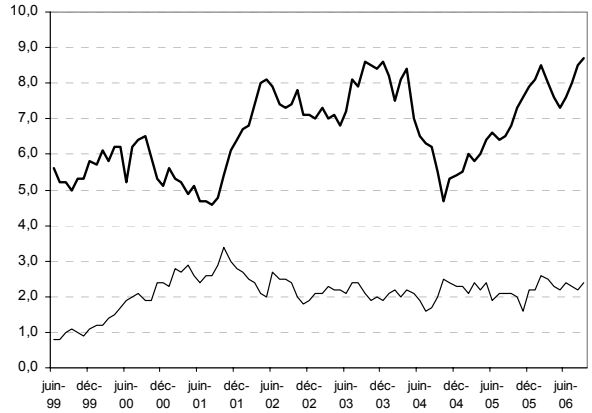


Figure 19 : Evolution de M3 (haut) **et de l'inflation** (indice IPCH, bas)

Cependant, même si la relation entre ces agrégats et l'inflation est plus stable, il n'en demeure pas moins que la croissance de M3 dépasse l'inflation de 5 points de pourcentage et que les deux séries sont loin de toujours fluctuer dans le même sens. Les banques centrales ont donc été conduites à abandonner les cibles d'agrégats monétaires et à adopter des cibles d'inflation (voir également l'annexe sur **la règle de Taylor** à la fin du prochain chapitre).

Pour comprendre la difficulté de la Banque centrale à réguler le rythme de la création monétaire, il nous faut maintenant décrire le processus de création monétaire.

3.3.3 Création monétaire dans un système bancaire hiérarchisé

Le système bancaire est un système hiérarchisé. Une seule banque qualifiée de Banque centrale (ou Banque de premier rang) dispose du pouvoir d'émettre des billets[41]. Les autres banques appelées banques commerciales (ou banque de second rang) ne peuvent créer qu'à travers leurs opérations de la monnaie scripturale.
Les banques commerciales créent de la monnaie lorsqu'elles distribuent des crédits à l'État ou aux agents non financiers non étatiques (ménages et entreprises), souscrivent à des émissions de titres, monétisent des devises ou achètent des actifs réels.
La création monétaire de la part des banques commerciales intervient donc dans trois cas : (i) lors d'un octroi de crédits aux agents non financiers titulaires d'un compte (détention par la banque d'une créance sur les agents non financiers), (ii) lors de l'acquisition de titres du Trésor Public (détention d'une créance sur l'Etat), (iii) lors de l'acquisition de devises auprès d'agents non financiers (détention d'une créance sur l'étranger).
Le crédit bancaire[42] est l'opération par laquelle une banque crédite le compte d'un agent non financier d'une somme déterminée pour une durée convenue, contre engagement pour l'agent de la rembourser ainsi que les intérêts à l'échéance du prêt.

[41] Le droit d'émettre de la monnaie a été progressivement réglementé par l'État. Ce dernier est le garant de la valeur de la monnaie. La Banque centrale a donc obtenu le monopole d'émission des billets. Elle s'est ainsi trouvé investie d'une responsabilité sociale : celle d'assurer la régulation de la quantité de monnaie dans l'économie afin que la valeur de la monnaie ne chute pas.

[42] **Exemple :** M. et Mme X obtiennent un crédit de la Banque Z de10 000 euros. Par souci de simplification, on fait abstraction des intérêts. Leur compte est donc crédité de 10 000 euros, au passif de la banque (la banque doit cette somme à M. et Mme X auxquels elle a accordé le crédit). Elle détient en contrepartie une créance sur

Lorsque les agents économiques se financent (en obtenant des crédits), ils détiennent de la monnaie « banque commerciale ». Les banques commerciales doivent pour leur part faire face aux besoins en billets et devises de leur clientèle, et respecter l'obligation réglementaire de constituer des réserves auprès de la Banque centrale. Elles doivent pour cela se refinancer afin d'obtenir la quantité de « monnaie banque centrale » nécessaire.

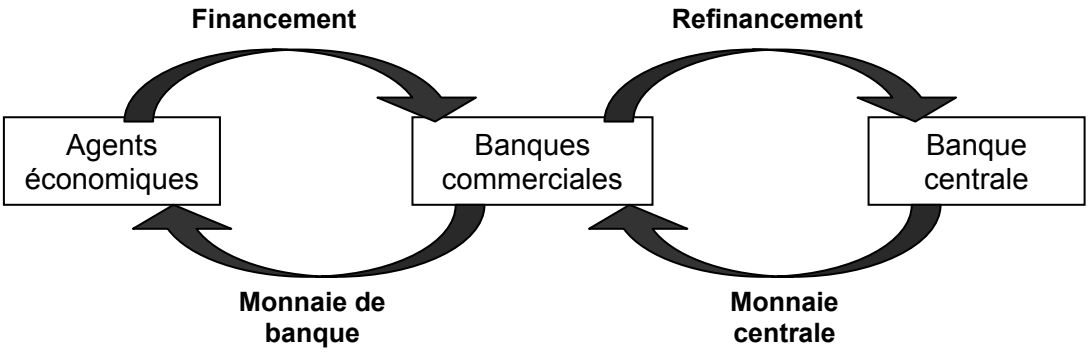


Figure 20 : Financement et refinancement dans un système bancaire hiérarchisé

Les banques se refinancent essentiellement de deux façons : (i) par des émissions d'obligations sur le marché financier, (ii) sur le marché monétaire (marché interbancaire ou marché des titres de créances négociables).
Ce sont les banques commerciales qui réalisent la majeure partie de la création monétaire à travers l'octroi de crédits. La Banque centrale joue cependant un rôle important à leur égard puisqu'elle contrôle la base monétaire et fixe le niveau des réserves obligatoires[43].

3.3.4 Le rôle de la banque centrale

Puisque la Banque centrale est la seule à pouvoir émettre des billets, les banques commerciales vont se refinancer auprès d'elle pour obtenir les billets demandés par leurs clients[44]. L'importance de la demande de billets dépend du taux de préférence des agents économiques pour la détention de billets. La Banque centrale contrôle ce qu'on appelle la base monétaire, mais les banques commerciales par leur activité d'octroi de crédits par exemple créent aussi de la monnaie à partir de cette base monétaire.
Toutes les sources de la création de monnaie scripturale, dans la mesure où elles se traduisent par une augmentation des dépôts des agents non financiers, génèrent des fuites en monnaie centrale pour les banques (accroissement des réserves obligatoires et retrait en billets de la part des agents, cf. l'annexe en fin de chapitre sur **le principe du multiplicateur de crédit**).
Lorsque les banques créent de la monnaie scripturale, elles subissent des pertes en monnaie centrale. Leur création monétaire est donc conditionnée par leur détention de monnaie centrale. Comme seule la Banque centrale émet de la monnaie centrale, elle peut éventuellement contrôler la création monétaire en régulant la quantité de monnaie centrale qu'elle met à disposition des banques commerciales.

La liaison entre monnaie centrale et pouvoir de création monétaire (par octroi de crédits de la part des banques commerciales) a fait l'objet de deux formulations théoriques différentes. La première, d'origine monétariste, estime que la création monétaire consécutive à un crédit est entièrement

---

M. et Mme X (ils doivent rembourser la somme empruntée à l'échéance). **Ce sont donc les crédits qui font les dépôts.** Le remboursement se traduit par une destruction de monnaie scripturale d'un montant de 10 000 euros et le passif de la banque diminue de 10 000 euros.
Au niveau macroéconomique, si la quantité de monnaie détenue par les agents non financiers s'accroît sur une période donnée, c'est que le montant des crédits accordés durant cette période est supérieur au montant des remboursements de crédits accordés auparavant.

[43] Les **réserves obligatoires** sont un instrument de la politique monétaire. Il s'agit généralement d'un pourcentage de leurs encours de crédits que les banques et autres établissements financiers doivent déposer auprès de la banque centrale, rémunérés ou non.

[44] Les banques disposent de titres provenant de leurs clients en contrepartie desquels elles ont accordé des avances (opérations d'escompte). Pour se procurer de la monnaie centrale auprès de la banque d'émission, les banques peuvent lui proposer ces titres (opération de réescompte dont le coût dépend du taux d'intérêt).

contrôlée par la Banque centrale qui, en fournissant plus ou moins de monnaie centrale induit une augmentation ou un ralentissement de la création monétaire. L'offre de monnaie de la part des banques dépend de l'offre de monnaie centrale (exogène et contrôlée par la Banque centrale) : c'est l'optique du **multiplicateur de crédit** (voir l'annexe sur le multiplicateur de crédit).
La deuxième approche prétend, au contraire, que ce sont les banques qui détiennent l'initiative quant aux crédits qu'elles accordent. Elles déterminent le montant de crédits accordés en fonction des besoins de l'activité économique. L'offre de monnaie est alors endogène. C'est l'optique dite du **diviseur de crédit**. La création de monnaie centrale dépend de la création de monnaie scripturale de la part des banques.

3.3.5 Les instruments de la Banque centrale

Pour réguler la quantité de monnaie en circulation dans l'économie, les autorités monétaires, on l'a vu au paragraphe précédent, peuvent agir sur les besoins des banques commerciales en monnaie centrale afin de limiter ou de faciliter la création monétaire.
La Banque centrale peut également intervenir sur le marché interbancaire (détermination des **taux d'intérêt directeurs**). La Banque centrale peut prendre l'initiative à travers des appels d'offre. Elle fixe périodiquement le taux d'intérêt et la quantité de monnaie qu'elle souhaite injecter au regard des besoins de financement des banques. Soit l'initiative vient des banques qui éprouvent un besoin urgent de refinancement entre deux appels d'offre. Cette intervention se traduit par des avances de monnaie centrale contre la prise en pension de titres qui devront être remboursés à l'échéance. Ces deux formes d'intervention déterminent les taux d'intérêt directeurs.
Lorsque l'initiative vient de la Banque centrale, le taux d'intérêt est plus faible (taux plancher). En revanche la prise en pension à l'initiative des banques se traduit par un taux plus élevé (taux plafond) (voir le graphique en introduction de ce chapitre pour visualiser l'évolution des deux taux).
Ces deux taux dirigent les taux du marché interbancaire (les taux ne peuvent pas être inférieurs au taux des appels d'offre ni supérieurs au taux de prise en pension puisque sinon les banques peuvent toujours avoir recours à la Banque centrale).
Les autorités monétaires peuvent également agir sur la création monétaire à travers la fixation du taux des réserves obligatoires (taux non significatif en France) ou jadis l'encadrement du crédit (plus d'utilisation après 1986).
Une dernière possibilité d'intervention de la Banque centrale est ce qu'on appelle la politique d'***Open Market***. L'intervention de la Banque centrale consiste à acheter ou à vendre des titres (souvent des bons du Trésor, c'est-à-dire des titres de court/moyen terme émis par l'Etat). En acquérant des titres, la Banque centrale injecte de la monnaie sur le marché et augmente l'offre de capitaux entraînant une baisse des taux. Inversement toute vente de titres de sa part revient à prélever des liquidités du marché et à diminuer l'offre de capitaux, ce qui provoque un hausse des taux (voir l'annexe en fin de chapitre **l'Open Market et le comportement des agents** pour une description plus fine des mécanismes à l'œuvre). Les pratiques d'*Open Market* sont cependant plus efficaces pour restreindre la création monétaire que pour stimuler l'activité (l'injection de liquidités par vente de titres et la baisse des taux n'ont aucun effet si les agents n'ont pas de besoin de financement).

Nous venons d'examiner les effets de l'octroi de crédits sur la création monétaire et les rôles joués par les banques commerciales et la Banque centrale dans le processus de création monétaire. Nous allons dans la prochaine section nous attacher à décrire plus précisément le processus de financement de l'économie, entendu que celui-ci ne passe pas uniquement par les crédits bancaires. Nous allons voir qu'une évolution forte a marqué le processus de financement au cours des dernières décennies.

3.4 Le processus de financement de l'économie

Les agents, qu'ils soient consommateurs ou entrepreneurs, doivent se procurer des ressources pour financer leur activité économique, la consommation ou l'investissement. Lorsque ces agents ne disposent pas d'une épargne suffisante pour leur activité, ils doivent avoir recours au crédit bancaire. Si les agents sont des entreprises, ils peuvent également avoir recours aux marchés financiers pour trouver les ressources nécessaires à leur activité.
Nous avons vu dans le chapitre introductif que l'économie est la science qui étudie la rareté. L'étude du mode de financement d'une économie participe à cette entreprise car les ressources sont bien souvent en quantités insuffisantes pour satisfaire l'ensemble des besoins de tous les agents (l'épargne disponible peut être insuffisante au niveau macroéconomique).

Le financement par le crédit est un premier mode de financement. Les ressources empruntées favorisent la croissance (augmentation de la demande lorsque les ressources empruntées sont utilisées à des fins de consommation, amélioration des capacités de production lorsque les ressources empruntées sont utilisées à des fins d'investissement). La croissance des revenus va permettre le remboursement des crédits souscrits par les agents. Le système de financement de l'économie par le crédit fonctionne donc lorsque la croissance est régulière. En cas de ralentissement de la croissance, les revenus ne sont plus suffisants pour rembourser les dettes contractées. Les agents doivent s'endetter de nouveau pour rembourser leurs dettes. Ce cercle vicieux mène au surendettement. A l'écueil du surendettement, s'ajoute un autre risque majeur d'un système de financement de l'économie principalement basé sur le crédit bancaire. La création monétaire qui découle des crédits octroyés par les banques a des conséquences inflationnistes (voir la section précédente).

Le financement par le marché des capitaux n'induit pas ce dysfonctionnement. Puisque les offreurs et les demandeurs de capitaux se rencontrent directement sur un marché, l'épargne disponible finance les investissements sans risque inflationniste. Cependant, le financement par les marchés financiers peut également présenter un dysfonctionnement majeur. La présence d'une épargne suffisante ne garantit pas que les entreprises investissent. Les ressources disponibles peuvent se diriger vers les marchés financiers sans finalité d'investissement productif. L'afflux d'épargne, et la demande de capitaux qui l'accompagne, rendent encore plus attractifs les placements boursiers puisqu'ils font grimper le cours des titres. De ce phénomène cumulatif peut émerger une bulle spéculative, c'est-à-dire une progression des marchés financiers sans lien avec l'économie réelle (déconnection sphère réelle / sphère financière)[45].

Le développement de l'un ou l'autre des modes de financement de l'économie engendre donc de possibles dysfonctionnements. Après avoir brièvement défini les besoins et les capacités des agents, nous allons examiner tour à tour ces deux modes de financement. Puis nous tenterons de comprendre pourquoi la France est passée d'une économie d'endettement (prépondérance des crédits bancaires) à une économie de marché de capitaux au cours des années quatre-vingts. Nous conclurons alors sur les conséquences économiques de cette nouvelle situation qu'est la prépondérance du financement par les marchés de capitaux[46].

3.4.1 Besoins et capacités de financement des agents

Le besoin de financement se définit comme une insuffisance d'épargne. Son montant correspond à la somme des ressources qu'un agent doit se procurer auprès d'autres agents.
Les entreprises ont des besoins de financement liés à leur activité de production, tout d'abord à court terme du fait des décalages entre les recettes et les dépenses (les entreprises achètent et produisent avant de vendre). Les différences entre les délais de paiement accordés par les fournisseurs et ceux accordés aux clients font croître ou diminuer ce besoin de financement. C'est un besoin durable qui apparaît à chaque cycle d'exploitation. La trésorerie de l'entreprise peut par ailleurs se détériorer engendrant un besoin sur le court terme. Ensuite à long terme, le besoin découle des investissements (Formation Brute de Capital Fixe). En effet, dans une perspective de longue période, l'entreprise doit remplacer et moderniser son appareil productif, développer ses capacités de production. Des ressources peuvent également être nécessaires du fait de l'activité financière de l'entreprise.
Les ménages ont également des besoins de financement à court terme et à long terme. A court terme, le besoin de financement des ménages correspond à des achats à « tempérament » (équipement électro-ménager, équipement Hi-Fi, voiture…). Sur le long terme, l'investissement dans l'immobilier nécessite souvent un financement, l'épargne disponible étant le plus souvent insuffisante pour faire face à ces dépenses importantes.

[45] L'éclatement d'une bulle financière engendre des pertes importantes et déstabilise les marchés (fuite des investisseurs vers la sécurité et la liquidité, c'est-à-dire vente à perte et dans l'urgence des actifs dont les cours chutent et rachat d'obligations d'Etat ou de titres plus liquides et aux rendements plus assurés). Cela n'est pas sans conséquences sur l'économie réelle car ces ventes à perte aboutissent à un effet richesse négatif (chute de la consommation pour les ménages dont le patrimoine a baissé suite à l'éclatement de cette bulle) et la déstabilisation des marchés financiers occasionne des difficultés pour les entreprises à lever des capitaux (chute de l'investissement).

[46] La distinction entre économie d'endettement et économie de marchés financiers remonte à Hicks (1974). Bien entendu, aucune économie n'a recours qu'à un mode unique de financement. Le système est toujours mixte. Lorsqu'on parle de passage d'une économie d'endettement à une économie de marchés financiers, on parle d'une évolution des proportions des ces différents modes de financement dans la totalité des financements en faveur du financement par les marchés.

L'Etat a des besoins de financement dans le long terme (développement de ses infrastructures) mais également à court terme pour financer son déficit budgétaire et régler les salaires des fonctionnaires avant la perception de l'impôt.
Les besoins de financement dépendent de différents autres paramètres : (i) lorsque la croissance est forte, la consommation et l'investissement sont élevés. Les entreprises et les ménages ont des besoins importants même si les revenus résultant de la forte activité les limitent, (ii) les besoins de financement des Etats dépendent du degré de l'intervention des pouvoirs publics dans l'économie.
En premier lieu, les agents assurent le financement de leurs besoins grâce à leurs ressources. La comparaison entre les ressources et les besoins permet donc de déterminer les agents structurellement en situation **de besoin net** ou en situation de **capacité de financement**.

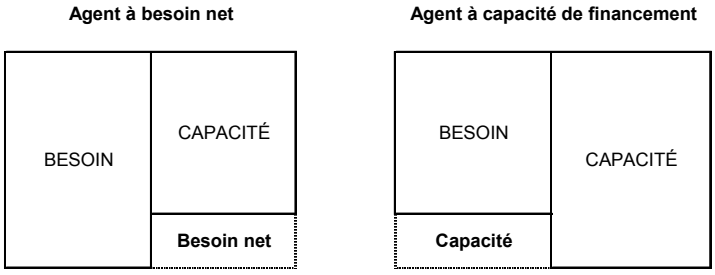


Figure 21 : Besoin net et capacité de financement

Pour combler ces besoins de financement, les ménages disposent de leur revenu, les entreprises dégagent des ressources de leur activité (autofinancement), les Etats se financent grâce aux rentrées fiscales. En général, au niveau macroéconomique, les ménages parviennent à épargner une partie de leur revenu sous différentes formes (thésaurisation monétaire, dépôts à vue, placements…). Ils dégagent donc structurellement une capacité de financement. La situation des entreprises a évolué. L'analyse des besoins ou capacités de financement des entreprises s'effectue à partir du **taux d'autofinancement**, qu'on détermine comme le ratio capacité de financement/investissement. Le taux d'autofinancement était inférieur à 100% au début des années 1980 puis il est passé depuis au-dessus du seuil des 100%. Les entreprises disposent donc elles aussi de capacités de financement depuis le milieu de la décennie quatre-vingts (en 1996 le taux d'autofinancement (en base 80) était encore de 112%).
L'Etat avait su maintenir un certain équilibre budgétaire jusque dans les années soixante-dix. La situation s'est alors dégradée (voir le chapitre Conjoncture et politiques économiques et le développement sur l'endettement de l'Etat). L'Etat est devenu structurellement en situation de besoin net de financement.
Tout cela est néanmoins global. Certains ménages se trouvent en situation de besoin net de financement. De même que certains états peuvent parvenir à des soldes budgétaires positifs (leur permettant de disposer d'une capacité de financement). Si les ressources disponibles dans l'économie s'avèrent insuffisantes, des capacités supplémentaires peuvent être obtenues par **création monétaire** à partir de **crédits bancaires**.

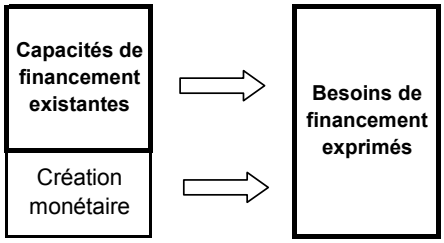


Figure 22 : Rencontre entre besoins et capacités de financement

3.4.2 Les deux formes de financement et le rôle de l'intermédiation

Il existe donc deux formes de financement qui engendrent deux types d'économie : (i) **l'économie d'endettement** et (ii) **l'économie de marchés financiers**. Dans une économie d'endettement, il est fait largement appel au crédit bancaire. Dans une économie de marchés financiers, les agents dégagent des capacités de financement suffisantes. La rencontre entre l'épargne et les besoins de financement s'effectue directement sur le marché financier.

Dans une économie d'endettement, le financement passe principalement par les banques qui servent d'intermédiaires entre les agents. Il y a **intermédiation bancaire**. Accorder un crédit nécessite en effet d'avoir des garanties quant à la capacité de remboursement de l'emprunteur. Les épargnants ne disposent pas des informations suffisantes pour avoir cette garantie car il existe une asymétrie d'information (puisque l'emprunteur a une idée assez précise de sa capacité de remboursement alors que le prêteur est soit totalement ignorant soit n'a qu'une idée imprécise de celle-ci). L'intermédiation bancaire présente l'avantage de résoudre ce problème d'asymétrie d'information puisque les banques exigent des emprunteurs qu'ils fournissent des éléments d'appréciation sur les projets à financer et sur leur capacité à rembourser[47].

Dans une économie de **marchés financiers**, la rencontre entre l'épargne et les besoins de financement s'effectue directement sur le marché financier. Il y a **désintermédiation**.
Le financement d'une économie n'est bien sûr jamais réalisé exclusivement par l'un ou l'autre des systèmes.
On constate cependant pour tel ou tel pays une prépondérance du financement par l'endettement alors que pour d'autres, le passage par les marchés est privilégié.
Le **taux d'intermédiation** représente la part prise par les institutions financières résidentes (établissements de crédit, OPCVM, sociétés d'assurance) dans l'enveloppe globale des financements accordés aux agents résidents non financiers. Lorsqu'on ne tient compte que des crédits distribués par ces institutions, on parle d'un taux d'intermédiation au sens étroit. Lorsque les placements de ces institutions en titres émis par les agents non financiers (actions, obligations, titres de créances négociables) sont également pris en compte, il s'agit d'un taux d'intermédiation au sens large. Le taux d'intermédiation au sens étroit est donc un indicateur, égal au total des crédits auprès des établissements de crédits sur l'ensemble des financements au bénéfice des agents non financiers qui permet de déterminer le degré d'intermédiation et donc la prépondérance ou non de l'économie d'endettement.

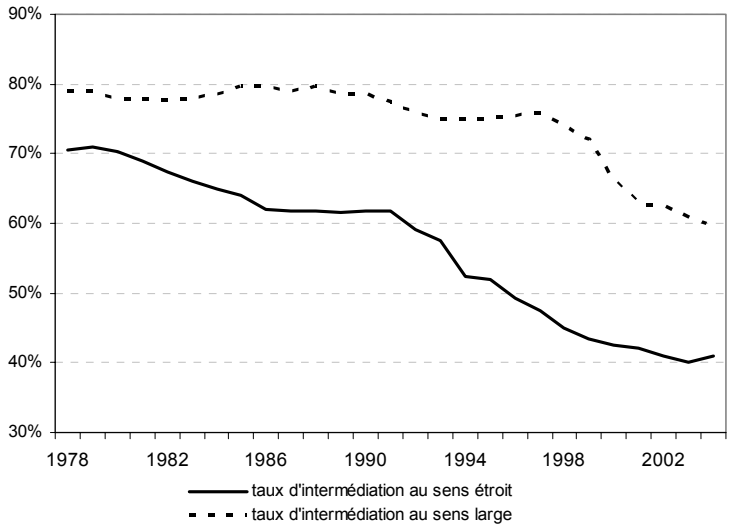


Figure 23 : Taux d'intermédiation en France
source : Comptes nationaux et Bulletin de la Banque de France

Le mode de financement d'une économie peut évoluer au cours du temps. Cela a par exemple été le cas de la France où le taux d'intermédiation au sens étroit est passé de 70% à un peu plus de 40% entre 1978 et 2004.

[47] Les banques ou établissements de crédits disposent de bases de données historiques sur le déroulement des remboursements des prêts par les emprunteurs, ainsi que des caractéristiques de ces derniers (situation familiale, niveau de revenu, ancienneté dans l'emploi, apport…). Ils peuvent ainsi associer un niveau de risque aux différentes populations d'emprunteurs (*scoring*) et pratiquer des politiques tarifaires différenciées (mise en place de primes de risque).

3.4.3 Le passage de l'économie d'endettement à celle de marchés financiers

3.4.3.1 Un peu d'histoire…

Au sortir de la seconde guerre mondiale, le financement par le crédit est la seule option envisageable car les investissements à réaliser sont considérables et l'épargne disponible est nettement insuffisante. L'Etat prend directement en charge l'intermédiation et réorganise le secteur bancaire (la Banque de France et l'ensemble des banques de dépôts sont nationalisées). L'Etat assure lui-même le financement de l'économie. Il accorde des crédits aux entreprises publiques et des avances aux banques afin que celles-ci puissent accorder des prêts à taux réduits. C'est donc un financement monétaire. Les banques jouent un rôle prépondérant dans cette économie d'endettement. Puisque les capacités d'autofinancement demeurent faibles durant la période des trente glorieuses et que l'épargne longue des ménages est insuffisante, l'Etat est omniprésent sur les marchés financiers. En 1974, la part des investissements financés par les banques, qui n'a cessé de croître sur cette période, est de l'ordre de 60%.

Mais, nous l'avons signalé dans l'introduction de ce chapitre, l'économie d'endettement n'est un système viable que si la croissance est soutenue. Or, dans les années soixante-dix la croissance forte des trente glorieuses cesse et l'économie française entre dans une phase de croissance faible. Par ailleurs, le financement monétaire par l'Etat puis le développement du crédit bancaire sont à l'origine d'une forte création monétaire qui tend à provoquer de l'inflation. Or, lorsque le taux d'intérêt réel est négatif, les agents sont fortement incités à s'endetter. C'est le cas dans les années soixante-dix où le taux d'inflation peut être supérieur au taux d'intérêt nominal (l'inflation est si forte qu'elle n'est pas entièrement intégrée dans les taux d'intérêt).

Alors que se produit le retournement de conjoncture, l'accélération forte de l'inflation incite les ménages à s'endetter plus encore. Cependant, le ralentissement de la croissance générant moins de revenus, les ressources destinées au remboursement des emprunts diminuent fortement. Lorsque les taux d'intérêt réels vont redevenir positifs et alourdir les charges financières, les agents vont se retrouver en situation de surendettement. Ce phénomène touche tous les agents économiques. La situation est cependant encore plus problématique pour l'Etat qui avait dû financer ses fortes interventions d'inspiration keynésienne durant les trente glorieuses. L'impact positif de la croissance sur les rentrées fiscales avait contribué à maintenir les équilibres budgétaires mais, à partir de 1973, on assiste à une détérioration très rapide des soldes budgétaires. Le phénomène est de plus cumulatif. Pour financer ces déficits, l'Etat n'a d'autre solution que de s'endetter. Les charges financières du déficit accroissent encore le déficit.

3.4.3.2 Le tournant de la lutte contre l'inflation

Or, au début des années quatre-vingts, la pensée monétariste s'est imposée. Il en résulte une volonté nouvelle des pouvoirs publics de lutter contre l'inflation. Pour cela, les taux d'intérêts réels vont croître fortement (afin de ralentir le rythme de l'octroi de crédits et donc les pressions inflationnistes). La forte croissance des taux d'intérêt réels va remettre en cause le financement des déficits publics par « création monétaire ». L'Etat se trouve ainsi obligé de recourir à de nouvelles sources de financement. Comme alternative au financement monétaire, l'Etat va se procurer ses ressources sur le marché des capitaux.

Les entreprises sont dans une situation relativement comparable à celle de l'Etat. Le renchérissement du crédit bancaire les contraint à trouver d'autres formes de financement. Cependant, la faiblesse de l'épargne longue à la fin des années soixante-dix fait que les marchés de capitaux ne sont pas encore une source de financement suffisante. Les interventions de l'Etat sur les marchés de capitaux vont donc se faire au détriment des entreprises (effet d'éviction).

Le passage de la situation d'économie d'endettement à la situation d'économie de marchés de capitaux s'est opéré progressivement. Il a impliqué notamment une modification importante du rôle des banques.

3.4.3.3 Le financement de l'économie par les marchés

Le marché des capitaux est le lieu de rencontre de l'offre et de la demande de capitaux. On distingue deux marché de capitaux : (i) le **marché monétaire** et (ii) le **marché financier**.

Sur le marché monétaire s'échangent des capitaux à court terme (échéances inférieures à sept ans). Sur le marché financier, se rencontrent l'offre et la demande de capitaux à long terme.

Sur le marché monétaire, on distingue le **marché interbancaire** et le **marché des titres de créances négociables**. Sur le marché interbancaire n'interviennent que les établissements de crédits, le Trésor et la Banque de France qui s'échangent pour des durées très courtes de la monnaie centrale (voir la

première partie de ce chapitre). Le marché des titres de créance négociables est ouvert aux banques, au Trésor et également aux entreprises. Les titres permettant de se financer sont les certificats de dépôts (banques), les bons du Trésor négociables et les billets de trésorerie (entreprises). Les entreprises peuvent ainsi se financer à court terme par l'intermédiaire du marché sans avoir recours au crédit bancaire (les montants minima des billets de trésorerie limitent cependant l'accès de ce marché aux entreprises de grande taille).
Sur le marché financier interviennent des agents (entreprises, Etat) qui ont des besoins de financement et des agents qui ont des capacités de financement (ménages, certaines entreprises). On distingue deux catégories de titres échangés : les **actions** et les **obligations**.
Les actions rendent leurs détenteurs propriétaires d'une partie de l'entreprise (c'est un droit de propriété). Les obligations que les entreprises émettent sont assimilables à des emprunts sur le long terme (droit de créance). Il convient bien entendu de distinguer sur l'ensemble des compartiments du marché, le marché primaire et le marché secondaire. Sur le marché primaire sont émis les nouveaux titres. Une fois les titres souscrits, leurs détenteurs peuvent les céder sur le marché secondaire.
L'ensemble des acteurs de l'économie a su trouver son intérêt dans le financement par le marché. L'Etat a trouvé une source de financement non inflationniste de ces déficits publics. Les entreprises se sont vues offrir une source de financement alternative au crédit bancaire pour leurs besoins de court et long terme, ainsi qu'une possibilité de croissance externe (acquisition de tout ou partie d'une autre entreprise grâce à des OPA (offre publique d'achat) ou OPE (offre publique d'échange)[48]).
Les intervenants sur ces marchés ont par ailleurs cherché à se protéger contre les risques inhérents au fonctionnement normal de ces marchés (variations de taux d'intérêt et des cours des titres qui peuvent générer des pertes importantes pour les agents). Deux marchés ont été créés pour permettre de réaliser des opérations à terme à un cours ou un taux prédéterminé : le **MATIF** (Marché à terme international de France) et le **MONEP** (marché des opérations négociables de Paris)[49].
L'offre de capitaux sur les marchés financiers provient de l'épargne disponible. Le fonctionnement des marchés nécessite donc une épargne abondante. Plus l'épargne est abondante et plus les marchés financiers sont une source de financement intéressante (la valeur des titres va s'accroître par application de la loi de l'offre et de la demande puisque l'offre de capitaux est aussi une demande de titres[50]).
Les agents vont placer leur épargne dans les titres financiers qui leur rapportent les meilleurs revenus. Pour les actions, le revenu dépend des profits réalisés par l'entreprise puisqu'une partie de ceux-ci est reversée sous forme de dividendes aux actionnaires. Une action attire donc les épargnants lorsque ceux-ci jugent que l'entreprise présente de bonnes perspectives de bénéfices et le cours de cette action grimpe[51].
Pour les obligations, le revenu est fonction du taux d'intérêt proposé. Plus le taux d'intérêt proposé pour une obligation est élevé par rapport aux autres obligations sur le marché et plus ce titre va être demandé. Son cours va dépasser sa valeur d'émission jusqu'à égaliser le taux de rendement des obligations du marché.
En période de hausse des taux d'intérêt, les obligations deviennent donc plus attractives et concurrencent les actions (puisque taux d'intérêt et cours des obligations sont inversement reliés).
En résumé, ce qui importe pour le développement des marchés financiers, ce sont les perspectives de profits des entreprises et le niveau des taux d'intérêt.

3.4.3.4 L'avènement d'une économie de marchés de capitaux

Pour l'économie française, les années quatre-vingts ont marqué le passage d'une économie d'endettement à une économie de marché de capitaux (les transactions sur les obligations ont été multipliées par trente et par dix sur les actions). Un vaste marché de capitaux est apparu offrant des possibilités de placement aussi bien sur le court que sur le long terme. Le système bancaire a dû

[48] Pour les dirigeants des entreprises, le risque de perte de contrôle de la société est apparu avec l'ouverture du capital au public. Les OPA sont parfois dites hostiles, c'est-à-dire non désirée par les gens en place (exemple MITTAL/ARCELOR). La modification de l'actionnariat modifie le pouvoir au sein des entreprises et oblige les dirigeants à adopter la « gouvernance » voulue par les actionnaires.
[49] Né d'une volonté de se couvrir contre les fluctuations des marchés, le MATIF est devenu au fil des années au même titre que le MONEP, un marché dédié quasi exclusivement à la spéculation. Sur le MATIF, les agents se couvrent contre les risques de taux d'intérêts. Le MONEP permet de se protéger contre les variations du cours des actions.
[50] On peut faire le même raisonnement sur l'émission d'une obligation. Le taux d'intérêt proposé est en partie déterminé par la confrontation de l'offre et de la demande de capitaux. Si l'offre est supérieure à la demande, l'émetteur de l'obligation (entreprises ou Etat) va pouvoir proposer des taux plus faibles à rendement égal (voir l'annexe en fin de chapitre sur **le lien inverse entre valeur d'une obligation et taux d'intérêt**).
[51] Voir l'annexe en fin de chapitre sur **valeur d'une action et condition de non arbitrage**.

évoluer durant ces années pour s'adapter à cette nouvelle donne (plus de spécialisations (banques de crédits, banque d'investissement…), règles identiques pour tous les établissements).
Les dépôts des ménages et des entreprises ont donc diminué car les agents ont souhaité voir leur épargne mieux valorisée et l'amélioration de l'autofinancement des entreprises leur a permis d'avoir moins recours au crédit bancaire. En 1980, les banques assuraient encore 70% du financement de l'économie (voir le graphique sur le taux d'intermédiation). Ce pourcentage diminue à 60% en 1990 avant de poursuivre sa chute durant la décennie quatre-vingt-dix. En conséquence, les ressources bancaires diminuent et les banques doivent assumer des risques d'insolvabilité plus importants car les entreprises les plus solides se financent directement sur les marchés. Devant la chute de leur rentabilité, les banques tentent d'abord de réorienter leurs activités vers l'immobilier (mais crise dans ce domaine au début de la décennie quatre-vingt-dix) puis sont forcées de s'internationaliser (mais cela ne fait qu'ajouter des risques liés au taux de change).
On pourrait donc penser que la nouvelle intermédiation, la rencontre directe entre les offreurs et les demandeurs de capitaux, a été fatale aux banques. Ce n'est pas en fait le cas (il n'y a qu'à constater la santé du secteur bancaire en France). En effet, il existe un certain nombre de freins à la rencontre directe entre offre et demande de capitaux : (i) les offreurs de capitaux ne disposent pas de toutes les informations sur les demandeurs de capitaux (problème d'asymétrie d'informations qui peut bloquer les échanges), (ii) les techniques d'intervention sur ces marchés sont de plus en plus complexes (développement des marchés dérivés). Des OPCVM[52] (Organismes de placements collectifs de valeurs mobilières) servent donc d'intermédiaires entre les épargnants et les demandeurs de capitaux. C'est une nouvelle forme d'intermédiation. Les banques possèdent 80% des OPCVM, émettent des titres sur le marché monétaire, prennent des participations dans les capital des entreprises. La distinction entre financement directe et financement intermédié est donc délicate tant le rôle des banques est devenu double.
Les banques, à travers les OPCVM servent donc d'intermédiaires entre l'épargne populaire et les marchés de capitaux, tandis que la demande est totalement désintermédiée (les grandes entreprises ou l'Etat s'adressent directement aux marchés).

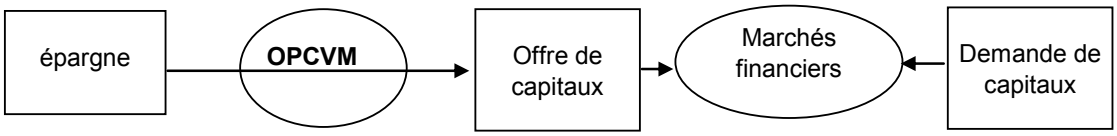


Figure 24 : Une nouvelle forme d'intermédiation : les OPCVM

Les économies, nous le verrons au chapitre 6 (Ouverture internationale et globalisation), ne sont plus fermées sur elles-mêmes. Les mouvements de capitaux diffusent les fluctuations des cours, et parfois les crises, d'une place financière à l'autre. La globalisation financière entraîne de nouveaux risques.

3.4.4 La globalisation financière et ses risques

L'avènement du financement de l'économie par les marchés de capitaux et la globalisation financière amènent de possibles dysfonctionnements et de nouvelles contraintes en matières économique et financière. Nous allons aborder dans cette dernière sous-section quelques éléments : (i) la possible dérive spéculative et la déconnection sphère réelle/sphère financière, (ii) la globalisation financière ; ses conséquences et les contraintes qu'elle impose.

3.4.4.1 La dérive spéculative

La **spéculation** est l'activité consistant à acheter (vendre) un titre en anticipant une hausse (baisse) de celui-ci afin de le revendre (racheter) dans un future relativement proche pour réaliser une plus-value.
Le phénomène spéculatif peut devenir déstabilisant lorsqu'il devient cumulatif et général. Puisque l'offre de capitaux accroît la valeur des titres, la hausse des titres incite d'autres épargnants à acheter des cours toujours plus élevés, en espérant la poursuite de la hausse (phénomène de mimétisme). L'achat de l'action ne repose plus sur le revenu que la détention de l'action est censée rapporter (le dividende) mais sur l'espoir de réaliser une plus-value par la revente. Les mécanismes des marchés financiers, tel que le règlement mensuel, entretiennent la spéculation puisqu'on peut acheter les titres

[52] Il existe deux types d'OPCVM : les SICAV (sociétés d'investissement à capital variable) et les FCP (Fonds communs de placement). Il s'agit dans les deux cas de la gestion par des spécialistes d'institutions financières d'un portefeuille de titres constitués à partir de fonds provenant de nombreux épargnants. La notion de portefeuille permet une diversification des placements et donc un plus faible risque. Cette « sécurité » a contribué à l'orientation de l'épargne populaire vers les marchés de capitaux.

sans forcément disposer des fonds nécessaires. On a vu par ailleurs que les marchés qui avaient été créés pour se couvrir contre les risques (MONEP, MATIF) étaient utilisés à des fins spéculatives puisqu'ils autorisent à jouer sur les cours ou les taux d'intérêts futurs[53]. Lorsque les revenus dégagés par les actions (dividendes) deviennent négligeables dans les revenus dégagés, on dit qu'on entre dans une spirale spéculative (l'exemple le plus récent, voir le graphique en introduction, est la bulle « internet » en 1999-2000).
Une bulle financière apparaît suite à l'agrégation des comportements mimétiques des agents. La valeur des cours s'éloigne progressivement de leur valeur fondamentale. Il y a déconnection entre sphère financière et sphère réelle. Ce genre de hausse est toutefois dangereux car il s'appuie sur des achats de titres à des cours de plus en plus élevés, financés par endettement ou en jouant sur les règlements à terme. L'éclatement de la bulle (le mouvement brutal à la baisse) s'explique également par le comportement mimétique des agents. Une information économique ou financière quelconque (voire banale) va conduire certains agents à anticiper une baisse des cours et donc à vendre leurs titres. Les autres agents leur emboîtent le pas et cèdent rapidement leurs actions pour éviter des pertes trop lourdes. Cette dérive spéculative et sa fin brutale a des répercussions sur l'économie réelle puisque les agents ayant perdu une partie de leur patrimoine vont revoir à la baisse leur consommation.

3.4.4.2 Les conséquences de la globalisation financière

On définit habituellement la globalisation financière à travers trois phénomènes concomitants, les 3D, qui sont intervenus dans les années quatre-vingts : (i) la **déréglementation** qui a autorisé les mouvements de capitaux entre pays (suppression du contrôle des changes en France en 1989 par exemple), (ii) la **désintermédiation** qui a accru le nombre de participants sur les marchés de capitaux (pas forcément des intervenants nationaux), (iii) le **décloisonnement** des marchés de capitaux (absence de séparation entre marché à court et long terme) qui permet que toutes les formes d'épargne rencontrent tous les besoins de financement dorénavant à l'échelle de la planète.

Les variations des cours des devises sont ainsi essentiellement déterminées par les mouvements de capitaux. Les différentes places financières mondiales sont extrêmement connectées (voir la figure ci-dessous qui met en regard l'évolution des capitalisations d'entreprises américaines (indice S&P500) et européennes (EUROSTOXX Broad)).

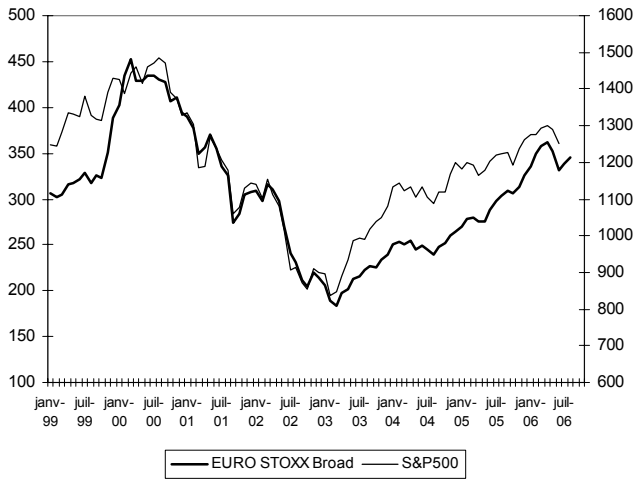


Figure 25 : Indices des valeurs américaines (S&P500) et européennes (EURO STOXX)

L'émergence d'un marché mondial des capitaux devrait être considérée comme positive. Les pays à épargne excédentaire peuvent ainsi contribuer au financement de pays structurellement en situation de besoin de financement. Les titres publics attirent les investisseurs étrangers et cela aide à la liquidité des marchés. Cela permet même d'équilibrer la balance des paiements de certains pays. Le déficit de la balance des transactions courantes américaine du fait des importations en provenance notamment de Chine est ainsi comblé par une entrée de capitaux, notamment chinois, venant s'investir sur des titres publics américains. C'est le problème des déficits jumeaux aux Etats-Unis (déficit public et déficit de la balance commerciale). Tant que ce mécanisme continue de fonctionner,

[53] Effet de démultiplication des gains lors de l'évolution du cours du sous-jacent (action, obligation, taux d'intérêt, taux de change) permis par l'utilisation de produits dérivés (options, warrants…).

cela alimente la croissance mondiale puisque cela stimule la consommation américaine et la croissance chinoise mais tout grippage de cet « échange de bon procédé » entre la Chine et les Etats-Unis (« j'achète tes produits, tu finances ma dette ») plongerait l'économie mondiale dans une grave crise (baisse du dollar, arrêt de la croissance chinoise).
L'émergence d'un marché mondial des capitaux rend également les opérations de fusions entre sociétés de nationalités différentes possibles. Cependant cette internationalisation n'est pas sans risque. L'arrivée de fonds étrangers (notamment les fameux « fonds de pension » américains gérants l'épargne retraite des salariés américains) dans le capital des entreprises françaises a profondément transformé la relation entre actionnaires et dirigeants des entreprises (mise en place d'une nouvelle gouvernance peu favorable à l'emploi). En effet, les nouveaux actionnaires exigent une transparence plus forte et plus contraignante dans la gestion des entreprises (information périodique, normes comptables nouvelles…). Ces nouveaux intervenants exigent également des taux de rendement à court terme extrêmement élevés auxquels les dirigeants doivent se résoudre. Cela induit le plus souvent une obligation de réduction des charges (notamment salariales) et donc aboutit à des licenciements. Cette focalisation sur le court terme et cette exigence de rendement élevé (15%) amènent les entreprises à abandonner certains investissements et à adopter des stratégies non durables.

En guise de conclusion, nous pouvons nous interroger sur les contraintes que font peser l'émergence d'une économie de marchés de capitaux et la globalisation financière sur la politique monétaire (que nous allons étudier dans le prochain chapitre). Il existe des taux d'intérêt à court terme (déterminés sur le marché monétaire et donc fortement liés aux niveaux des taux d'intérêt directeurs) et des taux d'intérêt à long terme (fixés sur le marché financier par confrontation entre la demande et l'offre de capitaux). Les taux d'intérêt à court et long terme sont cependant liés. Ainsi, la hausse des taux sur le marché monétaire n'est jamais sans conséquence sur les valeurs boursières. Une hausse des taux d'intérêt peut provoquer, si elle n'est pas comprise ou anticipée par les acteurs du marché, des évolutions marquées sur les marchés, voire une crise financière.

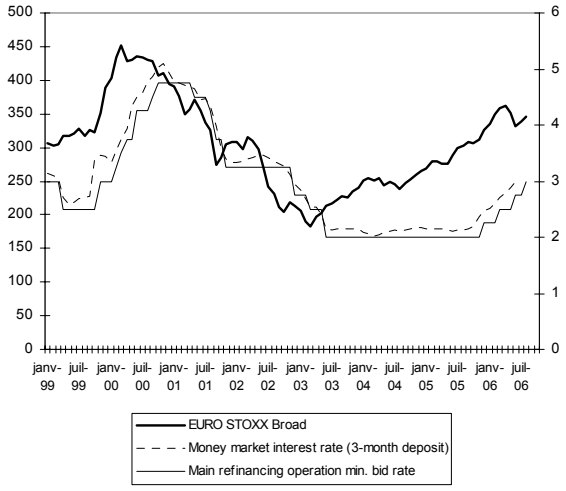


Figure 26 : Evolution du cours des actions et des taux d'intérêts

Cette nouvelle donne complique énormément l'action des banques centrales car celles-ci doivent intégrer cette contrainte complémentaire que constitue la pression des investisseurs. Elles se doivent donc de communiquer pour expliquer leurs décisions en matière de taux. Au-delà du relèvement ou de l'abaissement des taux d'intérêt directeurs, les intervenants auscultent les déclarations des banquiers centraux afin de se faire une idée sur leurs décisions futures.

3.5 Annexe

3.5.1 Les agrégats monétaires

Un agrégat monétaire est le regroupement d'un ensemble homogène d'actifs monétaire ou non monétaire. Les agrégats sont classés par ordre de liquidité décroissante.

- L'agrégat M1 comprend la monnaie fiduciaire (billets) et les dépôts à vue (y compris à la Poste et au trésor Public)
- L'agrégat M2 s'obtient en ajoutant à M1 des dépôts moins liquides (dépôts à terme, dépôts remboursables avec préavis)
- L'agrégat M3 comprend M2 majoré des exigibilités négociables des IFM (Insititutions Financières Monétaires) : pensions, titres d'OPCVM monétaires, instruments du marché monétaire.

			Parts dans M3
	Billets et pièces	514,8	6,6%
+	DAV	2 902,5	37,1%
=	**M1**	**3 417,4**	**43,7%**
+	DAT	1 109,1	14,2%
+	DAP	1 540,8	19,7%
=	**M2**	**6 067,3**	**77,6%**
+	Pensions	990,1	12,7%
+	Titres d'OPCVM monétaires	626,6	8,0%
+	IMM	131,1	1,7%
=	**M3**	**7 815,1**	**100,0%**

Tableau 2 : Agrégats monétaires de la zone euro en 2005 (en Milliards d'euros)

Dépôts à vue (DAV) : dépôts dont le retrait se fait sans préavis et instantanément. Les dépôts à vue forment la monnaie scripturale et permettent de s'acquitter d'une dette directement.
Dépôts avec préavis inférieur ou égal à 3 mois (DAP) : Dépôts d'épargne à vue qui présentent, à la différence des DAV, un caractère de transférabilité conditionnelle (Livrets A et Livrets bleus, Livrets d'Epargne Populaire, Livrets jeunes…)
Dépôts à terme (DAT) : Dépôts qui peuvent être convertis en espèces avant un terme fixe convenu au prix d'une révision totale ou partielle de la rémunération convenue. Seuls les DAT d'une durée initiale inférieure ou égale à 2 ans sont pris en compte dans les agrégats monétaires.
Pensions : Liquidités reçues en contrepartie de titres vendus à un prix donné dans le cadre d'un engagement de rachat des dits titres (ou de titres similaires) à un prix et à une date déterminés.
Titres d'OPCVM monétaires : Parts souscrites par les agents non financiers et émises par des organismes, société d'investissement à capital variable (SICAV) et fond commun de placement (FCP), effectuant la quasi-totalité de leurs placements en titres monétaires.
Instruments du marché monétaire (IMM) : Titres émis à court terme par les IFM et négociables en France. Les IMM émis par les IFM comprennent essentiellement les certificats de dépôts.

3.5.2 Le principe du multiplicateur de crédit

Soit H la quantité de « monnaie Banque centrale », constituée des billets B et des réserves R. Soit M la monnaie en circulation constituée des billets et des dépôts D.
0n considère que les agents non financiers ont l'habitude de détenir une partie de leurs encaisses (M) sous forme de billets. Soit b le rapport stable entre le volume de billets demandés et la quantité de monnaie à la disposition des agents non financiers (b=B/M). Les banques commerciales sont contraintes de mettre en réserves dans leur compte à la Banque centrale une proportion fixe de leur dépôt r (r=R/D).
Nous établissons ainsi la relation existant enter la monnaie centrale H et la masse monétaire M.

$$M = \frac{H}{b + r(1-b)}$$

Nous pouvons déduire de cette équation la valeur du multiplicateur reliant la quantité de monnaie en circulation à la quantité de monnaie centrale k = $1/\left(b + r\left(1-b\right)\right) > 1$ sachant que r et b sont relativement faibles et proches de 0). La monnaie « Banque centrale » est appelée la « base monétaire » puisque c'est à partir de cette base, lorsque s'appliquent les différents comportements des agents, qu'on déduit le volume de la masse monétaire. Réguler la base monétaire permet *in fine* de réguler la masse monétaire puisque les deux sont proportionnelles (proportionnalité donnée par le multiplicateur ci-dessus).

Lorsque la Banque centrale décide, par exemple, de réduire le taux de ses réserves obligatoires, les banques vont pouvoir utiliser ces réserves excédentaires pour accorder de nouveaux crédits (c'est leur métier). Soit ΔR les réserves excédentaires provoquées par cette décision de la Banque centrale. Soit ΔC les nouveaux crédits octroyés par les banques. L'apparition de réserves excédentaires provoquées par la Banque centrale donne naissance à un montant équivalent de nouveaux crédits dans un premier temps (ΔC =ΔR). Ces nouveaux crédits vont faire naître des fuites en billets (bΔR) et nécessiter des réserves obligatoires supplémentaires (r(1-b) ΔR), les dépôts nouveaux s'élevant à (1-b) ΔR.
L'excédent de réserves dont vont disposer les banques à l'issue de la première vague d'octroi de crédits sera ΔR-[b+r(1-b)] ΔR. Cet excédent de réserve peut alors être utilisé pour une seconde vague d'octroi de crédits. L'ampleur de la seconde vague de crédits est : ΔR=ΔR[(1-b)(1-r)]. La demande de billets de la part des agents sera d'un montant b[(1-b)(1-r)]ΔR et la nécessité de nouvelles réserves obligatoires r(1-b) [(1-b)(1-r)]. Il est aisé de constater que le montant des crédits nouveaux suit une progression géométrique de raison égale à [(1-b)(1-r)]. Ainsi le montant des crédits accordés au cours de la nième vague a-t-il pour expression : $[(1-b)(1-r)]^{n-1}\Delta R$. Au final, le montant de crédits octroyés par les banques grâce à cette baisse du taux de réserves obligatoires est donc : $\Sigma C=1/(1-(1-b)(1-r))\Delta R=1/(b+r(1-r))\Delta R=k\Delta R$.

Dans cette analyse en termes de **multiplicateur de crédit**, les banques commerciales peuvent accroître leur offre de crédits lorsque le montant de monnaie centrale dont elles disposent dépasse le minimum qui leur est nécessaire pour combler les fuites qu'elles subissent (retraits de billets et constitution de réserves obligatoires) suite à ces crédits.
Cependant, l'analyse en termes de multiplicateur de crédit pose plusieurs problèmes : (i) la demande de crédits peut être inférieure à ce que les banques pourraient accorder, étant données leurs réserves excédentaires de monnaie centrale, (ii) les banques peuvent utiliser différemment leurs excédents de monnaie centrale (reprise des titres préalablement refinancés auprès de la Banque centrale), (iii) en France, comme dans la zone euro, les banques sont généralement endettées vis-à-vis de la Banque centrale et ne détiennent donc pas d'excédent en monnaie centrale. La Banque centrale (prêteur en dernier ressort) est obligée de les refinancer. Ainsi, les banques commerciales créent de la monnaie, non pas en fonction de leurs excédents en monnaie centrale, mais en fonction de la demande de crédits (optique du **diviseur de crédit**).

3.5.3 L'*Open market* et le comportement des agents

Soit β=B/D le rapport que les agents non financiers souhaitent conserver entre les monnaies fiduciaires et scripturales détenues (les autres éléments et variables sont identiques à ceux du modèle présenté ci-dessus).
La valeur du multiplicateur reliant la quantité de monnaie en circulation à la quantité de monnaie centrale peut être exprimée sous la forme suivante M=(1+β)/(β+r)H. Comme r est compris entre 0 et 1, M est un mulitple de H. Les contraintes liées à la création de dépôts par les banques sont les retraits en billets par les agents et l'obligation de constituer des réserves obligatoires.
La Banque centrale décide de se porter acquéreur de titres (essentiellement des bons du trésor) sur le marché monétaire (politique d'« open market »). Ces titres publics étaient initialement détenus par agents financiers (banques) et non financiers (ménages, entreprises). Des agents financiers (banques) ou non financiers '(entreprises, ménages) vont participer à cette opération d'« open market ».
Pour comprendre la façon dont l'opération d'« open market » influence la quantité de monnaie en circulation, il convient de s'interroger sur les comportements des différents agents participants à cette opération.
Les banques qui participent à l'opération d'« open market » voient leur ratio dépôts/réserves (D/R=1/r) baisser par rapport à sa valeur initiale avant l'opération. La Banque centrale crédite en effet leurs comptes à la Banque centrale suite à la vente de titres, ce qui fait augmenter le montant de leurs

réserves. Pour retrouver leur rapport désiré (le niveau initial du ratio D/R), les banques commerciales vont octroyer des crédits avec leur marge de réserves ce qui aura comme effet d'accroître le montant des dépôts. Il y a création de monnaie. L'effet multiplicateur de crédits commence à fonctionner. Les dépôts alimentent de nouveaux crédits. Les effets de l'« *open market* » s'étendent à des établissements qui n'avaient pas initialement pris part à l'opération.
Les agents non financiers qui ont participé à l'opération voient leur rapport $1/\beta$=D/B s'accroître, car le crédit de la vente de leurs titres est porté à leurs comptes. Afin de ramener ce ratio à son niveau initial, ils vont substituer des billets aux dépôts et réaliser des achats.
C'est donc par les désajustements de D/R et D/B, et par les réactions des agents financiers et non financiers qui cherchent à retrouver leur rapport désiré que la création de monnaie centrale produit ses effets et explique l'expansion de la monnaie en circulation.

3.5.4 Evolution inverse de la valeur d'une obligation et des taux d'intérêt

Keynes s'est opposé aux classiques en matière de déterminant de la détention de monnaie. Pour lui, le **motif de transaction** ne saurait suffire à expliquer le comportement de détention d'encaisses monétaires. Keynes a mis en avant le motif de spéculation qui conduit les agents à arbitrer entre la détention de monnaie et la détention d'actifs financiers. Keynes s'est plutôt focalisé sur l'arbitrage entre détention d'actif à rendement fixe (les obligations) et monnaie. L'arbitrage entre deux placements dépend de leurs taux de rendements respectifs. Le taux de rendement de la monnaie est nul. L'agent décide donc d'acquérir des titres si leur taux de rendement est positif.
Quel est le taux de rendement d'une obligation ? On va considérer pour simplifier qu'il s'agit d'obligation « perpétuelle » rapportant un coupon de 1 euro chaque année. Pendant une année, l'obligation rapporte donc 1 euro placé au taux d'intérêt r. L'obligation a été acheté à un cours V_O et pourra être revendue en fin d'année au prix anticipé V_O^a. Si le prix de revente de l'obligation est supérieur au prix d'achat, l'agent réalise un gain en capital (sinon perte en capital). Le rendement total de l'obligation sera positif si le taux d'intérêt excède le taux de perte en capital potentiel.
Le prix de marché des obligations est également fonction des taux d'intérêt. La valeur du titre se fixe en effet au niveau de la valeur du flux des coupons futurs actualisés (si le prix était inférieur, des agents exploiteraient cette opportunité et achèteraient des titres ce qui ferait grimper le prix jusqu'à cette valeur).
L'**actualisation** correspond au calcul de la valeur présente du revenu futur procuré par l'obligation. Si l'obligation rapporte 1 euro l'année prochaine, il aurait fallu placer aujourd'hui la somme de 1/(1+r) euro au taux d'intérêt r pour disposer de ce même euro demain.
Le prix de l'obligation perpétuelle est ainsi : $V_O = \sum_{t=1}^{\infty} \frac{1}{(1+r)^t} = \frac{1}{r}$.
Si l'agent anticipe un taux d'intérêt r^a pour les périodes futures, le prix anticipé de la revente du titre est $V_O{}^a = \sum_{t=1}^{\infty} \frac{1}{(1+r^a)^t} = \frac{1}{r^a}$. Le gain anticipé en capital est donc $\Delta V^a = \frac{1}{r^a} - \frac{1}{r}$. Si l'agent anticipe une élévation du taux d'intérêt, il anticipe par conséquence une perte en capital. Le taux de rendement anticipé R^a de l'obligation perpétuelle (pour un euro investi) est : $R^a = r + \frac{\left(\frac{1}{r^a} - \frac{1}{r}\right)}{\frac{1}{r}} = r\left(\frac{1+r^a}{r^a}\right) - 1$. Ce taux de rendement est positif si et seulement si : $r > \frac{r^a}{1+r^a} = \bar{r}$ où $\bar{r}$ est le seuil de taux d'intérêt au-delà duquel l'agent préfèrera placer ses actifs sous forme de titres plutôt que de détenir de la monnaie. Le choix binaire d'un individu dépend de ses anticipations de taux d'intérêt. Lorsqu'on considère que les anticipations des agents sont dispersées (perceptions économiques différentes des individus) et que le taux d'intérêt anticipé ne dépend pas du taux d'intérêt courant, on peut conclure que plus le taux d'intérêt augmente et plus la proportion d'individus pour lesquels $r > \bar{r}$ devient importante. De plus en plus d'individus se tournent vers l'achat de titres et refusent de détenir de la monnaie.
La demande de monnaie est par conséquent une fonction décroissante du taux d'intérêt courant.
L'intuition est relativement simple. La valeur d'une obligation est décroissante avec le taux d'intérêt. Lorsque le taux d'intérêt est bas, les agents anticipent qu'il va remonter et donc que la valeur des obligaions va baisser. Pour ne pas avoir à subir une perte en capital, les agents préfèrent donc détenir

de la monnaie en attendant que les taux remontent et que la valeur des obligations chutent afin d'acquérir les titres quand les espérances de gain en capital sont les plus probables.

3.5.5 Valeur d'une action et condition de non arbitrage

La valeur d'une action d'une entreprise est la somme actualisée des profits futurs réalisés par cette entreprise. En appelant π_τ le profit réalisé à la date τ par cette entreprise, r le taux d'actualisation (qu'on confond au taux d'intérêt sans risque à la date t r_t) et V_t la **valeur de cette entreprise** à la date t (sa capitalisation boursière), et en considérant que le temps est continu, on peut écrire :

$$V_t = \int_t^\infty e^{-r_t(\tau - t)} \pi_\tau d\tau$$

En dérivant cette équation par rapport au temps on obtient la variation de la valeur de cette entreprise

$$\dot{V}_t = r_t e^{r_t t} \int_t^\infty e^{-r_t \tau} \pi_\tau d\tau - e^{r_t t} e^{-r_t t} \pi_t$$

Qu'on peut réécrire de la façon suivante ; $\dot{V}_t = r_t V_t - \pi_t$. D'où l'expression de la **condition** dite **de non arbitrage** :

$$\frac{\dot{V}_t}{V_t} + \frac{\pi_t}{V_t} = r_t$$

A l'équilibre, le gain potentiel en valeur ($\frac{\dot{V}_t}{V_t}$) plus le taux de dividende ($\frac{\pi_t}{V_t}$) égalise le taux d'intérêt sans risque (r_t) car les agents arbitrent entre ces deux types de placement. Bien entendu, une prime de risque peut s'ajouter au taux d'intérêt sans risque (arbitrage habituel en finance entre rentabilité d'un placement et risque associé).

Bon bien entendu dans les calculs ci-dessus, on ne tient pas compte de la prime de risque qui peut venir s'ajouter aux taux d'intérêt sans risque (or, l'investissement en actions est bien entendu risqué).

La progression importante de la valeur d'une action ($\frac{\dot{V}_t}{V_t} > 0$), à rendement et à taux d'intérêt donnés, ne peut ainsi être obtenue qu'au prix d'une élévation de la prime de risque puisque le taux de dividende devient de plus en plus faible. A un moment donné, l'importance de cette prime de risque fait prendre conscience aux investisseurs qu'ils ont survalorisé cette action et le retournement de tendance peut être brutal…

4 Conjoncture et politiques économiques

Quel sera le niveau du taux de croissance du Produit Intérieur Brut français[54] pour l'année prochaine ? Les instituts de conjoncture, les services économiques des grandes banques, la direction de la prévision du ministère de l'économie et des finances ont chacun leur propre idée sur cette question, souvent étayée par des modèles macroéconométriques qui permettent d'effectuer des prévisions en fonction d'hypothèse sur l'environnement économique et financier de l'année prochaine (cours du pétrole, niveau des taux d'intérêts, état de l'économie mondiale et particulièrement de l'économie américaine, cours €/$...). Les prévisions ne sont pas forcément convergentes car chaque prévision dépend des hypothèses retenues, du modèle utilisé, de l'*a priori* du conjoncturiste : 1,5% ?... 2% ?... 2,5% ? Beaucoup moins ? Encore plus ? Quel modèle simple utiliser pour vous-même vous faire une idée de la vigueur de la croissance ? Sait-on aussi prévoir les retournements et les récessions ?

Si croissance il y a, comment se répartira cette croissance sur les quatre trimestres de 2008 ? Quels seront les moteurs de cette croissance (la consommation des ménages, l'investissement des entreprises, le commerce extérieur) ?

Et si jamais, la croissance n'est pas au rendez-vous (morosité du climat des affaires, moindre consommation des ménages, ralentissement de la demande mondiale, effet sur l'économie de la crise des *subprimes* aux Etats-Unis …), le gouvernement a-t-il la possibilité de soutenir cette croissance par une politique appropriée ?

Ces questions sont les interrogations de la macroéconomie et de l'analyse macroéconomique conjoncturelle. Depuis l'analyse de Keynes et la synthèse néo-classique des idées keynésiennes (le modèle IS-LM, voir plus loin), on dispose de modèles relativement simples pour répondre en première approximation à ces questionnements.

Mais la croissance économique (mesurée par la croissance du PIB) n'est pas tout, d'autres indicateurs de la santé d'une économie et de la pertinence des politiques économiques menées dans cette économie existent. L'économiste Nicolas Kaldor a proposé de représenter la situation économique d'un pays à l'aide de ce qu'il a appelé le « carré magique ». La situation économique d'un pays sera jugée d'autant plus satisfaisante que la surface d'un quadrilatère dans un plan (taux de croissance, taux d'inflation (échelle inversée), taux de chômage (échelle inversée), Balance courante en point de PIB) est grande et proche du carré magique (situation la plus enviable: pas de chômage, pas d'inflation, une forte croissance et un excédent extérieur important). Une forte croissance, un faible chômage, une inflation maîtrisée et un solde extérieur positif sont donc les objectifs de la politique macroéconomique. On peut juger année par année de la pertinence et de l'efficacité de celle-ci dans un pays ou une zone. Les objectifs finals sont décomposables en objectifs internes (croissance, chômage, inflation) et en un objectif externe (équilibre des comptes extérieurs). On pourrait également opérer une partition entre objectifs de court et de long terme. Mais ce qui va nous intéresser dans ce chapitre, ce sont les fluctuations de PIB, leurs résultantes sur le chômage et l'inflation (nous traiterons au Chapitre 6 : Ouverture internationale et globalisation, de la politique économique en économie ouverte).

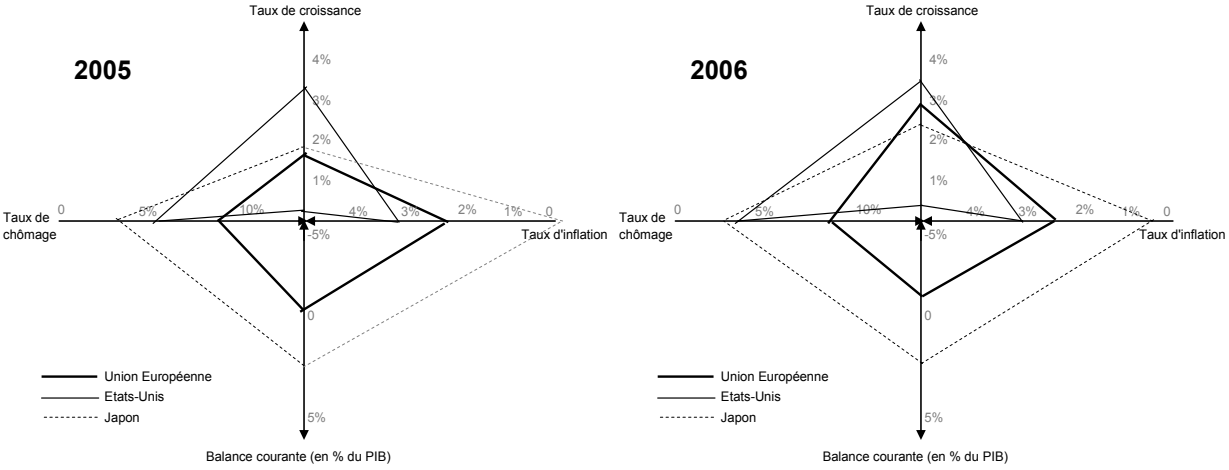


Figure 27 : « Carrés magiques de Kaldor », (années 2005 et 2006)

[54] Voir l'annexe **Le cadre de comptabilité nationale** en annexes générales pour la détermination du PIB.

Bien entendu le carré magique est difficilement atteignable. Selon les périodes, mais également selon les zones géographiques, certaines de ses composantes sont privilégiées par les autorités en charge de la politique économique. On le voit sur les figures ci-dessus représentant les « carrés magiques » des Etats-Unis, du Japon et de l'Europe. L'Europe n'apparaît pas comme la zone ayant réalisé le meilleur compromis entre les différents objectifs, ou menant la politique macro-économique la plus efficace en ce début de vingt-et-unième siècle…
La croissance européenne est relativement plus faible que celles de ses deux grands concurrents développés, Etats-Unis et Japon. L'arbitrage entre chômage et inflation (arbitrage sur lequel nous reviendrons dans ce chapitre, c'est l'un des passages obligés de la macroéconomie) semble privilégié le maintien d'une inflation faible au risque d'un chômage élevé (l'échec majeur de la politique économique européenne depuis deux décennies, voir le prochain chapitre)
La faible croissance en Europe, nous le reverrons au chapitre 7 (Les moteurs de la croissance : Education et innovation) a bien entendu certainement des ressorts structurels (d'où l'imprécation de la stratégie de Lisbonne, visant à renforcer le potentiel de croissance européen à long terme).
Mais ce qui nous intéresse dans ce chapitre, ce n'est pas le niveau de croissance tendanciel de long terme, ce sont les fluctuations de court terme et la façon dont la politique économique peut les accompagner ou les contrecarrer. Sur le graphique ci-dessous, on peut constater que la croissance du PIB français (en rythme annuel) mesurée par trimestre présente des variations importantes (des cycles temporels). Ces fluctuations conjoncturelles ont bien entendu des répercussions sur l'inflation, mais surtout sur le niveau du taux de chômage.

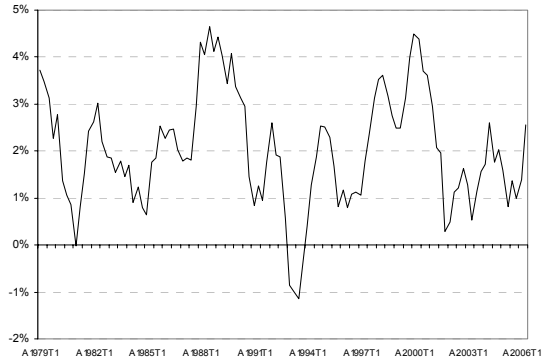


Figure 28 : Taux de croissance du PIB (rythme annuel)

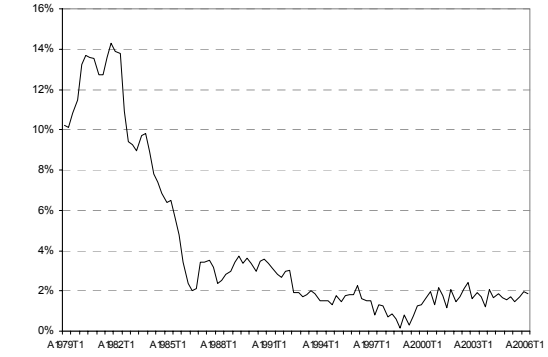


Figure 29 : Taux de croissance de l'indice des prix à la consommation (rythme annuel)

Depuis Keynes, et l'avènement d'une pensée économique s'autorisant à réfléchir en termes d'équilibre de sous-emploi (et donc prenant en considération un chômage involontaire à la différence du modèle classique), il est bien entendu que lorsque la croissance effective du PIB n'atteint pas sa croissance potentielle[55], il en résulte une dégradation du chômage. Au contraire, lorsque les capacités de production sont très sollicitées et que le taux d'utilisation des capacités de production des entreprises augmente, ces tensions aboutissent à des pressions inflationnistes.
Durant les trente glorieuses, ces deux phénomènes étaient exclusifs et la politique économique était relativement simple (un consensus avait même émergé entre économistes sur la pertinence de celle-ci, la fameuse « synthèse néo-classique ») : (i) en cas de récession (fort chômage, faible inflation), il fallait stimuler l'économie par les politiques adéquates (politiques budgétaire ou politique monétaire expansionnistes), (ii) en cas de croissance forte (entraînant des pressions inflationnistes dans un cadre de chômage faible), il fallait réussir à ralentir l'économie pour ne pas qu'une spirale inflationniste se mette en marche (utilisation de politiques budgétaire et monétaire restrictives).
Cet arbitrage possible entre inflation et chômage à travers la politique économique dite de « stop and go » décrite ci-dessus trouve sa représentation la plus aboutie dans la courbe de Phillips (courbe décroissante entre inflation et chômage, voir l'annexe sur **la courbe de Phillips** en fin de chapitre).
Cependant, à partir des années soixante-dix la politique économique va être rendue plus difficile par l'apparition du phénomène de stagflation (stagnation (chômage croissant) + inflation).

[55] On appelle croissance potentielle, la croissance qui peut être atteinte par une économie si celle-ci mobilise l'ensemble de ses facteurs de production disponibles, étant donné le stock de capital, le niveau de productivité et la croissance de la population active. Le problème européen en matière de croissance (voir le chapitre 7) est que l'on considère que la croissance potentielle ne cesse de diminuer. Auparavant, on considérait que le niveau de la croissance potentielle en Europe était de l'ordre de 2,5%/3%. Les plus sceptiques la voient s'établir aux alentours de 1% à moyen terme (du fait de la faible croissance démographique et de l'insuffisance des investissements en recherche…).

Les politiques économiques d'inspiration keynésienne qui s'avèrent inefficaces par rapport à ce phénomène vont ainsi prêter le flan aux critiques du courant monétariste (Friedman) puis des nouveaux classiques (Lucas, Barro…).
A partir des années quatre-vingts, on peut considérer que le consensus de la synthèse néo-classique a vécu. Deux visions théoriques s'affrontent en matière de diagnostic de la situation et d'efficacité des politiques économiques discrétionnaires[56].
On peut résumer ce débat à la perception différente qu'ont les deux courants antagonistes des fluctuations conjoncturelles (et donc du chômage qui l'accompagne). Pour les auteurs **Néo-keynésiens** qui tentent de sauver l'héritage du fondateur de la macroéconomie, les fluctuations sont un indice à grande échelle de la défaillance des marchés. Il existe des rigidités, tant sur le marché des biens que sur celui de l'emploi, qui empêchent des ajustements par les prix et les salaires. L'Etat doit intervenir pour réguler le marché. Au contraire, pour les **nouveaux classiques** et plus précisément les tenants de l'école des cycles réels, les fluctuations sont les réponses naturelles du marché aux changements technologiques. L'intervention de l'Etat à travers une politique économique discrétionnaire n'est pas nécessaire. Pire encore, elle est inutile et néfaste.

Alors que la situation décrite par les « carrés magiques » de 2005 et 2006 n'est pas totalement satisfaisante pour l'Europe (chômage élevé, croissance relativement modeste), on est en droit de s'interroger. Qui doit-t-on croire ? Ceux (les monétaristes, les nouveaux classiques) qui pensent que les politiques économiques discrétionnaires sont d'une part inefficaces à long terme (ou même à court terme pour les tenants des anticipations rationnelles), et d'autre part responsables de l'inflation des années soixante-dix, puis des déficits publics importants des années quatre-vingt dix ?
Ou bien ceux (les économistes néo-keynésiens) qui considèrent que les rigidités de prix et de salaires empêcheront toujours le rééquilibrage automatique par le marché et qui prônent donc en conséquence une intervention efficace de politique économique pour faire face à l'atonie de la demande ?

Durant la seconde partie des années quatre-vingts et les années quatre-vingt dix, la lutte contre l'inflation a été en Europe le principal objectif de la politique économique (politique de « désinflation compétitive », voir le graphique ci-dessus montrant comment l'inflation galopante du début des années quatre-vingts a progressivement été jugulée). Il est vrai que le niveau atteint par l'inflation au début des années quatre-vingts pouvait apparaître comme un problème économique majeur pour le pouvoir d'achat des salariés et l'exercice serein des anticipations des entreprises. Cette politique, fondée sur une politique monétaire relativement restrictive, a été couronnée de succès (l'inflation semble sous contrôle en Europe depuis le début des années quatre-vingt dix). Cette politique a pourtant été poursuivie en Europe alors que l'inflation ne semblait plus représenter une menace. Cela pour deux raisons : d'abord à cause de la réunification allemande (ayant obligé à maintenir des taux d'intérêt élevés) puis du fait de la mise sur pied de l'Union Economique et Monétaire (puisque la maîtrise de l'inflation semblait être une condition de l'avènement d'un euro fort, préalable imposé par les autorités allemandes réticentes à abandonner leur monnaie : le mark).
Après la lutte contre l'inflation, l'accent de la politique économique semble désormais être porté sur la réduction des déficits publics. Nous allons voir à l'aune de cette actualité et de cet objectif qui semble peu à peu s'imposer, si la politique budgétaire est encore possible en Europe et en France[57].
Les gouvernements ont-ils encore la latitude de mener des politiques discrétionnaires pour amoindrir les effets de la conjoncture ? Quels sont les instruments et les degrés de liberté dont-ils disposent ? Quelles en sont les contraintes ?

[56] Voir par exemple l'article de **G. Mankiw [1990]**, « A Quick Refresher Course in Macroeconomics », *Journal of Economic Literature*, vol.XXVII, pp.1645-1660, pour des précisions sur les écoles de pensées actuelles en macroéconomie, leurs programmes de recherche respectifs et leurs prescriptions en matière de politique économiques.
[57] Pour la politique monétaire, l'affaire est définitivement réglée… La BCE est indépendante et il n'y a donc plus de politique monétaire discrétionnaire déterminée par les Etats. La BCE a son propre objectif en matière de politique monétaire (la lutte contre l'inflation principalement), objectif qui n'est pas forcément en adéquation avec les politiques économiques menées par certains membres (les politiques français lancent ainsi régulièrement des appels en direction de la BCE pour que celle-ci pratique une politique monétaire moins contraignante afin que celle-ci accompagne leur propre politique de demande, mais en vain…).

4.1 La politique budgétaire

Dans les modèles économiques de base (voir les annexes sur la présentation du **modèle IS-LM** et du **modèle Offre globale-Demande globale** en fin de chapitre), la politique budgétaire constitue, avec la politique monétaire, l'un des principaux leviers de la politique économique de l'État. Elle consiste à utiliser certains instruments budgétaires (dépenses publiques, prélèvements fiscaux) pour influer sur la conjoncture économique.

Dans la réalité, les politiques budgétaires nationales apparaissent aujourd'hui en Europe comme le seul instrument pour faire face à des chocs économiques ponctuels, car les gouvernements nationaux ne peuvent plus recourir à la politique monétaire pour amortir les fluctuations conjoncturelles dans le cadre de l'Union économique et monétaire. Cependant, les contraintes fixées par le traité de Maastricht et le pacte de stabilité et de croissance limitent également l'utilisation de la politique budgétaire. Si cette mesure se justifie par la volonté d'éviter qu'un pays fasse porter les effets négatifs de sa politique budgétaire (inflation et donc hausse à terme des taux d'intérêts et du taux de change de l'euro) sur l'ensemble des pays de l'UEM, elle place les gouvernements européens dans une situation « kafkaïenne » en termes de politique économique : la politique monétaire n'est plus entre leurs mains (mais dans celles de la BCE, indépendante) et leur politique budgétaire est tellement encadrée qu'elle en devient difficilement utilisable.

Cette absence de marge de manœuvre dans la conduite de la politique économique semble répondre à une volonté d'encadrement des politiques discrétionnaires et signe la fin provisoire de la conduite des politiques économiques actives, mises en place à la suite des travaux keynésiens. A la politique économique discrétionnaire semble succéder des politiques de règle : pacte de stabilité ou politique monétaire de la BCE.

Il convient de noter que cela est un changement marquant, voire même un retour à la situation pré-keynésienne.

En effet, jusqu'à la crise des années 1930, le volume des dépenses de l'État n'était pas considéré comme une variable susceptible d'influencer le niveau d'activité de l'économie. L'analyse de Keynes a modifié cette conception en soulignant l'impact de la politique budgétaire sur le niveau d'activité économique d'un pays (voir l'annexe sur le **modèle IS-LM** en fin de chapitre). Lors d'une récession, l'ensemble des acteurs de l'économie ont tendance à diminuer leurs dépenses (les ménages consomment moins, les entreprises n'investissent plus), de sorte que l'on s'enfonce dans une crise de sous-emploi. Il n'est plus besoin de produire davantage puisque la demande de consommation est faible. Les entreprises ont donc moins besoin de main-d'œuvre et licencient. Ces licenciements entraînent une chute du pouvoir d'achat (si les anciens revenus salariaux ne sont pas totalement compensés par les indemnités chômage, si celles-ci existent) et donc une nouvelle baisse de la consommation. Seul l'Etat, par une politique volontariste d'investissements publics peut sortir l'économie de ce cercle vicieux et relancer la demande. Par conséquent, la plupart des pays développés ont mené, depuis les années 1930, des politiques de relance budgétaire lors des périodes de récession ou de moindre croissance. Ces politiques s'accompagnent de politique monétaire accommodante lorsque l'émission de monnaie n'avait pas encore été confiée à des Banques centrales indépendantes. Cela a permis une période de prospérité importante au sortir de la seconde guerre mondiale (les trente glorieuses en France). À compter de la crise consécutive au choc pétrolier de 1973, les économistes monétaristes et nouveaux classiques ont souligné les limites de la politique budgétaire, particulièrement les effets néfastes des déficits et de la dette publique ainsi que les conséquences inflationnistes de ces politiques.

L'accent apparaît désormais mis, au moins en Europe, sur le respect strict d'une orthodoxie budgétaire. A la suite du processus de désinflation compétitive des années quatre-vingts (victoire sur l'inflation…), succède donc une nouvelle lutte contre les déficits publics et le niveau élevé de la dette publique. Cette discipline budgétaire naissante est rendue nécessaire par l'existence de l'Union Economique et Monétaire, et elle semble correspondre à une prise en compte politique des critiques théoriques de l'efficacité des politiques économiques (critique de l'efficacité des politiques budgétaires lorsqu'on prend en compte les anticipations rationnelles des agents ; « équivalence Barro-Ricardo » : lorsque le gouvernement mène une politique budgétaire de relance, les agents anticipent les élévations d'impôts qui seront nécessaire dans le futur pour résorber la dette publique ; ils épargnent en prévision de la hausse d'impôt, plutôt que d'augmenter leur consommation, la politique budgétaire n'a plus l'effet multiplicateur keynésien[58] qu'on lui attribue habituellement).

[58] Voir l'annexe sur **le multiplicateur keynésien** en fin de chapitre.

4.2 L'endettement de l'Etat

La participation de la France à la monnaie unique européenne lui impose le respect d'une discipline économique et budgétaire. Les termes en ont été fixés par le traité de Maastricht en 1992 et le pacte de stabilité et de croissance, conclu à Amsterdam en 1997 et légèrement assoupli en 2005 à Bruxelles. Il s'agit d'obtenir des finances publiques saines, en évitant les déficits publics excessifs, notamment les déficits budgétaires. Le déficit public, à savoir le déficit cumulé des administrations publiques (État mais aussi collectivités territoriales et organismes de Sécurité sociale), est jugé excessif à partir d'un seuil de 3% du produit intérieur brut (voir annexe en fin de chapitre sur **pourquoi 3% ?**), qui peut toutefois être dépassé dans certaines circonstances (grave récession…).
Pourquoi ce garde-fou ? Parce que le solde budgétaire de l'État est souvent négatif : autrement dit, les recettes de l'État ne suffisent presque jamais à financer toutes ses dépenses. L'État connaît ainsi des besoins de financement qui peuvent trouver leur origine dans trois facteurs :

- le décalage en cours d'année entre le rythme de perception des recettes et le calendrier des dépenses, en raison notamment des avances de trésorerie consenties par l'État aux collectivités territoriales,
- le déficit budgétaire constaté en fin d'année,
- le remboursement de ses emprunts parvenus à échéance.

Pour couvrir ses différents besoins de financement, l'État s'endette, selon des modalités qui ont beaucoup varié au cours du XX^{e} siècle. Avant 1914, la dette de l'État était composée à plus de 80 % d'emprunts perpétuels ou de rentes amortissables à très long terme. L'inflation qui a sévi durant les deux guerres mondiales a considérablement réduit la valeur de ces rentes et, par contrecoup, la confiance que les épargnants accordaient aux titres d'État à long terme. Depuis la fin de la Seconde Guerre mondiale, la dette publique a pris la forme de bons du Trésor ou d'avances monétaires de la Banque de France (ces dernières ont été interdites par le traité de Maastricht), c'est-à-dire d'instruments de plus court terme. La dette de l'État était en 2003 composée à 98 % de titres négociables sur les marchés financiers, dont 64 % sous la forme d'obligations assimilables du Trésor (dette à long terme) et 21 % sous la forme de bons du Trésor (dette à court et à moyen termes)[59].
Les titres de la dette publique présentent une grande liquidité (ils sont facilement négociables) et constituent une référence pour les marchés financiers. Un certain degré d'endettement de l'Etat présente donc une utilité certaine car elle introduit le fond de liquidité indispensable au bon fonctionnement des marchés financiers.
Cependant, on constate une forte augmentation depuis trois décennies de la dette de l'Etat : de 55 milliards d'euros en 1980 (92 pour la dette publique), la dette de l'État est passée à 889 milliards d'euros en 2005 (1138 pour la dette publique), le budget de l'État étant chaque année en déficit depuis 1975 (voir graphique sur le déficit public depuis 1978…)[60].

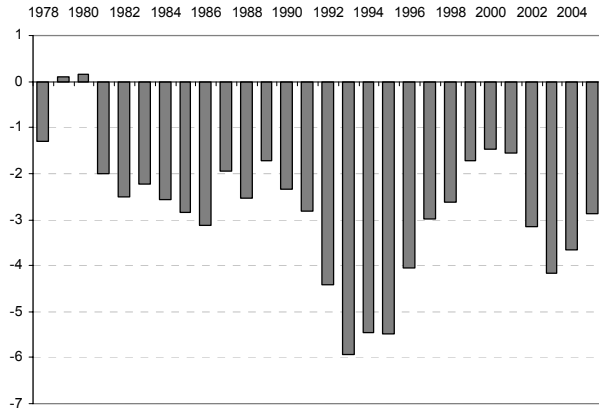


Figure 30 : Déficit public en% du PIB[61]

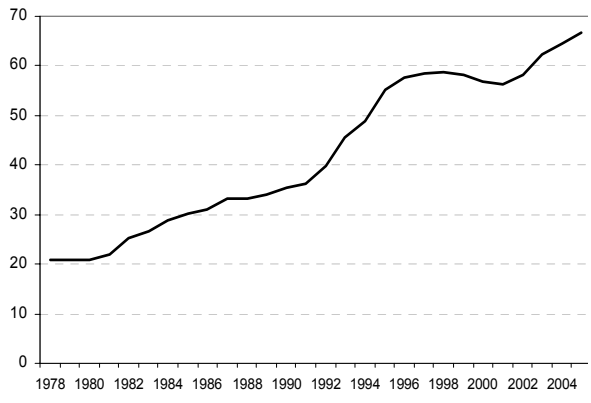


Figure 31 : Dette publique en% du PIB

[59] Voir le précédent chapitre sur la Monnaie et le Financement de l'économie pour un examen du changement de mode de financement de l'économie (passage d'une économie d'endettement à une économie de marchés financiers). L'Etat a été l'aiguillon de cette mutation lorsqu'il a du trouver un moyen de financement non-inflationniste au début des années quatre-vingt.
[60] Voir à ce sujet le rapport Camdessus. Le sursaut, *Rapport du groupe de travail présidé par* Michel Camdessus, La documentation française, 2004.
[61] Au sens de Maastricht : comprend les déficits de l'Etat, des divers organismes d'administrations centrales, des administrations locales et de sécurité sociale.

L'examen du montant de la dette en euros courants est toutefois moins pertinent que celui du ratio dette publique / PIB, car ce ratio est un critère essentiel pour vérifier si la dette de l'État est « soutenable » (voir annexe sur la **soutenabilité de la dette publique** en fin de chapitre). Ainsi, la dette de l'État peut très bien croître dans l'absolu tout en se réduisant relativement au PIB parce que l'inflation est importante et dévalorise l'endettement public en termes réels, ou parce que le rythme de croissance du PIB est soutenu. L'augmentation de la dette de l'État a progressé de 12,4% du PIB en 1980 à 52% du PIB en 2005. Les charges d'intérêt de la dette sont passées de moins de 5% du budget général de l'État en 1980 à plus de 15% aujourd'hui.
Si l'on tient compte de la dette des collectivités territoriales, de celle des administrations de Sécurité sociale et de celle des divers organismes d'administration centrale, la dette de l'ensemble des administrations publiques calculée selon les règles de comptabilité européennes atteint alors 66,6 % du PIB en 2005. Il s'agit-là d'un niveau supérieur au plafond prévu par les critères de Maastricht (60%), mais proche de la moyenne des pays de l'Union.
Ces chiffres ne donnent toutefois qu'une vision très partielle de l'évolution de la situation patrimoniale des administrations publiques. Ils correspondent à la dette publique « brute » des administrations publiques : n'en sont déduits ni les avoirs financiers (notamment leurs participations dans les entreprises publiques), ni la valeur de leurs autres actifs (immeubles, oeuvres d'art, difficiles à apprécier)[62].
En revanche, les chiffres précités ne tiennent pas compte de la dette « implicite » résultant des déséquilibres à venir des régimes publics de retraite par répartition[63]. Ainsi, les retraités actuels et futurs détiennent une créance morale sur l'État et la Sécurité sociale. Celle-ci n'est pas comptabilisée, parce qu'elle n'est pas encore matérialisée et parce que son montant dépendra des réformes à venir des régimes de retraite. Mais elle devrait peser lourdement sur les finances publiques.

Ces chiffres ne sont pas forcément inquiétants si cette dette a permis de financer des activités génératrices de croissance économique dans le futur (infrastructure, investissements, dépenses de formation…). En revanche si la dette a financé principalement des dépenses de consommation courante ou si la charge de la dette freine la croissance économique, cela devient problématique. Il convient pourtant de ne pas dramatiser le risque d'insolvabilité de l'État : contrairement aux ménages, qui doivent rembourser leurs dettes sous peine de voir leur héritage amputé, la continuité de l'État lui permet, en théorie, d'être perpétuellement endetté. Par ailleurs, le maintien d'un socle minimal de titres de dette publique est nécessaire au bon fonctionnement des marchés financiers modernes (voir chapitre sur **la monnaie et le financement**).
Après cette brève description de la situation budgétaire et patrimoniale de la France, situation souvent qualifiée de dramatique (alors qu'on a vu que si la situation est inquiétante, elle n'est pas encore désespérée…), nous pouvons maintenant aborder les effets des politiques économiques utilisant l'instrument budgétaire.

4.2.1 Stabilisateurs économiques

Les recettes et les dépenses publiques exercent spontanément une action contra-cyclique sur l'activité économique, c'est-à-dire qu'elles atténuent les aléas de la conjoncture économique. En effet, une partie des dépenses publiques sont mécaniquement liées à la conjoncture. C'est notamment le cas des dépenses d'indemnisation du chômage ou des prestations sociales versées sous condition de ressources, qui augmentent quand l'activité économique se dégrade (on considère ainsi que l'élasticité des dépenses publiques à la conjoncture est comprise entre 0,1 et 0,3). Aussi, lorsque l'activité économique ralentit, les dépenses publiques vont avoir tendance à s'accélérer tandis que les entrées de recettes ralentissent mécaniquement, ce qui provoque une détérioration du solde budgétaire. La détérioration de l'activité économique provoque donc un transfert de revenus des administrations publiques vers les agents privés (ménages et entreprises), ce qui atténue l'effet du ralentissement économique pour ces derniers[64].

[62] Il est probable que la prise en compte des actifs de l'État ne modifierait pas fondamentalement le diagnostic de la dégradation de sa situation patrimoniale.
[63] La commission Pébereau estime à un montant minimal de 400 milliards d'euros, les engagements « hors bilan » comme les retraites des fonctionnaires par exemple. Des finances au service de votre avenir, *Rapport du groupe de travail présidé par* Michel Pébereau, La Documentation française, 2005.
[64] À l'inverse, en période de forte expansion économique, les prélèvements fiscaux et sociaux augmentent mécaniquement, tandis que les dépenses diminuent, ce qui a tendance à freiner la croissance de la demande intérieure (mais engendre les fameuses « cagnottes fiscales »…).

Par conséquent, les recettes et les dépenses publiques fonctionnent comme des « stabilisateurs automatiques » puisqu'elles contribuent à amortir les variations conjoncturelles de l'activité économique. Cependant, ce mécanisme de stabilisation automatique ne fonctionne pleinement que si les ménages et les entreprises ne modifient pas leur comportement de consommation et de production, et si les taux d'intérêt ne sont pas affectés par la croissance de la dépense publique en période de récession.

4.2.2 Les politiques budgétaires volontaristes

En cas de forte dégradation de la conjoncture économique, les gouvernements peuvent être tentés de mener une politique budgétaire volontariste. Une telle politique consiste à soutenir l'activité économique à court terme, en faisant jouer le « multiplicateur keynésien » (voir annexe sur le **multiplicateur keynésien** en fin de chapitre). L'objectif de cette politique est de compenser la faiblesse des dépenses privées par un accroissement des dépenses publiques. En effet, une augmentation des dépenses publiques engendre des revenus supplémentaires qui sont pour partie consommés, pour partie épargnés, et pour partie récupérés par les administrations publiques sous la forme d'impôts et de cotisations sociales. La partie de ces revenus supplémentaires qui est consommée vient nourrir la demande intérieure adressée aux entreprises. Ces dernières peuvent dès lors augmenter leurs investissements, leurs emplois, et distribuer des revenus supplémentaires. Le surcroît de dépenses publiques provoque par conséquent un effet cumulatif qui stimule d'autant plus l'activité économique que les revenus sont peu épargnés, peu imposés, et que la demande de consommation s'adresse principalement aux entreprises nationales (sinon il y a éviction par l'extérieur). Cette politique peut augmenter la demande et rapprocher la production effective de la production de plein-emploi. Elle est donc utile si on considère qu'une part importante du chômage est « keynésien », c'est-à-dire un chômage résultant de l'absence de débouchés lorsqu'on est dans un équilibre de sous-emploi (voir l'annexe en fin de chapitre sur **le chômage keynésien**).

Les gouvernements peuvent également soutenir l'activité en réduisant les charges fiscales et donc en augmentant le revenu disponible des personnes privées. Cette politique stimule l'activité économique, mais dans une moindre proportion que la dépense publique, car une partie de ce revenu supplémentaire est épargnée par les ménages et les entreprises.

4.2.3 Les limites de la politique budgétaire

Une politique budgétaire volontariste peut néanmoins avoir des effets défavorables sur l'activité économique. C'est d'ailleurs en raison de ces effets défavorables qu'elle tend aujourd'hui à être de plus en plus critiquée.

Ainsi, les besoins de financement liés à l'accroissement des dépenses publiques provoquent généralement une hausse des emprunts de l'État et des taux d'intérêt, du fait de la demande supplémentaire adressée aux marchés de capitaux. Or, cette hausse des taux décourage une partie des achats des consommateurs financés par l'emprunt, et réduit les investissements des entreprises (« effet d'éviction »).

De plus, la demande supplémentaire permise par l'augmentation des dépenses publiques doit s'adresser en priorité aux producteurs nationaux. Or, compte tenu de l'ouverture croissante des économies, l'effet des politiques budgétaires expansionnistes s'est réduit au cours des dernières décennies et profite désormais davantage aux producteurs étrangers.

Par ailleurs, l'accumulation des déficits budgétaires vient gonfler l'encours de dette publique et augmente les charges futures de l'État. Or, plus un État est endetté, plus la charge de cette dette est élevée. Celle-ci pèse d'autant plus sur son budget qu'un niveau de dette important entraîne des taux d'intérêts élevés lorsque l'État veut contracter de nouveaux emprunts. À terme, le poids de la dette peut ainsi devenir insoutenable par rapport au niveau des recettes de l'État.

En outre, les délais liés à la prise de décision politique (le budget est voté une fois par an) font que les effets de la politique budgétaire peuvent être inadaptés à la conjoncture au moment où ils affectent les décisions des agents économiques.

De plus, certaines dépenses publiques sont particulièrement rigides, et donc, difficilement réversibles en cas de retournement de la conjoncture.

4.3 La politique monétaire

C'est la banque centrale qui a en charge la politique monétaire de la nation. Conséquence du traité de Maastricht, la Banque de France est devenue indépendante. Créée le 18 janvier 1800, la Banque de France s'est vu progressivement confier des missions de régulation du secteur financier en plus de son rôle initial d'institut d'émission (le privilège d'émission lui avait été accordé en 1803). Elle a ainsi été amenée à assurer gratuitement le service de caisse des comptables du Trésor puis à consentir à l'Etat des avances publiques. La Banque de France fût réorganisée par une loi du 24 juillet 1936 à la suite de la victoire électorale du Front populaire et nationalisée en 1945 (transfert du capital à l'Etat devenant effectif le 1er janvier 1946).

Le traité de Maastricht, puis l'introduction de l'euro et la création du SEBC, ont modifié en profondeur le statut et les activités de la Banque de France, en les inscrivant dans un cadre européen. Une loi du 4 août 1993 garantit l'indépendance de la Banque de France. C'est désormais la pierre angulaire du droit applicable à la banque centrale nationale, le cadre des missions fondamentales de la Banque de France[65].

Depuis la mise en place de l'euro, la France a donc abandonné sa souveraineté nationale en matière monétaire à la Banque centrale européenne (la BCE situé à Franckfort).

Indépendance puis…abandon de souveraineté ! Comment expliquer cette particularité associée à la conduite de la politique monétaire (la Banque centrale américaine, la Fed, est également indépendante) ? Nous avons vu au chapitre précédent le rôle de la monnaie et ses incidences sur l'économie réelle. Vous devez être donc en mesure de comprendre pour quelles raisons, il est devenu justifiable de ne plus confier la gestion de la politique monétaire aux gouvernements.

L'objectif essentiel de la politique monétaire est d'ajuster la quantité de monnaie en circulation aux besoins de l'économie. Mais, on l'a vu pour les économistes keynésiens, une augmentation de la quantité de monnaie en circulation peut permettre de relancer l'économie en dopant le pouvoir d'achat. La régulation de la création monétaire a une incidence sur la distribution du pouvoir d'achat. Lorsque cette régulation est entre les mains du politique, son utilisation à des fins électoralistes peut avoir des conséquences désastreuses[66]… Les dérives passées en matière d'inflation justifient donc l'indépendance des Banques centrales. Il convient cependant que la politique monétaire menée par un organisme indépendant soit efficace et non dogmatique, pour que celle-ci bénéficie à l'ensemble des citoyens.

4.3.1 Les objectifs et les instruments de la politique monétaire

Au cours des années quatre-vingts, on a assisté à une réorientation des politiques économiques. La priorité qui s'est imposée, particulièrement en Europe, et plus spécialement en France, fût et demeure celle de la lutte contre l'inflation. Si les mesures prises pour éradiquer la flambée des prix des années soixante-dix ont été couronnées de succès, attestant de l'efficacité de la politique monétaire (restrictive) durant les années quatre-vingts, il est plus difficile de saluer les conséquences de la poursuite des politiques visant à la stabilité monétaire en Europe durant les années quatre-vingt dix. Avec l'avènement de la monnaie unique en Europe, de nouveaux objectifs et de nouvelles contraintes se doivent d'être pris en compte dans la définition et la poursuite des politiques monétaires.

Nous allons donc décrire les objectifs et les instruments de la politique monétaire et indiquer quelles variables et quels objectifs de politique économique ont été privilégiés pour la politique monétaire depuis la fin de la seconde guerre mondiale.

---

[65] « *La Banque de France fait partie intégrante du Système européen de banques centrales, institué par l'article 8 du traité instituant la Communauté européenne, et participe à l'accomplissement des missions et au respect des objectifs qui sont assignés à celui-ci par le traité* ».

« *Dans ce cadre, et sans préjudice de l'objectif principal de stabilité des prix, la Banque de France apporte son soutien à la politique économique générale du Gouvernement* ».

« *Dans l'exercice des missions qu'elle accomplit à raison de sa participation au Système européen de banques centrales, la Banque de France, en la personne de son gouverneur, de ses sous-gouverneurs ou d'un autre membre du Conseil de la politique monétaire, ne peut ni solliciter ni accepter d'instructions du Gouvernement ou de toute personne* ».

[66] Songez à une relance monétaire quelques mois avant des élections qui permettrait de doper le pouvoir d'achat des ménages, la consommation et donc la croissance. Les conséquences heureuses de cette politique monétaire expansionniste (décrites précédemment auxquelles on peut ajouter une décrue du chômage) sont visibles avant l'élection mais les conséquences néfastes (l'inflation et la remontée à termes du chômage vers son taux naturel) n'interviennent qu'après l'élection. La tentation est donc grande. Le passage à l'acte fût régulier, par exemple aux Etats-Unis…

La politique monétaire est partie intégrante de la politique économique prise dans son ensemble. Ses objectifs sont donc ceux de la politique économique (objectifs contenus dans le carré magique de Kaldor, *cf.* l'introduction de ce chapitre).
Les finalités générales de la politique économique peuvent néanmoins être déclinées à travers un second niveau d'objectifs, propres à la politique monétaire : les objectifs intermédiaires. On distingue trois objectifs intermédiaires pour la politique monétaire : (i) un objectif de progression d'un agrégat monétaire, (ii) un objectif de taux d'intérêt, (iii) un objectif de taux de change.
La spécificité de la politique monétaire va se traduire par le choix d'une variable de politique monétaire comme instrument. Pour que ce choix s'avère pertinent, la variable monétaire choisie doit permettre d'atteindre un ou plusieurs des objectifs finaux de la politique économique (un ou plusieurs éléments du « carré magique »). La variable monétaire choisie doit être quantifiable et mesurable afin que les autorités puissent apprécier si l'objectif qu'elles se sont fixées est atteint ou non. Enfin, bien entendu, les autorités doivent pouvoir agir sur cette variable monétaire qui constitue l'instrument de leur politique monétaire.

On peut difficilement atteindre plusieurs objectifs en même temps. Aussi, un choix et des priorités s'imposent. Selon les époques, le contexte économique et les considérations théoriques ont déterminé l'objectif prioritaire de la politique économique et donc la variable instrumentale de la politique monétaire.

4.3.1.1 La pensée Keynésienne prédomine durant les « trente glorieuses »

Pour Keynes, on l'a vu dans le chapitre précédent, la monnaie joue un rôle actif dans l'économie. Les politiques économiques de l'après seconde guerre mondiale ont été fortement inspirées par les travaux de Keynes. Les décisions en matière de politique monétaire étaient alors laissées à la discrétion des autorités en fonction de la conjoncture économique. Les objectifs de la politique monétaire dépendent du niveau d'emploi et du niveau d'inflation[67]. Le taux d'intérêt est l'instrument et l'objectif intermédiaire privilégié de la politique monétaire par les autorités durant les trente glorieuses (politique de « *stop and go* »).
Lorsque la croissance est faible (chômage), la baisse des taux d'intérêt va relancer les investissements et donc stimuler la croissance. Au contraire lorsque la croissance est trop forte (démarrage de l'inflation), la hausse des taux d'intérêt va freiner les investissements et permettre de contenir l'inflation.

4.3.1.2 L'école Monétariste

Avec le phénomène de stagflation des années soixante-dix, les politiques « keynésiennes » de « *stop and go* » se retrouvent sous le feu des critiques. Pour Milton Friedman, la cause de l'inflation est un accroissement anormalement élevé de la quantité de monnaie en circulation. Friedman et les monétaristes condamnent l'utilisation des politiques monétaires expansionnistes qu'ils jugent inefficaces à long terme et source d'inflation (voir l'annexe sur **la courbe de Phillips** et la critique de Friedman de l'arbitrage inflation-chômage sous-jacent à cette relation en fin de chapitre). Pour les monétaristes, l'objectif final de la politique monétaire est la stabilité des prix et l'objectif intermédiaire retenu est le contrôle de l'évolution d'un agrégat monétaire[68]. Cet objectif intermédiaire s'inscrit donc dans le cadre de la croyance en la pertinence de la théorie quantitative de la monnaie.
Pour Friedman et les monétaristes, si les conséquences de la politique monétaire expansionniste ne sont que de courte durée et préjudiciables sur le long terme (inflation), c'est avant tout parce que les agents sont victimes d'une « illusion monétaire ». Lorsque la monnaie en circulation s'accroît de manière non anticipée par les agents, les encaisses nominales détenues par ces derniers augmentent (*M*). Ils ont l'impression d'une hausse de leur pouvoir d'achat et accroissent leur demande. Ceci a pour conséquence de faire augmenter les prix, ce qui va contracter la demande. Au final, les agents

[67] Keynes : « Tant qu'il existe du chômage, l'emploie varie proportionnellement à la quantité de monnaie, lorsque l'économie connaît le plein emploi, les prix varient proportionnellement à la quantité de monnaie ».

[68] De ses études sur la vitesse de circulation de la monnaie, Friedman conclut que si elle peut varier à court terme, elle est relativement stable dans le long terme. Ainsi, si on considère que la vitesse de circulation de la monnaie est constante, l'équation quantitative de la monnaie devient, en taux de croissance : $\Delta p/p + \Delta T/T = \Delta M/M$. La politique de suivi de la progression d'un agrégat monétaire consiste donc à se donner un objectif d'inflation (moins de 2% par exemple pour la BCE, $\Delta p/p<2\%$), à faire une prévision sur l'évolution de l'activité (grosso modo, estimer le taux de croissance du PIB pour l'année suivante, disons 3% un bon taux de croissance potentiel pour l'UE, $\Delta T/T=3\%$) et d'en déduire le seuil de progression de l'agrégat monétaire en deçà duquel il n'y a pas de raison de renforcer la rigueur de la politique monétaire. Ici $\Delta M/M<5\%$ donc… L'agrégat monétaire dont la progression est ciblée et mesurée est le plus souvent l'agrégat M3 (voir le chapitre précédent).

finissent avec le même niveau d'encaisses réelles (*M/p*) qu'initialement. Leur pouvoir d'achat a été progressivement ponctionné par l'inflation. A chaque nouvelle expansion de la politique monétaire, les agents vont de mieux en mieux anticiper ses conséquences (anticipations adaptatives) et une partie de plus en plus importante de la monnaie injectée dans l'économie sera épargnée pour faire face à la future montée de l'inflation (effet de reconstitution des encaisses réelles[69]).

4.3.1.3 La Nouvelle école classique et les politiques de règle

En raisonnant avec des anticipations adaptatives (je prévois demain en examinant l'état d'aujourd'hui), les monétaristes avaient montré l'inefficacité à long terme des politiques monétaires expansionnistes et donc la nécessité d'abandonner les politiques discrétionnaires en matière monétaire. En utilisant l'hypothèse d'anticipations rationnelles[70], les nouveaux classiques vont aller encore plus loin dans leur condamnation des politiques monétaires expansionnistes. Elles sont inefficaces tant à court terme qu'à long terme. Lors d'un accroissement de la monnaie en circulation par les autorités, les entrepreneurs revoient leur prix à la hausse. Les salariés qui anticipent rationnellement cette hausse de prix, réclament immédiatement une hausse du salaire nominal de façon à conserver un même niveau de salaire réel (salaire nominal divisé par les prix) et donc de pouvoir d'achat. Puisque le salaire réel ne bouge pas, la demande de travail des entreprises reste inchangée. Il n'y a aucun effet réel de cette relance.

La nouvelle école classique (Lucas, Kydland et Prescott…) insiste donc sur la nécessité de mettre en œuvre des politiques monétaires sur le long terme et de les annoncer afin d'offrir un cadre stable aux anticipations des agents. La manipulation des variables monétaires de manière irrégulière perturbe en effet les anticipations des agents sans conséquence sur l'économie réelle.

Selon l'école des anticipations rationnelles (autre nom de la nouvelle école classique), la règle à suivre pour les autorités monétaires est donc d'établir une règle crédible et de s'y soumettre (la règle peut être comme chez les monétaristes le suivi de la progression d'un agrégat monétaire afin de contrôler la stabilité des prix). Une des conséquences importantes est donc que l'organisme le plus pertinent pour mener la politique monétaire et un organisme indépendant (justification à l'indépendance des Banques centrales), seul capable d'assurer la cohérence temporelle des décisions de politique monétaire (si la politique monétaire demeure du domaine du politique, tout changement de majorité peut occasionner une remise à plat de la politique monétaire et un changement de la règle…).

On doit également souligner que la Banque centrale se doit d'obtenir le plus rapidement possible une crédibilité de son action de la part des agents. Il faut convaincre les agents que la règle est intangible et qu'on ne pliera pas devant un autre objectif (d'où le dogmatisme des Banques centrales en matière de lutte contre l'inflation). Pour cela, la Banque centrale peut devoir aligner sa politique sur celle d'une Banque centrale dont la réputation en matière d'orthodoxie monétaire est déjà bien établie. Ce fût le cas pour la Banque de France dans les années quatre-vingts, qui rompait avec une vieille tradition inflationniste bien ancrée dans les anticipations des agents, et dût suivre la politique de la Bundesbank, pour mettre en place sa politique du « franc fort » (politique de désinflation compétitive). Un autre objectif de politique intermédiaire voit alors le jour. C'est le taux de change puisqu'on veille à la parité de la devise nationale avec celle du pays dont on veut suivre la politique monétaire.

---

[69] Il existe deux effets de l'inflation sur la consommation des ménages : (i) l'effet de « **fuite devant la monnaie** ». En cas d'inflation les ménages peuvent choisir de consommer immédiatement leur revenu pour éviter d'avoir à payer leurs biens plus chers demain, (ii) l'effet de « **reconstitution des encaisses réelles** ». Comme les prix vont augmenter du fait de l'inflation, les agents épargnent aujourd'hui pour être à même de consommer demain. A court terme, l'effet de fuite devant la monnaie peut l'emporter, mais à long terme l'effet de reconstitution des encaisses réelles est plus important de telle sorte que l'inflation a pour conséquence de diminuer la consommation et d'augmenter l'épargne.

[70] La théorie des anticipations rationnelles a initialement été développée par Muth (1961). Lorsque les agents formulent des anticipations rationnelles, ils utilisent au mieux toutes les informations dont ils disposent et sont capables d'apprécier les conséquences des évènements économiques qui se déroulent. Ils ont donc en tête le bon « modèle » de l'économie et savent l'utiliser pour prévoir les conséquences de leur décision. Un agent économique ne prévoit plus le prix d'un bien en examinant le prix de ce bien aujourd'hui et en extrapolant. Il connaît les déterminants économiques de l'offre et de la demande de ce bien. En fonction des informations dont il dispose sur l'évolution des variables économiques qui influencent l'offre et la demande, il est à même de prévoir le futur équilibre et donc le niveau de prix qui en résultera. Bien entendu, les agents peuvent se tromper dans leurs anticipations, notamment parce que l'information sur l'évolution des variables économiques n'est pas forcément parfaite. Mais comme les erreurs de prévision des agents sont réparties de manière « normale », les agents anticipent en moyenne le bon niveau de prix lorsqu'ils formulent des anticipations rationnelles.

4.3.2 La politique de lutte contre l'inflation

Depuis le milieu des années quatre-vingts, on a assisté à une réorientation des politiques économiques avec pour nouvelle priorité la lutte contre l'inflation. Les politiques monétaires ont donc logiquement été privilégiées dans ce nouveau contexte. Les mesures prises pour éradiquer l'inflation ont été couronnées de succès. En France depuis le début des années quatre-vingt-dix, l'inflation a toujours été contenue en dessous de 2,5% (voir le graphique ci-dessous montrant l'évolution trimestrielle de la croissance de l'indice des prix à la consommation).

En France, la lutte contre l'inflation s'est imposée comme une priorité à partir de 1983. La France avait vécu auparavant sur une longue tradition inflationniste, les autorités monétaires ayant fait preuve d'un certain laxisme en matière de lutte contre la hausse des prix. Le « tournant de la rigueur » fait suite en 1983 à l'échec de la relance budgétaire du gouvernement Mauroy (1982). La hausse importante du niveau des prix, comme c'était alors le cas (cf. graphique) entraîne des effets déstabilisants à la fois pour les entreprises et pour les ménages. L'inflation nuit à la compétitivité des entreprises exportatrices. Par ailleurs, dans un contexte de rigueur salariale, l'inflation pénalise les salariés dont le revenu stagne (cela rogne sur leur pouvoir d'achat).

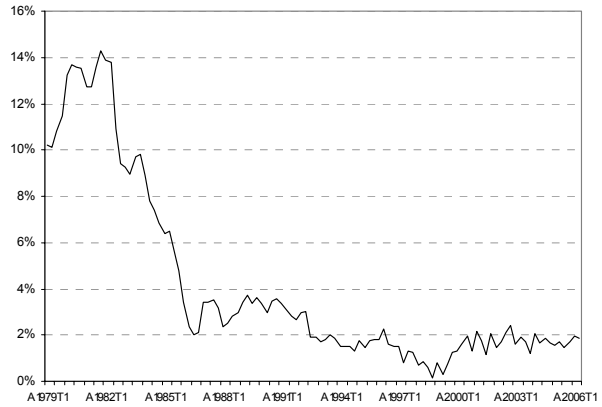


Figure 32 : Evolution de la croissance l'indice des prix à la consommation
(1979-2006)

La crédibilité de la Banque de France a été difficile à construire. Les autorités françaises ont mis en œuvre durant toute la décennie quatre-vingts et le début des années quatre-vingt-dix une politique dite de « **désinflation compétitive** » (dite aussi politique du « franc fort »). Dès 1983, les taux d'intérêts réels (qui étaient encore nuls ou négatifs quelques années auparavant) sont portés à 7%. Cette hausse va contribuer à limiter la création monétaire et donc l'inflation. L'objectif de la Banque de France était de parvenir à « importer » la crédibilité de la Bundesbank en maintenant la parité du franc vis-à-vis du Deutsche Mark. L'élévation du taux de change permet également de réduire les hausses de prix. La baisse relative des prix des produits importés provoque une désinflation importée. Les coûts pour les entreprises importatrices sont ainsi réduits. Au niveau des firmes exportatrices, c'est un changement radical qui va se mettre en place. Les firmes françaises qui avaient jusque là bâti leurs avantages concurrentiels à l'échelle internationale sur le facteur prix (aidées le plus souvent par les dévaluations du franc) vont perdre cet avantage ce qui va les forcer à rechercher une nouvelle forme de compétitivité, passant par une meilleure maîtrise des coûts et une amélioration de la qualité des produits proposés (mise en place d'une compétitivité hors-prix). Les efforts vont se trouver récompenser à partir du milieu des années quatre-vingts, avec d'une part la maîtrise de l'inflation et d'autre part un solde extérieur positif (deux objectifs du carré magique !).

La politique monétaire n'explique cependant pas à elle seule cette baisse de l'inflation à partir du milieu des années quatre-vingts même si son rôle a été prépondérant. D'autres facteurs ont facilité cette baisse : (i) la stabilité des salaires (désindexation du niveau des salaires sur le niveau des prix décidée au début des années quatre-vingts), (ii) faiblesse du prix des matières premières (contre-choc pétrolier au milieu des années quatre-vingts).

La question que l'on peut se poser est alors la suivante. Etant donné la faiblesse du taux d'inflation au cours des années quatre-vingt-dix, comment peut-on expliquer la poursuite de la politique monétaire restrictive ?

La réponse est que le contexte économique des années quatre-vingt-dix a transformé les enjeux de la politique monétaire. De nouveaux enjeux et de nouvelles contraintes ont du être pris en compte dans le cadre de la politique monétaire du fait de la mise en place en Europe de la monnaie unique.

4.3.3 La construction de l'UEM et les contraintes associées

Au cours des années quatre-vingt-dix, la raison essentielle de la persistance des politiques monétaires rigoureuses est la construction de l'Europe monétaire. En 1991, les Etats européens décident de mettre en œuvre la monnaie unique dès 1999. Ils se contraignent donc à une harmonisation de l'espace monétaire européen. L'harmonisation va se construire autour des critères de convergence, qui représentent des objectifs à atteindre pour que les pays candidats puissent participer au projet de l'euro.

Critères	Objectifs
stabilité des prix	taux d'inflation ne devant pas dépasser de plus de 1,5% la moyenne des taux des trois pays membres les plus performants
taux d'intérêt	taux d'intérêt de moyen et long terme ne dépassant pas de plus de 2% les taux des trois pays membres les moins inflationnistes
déficit public	déficit public ne dépassant pas 3% du PIB
dette publique	dette publique ne dépassant pas 60% du PIB
taux de change	devise étant restée au moins depuis deux ans dans le Système Monétaire Européen

Tableau 3 : Les critères de convergence

Les objectifs assignés par les critères de convergence vont dans le sens d'une poursuite des efforts en matière de lutte contre l'inflation et de stabilité des changes. Certains critères vont dans le sens d'un abandon, ou d'une contrainte dans l'utilisation, des politiques budgétaires actives (voir l'annexe en fin de chapitre sur **pourquoi 3% ?**).

Cependant, si la politique monétaire restrictive demeure relativement efficace pour lutter contre l'inflation, elle limite l'investissement et la relance de l'activité. L'impression qui ressort de cette période est donc celle d'un sacrifice de la croissance sur l'autel de la lutte contre l'inflation.

Le conseil européen d'Amsterdam en 1997 a mis sur pied le **Pacte de stabilité** afin que, une fois l'euro mis en place, les dérapages budgétaires des pays membres de l'UEM n'affaiblissent pas la monnaie unique ou ne représentent pas une source potentielle d'inflation. Les troisième et quatrième critères de convergence ont ainsi été réaffirmés. Malgré les dérapages des déficits publics de la France et de l'Allemagne en 2003 et 2004, le conseil européen de Bruxelles en mars 2005 n'a pas assoupli ces critères du pacte de stabilité, même si il a introduit une plus grande souplesse d'application en cas de récession. Il ne faut cependant pas en conclure que le Pacte de stabilité a été assoupli, puisqu'il ressort de ce Conseil européen de 2005 une incitation a dégager des soldes publics positifs en période de croissance forte.

4.3.4 La politique monétaire dans un environnement complexe

La politique monétaire présente un certain nombre d'avantages et son importance s'est affirmée au cours de la dernière décennie. C'est une politique d'une grande flexibilité puisqu'elle peut être réorientée plusieurs fois dans l'année (alors que le budget fait l'objet d'un vote annuel même si un correctif est possible). La politique monétaire ne semble pas entraîner de mesures lourdes pour les ménages à l'inverse de la politique fiscale (hausse des impôts !). On a cependant vu dans ce chapitre, que la pertinence d'une politique monétaire impliquait une action durable dans le même sens (politique de règle) de sorte que sa flexibilité n'est peut-être que potentielle.
Par ailleurs, l'ouverture des économies à la circulation des capitaux modifie considérablement le cadre et l'efficacité des politiques monétaire (voir les conclusions du **modèle IS-LM-BP**, modèle IS-LM en économie ouverte en annexe du chapitre Ouverture internationale et globalisation).
Deux questions méritent donc d'être posées en conclusion de cette présentation de la politique monétaire : (i) Comment les autorités monétaires peuvent-elles conserver leur crédibilité tout en s'adaptant à la conjoncture ?, (ii) Comment la politique monétaire peut-elle parvenir à concilier un objectif externe (la stabilité des changes) et un objectif interne (la stabilité des prix nationaux) ?
Nous nous interrogerons enfin sur les difficultés de mener une politique monétaire au niveau européen.

4.3.4.1 La règle de Taylor

En matière de crédibilité des politiques monétaire et d'adaptabilité à la conjoncture, Taylor (1993) a mis en évidence la possibilité de mener des politiques de règles actives (voir l'annexe en fin de chapitre sur **la règle de Taylor**). Les règles actives présentent l'avantage de pouvoir modifier l'orientation des politiques monétaires de façon presque automatique. La cohérence temporelle est ainsi conservée. Les agents connaissent le mode de fonctionnement des autorités monétaires et intègrent celui-ci dans leurs anticipations. Les autorités peuvent ainsi intégrer d'autres objectifs que la seule surveillance de l'inflation. La règle de Taylor (qui semble bien retracer le fonctionnement de la Fed ou même de la BCE) se propose ainsi de déterminer l'évolution des taux d'intérêt à court terme en fonction du niveau du taux d'inflation (par rapport au taux d'inflation cible) mais également en fonction du taux de croissance (plus précisément en fonction de son niveau par rapport au taux de croissance potentielle). La crédibilité de la Banque centrale est ainsi préservée et son action ne se limite plus à la seule poursuite d'un objectif monétaire de lutte contre l'inflation.

4.3.4.2 Le triangle d'incompatibilité de Mundell

L'économiste Mundell a mis en exergue, dans les années soixante, les incompatibilités qui existent entre les différents objectifs que peuvent se fixer les autorités monétaires (indépendance de la politique monétaire nationale, stabilité du taux de change, liberté des mouvements de capitaux). Seuls deux de ces objectifs peuvent être atteints simultanément.

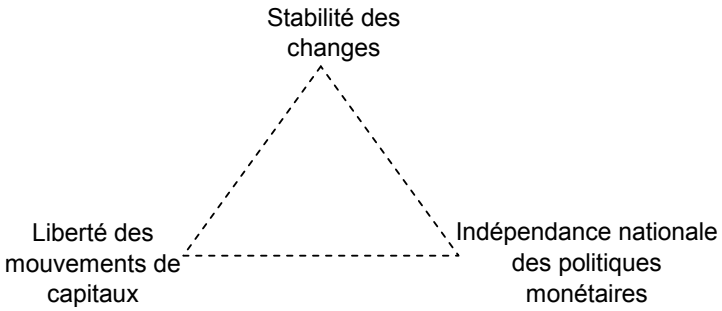


Figure 33 : Le triangle des incompatibilités de Mundell

La libre circulation des capitaux constituant désormais un acquis et donc une contrainte, les autorités monétaires se doivent d'arbitrer entre la stabilité des changes et la liberté de pratiquer la politique monétaire souhaitée. Au début de la décennie quatre-vingt-dix (1992, 1993), le Système Monétaire Européen a ainsi été secoué par plusieurs crises monétaires du fait de la difficulté pour certains pays de conserver leur stabilité des changes dans un contexte où la réunification allemande avait amené les autorités monétaires de ce pays à conserver des taux d'intérêts relativement élevés. La lire puis la Livre durent sortir du SME. En août 1993, le France n'a dû son maintien dans le SME qu'au prix d'une extension des marges de fluctuations autorisées entre les monnaies.

Depuis l'avènement de l'euro, il semble que les autorités monétaires européennes aient fait le choix d'une certaine indépendance de la politique monétaire de sorte que l'objectif de stabilité externe de l'euro ne semble pas une priorité. Face aux fluctuations du dollar, les autorités européennes ne semblent pas vouloir réagir et privilégient l'obtention de leur objectif interne (l'absence d'inflation).

4.3.4.3 La politique monétaire européenne

Avec l'arrivée de l'euro, la politique monétaire relève désormais de la Banque centrale européenne. L'objectif confié à la Banque centrale européenne est la stabilité des prix. Pour réaliser ses interventions, la Banque centrale européenne intervient sur le marché monétaire par l'intermédiaire des banques centrales des différents pays. Les décisions en matière monétaire sont prises par le conseil des gouverneurs au sein duquel siègent les gouverneurs des banques centrales de la zone euro.

Il est cependant difficile de mener une politique monétaire européenne unique dans une zone euro qui est loin d'être homogène : (i) il existe des décalages sur le plan conjoncturel (ainsi par exemple en 2003 la croissance était nulle en Allemagne et de 0,5% en France alors qu'elle était relativement forte en Irlande et aux Pays-Bas), (ii) sur le plan structurel, des différences subsistent dans les niveaux de

développement (le Portugal et la Grèce présentent encore un certain retard en termes de revenu par habitant[71]), (iii) certaines économies sont fortement spécialisées.
Désormais, avec la monnaie unique, lors d'un choc économique touchant avec une intensité variable les différents pays de la zone euro (choc macroéconomique asymétrique), les pays ne disposent plus du taux de change pour atténuer les effets de ces crises.
Ainsi, l'Europe n'est pas une **Zone Monétaire Optimale**, au sens où l'économiste Mundell a défini une zone qui bénéficierait de manière optimale d'une monnaie commune.
Une zone monétaire est dite optimale lorsqu'elle remplit deux conditions :

- **Existence d'un faible degré d'asymétrie** entre les chocs subis par les régions de la zone. Les chocs doivent être ressentis de façon homogène permettant de mettre en œuvre une politique monétaire unique (faible spécialisation des régions ou pays formant la zone).
- **Existence d'une mobilité importante des facteurs de production**, de façon à permettre les rééquilibrages en cas de choc. Le déplacement de main d'œuvre des zones où le chômage est élevé vers les zones où l'activité est forte permet alors de résoudre les problèmes d'emploi et évite les démarrages d'inflation dans les pays dont l'économie fonctionne à plein régime.

Or, dans la zone euro, ces deux conditions sont loin d'être remplies. Les économies européennes connaissent encore des spécialisations fortes et différentes de sorte que les chocs exogènes sont ressentis avec une intensité variable. De plus, la mobilité des salariés en Europe reste faible (barrières culturelles, linguistiques ou autres…). L'hétérogénéité de la zone euro constitue donc un obstacle majeur à la définition et à la conduite d'une politique monétaire commune.
L'absence d'un véritable gouvernement économique européen, en face de la BCE, interdit par ailleurs une bonne articulation des politiques économiques (budgétaire et monétaire) à l'échelle européenne.

4.4 Conclusion

La performance de l'Europe, notamment celle de la France dont la situation est tout à fait représentative du mal européen, en matière de promotion et de stabilisation de la croissance est relativement mauvaise en regard de celle des Etats-Unis. Le graphique ci-dessous[72] permet de juger de cette piètre performance. En abscisse figure l'écart-type de la croissance annuelle et en ordonnées la croissance moyenne annuelle par décennie (des années 60 jusqu'au début des années 2000).

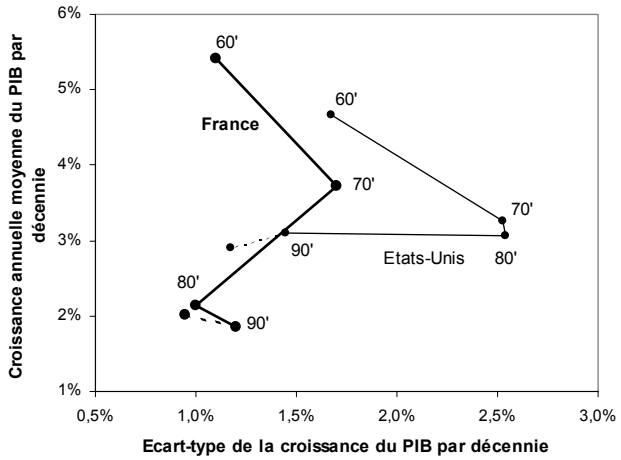


Figure 34 : Arbitrage croissance – stabilité : France et Etats-Unis

On constate immédiatement que depuis les années quatre-vingt dix, la performance des Etats-Unis en matière de croissance est meilleure que celle de la France. Depuis les années soixante-dix, aux Etats-Unis, la volatilité de la croissance s'est réduite tandis que les performances en matière de croissance ne cessaient de progresser. La politique économique aux Etats-Unis semble donc avoir permis d'augmenter le potentiel de croissance tout en réduisant les fluctuations. L'Europe et la France ont, durant la phase de rattrapage (années soixante et soixante-dix) eu l'avantage d'avoir une croissance à la fois plus forte et moins volatile que celle des Etats-Unis. Mais depuis la crise des années

[71] Les PIB/habitant en 2005 en euros de l'Allemagne et de la France sont de l'ordre de 27 000€, celui du Portugal de 14 000€.
[72] J'emprunte l'idée de cette représentation à Ph. Aghion et J. Pisany-Ferry, co-rédacteurs du « rapport Sapir sur la croissance en Europe » (2003), dans leur « réponse à leurs détracteurs » parue dans une lettre de l'OFCE (2003).

soixante-dix les performances en termes de croissance sont de plus en plus faibles, sans pour autant que la réduction de la volatilité de la croissance aille de paire avec cet amoindrissement.

Cette situation est donc révélatrice d'un échec relatif du *policy-mix* qui est mené en Europe (il convient de remarquer que la politique budgétaire est encore décidée au niveau national tandis que la politique monétaire est centralisée et indépendante, et qu'elle ne semble pas devoir être un adjuvant des politiques budgétaires nationales, bien au contraire). Si, comme nous l'avons vu dans ce chapitre, une partie des piètres performances de l'Europe en matière de croissance peuvent être imputées à la mise en place de l'Union Economique et Monétaire dans une zone qu'on ne peut pas encore qualifier d'optimale au sens de Mundell, il n'en demeure pas moins que l'accompagnement de la croissance par des politiques économiques adéquates semble faire défaut car la croissance européenne ne semble plus à même d'être, depuis deux décennies, suffisamment forte et durable pour pouvoir faire reculer significativement le chômage (ce que nous allons voir dans le prochain chapitre).

Il ne faut pas négliger, dans ce débat sur l'efficacité de la politique économique, les enjeux des politiques structurelles et leurs interactions avec les politiques plus conjoncturelles. Puisque l'efficacité des politiques fiscales ou budgétaires est sujette à caution, l'Europe doit reporter ses efforts sur des politiques permettant de renforcer à long terme son potentiel de croissance. L'Europe prend-t-elle ce chemin ? C'est possible. Ainsi, le sommet de Bruxelles en mars 2005 qui a vu l'Allemagne et la France obtenir un relatif assouplissement du Pacte de Stabilité avait, on s'en souvient moins, un deuxième ordre du jour : l'examen à mi-parcours de la stratégie de Lisbonne. Il ressort de ce sommet que les procédures pour déficits budgétaires excessifs pourraient être moins répressives lorsque les raisons de ce déficit sont des dépenses d'investissement visant à préparer l'avenir et particulièrement à développer de potentiel d'innovation en Europe. La réforme du Pacte de stabilité, si elle réaffirme l'importance du respect de l'orthodoxie budgétaire en Europe (elle engage notamment les gouvernements des différents pays européens à dégager des excédents en période de conjoncture favorable, et pas seulement à respecter le critère de 3% de déficit par rapport au PIB en période de récession), semble donc prendre en compte la nécessité de laisser une marge de manœuvre aux autorités en charge de la politique économique dans les différents pays européens, pour que ne soit pas hypothéqué l'avenir de la croissance européenne au profit d'un équilibre budgétaire de court terme dont la pertinence économique n'est pas partagée.

4.5 Annexes

4.5.1 Le modèle IS-LM

Déterminer l'équilibre macroéconomique, c'est appréhender la manière dont se forme au niveau global de l'économie les variables fondamentales, que sont le niveau de production, le niveau des prix et des salaires, le taux d'intérêt et le taux de chômage. Il convient pour cela de décrire les déterminants de la demande et de saisir les mécanismes par lesquels ils interagissent avec les facteurs d'offre.

Nous pouvons utiliser pour ce faire le modèle IS/LM qui correspond à l'extension proposée par Hicks (1937) de la pensée de Keynes contenue dans la Théorie générale l'emploi, de la monnaie et de l'intérêt (1936). Hicks, aidé par la suite par Hansen, a particulièrement mis en lumière l'un des apports de Keynes : l'analyse conjointe du marché de la monnaie et du marché des biens[73].

On considère que les prix sont fixes à court terme (le modèle IS/LM est donc un modèle à prix fixes). Lorsqu'on se situe à un horizon pour lequel il devient irraisonnable de penser que les prix demeurent fixes, c'est le modèle Offre Globale/Demande Globale qu'il convient de prendre en compte (celui-ci est présenté plus loin en annexe de ce chapitre). Le modèle IS/LM par d'un point de vue de comptabilité nationale et détermine le niveau de production comme la somme des dépenses (voir l'annexe sur **le cadre de comptabilité nationale** en Annexes générales). On part de l'équilibre ressources-emplois en économie fermée. Les déterminants du PIB sont donc la consommation, C, l'investissement privé, I, et l'investissement public, G. L'équilibre ressources-emplois en économie fermée (sans tenir compte des importations et des exportations donc) s'écrit : Y = C + I + G (production = consommation + investissements privé et public). Les deux composantes C et I vont faire l'objet d'une modélisation, c'est-à-dire que l'on va déterminer des équations de comportement (équation de consommation, équation d'investissement) qui permettront par la suite de déterminer la demande agrégée en fonction des variables et paramètres exogènes (G, les dépenses publiques sont ainsi considérées comme exogènes).

Le principal déterminant de la demande agrégée est la consommation des ménages. Les ménages arbitrent entre consommation courante et consommation future (épargne) en fonction du taux d'intérêt. Soit Y_D le revenu disponible des ménages : Y_D=Y-T où T sont les impôts nets des subventions. La fonction de consommation des ménages est : $C = C\left(\underset{+}{Y_D}, \underset{-}{r}\right)$ où r est le taux d'intérêt. La consommation dépend positivement du revenu et négativement du taux d'intérêt. La hausse des taux incite à épargner et donc à réduire la consommation.

L'investissement mesure les dépenses brutes des entreprises pour accroître leur stock de capital. C'est une dépense courante qui vise à faire bénéficier l'entreprise, dans le futur, d'une meilleure capacité de production et donc de profits plus importants. La fonction d'investissement est : $I = I\left(\underset{+}{Y}, \underset{-}{r}\right)$. Le niveau des ventes courantes et prévues incite l'entreprise à investir. Le niveau élevé des taux d'intérêt freine au contraire l'investissement car les entreprises sont souvent obligées de s'endetter pour investir et le taux de rendement du capital doit être supérieur au taux d'intérêt pour que l'opération soit rentable. L'investissement dépend donc positivement du revenu et négativement du taux d'intérêt.

Enfin les dépenses gouvernementales, que l'on notera G, sont la troisième composante de la demande agrégée. Celles-ci sont considérées comme exogènes.

La **demande agrégée** est donc : $Z = Z(Y - T, r, G) = C(Y - T, r) + I(Y, r) + G$.

Elle dépend positivement du revenu et négativement du taux d'intérêt puisque consommation et investissement augmentent et diminuent respectivement en cas de hausse de ces variables.

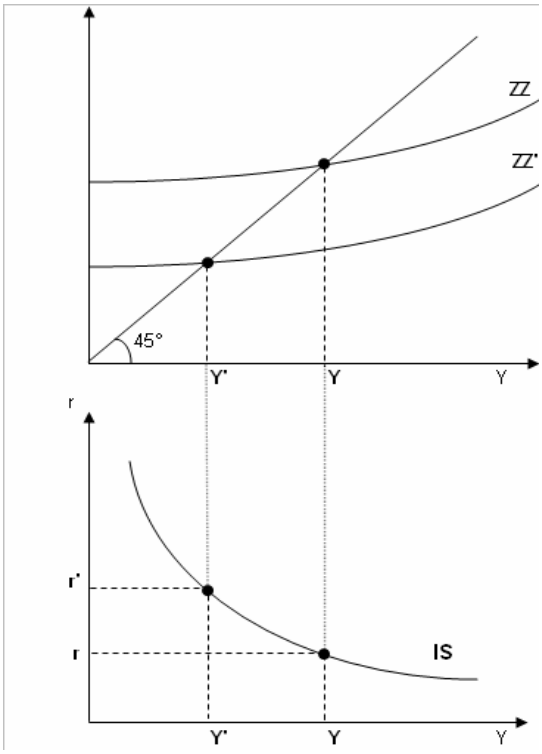
Soit Q la production totale vendue par les entreprises. On a nécessairement Q=Z. De plus dans le cadre macroéconomique retenue, on sait également que l'offre vendue Q est égale au revenu Y des agents puisque la production est la source des revenus distribués (voire l'annexe sur la

[73] Il convient cependant de noter que les keynésiens « purs » considèrent le modèle IS-LM comme une vulgate erronée de la pensée de Keynes (pas de prise en compte des anticipations, réflexion en termes d'équilibre, séquentialité et temporalité pas forcément pertinentes…). Keynes lui-même avait émis des réserves auprès de Hicks quant à la pertinence de cette mise en équation de sa Théorie générale. Il n'en demeure pas moins que l'aspect pédagogique de cette représentation a fait sa popularité et qu'elle reste, encore aujourd'hui, la pierre angulaire de l'enseignement de la macroéconomie.

correspondance entre différentes approches pour le calcul du PIB). L'égalité de l'offre et de la demande s'écrit donc :

$$Y = C(Y - T, r) + I(Y, r) + G$$

Le graphique sur la page suivante décrit la construction de la courbe IS dans le plan (Y,r). Chaque point de la courbe IS est l'intersection de l'offre et de la demande de biens (intersection de la courbe ZZ (la demande) et de la droite à 45° puisque l'offre est contrainte par la demande du fait de l'absence de débouchés).
La courbe ZZ' est obtenue pour un niveau plus élevé du taux d'intérêt (r'>r). La courbe ZZ'(r') est plus basse que la courbe ZZ(r) dans le plan (Y,r) car les différents déterminants de la demande (consommation et investissement) dépendent négativement du taux d'intérêt.



La courbe IS présentée sur la figure ci-contre a été construite pour des valeurs données de T et de G. Tout changement dans la politique fiscale ou budgétaire provoquera un déplacement de la courbe IS. Tout facteur qui, à taux d'intérêt donné, fait baisser le niveau de production d'équilibre (hausse des impôts, baisse des dépenses publiques) entraîne un déplacement de la courbe IS vers la gauche.

Figure 35 : Construction de la courbe IS

La courbe IS représente donc les couples (Y,r) qui permettent l'équilibre sur le marché des biens et services.
Pour compléter la description de l'économie, on peut présenter l'équation de demande de monnaie suivante : $\frac{M}{P} = L\underset{+}{(Y}, \underset{-}{r})$. Une hausse du revenu nominal augmente la demande de monnaie. Une hausse du taux d'intérêt réduit au contraire la demande de monnaie (motif de spéculation introduit par Keynes, voir l'annexe sur **le motif de spéculation** à la fin du chapitre précédent).
Soit M^s l'offre de monnaie nominale de la Banque centrale en face de cette demande de monnaie.
L'équilibre sur le marché de la monnaie implique que plus le revenu est élevé, plus la demande de monnaie est forte et donc plus le taux d'intérêt d'équilibre est élevé. Cette relation croissante entre revenu et taux d'intérêt est la courbe LM.

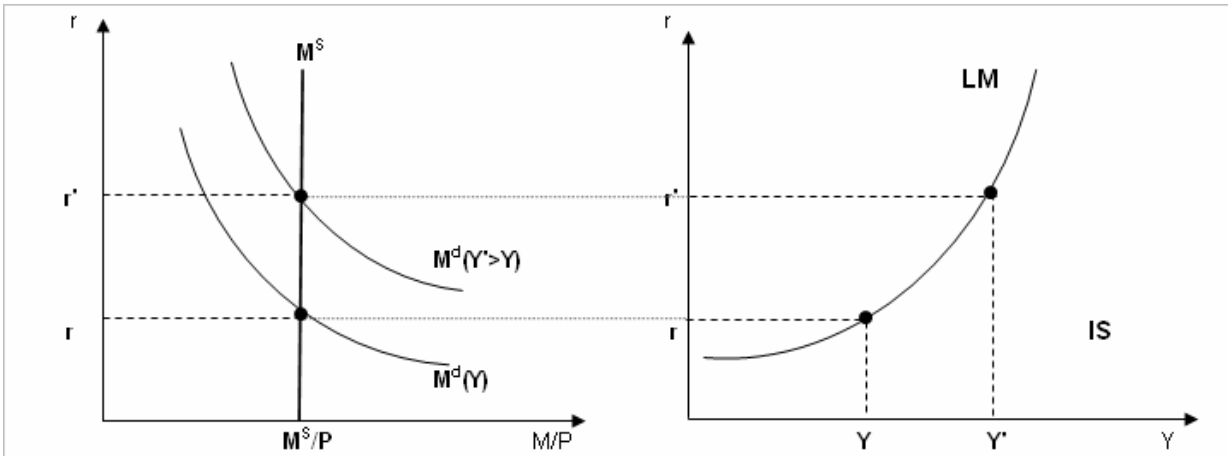


Figure 36 : Construction de la courbe LM

Bien entendu, la courbe LM est représentée ci-dessus pour une valeur donnée de l'offre nominale de monnaie par la Banque centrale. A prix donné, si l'offre réelle de monnaie augmente (de M^s /P à $M^{s'}$/P), cela entraîne, pour un revenu Y donné, une baisse du taux d'intérêt d'équilibre (déplacement de la courbe LM vers le bas). Inversement, une baisse de l'offre de monnaie entraîne une hausse du taux d'intérêt (déplacement de LM vers le haut).

Nous pouvons à présent étudier l'équilibre IS/LM. A chaque période, l'offre de biens est égale à la demande de biens et l'offre de monnaie est égale à la demande de monnaie (les relations IS et LM sont toutes deux vérifiées).

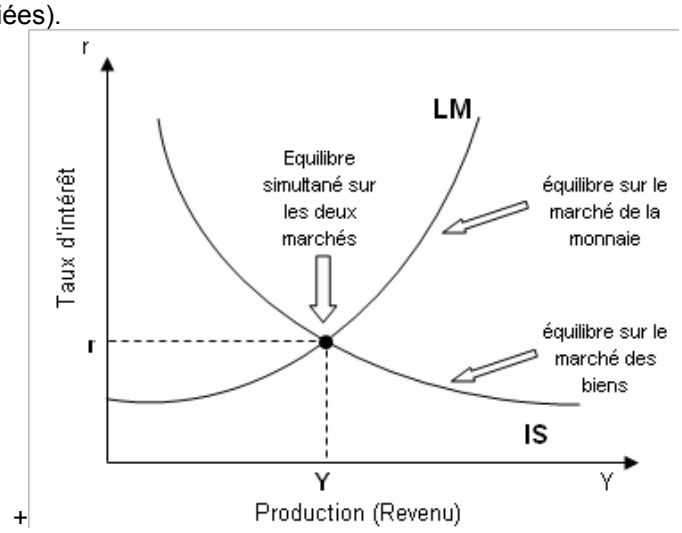


Figure 37 : L'équilibre IS-LM

L'équilibre sur le marché des biens implique une relation décroissante entre production et taux d'intérêt. L'équilibre sur le marché de la monnaie implique une fonction croissante entre la production et le taux d'intérêt. Le seul point auquel à la fois le marché des biens et le marché de la monnaie sont équilibrés est le point d'équilibre du modèle IS/LM.

Afin d'examiner à quoi peut bien être utile le modèle IS-LM, nous pouvons l'utiliser pour procéder à l'analyse de deux politiques économiques : une consolidation fiscale ou contraction budgétaire (baisse des impôts et/ou restriction des dépenses publiques décidée(s) par le gouvernement) et une expansion monétaire (accroissement de l'offre de monnaie à l'initiative de la banque centrale). Auparavant nous analysons le principe du multiplicateur, primordial pour comprendre certains mécanismes économiques à l'œuvre dans les ajustements qui suivent ces politiques.

4.5.1.1 Le principe du multiplicateur

Faisons l'hypothèse que la consommation et l'investissement ne dépendent pas du taux d'intérêt. Prenons une fonction affine pour la consommation en fonction du revenu disponible ($C = C_0 = c(Y - T)$). La courbe IS s'écrit dans ce cas : $Y = C_0 + c(Y - T) + I + G$. Comme l'investissement ne dépend pas du taux d'intérêt, cette seule équation permet de déterminer le niveau de production d'équilibre et LM ne sert qu'à déterminer le taux d'intérêt (modèle proche de celui qui sous-tend l'analyse de Keynes dans la Théorie générale). Le niveau de production d'équilibre est :

$Y^* = \frac{C_0 + I + G - cT}{1 - c}$. Puisque 0<c<1, 1/(1-c) est supérieur à 1. On appelle 1/(1-c) le **multiplicateur**. Pourquoi ? Parce que si, par exemple, la dépense autonome C_0 augmente de un milliard d'euros, le multiplicateur prédit que la production va augmenter de plus que un milliard d'euros (c est de l'ordre de 0.6 de sorte que l'effet multiplicateur est aux alentours de 2.5). Une augmentation des dépenses publiques ou de l'investissement aura le même effet d'entraînement sur l'économie. Bien entendu, l'augmentation de la production et des revenus va augmenter la demande de monnaie et, à offre de monnaie constante, cela va aboutir à une hausse des taux d'intérêt ce qui réduit dans la réalité la portée de l'effet multiplicateur (ce mécanisme est appelé « effet d'éviction »). Lorsqu'on prend en compte le bouclage avec le marché de la monnaie et avec l'extérieur (autre possibilité de fuite du circuit), l'effet multiplicateur est de l'ordre de 1.1-1.2 en France.

Quel est le phénomène économique sous-jacent à cet effet multiplicateur ? Lorsque la demande augmente, la production et le revenu augmentent. Reprenons notre exemple d'une augmentation de la dépense autonome ou des dépenses publiques de un milliard d'euros. Cette augmentation du revenu engendre une augmentation de la consommation qui entraîne une nouvelle augmentation de la demande et ainsi de suite… La première augmentation est de un milliard d'euros, la seconde de c milliards d'euros, la troisième de c² milliards d'euros et la somme de cette suite géométrique de raison c est donnée par 1/(1-c). Une augmentation initiale de la production engendre une succession de vagues d'augmentations de la production. Le multiplicateur correspond à la prise en compte de l'ensemble de ces vagues.

4.5.1.2 Effet d'une politique budgétaire restrictive

Un gouvernement, ayant pour objectif la réduction du déficit budgétaire, décide d'augmenter les impôts ou de réduire les dépenses publiques (ce qui importe c'est que G-T diminue). Soit E(Y,r) l'équilibre avant cette politique. La hausse des impôts va diminuer le revenu disponible des ménages, ce qui fait baisser la consommation et, par l'effet multiplicateur, la production. Le niveau de la production est également diminué si le gouvernement abaisse ses dépenses publiques conjointement. La courbe IS se déplace vers la gauche, de IS à IS'. Que se passe-t-il du côté des marchés financiers et de la monnaie (courbe LM) ? La courbe LM n'est pas influencée par cette politique (G et T n'entre pas dans l'équation qui définit l'équilibre LM). Le taux d'intérêt pour lequel l'offre et la demande de monnaie s'égalisent demeure identique. LM ne bouge pas. Ainsi, quand la courbe IS bouge, l'économie se déplace le long de la courbe LM de E à E'. Le nouvel équilibre se situe au point E'(Y',r'). Le taux d'intérêt a baissé de r à r'.

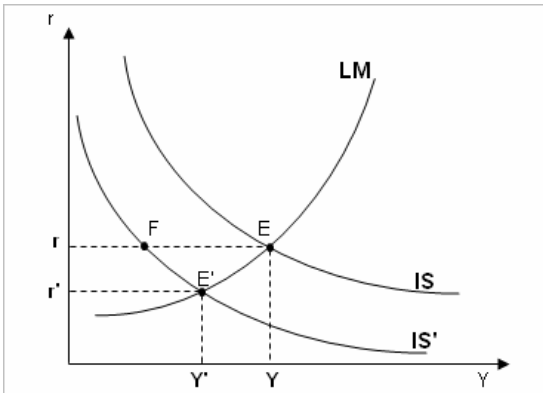


Figure 38 : Effet d'une politique budgétaire restrictive

La baisse du revenu a fait diminuer la demande de monnaie, ce qui a engendré la baisse du taux d'intérêt. Cette baisse du taux d'intérêt a pu compenser l'effet récessif de cette mesure de consolidation fiscale et de restriction budgétaire en relançant l'investissement et donc la demande de biens (moindre effet d'éviction). Sans cette baisse du taux l'intérêt, c'est-à-dire en n'intégrant pas l'effet des marchés financiers et en ne se focalisant que sur le marché des biens (passage de IS à IS'), le recul de la production aurait été plus important (point F sur la figure ci-dessus).

La consommation est réduite du fait de la diminution du revenu par deux canaux (hausse des impôts et effet multiplicateur négatif). Pour l'investissement, deux effets s'opposent. D'un côté la baisse de la production incite à moins investir. De l'autre, la baisse du taux d'intérêt encourage l'investissement. On ne peut donc rien conclure sur l'effet d'une telle mesure sur l'investissement (cela dépend de la

forme de la fonction d'investissement qu'il faudrait estimer par l'économétrie). La baisse du déficit public n'est donc pas forcément le gage d'une relance de l'investissement…

4.5.1.3 Effet d'une politique monétaire expansionniste

Supposons que la Banque centrale augmente la quantité de monnaie en circulation par une politique d'*Open Market* (telle que décrite dans le chapitre sur la monnaie et le financement de l'économie). Elle achète des obligations assimilables au Trésor (OAT) afin d'injecter des liquidités dans l'économie. On suppose dans le modèle IS/LM que les prix sont fixes et donc cette hausse de la masse monétaire nominale M entraîne une hausse proportionnelle de la masse monétaire réelle M/P.
L'offre de monnaie n'affecte pas la courbe IS qui reste inchangée. La courbe LM va en revanche être déplacée. Une hausse de la masse monétaire déplace LM vers le bas et la gauche (baisse du taux d'intérêt pour un niveau de revenu donné). Comme le taux d'intérêt s'ajuste immédiatement suite à cette mesure monétaire (alors que le niveau de production s'ajuste lentement), l'économie saute du point E au point F avant de se déplacer lentement le long de LM' vers l'équilibre final E'(Y',r').

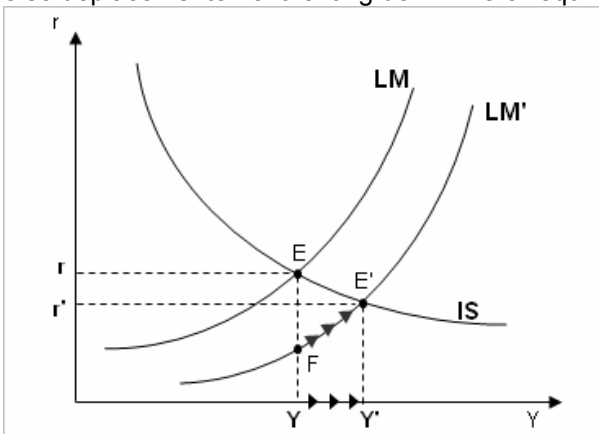


Figure 39 : Effet d'une politique monétaire expansionniste

La hausse de la masse monétaire en circulation entraîne donc une baisse du taux d'intérêt, ce qui a pour conséquence de relancer l'investissement et par effet multiplicateur d'entraîner une hausse de la production. La hausse du revenu engendre une demande de liquidités plus importante ce qui fait ré-augmenter légèrement le taux d'intérêt. Le revenu a augmenté et les impôts sont inchangés de sorte que la consommation augmente. La demande est plus élevée et le taux d'intérêt a diminué donc l'investissement augmente.
Les deux exemples ci-dessus ont décrit les politiques budgétaire et monétaire prises isolément. Dans la réalité, les deux outils sont utilisés conjointement. C'est ce que l'on appelle le ***policy-mix***. Le policy-mix peut résulter de deux politiques visant le même but (une relance budgétaire sans effet d'éviction de l'investissement privé) ou bien de tensions et de désaccords entre le gouvernement, menant la politique budgétaire, et la banque centrale régulant l'offre de monnaie Souvenez-vous que celle-ci est indépendante et qu'elle s'accroche, par exemple en Europe, à un dogme très « monétariste » en matière d'émission de monnaie. Si un certain nombre de gouvernements européens décidaient de mener des politiques budgétaires de relance (relance concertée), la BCE pourrait vouloir neutraliser les effets inflationnistes de cette politique et être ainsi amenée à mettre en place une politique monétaire restrictive, ce qui restreindrait beaucoup la portée de la politique budgétaire expansionniste des gouvernements européens….

4.5.2 Le chômage keynésien

Une des hypothèses de Keynes est que les entreprises produisent en fonction de la demande qu'elles anticipent. Le comportement d'offre de l'entreprise résulte du programme de maximisation du profit sous contrainte technologique (cf. l'annexe sur le programme du producteur), sachant que les prix et les salaires sont fixes : $\underset{N^d}{Max}\ \Pi\left(N^d\right) = pY - wN$ s.c. $Y = F\left(N\right)$. L'entreprise détermine sa demande de travail de façon à ce que la productivité marginale égalise le salaire réel : $F'\left(N^d\right) = \frac{w}{p}$

de sorte que $N^d = F'^{-1}\left(\frac{w}{p}\right)$ et l'offre globale optimale des entreprises est donc : $Y^{opt} = F\left(N^d\right)$.

Cependant, dans le modèle IS-LM, l'offre globale des entreprises ne coïncide pas toujours avec cette offre optimale, car la demande globale pour ce niveau de revenu peut être inférieure. Les entreprises sont ainsi contraintes sur leurs débouchés et n'écoulent pas leur offre optimale. Le revenu d'équilibre est déterminé par le niveau de la demande globale et les profits des entreprises ne sont pas maximisés puisqu'elles produisent moins. Comme le modèle IS-LM fait l'hypothèse que la production est toujours inférieure à $Y^{opt}\left(\frac{w}{p}\right)$, le niveau de production qui maximise le profit, les entreprises accepteront toujours de servir une augmentation de la demande puisque cela améliore leurs profits.

En outre, on suppose que le revenu d'équilibre « keynésien », Y^K, est inférieur au revenu de plein emploi car la demande de travail est inférieur à l'offre de travail. L'emploi se fixe donc au niveau de la demande de travail. Pour résumer : $N^K < N^o \Leftrightarrow F\left(N^K\right) < F\left(N^O\right) \Leftrightarrow Y^K < Y^{PE}$

La situation de chômage keynésien ($N^d<N^o$) peut être appréciée sur la figure suivante.

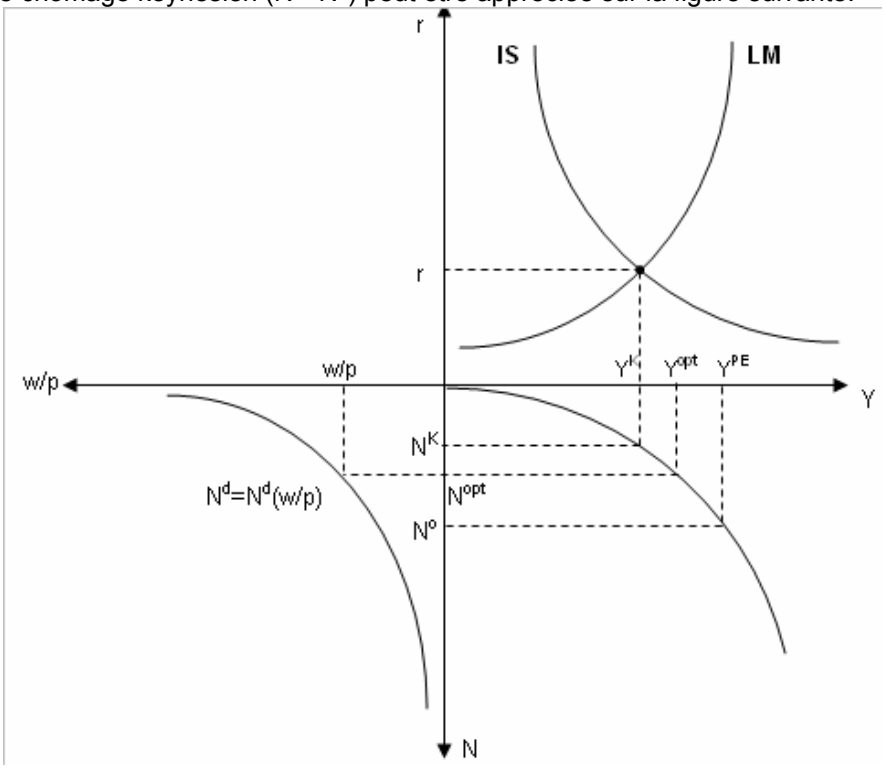


Figure 40 : Contrainte de débouchés et chômage keynésien

Le niveau effectif du revenu du modèle IS-LM est donc doublement contraint : contrainte de plein emploi des capacités de production et contrainte de rentabilité des entreprises ($Y^K \leq Y^{PE}$ et $Y^K \leq Y^{opt}\left(\frac{w}{p}\right)$).

Les entreprises sont contraintes car elles ne sont pas sur leur courbe de demande de travail (leur profit n'est donc pas maximum car la productivité marginale du travail est supérieure au salaire réel en vigueur).

Les politiques de relance de la demande (« relance keynésienne ») ne sont efficaces que jusqu'à ce que la production atteigne le niveau optimal pour les entreprises (Y^{opt}). Etant donné le salaire réel w/p rigide, les entreprises n'accepteront pas de produire plus que le niveau Y^{opt}. Au-delà de ce niveau de production, le chômage n'est plus imputable à une insuffisance de la demande mais à une insuffisance de l'offre rentable (chômage classique). Dans ce cas, la politique économique qui viserait à restaurer le plein emploi devrait restaurer la rentabilité des firmes en diminuant le coût du travail (le salaire réel w/p), par exemple en réduisant les cotisations sociales patronales ou bien en faisant en sorte que les salaires augmentent moins vite que les prix (voir le prochain chapitre sur le chômage).

4.5.3 Le modèle offre globale/demande globale

Le modèle IS-LM présenté ci-dessus considérait les prix comme fixes. L'hypothèse qui différencie le modèle offre globale/demande globale du modèle IS-LM est que le niveau général des prix va être supposé parfaitement flexible, ce qui va permettre la réalisation d'un équilibre entre offre et demande sur le marché des biens et services.

Il convient donc de définir une fonction d'offre globale puisque, à la différence du modèle IS-LM où les firmes sont contraintes par leurs débouchés et s'ajustent à la demande, elles vont pouvoir dans ce modèle choisir leur niveau optimal de production.

L'équation d'offre globale décrit les effets de la production agrégée sur le niveau des prix. Elle découle du comportement de maximisation des firmes et de l'équilibre sur le marché du travail. En prenant les comportements de fixation des salaires et des prix suivants :

$$\begin{aligned} w &= p^a F(u,z) \\ p &= (1+\mu)w \end{aligned} \quad \text{d'où il résulte } p = p^a(1+\mu)F(u,z)$$

Où w est le salaire nominal, p^a le niveau anticipé des prix, u le taux de chômage, z une variable composite tenant compte de l'ensemble des autres facteurs affectant la détermination des salaires.

Le niveau des prix et égal au salaire nominal multiplié par le taux de marge.

En retenant une fonction de production simple Y=N, on va pouvoir exprimer le taux de chômage en fonction du niveau de production. De par la définition du chômage, on sait que :U = L - N (U :nombre de chômeurs, L :population active, N : nombre de travailleurs employés). Le taux de chômage est donc : $u = \frac{U}{L} = 1 - \frac{N}{L} = 1 - \frac{Y}{L}$.

La fonction d'**offre globale** entre le niveau des prix, le niveau des prix anticipé et le niveau de production est donc : $p = p^a(1+\mu)F\left(1 - \underset{-}{\frac{Y}{L}}, z\right)$. La hausse de la production agrégée entraîne une hausse du niveau des prix car : *(i)* la hausse de l'activité entraîne une hausse de l'emploi, *(ii)* la hausse de l'emploi entraîne une baisse du taux de chômage, *(iii)* la baisse du taux de chômage entraîne une pression à la hausse sur les salaires, *(iv)* la hausse des salaires renchérit les coûts des entreprises qui augmentent en retour leur prix.

Soit Y^e le niveau de production d'équilibre de long terme. Si $Y<Y^e$ alors $p^a>p$, et inversement. Une hausse du niveau anticipé des prix déplace la courbe d'offre vers le haut.

La relation de **demande globale** est la relation qui traduit l'équilibre sur le marché des biens et de la monnaie (l'équivalent de l'équilibre IS-LM mais avec des prix flexibles). Elle peut être représentée par l'équation suivante : $Y = Y\left(\underset{+}{\frac{M}{p}}, G, \underset{-}{T}\right)$, issue de l'équilibre du modèle IS-LM. Le niveau de production globale est décroissant du niveau du prix. Des changements dans la politique monétaire, budgétaire ou fiscale déplacent la courbe de demande globale.

La figure ci-dessous représente l'équilibre (E) entre l'offre globale (OG) et la demande globale (DG). Au point E, tous les marchés (bien, monnaie, travail) sont équilibrés. Il y a équilibre du marché du travail, car E est sur OG. Il y a équilibres des deux autres marchés car E est sur DG. Lorsque le niveau de production est à son niveau d'équilibre, le niveau des prix est égal au niveau anticipé. Les niveaux d'équilibre de la production et des prix de court terme sont donnés par Y et p. L'équilibre (de court terme) ne correspond pas forcément au niveau d'équilibre Y^e (de long terme). Tout cela dépend de la position de la courbe d'offre globale (et donc du niveau anticipé des prix p^a) et de la position de la courbe d'offre globale (et donc de M, G et T, c'est-à-dire de la politique économique).

Sur la figure ci-dessous, on examine le cas où $Y<Y^e$ (le niveau de production est inférieur à son niveau d'équilibre et le taux de chômage est au-dessus de son niveau « structurel »). A court terme, la production peut différer donc de son niveau d'équilibre.

Comment l'économie rejoint-elle son point d'équilibre de long terme ? Cela dépend de la façon dont les agents forment leurs anticipations. Prenons l'hypothèse simpliste d'anticipations adaptives : $p(t)^a = p(t-1)$. Les agents anticipent que les prix de l'année à venir sont ceux qu'ils constatent actuellement.

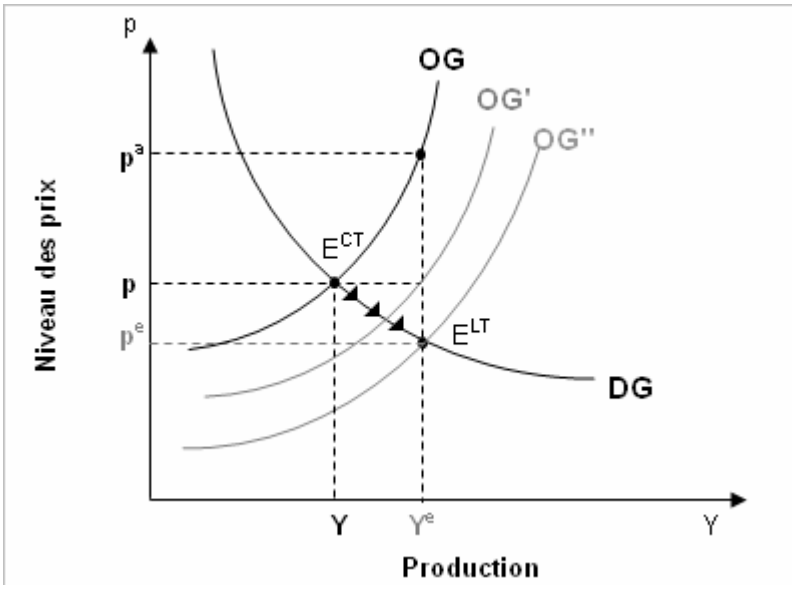


Figure 41 : Le modèle Offre Globale - Demande Globale

Partons de l'équilibre E^{CT}. On constate que l'anticipation du niveau des prix de la part des agents s'est révélée fausse. Le niveau des prix p(t) est moins élevé que le niveau anticipé $p^a(t)$ (moindre inflation ou déflation). La production Y(t) est en dessous de son niveau d'équilibre Y^e.
Que se passe-t-il en (t+1) ? La courbe de demande globale n'est pas influencée par les anticipations de prix. En revanche la courbe d'offre globale va se déplacer. Le nouveau prix anticipé est égal au prix constaté (du fait des anticipations adaptatives que l'on a postulées). La nouvelle courbe d'offre doit passer par le point auquel la production est égal à son niveau d'équilibre Y^e et le niveau des prix à la nouvelle anticipation (courbe OG'). A mesure que les anticipations sont reformulées on va s'approcher de la situation de long terme ou le niveau des prix anticipé correspond au niveau des prix d'équilibre p^e. Aussi longtemps que le niveau de production est inférieur à son niveau d'équilibre, le niveau des prix diminue et la courbe d'offre se déplace vers le bas. L'économie se déplace le long de la courbe de demande globale jusqu'au point E^{LT}. Lorsque $p^a=p^e$, il n'y a plus de pression à la baisse sur les prix et l'économie se stabilise sur son niveau de production Y^e.

4.5.4 La règle de Taylor

Il n'y pas forcément de relation étroite entre croissance monétaire et inflation car il peut exister des modifications dans la demande de monnaie (voir la partie sur les agrégats monétaires du chapitre Monnaie et Financement de l'économie). Même lorsqu'on retient un agrégat plus large (M2 ou M3) le lien entre progression de l'agrégat monétaire et inflation n'est pas très stable. Ces difficultés ont conduit les banques centrales à abandonner les cibles d'agrégats monétaires pour des cibles d'inflation.
Cependant, une fois déterminée la cible d'inflation, la question est de savoir de combien la banque centrale peut s'autoriser à dévier de celle-ci. Le fait qu'à court terme, la croissance monétaire ait des effets à la fois sur le produit et l'inflation suggère qu'il ne faut pas tenter de réduire trop brutalement l'inflation.
La question du rythme que doit adopter la Banque centrale pour atteindre sa cible a été traitée par J. Taylor (1993). Il a proposé une règle que la banque centrale peut vouloir suivre. Cette règle de Taylor est la suivante :

$$r = r^* + a(\pi - \pi^*) - b(u - u_n)$$

Avec π le taux d'inflation et π^* la cible d'inflation, r le taux d'intérêt nominal et r* le taux d'intérêt nominal visé (c'est-à-dire compatible avec le taux d'inflation π^* à moyen terme), *u* est le taux de chômage courant et u_n le taux de chômage structurel (NAIRU pour *Non Accelerating Inflation Rate of Unemployement*). Les coefficients *a* et *b* sont positifs.
Lorsque l'inflation courante est égale à la cible d'inflation ($\pi=\pi^*$) et que le taux de chômage est égal au taux de chômage naturel, la banque centrale doit choisir comme taux d'intérêt nominal de court

terme r= r*. L'économie demeure alors sur son sentier de croissance avec une inflation égale à sa cible et un taux de chômage égal au taux de chômage naturel.

Lorsque l'inflation est supérieure à la cible d'inflation choisie par la Banque centrale ($\pi > \pi*$), celle-ci doit augmenter son taux d'intérêt au dessus du taux d'intérêt cible de moyen terme. Le taux d'intérêt supérieur aura pour conséquence d'élever le chômage ce qui fera diminuer l'inflation. Le coefficient *a* reflète donc l'aversion relative de la banque centrale pour l'inflation par rapport au chômage. Plus le coefficient *a* est élevé et plus est importante la réaction de la banque centrale face à une élévation de l'inflation au dessus de sa cible. Taylor considère par ailleurs que le coefficient a doit être supérieur à 1 de manière à ce qu'une augmentation du taux d'intérêt corresponde à une augmentation du taux d'intérêt réel (et pas seulement nominal).

Lorsque le taux de chômage est supérieur au taux de chômage naturel $u > u_n$, la banque centrale doit baisser son taux d'intérêt nominal à court terme. Plus le coefficient b est élevé et plus la banque déviera de sa cible d'inflation pour ramener le taux de chômage vers son niveau structurel.

Même si Taylor reconnaît que la politique monétaire d'une banque centrale répond à d'autres objectifs et contraintes, il affirme que cette règle est une façon utile de se représenter la politique monétaire. Le choix d'une cible de taux d'inflation et l'arbitrage entre inflation courante et chômage courant semble correspondre au comportement des banques centrales (la Fed par exemple) sur les dernières décennies.

4.5.5 Le modèle IS-LM…sans LM

Le modèle IS-LM considère que la banque centrale (ou les pouvoirs publics qui ont en charge la politique monétaire) détermine directement le niveau de la masse monétaire (l'offre de monnaie). On a vu dans le chapitre précédent qu'il n'en était rien et que la banque centrale ne contrôle au mieux que la base monétaire, pas la masse monétaire. Ce chapitre nous a par ailleurs montré que l'instrument privilégie par les autorités monétaires pour leurs interventions était bien plus le taux d'intérêt que le suivi d'un agrégat monétaire. Ainsi, le modèle IS-LM, pour autant qu'il soit didactique, pêche par le fait qu'il considère le taux d'intérêt comme une endogène alors qu'il semble que ce soit une exogène du modèle, un instrument. Certaines recherches récentes en macroéconomie ont ainsi cherché à reconstruire le modèle IS-LM en substituant à la droite LM, une droite représentant la fixation du taux d'intérêt en réponse à l'application d'une règle de Taylor. Nous ne présentons pas ces modèles ici mais l'élève ingénieur intéressé pourra se reporter aux articles suivants :

G. Abraham-Frois (2003), « Pour en finir avec IS-LM : quelques propositions pour simplifier l'enseignement de la macroéconomie en premier cycle », *Revue d'économie politique*, vol. 113, n°2, pp.155-170.

J.-P. Pollin (2003), « Une macroéconomie sans LM : quelques propositions complémentaires », *Revue d'économie politique*, vol. 113, n°3, pp.273-293.

P.-Y. Hénin (2003), « En finir avec IS-LM?: remarque sur une contribution de Gilbert Abraham-Frois », *Revue d'économie politique*, vol. 113, n°4, pp. 431-434

P. Villieu (2004), « Une macroéconomie sans LM? Un modèle de synthèse pour l'analyse des politiques conjoncturelles » *Revue d'économie politique*, vol. 114, n°3, pp. 289-322.

4.5.6 La courbe de Phillips

En 1958, l'économiste néo-zélandais A.W. Phillips a proposé une estimation de la relation liant les taux de variation des salaires nominaux au taux de chômage observé en Grande-Bretagne (sur la période 1861-1957). Cette relation empirique devint un enjeu théorique important car elle fournissait l'équation manquante de la théorie keynésienne, celle qui permettait de passer de la macroéconomie statique (modèle IS-LM à prix fixes) à la macroéconomie dynamique (modèle offre globale / demande globale).

4.5.6.1 La relation de Phillips originelle (1958)

L'analyse de Phillips était exclusivement un travail empirique[74]. Phillips établissait, sur données britanniques, une relation statistique inverse entre le taux de croissance des salaires nominaux et le taux de chômage. Cette relation empirique reliait donc une évolution (celle des salaires) à un niveau (celui du taux de chômage). Le résultat de Phillips était le suivant : Une augmentation du chômage occasionne une baisse du taux de croissance du salaire nominal. La logique économique derrière ce résultat est que, lorsque la conjoncture est défavorable (élévation du taux de chômage), le pouvoir de négociation du patronat est plus important dans les négociations salariales, ce qui aboutit à une certaine retenue dans la progression des salaires. Pour la période 1861-1913, pour laquelle les résultats sont les plus significatifs, l'équation estimée est la suivante :

$$\ln(w_{t+1}) - \ln(w_t) = 9{,}64.u_t^{-1{,}39} - 0{,}9$$

Lorsque le taux de chômage dépasse un certain seuil du taux de chômage ($\bar{u}$ = 5,5%), le salaire nominal diminue. Ce mécanisme pourrait être saisi par la relation théorique suivante :

$$\frac{\dot{w}_t}{w_t} = \lambda(u_t - \bar{u}) \qquad \text{avec } \lambda > 0$$

Cette équation et la présence d'un « effet seuil » traduisent l'existence d'un certain degré de rigidité du salaire nominal. Le salaire nominal ne s'ajuste que progressivement et lentement au déséquilibre sur le marché du travail. L'existence de cette rigidité peut permettre de justifier l'existence d'un taux de chômage positif à l'équilibre sur le marché du travail ($\bar{u}$ est le taux de chômage « naturel » ou taux de chômage de plein emploi, c'est le chômage qui subsiste en raison d'inadéquations entre offre et demande de travail ou du fait de la période nécessaire à la recherche d'emploi (chômage frictionnel)).

4.5.6.2 La « courbe de Phillips » version Solow et Samuelson (1960)

Avec Solow et Samuelson[75], chevilles ouvrières de la synthèse néo-classique, la relation de Phillips va devenir « l'équation manquante de la théorie keynésienne » : la relation entre chômage et variation des salaires va devenir une relation entre le taux de chômage et le taux d'inflation.

Pour établir une telle relation entre chômage et inflation, il est nécessaire d'établir un certain nombre d'hypothèses supplémentaires.

On suppose que le taux de marge des entreprises (que l'on notera TM) est constant, de sorte que le taux de variation des prix est égal au taux de variation des coûts salariaux diminués du taux de croissance de la productivité du travail (on notera μ le taux de croissance de la productivité). Le « coin fiscal » est nul de sorte que les coûts salariaux sont égaux aux salaires versés.

$$p_t = (1 + TM)\left(\frac{w_t L_t}{Q_t}\right)$$

En passant cette expression en taux de croissance, on obtient une relation entre le taux d'inflation, qu'on notera π_t, le taux de croissance des salaires nominaux et le taux de croissance de la productivité du travail :

$$\pi_t = \frac{\dot{p}_t}{p_t} = \frac{\dot{w}_t}{w_t} - \mu$$

[74] **Phillips A.W. (1958)**, « The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957 », *Economica*, vol.25.

[75] **P. Samuelson et R. Solow (1960)**, « Analytical aspect of anti-inflation policy », *American Economic Review*, mai.

En reprenant l'équation théorique sous-jacente à la relation de Phillips ($\frac{\dot{w}_t}{w_t} = \lambda(u_t - \bar{u})$), on obtient une relation entre inflation et chômage qui s'écrit de la manière suivante :

$$\pi_t = \lambda(u_t - \bar{u}) - \mu$$

Le passage d'une relation à l'autre est tellement facile (décalage vers le bas de l'échelle mesurant l'inflation d'un pourcentage égal au taux de croissance de la productivité du travail), que l'habitude a été prise de qualifier également de « courbe de Phillips » la relation inflation – chômage.

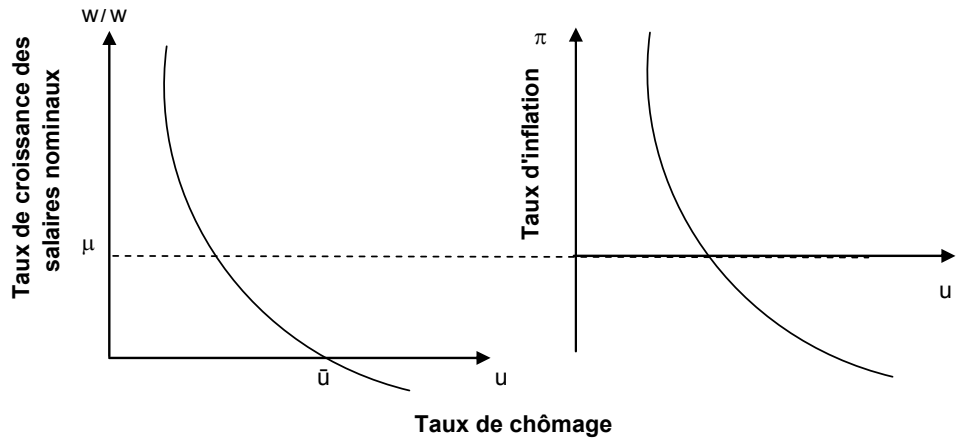


Figure 42 : Passage de la relation de Phillips originelle à la courbe inflation-chômage

Solow et Samuelson estiment, en 1960 sur données américaines, que pour avoir une inflation nulle, les salaires doivent croître au taux d'accroissement de la productivité (environ 2,5% à cette époque). Le taux de chômage correspondant est alors de 5,5%. En acceptant une inflation de 5%, on peut réduire le chômage de 3%. Solow et Samuelson demeurent cependant relativement prudents sur la stabilité de la courbe de Phillips. Cependant, les responsables politiques n'ont pas forcément la prudence des deux économistes. Ils pensent trouver dans la relation de Phillips la preuve qu'il existe un arbitrage possible entre inflation et chômage (« il est possible d'acheter un peu moins de chômage avec un peu plus d'inflation »).

Cependant, on va s'apercevoir rapidement que, hormis durant la période 1961-69 où la courbe présente un profil conforme à ce que prévoyait la théorie, la relation de Phillips n'a pas la stabilité que certains escomptaient et que l'arbitrage de politique économique est donc bien illusoire…

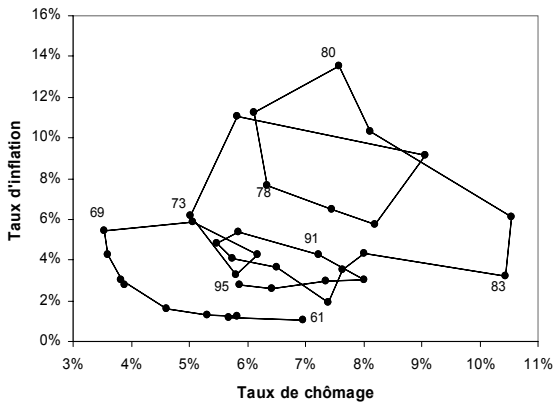


Figure 43 : Courbe de Phillips – Etats-Unis (1961-95) (données source OCDE)

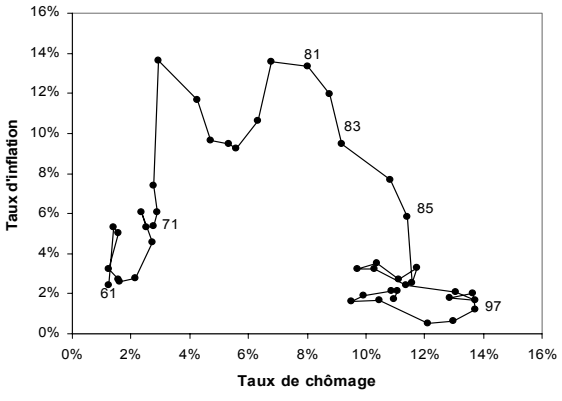


Figure 44 : Courbe de Phillips – France (1961-2005) (données source OCDE)

Pour la France, la seule période où la relation de Phillips semble vérifiée est le début de la décennie 80 (période où la politique de « désinflation compétitive va être menée »).

Si les changements dans les salaires nominaux ne dépendent pas uniquement du taux de chômage, c'est que la hausse des prix peut jouer également un rôle.

Lorsqu'il y a indexation des salaires sur les prix (comme ce fût le cas en France jusqu'en 1983…), l'augmentation passée des prix affecte la hausse future des salaires. Même lorsqu'il n'y a pas d'indexation, les salariés sont enclins à demander des hausses de salaires en cas de hausse des prix afin de ne pas perdre en terme de pouvoir d'achat.

Les mécanismes décrits ci-dessus sont ceux à la base de ce qu'on appelle une « **spirale inflationniste** » (La hausse des prix entraîne des demandes d'augmentation des salaires. Les entreprises pour conserver leur marge et leurs profits augmentent alors de nouveau leurs prix et ainsi de suite…).
Il convient par ailleurs de raisonner en termes d'anticipations du niveau des prix puisque les agents négocient leurs hausses de salaires en fonction de la hausse qu'ils prévoient pour le niveau des prix.
Ceci explique les dérives constatées des courbes de Phillips.

4.5.6.3 La critique de la courbe de Phillips par Phelps (1967) puis Friedman (1968)

Phelps puis Friedman vont apporter le raisonnement théorique qui permet de réfuter le possible arbitrage entre inflation et chômage. Pour ces auteurs, cet arbitrage n'est que temporaire car l'effet d'une mesure gouvernementale (de type politique monétaire expansionniste par exemple) n'a d'effet que parce qu'elle n'avait pas été anticipée par les agents. Lorsque la hausse des prix s'avère durable, les agents intègrent celle-ci dans leurs anticipations et réclament des hausses de salaires ce qui va, bien entendu, engendrer une baisse de la demande de travail des firmes et une hausse du chômage. Une fois que les agents ont intégré cette hausse de prix, la courbe de Phillips « disparaît » et on revient donc vers le taux naturel de chômage.

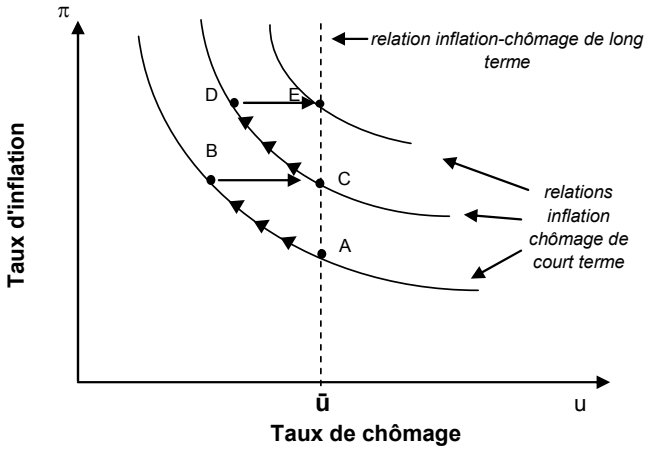


Figure 45 : Dérive de la relation inflation – chômage

Le raisonnement de Friedman est décrit dans le graphique ci-dessus sur lequel plusieurs courbes de Phillips différentes sont tracées. Supposons que l'on soit initialement au point A qui se situe au niveau du taux de chômage « naturel » (le NAIRU) et que le gouvernement souhaite, par une politique de relance de la demande, faire descendre le chômage en dessous de ū. La politique est dans un premier temps efficace (point B) car les salaires présentent une certaine inertie (les prix augmentent plus vite que les salaires donc le salaire réel chute ce qui aboutit à une augmentation de la demande de travail des entreprises et donc à une baisse du chômage).
Les salariés se rendent cependant compte que leur pouvoir d'achat a chuté et ils vont donc demander des augmentations de leurs salaires nominaux afin de retrouver le niveau initial de leur pouvoir d'achat (retour au point C, au niveau de ū, le taux de chômage naturel). Le système ne revient pas au point initial puisque l'inflation n'a pas diminué. Si le gouvernement souhaite d'aventure retenter une politique de relance, l'effet initial sera celui d'un passage de C à D sur la nouvelle courbe de Phillips, plus haute dans le plan (π,u) que la première.
A court terme, Friedman et Phelps reconnaissent bien la possibilité d'une arbitrage entre inflation et chômage mais celui-ci n'est que transitoire car à long terme l'arbitrage n'existe plus : la courbe de Phillips de long terme est verticale (trait pointillé sur notre schéma). Le seul souvenir qui demeure à long terme des politiques de relance, c'est le niveau élevé de l'inflation. Pour Friedman, le choix n'est donc pas entre inflation et chômage, mais entre chômage aujourd'hui et chômage demain.

Si Friedman et les monétaristes reconnaissent encore la possibilité d'un arbitrage à court terme entre inflation et chômage et donc l'efficacité même transitoire des politiques de relance, Lucas et les nouveaux classiques, usant de l'hypothèse « d'anticipations rationnelles » vont parvenir à un résultat encore plus tranché. Comme les agents anticipent rationnellement et non plus de manière adaptative les conséquences des politiques économiques, ils vont anticiper immédiatement la hausse des prix conséquente à ces relances et demander aussitôt des augmentations de salaires. Le salaire réel n'est donc pas influencé par les politiques économiques de relance et celles-ci n'ont donc aucun effet sur le chômage. Pour les nouveaux classiques, la courbe de Phillips n'est pas seulement verticale à long terme, elle l'est également à court terme. L'arbitrage inflation - chômage disparaît même à court terme

pour les théoriciens des « anticipations rationnelles » et donc la politique économique n'a aucun effet, même transitoire.

4.5.7 Soutenabilité de la dette publique

Soit $B_p(t)$ le déficit budgétaire primaire (différence entre les dépenses publiques primaires G(t) (hors charges de la dette) et les recettes publiques T(t)) : $B_p(t) = G(t) - T(t)$. Le déficit budgétaire total B(t) est le déficit budgétaire primaire augmenté des charges d'intérêt sur la dette existante. Soit r(t) le taux d'intérêt et D(t-1) le niveau de la dette publique en début de période.

$$B(t) = B_p(t) + r(t)D(t-1)$$

Si on suppose qu'il n'y a pas de financement du déficit par création monétaire et que les remboursements de la dette publique sont exceptionnels (ressources des privatisations par exemple), l'équation d'accumulation de la dette publique est la suivante :

$$D(t) = D(t-1) + B(t) = B_p(t) + (1 + r(t))D(t-1)$$

On peut exprimer cette équation sous la forme suivante en divisant les deux côtés par le PIB :

$$\frac{D(t)}{Y(t)} = \frac{B_p(t)}{Y(t)} + (1 + r(t))\frac{D(t-1)}{Y(t-1)}\frac{Y(t-1)}{Y(t)}$$

Soit g le taux de croissance du PIB , d(t)=D(t)/Y(t) le ratio dette publique/PIB et $b_p(t)=B_p(t)/Y(t)$ le ratio déficit publique primaire/PIB. L'équation ci-dessus peut être réécrite de la façon suivante :

$$d(t) = b_p(t) + \frac{1+r}{1+g}d(t-1)$$

C'est donc la relation taux d'intérêt- taux de croissance qui est la clé de l'évolution de la dette. Pour un déficit primaire donné, le ratio dette/PIB aura tendance à augmenter si le taux d'intérêt est supérieur au taux de croissance, diminuer ou stagner s'il est inférieur ou égal. Le taux de croissance assure un allègement de la charge de la dette (le problème de l'équilibre budgétaire est donc différent pour une nation et pour un ménage, sauf à envisager que le revenu du ménage lui assure un enrichissement exponentiel). Il faut cependant prendre en compte le déficit primaire, qui peut obliger à emprunter encore et accroître davantage la dette. Un excédent du solde primaire permet au contraire un stabilisation ou une diminution de la dette.

En France, la dette publique représentait en 1980 20% du PIB, 40% en 1992, 56.8% en 2001 et 65.6% début 2005. Cette dette (environ 17000 euros par français) est-elle soutenable ?

La condition de stabilité de la dette est obtenue en imposant d(t)=d(t-1) dans l'équation précédente.

$$b_p = \frac{g-r}{1+g}d$$

Si le taux d'intérêt est supérieur au taux de croissance (cas le plus fréquent), le solde primaire doit être négatif pour assurer la stabilité de la dette : la stabilité de la dette passe par un excédent budgétaire.

4.5.8 Le pacte de stabilité : pourquoi 3% ?

Le traité de Maastricht en 1992 prévoyait, parmi ses critères de convergence, une stabilisation du ratio dette publique/PIB au niveau de 60%. Ce critère de convergence devait éviter que des pays candidats à la monnaie unique aient une politique budgétaire laxiste qui eut affaibli la future monnaie unique. Ce ratio a été maintenu durablement par son inscription au pacte de stabilité et de croissance défini par le conseil européen réuni à Amsterdam en 1997. En mars 2005, un conseil européen réuni à Bruxelles a introduit des modifications importantes concernant ces critères du Pacte de Stabilité que sont : (i) ratio dette publique/PIB<60% et (ii) ratio déficit public/PIB<3%.

Mais au fait, pourquoi 3% pour le ratio déficit/PIB sachant qu'en 1991 60% était la moyenne tendanciellement acceptable communément admise pour le ratio dette/PIB ?

Les taux d'intérêt étaient à cette époque de l'ordre de 10%. Le taux de croissance (en valeur) était en moyenne de 5% en Europe durant la dernière décennie. De sorte que : $b_p = \frac{0.05 - 0.1}{1.05}0.6 = -0.03$.

Evident, non ? Il s'agit d'un excédent primaire de 3% de points de PIB mais comme l'encours de dette de 60% du PIB génère une charge d'intérêt, et donc un solde déficitaire de 0.6*0.1=-0.06, on obtient un déficit budgétaire égal à la somme de l'excédent primaire (+0.03) et de la charge de la dette (-0.06) de -0.03 points de PIB. C'est ainsi qu'a été défini le déficit public jugé compatible avec les accords de Maastricht puis du Pacte de Stabilité et de croissance.

5 Le chômage

En Europe, particulièrement en France, une part importante de la population active est dépourvue d'emploi (8% en cet été 2007). Cette situation est alarmante car le chômage de masse dans certains grands pays européens, notamment en France, est un phénomène persistant et ne peut être uniquement imputé à des dérèglements conjoncturels. Les politiques économiques depuis deux décennies ont pourtant souvent été tournées vers la lutte contre le chômage. Ainsi, le Président F. Mitterrand déclarait déjà en 1993 : « *en matière de chômage, on a tout essayé* ». Il est vrai qu'un certain nombre de mesures pour l'emploi ont été mises en place, avant mais également après cette phrase de dépit (hausse puis…baisse des dépenses publiques ; réduction de la durée du travail, utilisation massive des préretraites ; augmentation du nombre de diplômés du baccalauréat ; développement des travaux d'utilité collective ; promotion des emplois jeunes ; diminution des charges sociales des entreprises…). Et pourtant le chômage reste à un niveau élevé.

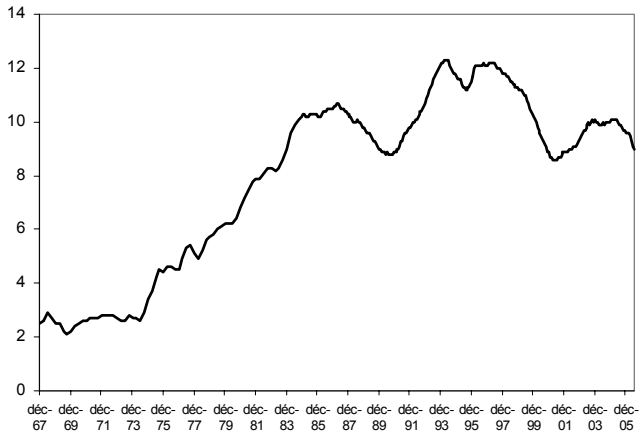


Figure 46 : Taux de chômage en France (déc.67-juin2006) en %

En arrière-plan de ces politiques, on peut distinguer deux grandes théories: (i) la théorie keynésienne qui part du postulat que le chômage résulte d'une insuffisance de la demande (et donc des débouchés pour les entreprises) et qu'il est donc involontaire, (ii) la théorie classique qui pense que la persistance du chômage provient de l'absence de flexibilité des salaires car sinon, le marché permettrait automatiquement l'ajustement de l'offre et de la demande de travail. Cette approche classique en termes d'équilibre ne laisse la place qu'à du chômage dit volontaire (les agents ne veulent pas travailler au niveau de salaire proposé).
La théorie keynésienne présente l'inconvénient de ne pas préciser les origines des rigidités nominales qui occasionnent l'insuffisance de débouchés de sorte qu'on peut raisonnablement se demander si ces rigidités ont une raison de perdurer dans le long terme. L'approche keynésienne semble ainsi peu adaptée pour comprendre la persistance du chômage[76]…
Cependant, la théorie classique ne fournit pour sa part pas réellement d'explication aux dysfonctionnements du marché du travail. Si elle situe le problème dans l'absence de flexibilité à la baisse des salaires, la théorie classique ne donne pas de pistes pour appréhender l'origine de la rigidité des salaires (étude nécessaire des comportements des acteurs sociaux, de leur environnement institutionnel…).

Les échecs des théories « traditionnelles » à expliquer le fonctionnement du marché du travail débouchent sur l'obligation de recourir à des modélisations non concurrentielles du fonctionnement du marché du travail. Dans ces théories plus récentes (théorie du salaire d'efficience, théorie des négociations salariales, théorie de le recherche d'emploi,…) les salaires ne sont plus déterminés par

[76] Certains travaux empiriques montrent par ailleurs que le taux de chômage observé est peu différent du taux de chômage que l'on qualifie de naturel (le NAIRU : *Non Accelerating Inflation Rate of Unemployment*, voir l'annexe sur la **courbe de Phillips** du chapitre précédent) ce qui signifie que le chômage résulte pour partie de dysfonctionnements du marché du travail et non d'une insuffisance conjoncturelle de la demande.

un processus concurrentiel abstrait, comme dans le modèle classique traditionnel, mais par les agents économiques eux-mêmes (négociations collectives, politique de ressources humaines…) sur des bases rationnelles. Le salaire apparaît supérieur au salaire qui résulterait de l'équilibre concurrentiel entre offre et demande de travail, ce qui conduit à la présence d'un chômage d'équilibre.
Nous allons cependant constater en examinant quelques faits stylisés du marché de l'emploi, qu'il est également nécessaire de prendre en compte la dimension dynamique des comportements (notamment le processus incessant de créations et de destructions d'emplois) pour ne pas se tromper de remède, c'est-à-dire pour choisir la politique de l'emploi la plus efficace.

5.1 Quelques définitions

La définition du chômage repose sur celle du Bureau International du Travail (BIT). Pour être considéré comme chômeur, trois conditions sont nécessaires : (i) être sans emploi, (ii) être disponible pour occuper un emploi, et (iii) être à la recherche d'un emploi.
La définition du Bureau International du Travail ne fait référence à aucun critère juridique ou institutionnel tel que la perception d'une allocation ou l'inscription sur une liste administrative. Le chômage est plus une situation de fait que de droit. Cette définition repose à la fois sur un critère de situation (être sans travail). Elle repose également sur un critère plus subjectif (être à la recherche d'un travail). Enfin, seul le travail rémunéré est pris en compte (activité salariée ou non salariée (indépendant)). Le chômage n'est donc pas une absence de travail, mais d'emploi rémunéré.
Le schéma ci-dessous permet en fonction de la réponse à deux critères (occuper un emploi, être à la recherche d'un emploi) de définir les trois situations possibles pour une personne appartenant à la population en âge de travailler : actif occupé, chômeur ou inactif. La population active est la somme des actifs occupés et des chômeurs. Le taux de chômage est déterminé par le ratio du nombre de chômeurs à la population active. Le taux d'activité est le rapport de la population active (chômeurs + actifs occupés) à la population en âge de travailler (15-64 ans). Le taux d'emploi est obtenu par le ratio des actifs occupés à la population en âge de travailler.

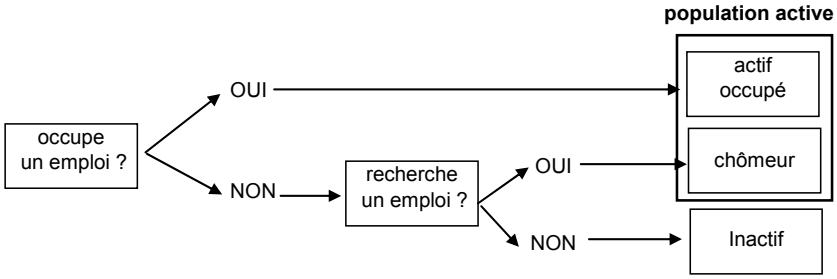


Figure 47 : Chômeur, actif occupé, inactif…

En France, la population active était de 27,6 millions d'individus en 2005, le nombre de chômeurs (au sens du BIT) d'un peu plus de 2,7 millions. Le taux d'activité des femmes était de 64,8%, celui des hommes de 74,5%, c'est-à-dire que la population en âge de travailler est de l'ordre de 40 millions d'individus. La France est l'un des pays qui connaît le plus faible taux d'activité, et cela malgré une participation croissante des femmes au marché du travail. La faiblesse du taux d'activité en France est en partie imputable à la faiblesse du taux d'emploi chez les « seniors » (le taux d'emploi des 55-64 ans était de 37,3% en France en 2004 contre 69,1% en Suède ou 56,8% au Royaume-Uni, 41,8% en Allemagne).
Le nombre de chômeurs constitue le stock de chômage à un instant donné, mais ce stock varie d'une période à l'autre en fonction des flux d'entrées et de sorties, qui sont, on va le voir dans la prochaine section, importants.
Le schéma ci-dessous indique d'une part les frontières floues entre inactivité, chômage et emploi. Un certain nombre de personnes sont dans une situation intermédiaire. On voit apparaître des situations de sous-emploi lorsque des personnes occupent des emplois (à temps partiel) mais cherche à travailler encore plus. D'autre part, le schéma indique les différents flux entre les trois catégories définies. On constate donc immédiatement que la création d'un emploi ne s'accompagne pas forcément la diminution d'une unité du stock de chômeurs. En effet, il existe ce qu'on appelle le **phénomène de flexion du taux d'activité**. Les emplois créés ou proposés peuvent attirer des personnes qui jusque-là n'étaient pas considérées comme chômeurs (des inactifs voyant la situation s'améliorer sur le marché du travail ou une opportunité intéressante décident alors de devenir actifs employés). Lorsque des emplois sont créés, certains vont être occupés par des anciens chômeurs d'autres par des inactifs, ce qui contribue à faire augmenter le taux d'activité.

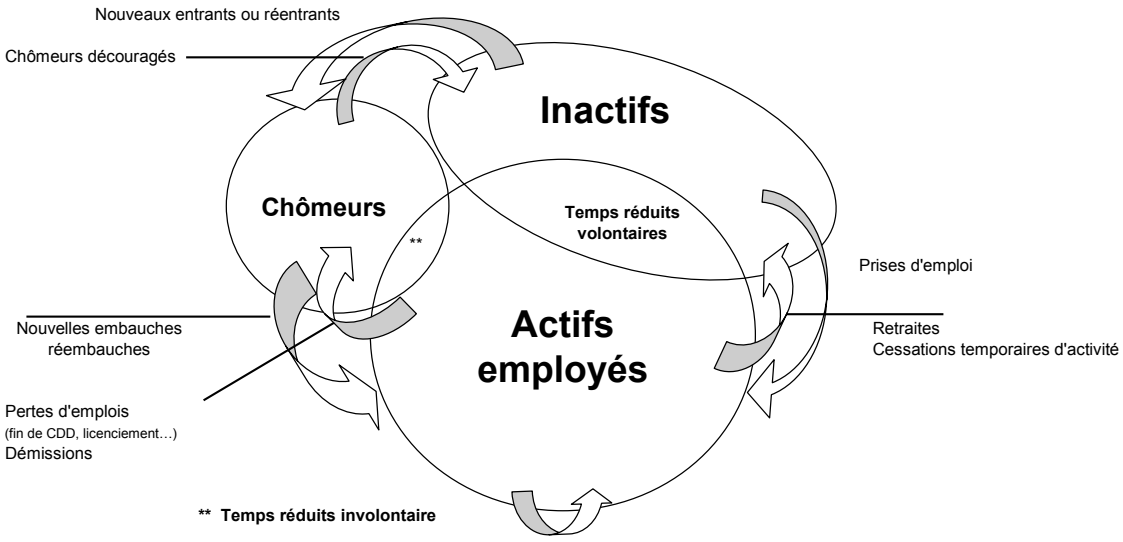


Figure 48 : Stocks et flux de chômage, d'actifs employés et d'inactifs

Pour mieux cerner la réalité du chômage, il convient également de s'intéresser à l'ancienneté moyenne des chômeurs, afin de savoir si le chômage est plutôt un chômage de longue durée (synonyme de déqualification et de précarité) ou plutôt un chômage conjoncturel ou frictionnel.

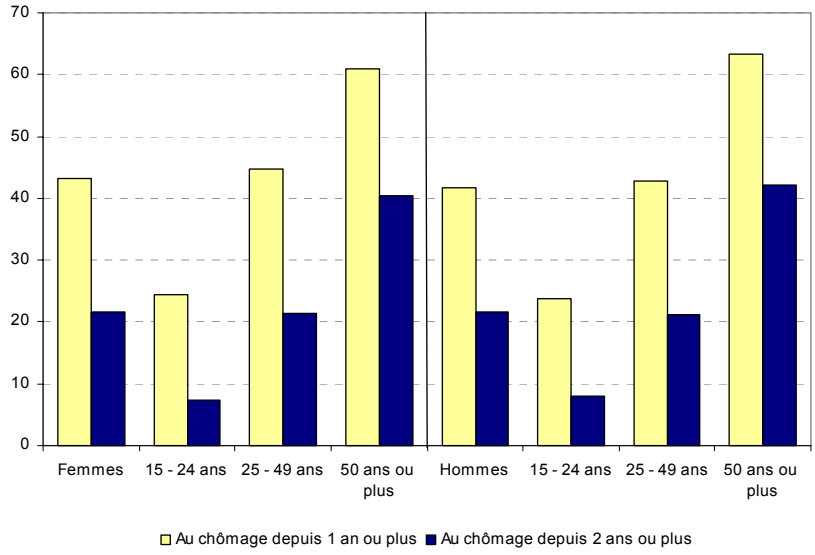


Figure 49 : Proportion du chômage de longue durée parmi les chômeurs selon l'âge et le sexe (France, 2005)

Cette prise en compte de l'ancienneté du chômage permet de se rendre compte de l'inégalité face au risque de chômage et du cloisonnement du marché du travail selon les caractéristiques des individus.
En définissant la **vulnérabilité** sur le marché du travail comme le risque de tomber au chômage (ratio du nombre de personnes au chômage depuis moins d'un mois à la population active occupée) et l'**employabilité** comme la probabilité de sortir du chômage (ratio du nombre de chômeurs ayant moins d'un an d'ancienneté au nombre total de chômeurs), on constate la différence de situation entre les travailleurs jeunes et les travailleurs âgés. Les jeunes se caractérisent par une forte vulnérabilité mais également par une forte employabilité. Au contraire les travailleurs âgés s'illustrent par une faible vulnérabilité (certaines dispositions législatives y veillent[77]) mais également par une faible employabilité. On constate ainsi sur la figure ci-dessus la proportion forte de chômeurs de longue durée parmi les chômeurs de plus de 50 ans, alors que cette proportion est beaucoup plus faible chez les 15-24 ans.

[77] La contribution Delalande est une taxe sur le licenciement des travailleurs âgés, donc les effets sont controversés puisque la perspective de devoir acquitter cette contribution peut dissuader l'embauche de travailleurs âgés.

5.2 La réalité du marché du travail

Nous allons présenter rapidement quatre caractéristiques empiriques importantes du marché du travail :

- Le **chômage** en Europe résulte davantage de l'**insuffisance de la création d'emplois** que de l'accroissement de la population active. Le taux d'emploi n'a cessé de baisser en Europe durant les trois décennies 60'/70'/80' alors qu'il augmentait aux Etats-Unis, et que la tendance s'inversait à partir des années 80 au Japon. En France la population active s'est accrue du fait de l'évolution démographique et de la montée du taux d'activité des femmes. Mais la croissance des emplois n'a pas été suffisante pour absorber ce surcroît de la population active, d'où la montée du chômage. En partant de l'égalité suivante : **PAT x ta = U + N** (avec PAT : population en âge de travailler, ta : taux d'activité, U : nombre de chômeurs, N : nombre d'emplois), on obtient l'expression suivante pour la variation du taux de chômage (tc) : $\Delta tc = g_{PAT} + g_{ta} - g_N$ (où g_X est le taux de croissance de la variable X).

Taux annualisés	Δtc	g_{PAT}	g_{ta}	g_N
1960-69	-0,1	0,5	-0,4	0,2
1970-79	0,3	0,7	0,0	0,4
1980-89	0,4	0,7	0,1	0,5

Emploi, activité et chômage en Europe

Taux annualisés	Δtc	g_{PAT}	g_{ta}	g_N
1960-69	-0,2	1,6	0,1	1,9
1970-79	0,3	1,8	0,8	2,3
1980-89	0,0	1,0	0,7	1,7

Emploi, activité et chômage en Amérique du Nord

Taux annualisés	Δtc	g_{PAT}	g_{ta}	g_N
1960-69	-0,1	1,8	-0,4	1,5
1970-79	0,2	1,0	0,0	0,8
1980-89	0,0	0,9	0,2	1,1

Emploi, activité et chômage au Japon

Le taux d'emplois constitue un indicateur de la création nette d'emplois rapportée à la population en âge de travailler. Jusque dans les années quatre-vingt, on a assisté à une baisse du taux d'emploi à la fois en Europe et au Japon mais ensuite seul l'Europe a continué à avoir une croissance négative de son taux d'emploi ($g_N - g_{PAT}$).

Un regard sur l'évolution comparée des créations nettes d'emplois et de l'augmentation nette de la population active entre trois pays européens sur la période 1985-2004, montre bien que le problème de chômage à la fois en France et en Allemagne résulte de l'incapacité à créer des emplois. Le Royaume-Uni avec une croissance de la population active comparable à celle de la France a su créer les emplois nécessaires pour faire chuter son taux de chômage.

1985-2004	Créations nettes d'emplois	Augmentation nette de la population active	Variation du taux de chômage	Taux de chômage 2004
France	3,1	3,5	0,4	9,7
Allemagne	2,2	3,5	2,6	9,6
Royaume-Uni	4,6	3,9	-2,4	4,6
Japon	4,7	6	1,8	4,7
Etats-Unis	29,6	30	-1,1	5,5

- Les études empiriques consacrées à l'influence du chômage sur le salaire réel concluent à une élasticité du taux de salaire réel par rapport au chômage négative mais faible (élasticité de l'ordre de -0,1). Or d'après le modèle néo-classique traditionnel (modèle offre/demande de travail et ajustement automatique par le prix (le salaire réel) considéré comme flexible), cette élasticité devrait être importante de manière à permettre les ajustements sur le marché de l'emploi. Les résultats empiriques suggèrent que l'offre de travail des femmes est faiblement croissante par rapport au salaire réel tandis que l'offre de travail des hommes n'en dépend pas significativement. Au niveau agrégé, on a donc une **très faible élasticité de l'offre de travail par rapport au salaire**. Nous verrons que cette caractéristique oblige à prendre en compte des fonctionnements non-concurrentiels du marché du travail et de la fixation des salaires.

- Le marché de l'emploi est caractérisé par d'**importantes réallocations d'emplois et mouvements de main-d'œuvre**. En matière de réallocations d'emplois en France, à l'échelle de la nation, chaque année, environ 15% des emplois disparaissent et chaque année environ 15% d'emplois nouveaux apparaissent (chaque jour ouvrable 10 000 emplois disparaissent et *grosso modo*, 10 000 emplois sont créés !). Les flux de main-d'œuvre sont nettement supérieurs à ceux des emplois, car en plus des entrées et des sorties associées aux créations et destructions d'emplois, s'ajoutent les remplacements sur un même poste de travail. En matière de mouvement de main-d'œuvre, chaque jour ouvrable, alors que 10 000 emplois sont détruits, 30 000 personnes quittent leur emploi et… 30 000 personnes en retrouvent un. En moyenne, les entreprises qui créent un emploi, embauchent trois travailleurs et se séparent de deux autres. Plus étonnant, les entreprises qui détruisent des emplois continuent d'embaucher. La perte d'un poste de travail, se traduit en moyenne par deux embauches et trois départs. 70% des embauches se font en CDD et seulement un tiers est transformé en emploi stable. Les réallocations d'emplois ont principalement lieu entre établissements appartenants au même secteur. Avec un découpage du système productif en 600 secteurs, les mouvements d'emplois entre secteurs représentent moins de 20% du total des réallocations.

	Taux de rotation	Taux d'entrée	Taux de sortie	Taux d'évolution
Industrie	19,1	18,3	20,0	-1,6
Construction	22,5	22,9	22,0	0,9
Transports, commerces, services	52,1	52,5	51,8	0,7
Ensemble	**40,4**	**40,5**	**40,4**	**0,0**

Tableau 4 : Mouvements de main d'œuvre par grands secteurs d'activité (2004)

En % de la population active dans ces secteurs. Le taux de rotation est la moyenne des taux d'entrée et de sortie

- Ce sont généralement **les travailleurs les moins qualifiés** qui **sont le plus victimes du chômage.** L'accroissement du taux de chômage des travailleurs les moins qualifiés est dû à une modification de la structure des emplois (voir le tableau ci-dessous). En France, le taux de chômage des ouvriers et des employés est passé de moins de 2% en 1968 à 13% en 1992, tandis que le taux de chômage des cadres et professions intellectuelles supérieures passait de 1 à 3% sur la même période. Or, malgré l'augmentation du nombre de cadres et de professions intermédiaires les deux CSP les plus nombreuses en France restent les employés et les ouvriers. Le développement du progrès technique dans les différentes activités productives impose le recours de plus en plus fréquent à des machines qui remplacent la main d'œuvre non-qualifiée et qui requièrent du personnel qualifié pour leur entretien et leur utilisation. Les activités industrielles qui employaient une forte main-d'œuvre peu qualifiée ont eu tendance à disparaître (mines, textile…), de telle sorte que parmi la population ouvrière le nombre d'ouvriers qualifiés a diminué au profit des ouvriers qualifiés (voir tableau ci-dessous).

	emplois en milliers				évolution 1982-2004	
	1982	1990	1999	2004	%	en milliers
Cadres et professions intellectuelles supérieures	1 860	2 603	2 996	3 566	+ 92	1 706
Professions intermédiaires	3 784	4 464	5 268	5 771	+ 53	1 987
Employés	5 502	5 899	6 599	7 114	+ 29	1 612
Ouvriers	7 044	6 546	5 794	6 127	- 13	- 917

	nombre en milliers				parts en %	
	1982	1990	1999	2004	1982	2004...
Ouvriers qualifiés	3 686	3 725	3 497	3 981	52	65
Ouvriers non qualifiés	3 069	2 586	2 163	1 865	44	30
Ouvriers agricoles	269	236	245	281	4	5

Tableau 5 : Déformation de la structure des emplois en France (1982-2004)

Progression des CSP qualifiées (tableau du haut)

Progression de la part des ouvriers qualifiés parmi les emplois ouvriers (tableau du bas)

En résumé, quatre faits stylisés ont été mis en évidence qui devraient pourvoir être décrits ou pris en considération par un bon modèle explicatif de la situation sur le marché de l'emploi. Ces quatre faits stylisés sont les suivants : (i) l'importance du chômage en Europe ne provient pas d'une croissance de la population active plus rapide mais d'une insuffisance de la création d'emplois, (ii) le salaire réel est peu sensible aux variations du taux de chômage et l'offre de travail varie peu avec le salaire réel, (iii) les mouvements de main-d'œuvre et les réallocations d'emploi sont très importants, et (iv) le chômage touche prioritairement les travailleurs les moins qualifiés.
Nous allons maintenant voir de quelles façons les recherches les plus récentes dans le domaine de l'économie du travail ont cherché à expliquer la rigidité des salaires réels et l'impact des mouvements de main-d'œuvre sur le chômage.

5.3 Les représentations du fonctionnement du marché du travail

5.3.1 La représentation néo-classique du marché du travail

Dans le modèle traditionnel[78] (économie néo-classique) l'offre et la demande de travail s'ajustent grâce à la flexibilité parfaite du salaire réel[79]. En l'absence de contraintes pesant sur le salaire w, celui-ci est toujours en mesure d'égaliser offre et demande de travail, de telle sorte que le modèle néo-classique traditionnel décrivant le fonctionnement du marché du travail est dans l'incapacité d'expliquer l'existence de **chômage involontaire**, c'est-à-dire d'une situation dans laquelle des agents désireux de travailler pour un salaire plus faible ou égal au salaire d'équilibre ne parviennent pas à trouver d'emploi.
A l'équilibre du marché du travail, il y a plein emploi. La productivité marginale du travail est égale au salaire réel (programme du producteur) et il y a également égalité entre taux marginal de substitution consommation/loisir et salaire réel (programme du consommateur).
Le seul chômage envisageable dans cette « vision néo-classique » est du **chômage volontaire**, c'est-à-dire une situation dans laquelle le travailleur refuse d'offrir son travail en dessous d'un taux de salaire minimal, que l'on qualifie de salaire de réservation. Seule la rigidité des salaires en dessous du salaire de réservation peut donc expliquer l'émergence de sous-emploi de nature volontaire (voir la figure ci-dessous).

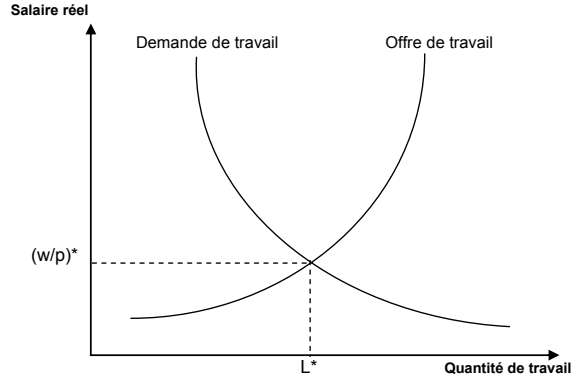


Figure 50 : Equilibre du marché du travail en concurrence parfaite

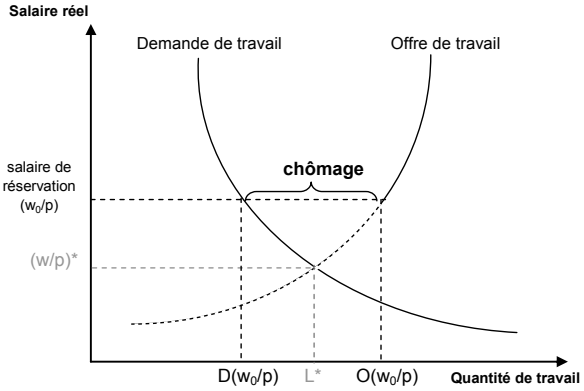


Figure 51 : Chômage volontaire

Une explication possible à la rigidité salariale est l'existence d'indemnités versées aux travailleurs sans emplois. Le salaire de marché doit donc compenser la désutilité du travail, ce qui explique que les ménages ne désirent pas offrir de travail pour des salaires inférieurs au salaire de réservation (d'où la croisade des plus libéraux contre le SMIC, responsable à leurs yeux d'une partie du chômage). Cette explication ne tient cependant pas compte de la rationalité individuelle. Le salaire de réservation est une hypothèse plutôt *ad hoc* qui n'a pas de justification.
Les économistes « classiques » ont cherché un fondement à l'existence du salaire de réservation dans le processus de recherche d'emploi (**théorie de la prospection d'emploi**). Le sous-emploi qui résulte de la recherche d'emploi a un caractère de chômage frictionnel (il correspond au taux de chômage « naturel » de Friedman, *cf.* l'annexe à la fin du précédent chapitre sur **la courbe de Phillips**). Le chômeur possède des informations imparfaites sur les caractéristiques des emplois

[78] Voir l'annexe **le modèle néo-classique du marché du travail** en fin de chapitre.
[79] Pour comprendre pourquoi le prix du travail est le salaire réel, voir l'annexe sur la détermination des demandes de facteurs (**le programme du producteur**) en annexes générales.

vacants (en matière de salaire et de stabilité de l'emploi). A chaque période successive de sa recherche d'emploi, il reçoit des offres et il calcule l'espérance d'utilité intertemporelle associée à un emploi offert (on fait l'hypothèse que le chômeur possède la distribution des salaires dans l'économie). Le chômeur réalise donc un arbitrage à chaque période entre le gain si il accepte l'emploi proposé et le gain associé au refus de cet emploi (l'espoir de trouver un emploi mieux rémunéré à la prochaine période). La décision prend la forme d'un salaire à partir duquel le chômeur accepte un emploi. C'est le salaire de réservation. On voit bien comment le refus d'un emploi à un salaire inférieur à ce « salaire de réservation » est une décision rationnelle puisque le chômeur se base sur un arbitrage qui maximise son utilité.

L'un des travers les plus importants du modèle traditionnel néo-classique du marché du travail est le paradoxe suivant. Les études empiriques ont montré (*cf.* notre deuxième fait stylisé) que l'offre de travail est verticale (peu sensible au salaire réel). En période de récession, lorsque la demande de travail baisse, on devrait donc assister à une chute importante du taux de salaire réel. Or ce n'est pas le cas (durant les années quatre-vingts, les salaires réels ont augmenté avec la montée du chômage aux Etats-Unis par exemple). Comment expliquer cela ? Comment concilier les changements observés en matière d'emploi et de salaire ?

La seule interprétation possible est que la courbe de demande de travail se déplace sans variation du salaire réel (pas de mouvement de la courbe d'offre), c'est-à-dire qu'il existe une certaine rigidité du salaire réel et que cette rigidité du salaire réel est source de chômage.

5.3.2 « Chômage classique » et « chômage keynésien »

Pour Keynes, le chômage ne résulte pas de dysfonctionnement sur le marché du travail mais d'une insuffisance de débouchés que les firmes anticipent[80]. Les solutions pour rétablir le plein-emploi sont donc à trouver dans une stimulation de la demande plutôt que dans une baisse des salaires qui ne peut au contraire qu'aggraver la situation en rognant sur le pouvoir d'achat des salariés. Les « théoriciens du déséquilibre », dont l'économiste français E. Malinvaud, ont proposé une représentation de l'économie à prix et salaires fixes, permettant de dégager une typologie du chômage. L'offre et la demande diffèrent sur les marchés (du travail et des biens) et le marché est dans l'incapacité d'assurer le retour à l'équilibre du fait de la rigidité des prix (situation de « déséquilibre » ou « d'équilibre à prix fixes » puisque ces situations sont stables). Pour E. Malinvaud, deux cas principaux de chômage peuvent se présenter, correspondant à deux situations de déséquilibre.

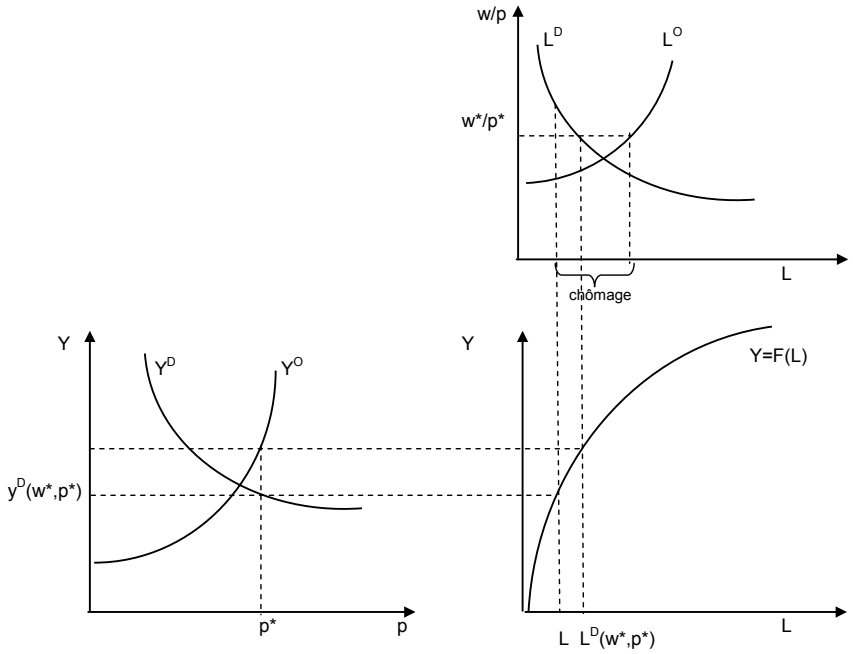


Figure 52 : Le chômage « keynésien »

Le chômage « keynésien » est caractérisé par le double déséquilibre suivant. Sur le marché des biens et services, l'offre est supérieure à la demande (cas d'insuffisance de la demande effective). Sur le

[80] Voir l'annexe dans le chapitre Conjoncture et politiques économiques sur **le chômage keynésien**.

marché du travail, l'offre est supérieure à la demande puisqu'en raison de la surproduction, les entrepreneurs réduisent leur demande de travail.
Ce sont donc les entreprises qui sont contraintes sur le marché des biens. La règle du « côté court » (on ne peut obliger quelqu'un à acheter un bien qu'il ne désire pas ou qu'il ne veut pas acheter à un prix qu'il considère comme trop élevé, ou inversement on ne peut forcer quelqu'un à vendre un bien à un prix qu'il considère trop bas, de sorte que le côté le moins enclin à l'échange détermine l'équilibre du marché) implique donc une contrainte de débouchés subie par la firme. L'insuffisance de la demande aboutit à une demande d'emploi inférieure au niveau de plein emploi (entendu comme l'équilibre entre les courbes d'offre et de demande de travail). L'entreprise embauche donc une quantité de travail inférieure à cette demande de travail puisque l'entreprise ne va embaucher que la quantité juste suffisante pour produire la demande qui lui est adressée (quantité L sur la figure ci-dessus).
Il y a donc un report du déséquilibre du marché des biens (où c'est l'entreprise qui est rationnée) vers le marché du travail, ce qui réduit encore le niveau d'emploi effectif en dessous de la demande de travail.

Le chômage « classique » est également caractérisé par un double déséquilibre. Sur le marché des biens et services, l'offre est inférieure à la demande (cas où les entrepreneurs réduisent leur offre en raison de coûts trop élevés). Sur le marché du travail, l'offre est supérieure à la demande. La demande de travail des entrepreneurs est réduite du fait du coût de la main d'œuvre jugé trop élevé.

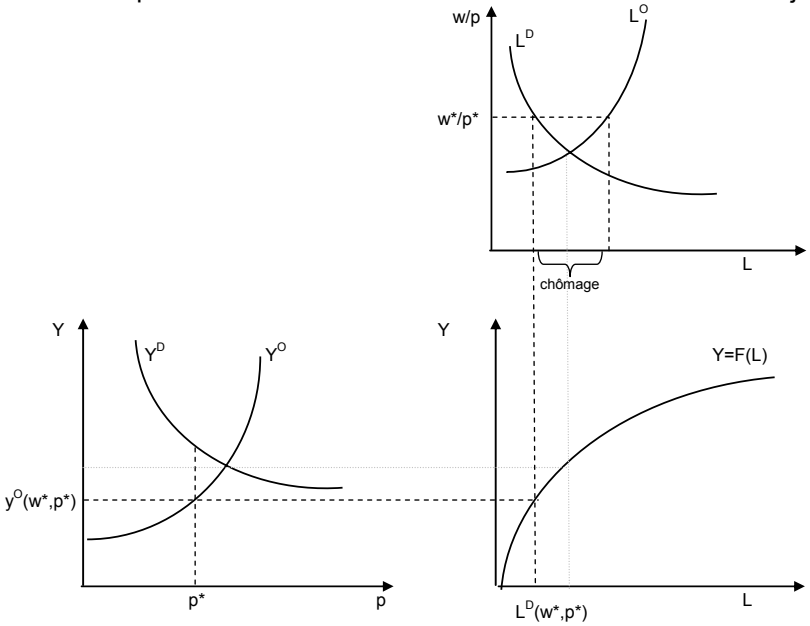


Figure 53 : Le chômage « classique »

Le cas du chômage classique correspond donc à un double rationnement des ménages qui sont contraints sur le marché des biens ($Y^D>Y^O$) et sur le marché de l'emploi ($L^O>L^D$). Si on autorisait la flexibilité des prix (c'est-à-dire que l'ajustement du salaire réel était possible, sous la forme d'une baisse des salaires nominaux et/ou d'une augmentation du prix des biens), la baisse du coût du travail réduirait la demande de biens tout en augmentant la demande de travail.

Le chômage classique peut être résorbé par un retour à la flexibilité des prix tandis que la résorption du chômage keynésien passe par la mise en œuvre d'interventions publiques destinées à augmenter la demande. La théorie du déséquilibre conclut donc à la validité des solutions keynésiennes lorsque le chômage est dû à l'insuffisance de la demande. L'analyse des théoriciens du déséquilibre attire donc l'attention des gouvernements sur le risque d'erreur de diagnostic en matière de chômage.
On va retrouver les notions de rigidité et de chômage d'équilibre dans les théories les plus récentes cherchent à lier l'existence du chômage à des mécanismes non concurrentiels sur le marché de l'emploi.

5.3.3 Le chômage d'équilibre et les approches récentes

Un paradoxe du modèle traditionnel néo-classique de représentation du marché du travail est la difficulté de celui-ci à décrire les changements observés en matière d'emploi et de salaire. Lorsque la demande de travail de la part des entreprises baisse (en période de récession), le modèle néo-classique traditionnel prévoit que le salaire réel va diminuer. Or, on ne constate pas dans la réalité de diminution du salaire concomitante à la chute de la demande de travail. Ainsi le pouvoir d'achat (et donc les salaires réels) a progressé en France durant les années quatre-vingts alors que le chômage ne cessait de grimper.

Comment cela peut-il s'expliquer ? L'ajustement salarial est trop faible ou inexistant pour permettre l'égalisation de la demande et de l'offre de travail. Les théories récentes en matière de marché du travail ont cherché à expliquer cette rigidité du salaire réel qui explique la persistance du chômage à l'équilibre.

Deux explications complémentaires ont émergé. D'une part, les entreprises ne peuvent pas baisser les salaires en raison du mode même de fixation des salaires et du fait de la pression syndicale. D'autre part, l'entreprise peut elle-même être amenée à proposer des salaires se situant au-dessus du salaire courant d'équilibre afin de motiver ses salariés. C'est ce qu'on appelle un salaire d'efficience.

5.3.3.1 Les modèles de négociations salariales et de salaire d'efficience

Le **modèle** théorique **de négociations salariales** et le **modèle de salaire d'efficience** sont présentés en annexe de ce chapitre. Revenons sur leur conclusion. La **théorie du salaire d'efficience** met l'accent sur l'asymétrie d'information qui existe entre un salarié et l'entreprise qui l'embauche. L'entreprise ne connaît pas la productivité de ce travailleur. Elle ignore si ce sera un salarié dévoué ou un « tire au flanc ». Le salarié lui a une certaine désutilité au travail (il a du abandonné du temps de loisir pour offrir du travail, *cf*. le modèle traditionnel néo-classique). Il peut donc être tenté de ne pas offrir l'effort maximum lorsqu'il est au sein de l'entreprise afin de gagner le même salaire mais sans la désutilité du travail. Le moyen pour l'entreprise d'inciter ce salarié à fournir l'effort maximal est de lui proposer un salaire supérieur à celui qu'il peut trouver dans les autres entreprises. Si le salarié est surpris à ne rien faire et qu'il perd son travail, il va donc perdre ce supplément de salaire qu'il avait par rapport aux salariés travaillant dans les autres entreprises ne pratiquant pas le salaire d'efficience. Il va donc fournir l'effort requis par l'entreprise. Si un nombre important d'entreprises agit de la sorte afin de motiver leurs salariés, le salaire d'équilibre qui en résulte est supérieur à celui qui permettrait l'équilibre de l'offre et de la demande de travail. Il y a donc du chômage mais c'est un chômage involontaire. Il existe certainement des chômeurs prêts à travailler pour un salaire inférieur à celui qui est pratiqué, mais il n'est pas optimal pour l'entreprise de les embaucher, car si elle les paye à un salaire inférieur au salaire d'efficience, ils ne seront pas incités à fournir l'effort maximum. On voit donc apparaître un chômage à l'équilibre qui résulte de comportement de maximisation de la part des agents. La rente informationnelle du salarié lui permet d'obtenir un supplément de salaire mais cela au détriment de l'emploi.

La **théorie de la négociation salariale** part du constat qu'une majorité des contrats de travail résulte de conventions collectives au niveau de la branche (85% des contrats de travail en France). Les contrats de travail sont donc négociés entre les institutions représentatives que sont les syndicats et le patronat. Le résultat de ce modèle (voir les détails en annexe) est que les salariés vont trouver un intérêt à l'action syndicale dès lors que celle-ci permet d'élever le salaire au-dessus du niveau du salaire d'équilibre constaté sur le marché du travail. Les entreprises qui négocient avec les syndicats le niveau des salaires conservent cependant le pouvoir dans la détermination de la quantité de travail qu'elles vont utiliser. La hausse des salaires au-dessus du niveau qui permettrait d'équilibrer l'offre et la demande de travail, va donc se traduire par une moindre demande de travail de la part des entreprises. La théorie des négociations salariales montre donc que si l'action syndicale est bénéfique pour les « *insiders* » (les travailleurs en place dans les entreprises), elle se fait au détriment des « *outsiders* » (les chômeurs)[81].

[81] Pour plus de précisions sur le modèle de Lindbeck et Snower (1986), voir Lindbeck A., Snower D. (2001), « Insiders versus Outsiders » *Journal of Economic Perspectives*, vol. 15(1), pp. 165-188. Il convient cependant d'avoir à l'esprit que la mission des syndicats n'est pas seulement l'obtention de salaires élevés. Le syndicat défend également les intérêts des salariés dans de nombreux domaines, veille à la sécurité, à l'hygiène dans l'entreprise, au respect des conditions de travail… De plus, la hausse des salaires n'est pas forcément défavorable à l'emploi si les suppléments de salaires versés aboutissent à une plus forte demande adressée aux entreprises.

La formation non concurrentielle des salaires, telle que décrite dans les théories du salaire d'efficience et de la négociation salariale, conduit généralement à du chômage. Les hypothèses des théories du salaire d'efficience et de la négociation salariale aboutissent à une relation décroissante entre le salaire et le chômage. La courbe WS du modèle WS-PS résume cette relation (voir en annexe la présentation du **modèle WS-PS**) décroissante entre taux de chômage et taux de salaire. Confrontée à la courbe déterminant le comportement de l'entreprise (la courbe PS), elle permet de déterminer les taux de chômage et de salaire d'équilibre. On peut alors examiner les effets de différentes mesures en matière de chômage et de salaire.

5.3.3.2 Le modèle d'appariement et la courbe de Beveridge

Le recrutement est un processus coûteux pour les entreprises. Il convient donc de prendre en compte ces coûts de recrutement qui peuvent s'avérer importants puisque les mouvements de main-d'œuvre sont nombreux (cf. le troisième fait stylisé décrit en début de chapitre). Par ailleurs les entreprises ne peuvent pas instantanément pourvoir leur emploi même lorsqu'il existe des chômeurs. Les coûts de recrutement ou les entraves à l'appariement entre les chômeurs et les emplois vacants peuvent constituer une source importante de chômage (voir l'annexe sur **la courbe de Beveridge** et celle sur **coût d'appariement et chômage d'équilibre**). Il convient en effet de distinguer les effets sur le chômage des chocs économiques agrégés (modifications des variables économiques) des chocs de réallocation (les modifications dans le processus d'appariement). Le modèle d'appariement, décrit en annexe, confronte la courbe de Beveridge, qui décrit l'efficacité du processus d'appariement entre emplois vacants et chômeurs, et la courbe VS qui décrit le comportement d'offre d'emplois par des entreprises confrontées à des coûts de recrutement. Ce modèle montre bien que lorsque des entraves à un fonctionnement efficace du marché de l'emploi existent, des chocs de réallocation peuvent se surajouter aux chocs agrégés et occasionner la persistance du chômage même lorsqu'il y a amélioration sur le plan conjoncturel.
Les dégradations de l'efficacité du processus d'appariement semblent avoir joué un rôle significatif dans l'augmentation du chômage en Europe.
La prise en compte de la difficulté de l'appariement entre offre et demande de travail permet ainsi de dégager une typologie du chômage qui montre bien la différence de situation entre les Etats-Unis et l'Europe. Aux Etats-Unis, l'effet des chocs agrégés est dominant à court terme, expliquant en grande partie les fluctuations du chômage mais, dans le long terme, seul persiste l'impact des chocs de réallocation. Au contraire, en France et en Allemagne, la dérive cumulative du chômage s'explique par les chocs agrégés tandis que les chocs de réallocation ont également des effets à court terme. Exprimé différemment, cela signifie que le chômage est plutôt « conjoncturel » aux Etats-Unis. Il suffit d'un retour de la croissance pour qu'il diminue. En Europe (particulièrement en France), le chômage est de nature plus « structurelle » et il est régulièrement aggravé par la conjoncture. Nous allons voir que cette distinction est importante lorsqu'il s'agit de choisir les remèdes efficaces en matière de lutte contre le chômage.

5.4 Les politiques de l'emploi

5.4.1 Un bref état de la situation

Okun, conseiller économique du Président Kennedy, avait dans les années soixante mis en lumière le lien existant entre les variations du taux de chômage et celles du taux de croissance. Ainsi, aux Etats-Unis, dans les années soixante, une baisse de 1% de chômage nécessitait une augmentation de 3% du taux de croissance par rapport au taux tendanciel (cette hausse pouvant s'étaler sur plusieurs années). Cela reste encore le cas pour les pays anglo-saxons. Cette donnée est cependant plus élevée pour l'Allemagne et plus encore pour la France pour laquelle on estime qu'un supplément de croissance de 5% est nécessaire par rapport au taux tendanciel pour faire baisser le chômage d'1 %. La différence, comme nous l'avons signalé dans le précédent paragraphe, tient en partie à la nature plus structurelle du chômage français. Le chômage « structurel » est lié aux inerties naturelles ou institutionnelles du marché du travail. Au-delà même de la croissance plus forte qu'ont connu les Etats-Unis sur la dernière décennie, et au contenu plus élevé en emploi de celle-ci, cette différence

dans la nature du chômage explique l'incapacité à faire diminuer le chômage en France même en période de croissance forte[82].
Le niveau de chômage « naturel », que l'on qualifie de NAIRU, dépend des caractéristiques structurelles de l'économie et des diverses rigidités qui peuvent exister. Le niveau des capacités de production (qui résultent de l'investissement), le degré de rigidité à la baisse des salaires réels, le niveau de l'indemnisation du chômage, le caractère protecteur de la législation du travail, le niveau du salaire minimum, l'adaptabilité de la population active aux besoins des entreprises sont les éléments qui vont influencer le niveau du chômage structurel. C'est donc en réformant le marché du travail, en permettant l'abandon de certaines rigidités existantes que l'on pourra améliorer la situation en matière de chômage.
Le NAIRU n'a cessé de baisser aux Etats-Unis pour se situer aujourd'hui entre 4 et 5%. Au contraire, pour la période 2002-2004, le NAIRU se situe à 7% en Allemagne et à 9% en Italie et en France (où il n'était encore que de 6% il y a vingt ans). L'augmentation du chômage en Europe s'explique donc par la moindre création d'emplois causée elle-même par la conjonction de politiques de croissance restrictives (voir dans le chapitre précédent la partie sur la politique monétaire) et l'existence de rigidités sur le marché du travail.

Le marché du travail européen renferme donc plus de rigidités que celui des Etats-Unis : syndicats puissants (malgré la faiblesse de la proportion de la main-d'œuvre syndiquée), législation plus restrictive sur le licenciement, indemnisation plus généreuse du chômage. Le modèle WS-PS (voir l'annexe sur **le modèle WS-PS**) nous permet de voir combien ces éléments sont de nature à élever le taux de chômage d'équilibre. Il convient cependant de relativiser cette conclusion, car il existe des pays européens qui ont su concilier une faible flexibilité (protection sociale forte) et un faible chômage (Suède, Danemark, Pays-Bas…). C'est ce qu'on appelle le « modèle nordique ». Les tentatives de réformes les plus récentes du marché de l'emploi ont ainsi été tentées d'aller chercher dans ces modèles « suédois » ou « danois » des éléments de solution au chômage de masse (suivi personnalisé des chômeurs par exemple).

Par rapport à la moyenne des pays européen, la France présente une législation plus complexe sur le droit de licenciement (plus forte rigidité) et un recours plus important aux emplois temporaires (plus forte flexibilité entraînant cependant une plus forte précarité pour les titulaires de ce type de contrats).
Les rapports et propositions qui ont vu le jour en France ont donc tenté à la fois d'introduire un certain degré de flexibilité (avec l'espoir que cela pourrait à terme permettre une diminution du taux de chômage structurel) tout en tentant d'améliorer la sécurité des parcours professionnels.

Rapport du conseil d'Analyse Economique **Blanchard et Tirole (2004)**	Proposition d'un régime d'incitations financières négatives pour les entreprises afin de réduire le recours au licenciement, en contrepartie d'un assouplissement des conditions juridiques des licenciements.
Rapport de Virville (2004)	Création d'un contrat de mission (pour la mise en œuvre de contrats de projet) de quelques mois ou de quelques années.
Rapport Camdessus (2004) *« Vers une nouvelle croissance pour la France »*	Suggestion d'une fusion du CDI et du CDD et de la création d'un contrat de travail unique (avec des droits progressifs en fonction de l'ancienneté dans l'entreprise).
Rapport Cahuc et Kramarz (2004) *« De la précarité à la mobilité, vers une sécurité sociale professionnelle »*	Proposition de contrat de travail unique + participation financière des entreprises au reclassement des salariés, effectué par la puissance publique. Renforcement de l'accompagnement public des chômeurs.

Tableau 6 : Quelques rapports récents sur l'emploi

Enfin, pour achever la description de la situation contemporaine du chômage en Europe et particulièrement en France, il convient de relever que les coûts salariaux dans les pays européens sont des éléments importants de rigidité et une cause majeure de l'insuffisance de la création d'emplois. Tous les pays de l'OCDE disposent d'un système de salaire minimum mais la hauteur de celui-ci par rapport au salaire moyen diffère selon les pays (35% aux Etats-Unis, 55% en France). Le

[82] Le chômage structurel élevé bride la croissance puisque les capacités de production viennent buter sur la barrière des 8 à 9% de chômage « structurel » (comme en 1989-90 et en 2000-01) lors des derniers épisodes de croissance forte.

salaire minimum a des justifications sociales évidentes mais réduit l'emploi des moins qualifiés. On considère que le SMIC, en France, avant les mesures de compensation de charges sociales de la dernière décennie, était à l'origine d'environ 650 000 chômeurs « classiques »[83].

5.4.2 La baisse du temps de travail : bonne ou mauvaise solution ?

L'idée couramment répandue est qu'une baisse du taux d'activité est une bonne réponse au chômage. Les différentes politiques de l'emploi qui se sont appuyées sur ce précepte ont donc eu pour conséquence de faire augmenter la population inactive. Or, c'est dans les pays où les taux d'activité sont les plus élevés que l'on crée le plus d'emplois et que les taux de chômage sont les plus faibles. Ceci est une divergence importante entre la situation sur le marché de l'emploi dans certains pays d'Europe et aux Etats-Unis.

En 1973, les taux d'activités étaient identiques dans la « zone euro » et aux Etats-Unis (65%). En 2003, malgré l'augmentation du travail féminin, le taux d'activité européen n'est que de 70% quand le taux d'activité aux Etats-Unis est de 76%. Le graphique ci-dessous illustre cette baisse de la participation au marché du travail en France

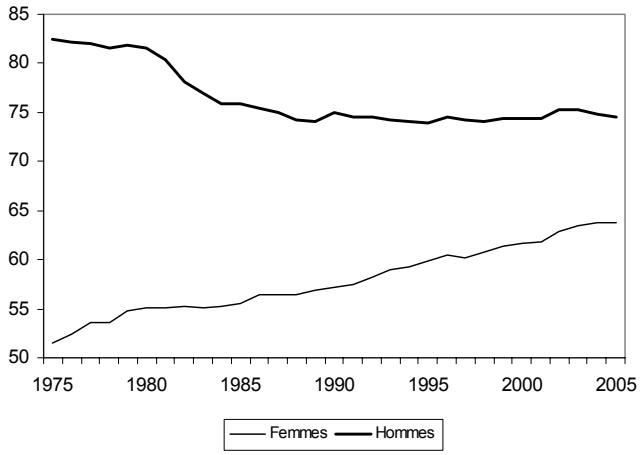


Figure 54 : Taux d'activité

Or l'idée que le travail est de plus en plus rare et qu'il doit être partagé n'est pas vraiment fondée économiquement. De nombreux exemples (rapatriement des français d'Algérie en 1962, exode des cubains à Miami en 1980) montrent qu'au contraire, si le capital est capable de s'ajuster, une augmentation même forte et rapide de la population active, n'a pas d'effet significatif sur le taux de chômage à moyen terme.

Les politiques qui visent limiter la population active (politiques « malthusiennes » du marché du travail), en considérant que le nombre d'emplois est donné, ne paraissent donc pas fondées et peuvent même s'avérer dangereuses, si on pousse ce raisonnement au plus loin[84].

La France a mis en place à la fin de la dernière décennie (lois Aubry de 1998 et 2000), des lois visant à abaisser la durée légale hebdomadaire du travail à 35 heures (assorti d'un régime rigoureux en matière d'heures complémentaires). Des mécanismes d'aides financières aux entreprises avaient été mis en place.

Certains voyaient dans cette réforme un « remède miracle » au chômage, d'autres parlent d'une « anomalie historique ». Si nous venons de souligner combien la thèse « malthusienne » d'une réponse au chômage par une baisse du taux d'activité paraissait peu fondée, il convient cependant d'ajouter que la baisse du temps de travail procède d'une évolution séculaire. C'est la vocation même du progrès technique que de permettre de produire plus en travaillant moins grâce à l'augmentation de la productivité du travail.

[83] En 1999, B. Salanié avait estimé, à l'aide d'une maquette de l'économie française qu'il avait construite, qu'une baisse de 10% du SMIC net permettrait d'obtenir une augmentation de 0,5% du PIB et de 1,3% d'emploi.

[84] En effet cela signifie qu'il faut : allonger la durée de scolarité ou la période de formation chez les jeunes (plutôt une bonne chose), généraliser les préretraites ou abaisser l'âge de la retraite (gros problème de financement des retraites qui pèsera à termes sur l'emploi), pratiquer une politique familiale visant à retirer les femmes de la population active (un peu rétrograde à l'heure de la parité, non ?), faire repartir chez eux les immigrés (idiot, surtout lorsque la démographie européenne doit aboutir dans quelques années à un manque de main d'œuvre).

Les résultats de la réduction du temps de travail en matière d'emploi sont les suivants. On estime entre 300 000 et 350 000 le nombre d'emplois cumulés créés par cette mesure, mais cela ne tient pas compte des effets induits macroéconomiques, dont certains sont négatifs.
Il apparaît que cette réforme a été un bon soutien conjoncturel (hausse de la consommation, des loisirs), mais que l'arbitrage en faveur de l'emploi et au détriment des salaires a lourdement pénalisé les travailleurs à faibles salaires. Le plafond imposé aux capacités de production des grandes entreprises début 2000 est néanmoins survenu à un moment où des contraintes d'offre apparaissaient dans l'économie et a peut-être précipité l'entrée en récession de l'économie française. Le délai supplémentaire accordé aux PME était trop limité (ne leur permettant pas de s'ajuster). Cette réforme, sur le plan structurel, ne s'avère donc pas forcément bonne. Les 300 000 emplois créés ne représentent qu'une part minoritaire des créations d'emplois en France entre 1997 et 2001 (deux millions d'emplois créés) et les contraintes supplémentaires imposées aux entreprises par cette mesure expliquent peut-être la piètre performance récente en matière de créations d'emplois.

5.4.3 Vers des réformes plus structurelles.

Depuis quelques années en Europe, des politiques plus « structurelles » semblent mises en œuvre. Ces politiques visent à stimuler la demande de travail des entreprises et l'incitation au travail, pour renforcer l'offre des salariés. Ces mesures sont appelées à montrer leur plein effet que dans la durée. Elles concernent principalement la **baisse du coût du travail** et la mise en place d'**incitations au travail**.

5.4.3.1 La baisse du coût du travail

La baisse du coût du travail semble nécessaire pour les catégories de travailleurs les plus touchées par le chômage. Il semble qu'un consensus ait émergé sur ce point. Les mesures ont été amorcées par la gauche en 1992 (gouvernement Bérégovoy), amplifiées par le gouvernement Balladur en 1993 puis le gouvernement Juppé en 1995 (celui-ci a amené les exonérations de charges patronales jusqu'à 1,3 SMIC). Le gouvernement Jospin (dans le cadre de la mesure d'accompagnement à la réduction du temps de travail) a porté le seuil des exonérations à 1,8 SMIC en 1999. Le gouvernement Raffarin (2002) avait poursuivi dans cette voie, avec une tentative de simplifications de ces baisses de charges pour compenser l'effet de l'harmonisation du SMIC.
Afin de ne pas altérer le pouvoir d'achat des salariés, la baisse du coût du travail passe par une réduction du « coin fiscal ». Le « coin fiscal » est la différence entre les coûts supportés par l'employeur et le revenu correspondant dont disposent effectivement les travailleurs. En 2003, le coin fiscal était de l'ordre de 50% en France, en Allemagne et en Italie ; de 30% au Canada, au Royaume-Uni et aux Etats-Unis ; et de 20% au Japon. L'importance du coin fiscal peut être considéré comme un frein important à l'emploi. La manière dont sont utilisés les prélèvements obligatoires a également une influence sur le taux de chômage[85].
Les mesures d'exonération de charges sur les bas salaires ont été coûteuses pour les finances publiques mais les résultats sont relativement probants. Le nombre d'emplois non qualifiés créés ou préservés, du fait de ces mesures, est estimé entre 250 000 et 490 000 (chiffre à rapporter aux 650 000 chômeurs « classiques » estimés auparavant). La part des emplois non qualifiés a ainsi cessé de baisser depuis le milieu des années quatre-vingt-dix.
Les réformes du gouvernement de Villepin avaient par ailleurs tenté d'assouplir le contrat de travail afin de compléter les réformes de baisse du coût du travail par des réformes visant à développer la demande de travail des entreprises en leur autorisant plus de flexibilité (si le CNE a bien été mis en œuvre pour les entreprises de moins de 20 salariés, le CPE (contrat première embauche) n'a en revanche jamais pu être appliqué).

5.4.3.2 La mise en place d'incitations au travail

Lorsque le chômage est indemnisé à des niveaux voisins du salaire d'activité, l'incitation à la recherche d'emploi peut être faible et la durée du chômage peut devenir longue. Les allocations chômages ont été développées lorsque le chômage était encore faible de sorte que les régimes

[85] Si les chômeurs bénéficient prioritairement de ces transferts (augmentation des indemnités chômage par exemple) cela va déplacer à la fois la courbe WS vers le haut et la courbe PS vers le bas. Le chômage sera ainsi grandement accru (voir l'annexe en fin de chapitre sur **le modèle WS-PS**). Si les transferts sont effectués principalement entre salariés (dépense de santé, financement des retraites…), la relation WS n'est pas modifiée et la redistribution n'amplifie pas l'impact négatif du coin fiscal sur l'emploi.

d'indemnisation pouvaient être relativement « généreux ». Les différentes caractéristiques d'un régime d'indemnisation du chômage sont les suivantes : (i) durée du travail préalable nécessaire pour pouvoir bénéficier de l'indemnisation, (ii) durée et dégressivité des indemnités, (iii) prise en compte ou non de la situation familiale.
Le tableau suivant montre le pourcentage d'indemnisation d'un travailleur au début de son chômage lorsque celui-ci percevait auparavant le salaire moyen du pays considéré :

>75%	Danemark, Finlande, Pays-Bas, Suède
>70%	Allemagne, Espagne, France, Norvège
>65%	Canada
>60%	Belgique, Etats-Unis, Irlande, Royaume-Uni
<60%	Italie, Japon

Tableau 7 : Taux d'indemnisation du chômage

Ces taux peuvent paraître élevés car ils ne tiennent pas compte de l'indemnisation du chômage de longue durée. L'OCDE a ainsi calculé des indices généraux de remplacement qui sont la moyenne des différentes situations possibles. Le classement est relativement similaire mais quelques différences apparaissent tenant notamment au fait que certains pays (Allemagne, Suède) ont mis en place des mesures destinées à inciter les chômeurs à retrouver rapidement un emploi.
Les Pays-Bas, le Danemark et la Finlande ont des taux de remplacement généraux élevés (>45%). La France et l'Italie ont des taux de remplacement entre 30 et 40%. L'Allemagne et la Suède ont des situations intermédiaires. Les pays dont les taux de remplacement généraux sont les plus faibles sont les Etats-Unis, le Japon et le Royaume-Uni.
Les régimes d'indemnisation ont été rendus plus sélectifs ces dernières années, le plus souvent pour des raisons financières : (i) montant des prestations réduit (Royaume-Uni), (ii) caractère dégressif dans le temps (Allemagne), (iii) Contrôle renforcé du caractère effectif de la recherche.
La question centrale est ici celle du lien qui existe entre l'indemnisation du chômage et la désincitation au travail. Les études empiriques ne permettent pas de trancher. Il n'y a pas d'effet marqué du niveau des mesures d'indemnisation sur la durée et le niveau du chômage. On peut également remarquer que les systèmes d'indemnisation étaient plus généreux dans les années soixante-dix et que la plus grande sélectivité des indemnisations s'est accompagnée d'une montée du chômage.
Les dispositifs d'accompagnement à la recherche d'emplois et de sanctions (lorsque les recherches sont insuffisantes) semblent en revanche efficaces.
En France, le PARE (programme d'aide à la recherche d'emploi) a été mis en place en 2000. Il s'agit d'une aide à la formation et à la reconversion, ainsi qu'une prise en charge individuelle du chômeur. Enfin, la réforme de janvier 2004 a réduit les durées maximales d'indemnisation au titre de l'assurance chômage (de 30 à 23 mois pour les moins de 50 ans, de 45 à 36 mois pour les 50-55 ans).

5.4.3.3 Minima sociaux et « trappe à pauvreté »

La plupart des pays ont mis en place des systèmes de revenu minimum (RMI en France par exemple). Le revenu minimum, si il a bien entendu une justification sociale évidente, présente cependant un effet pervers car il peut s'avérer une véritable « trappe à pauvreté ». La reprise d'un travail faisant perdre le bénéfice d'allocations publiques qui accompagnent le revenu minimum, le bénéficiaire d'un revenu minimum peut ne pas être incité à sortir de cette précarité. Afin d'encourager la reprise d'emploi pour les rémunérations les plus faibles, la prime pour l'emploi (PPE) a été créée en 2001. A l'origine, son montant était proportionnel aux revenus du travail de 0,3 à 1 SMIC puis dégressif jusqu'à s'annuler pour 1,4 SMIC. Ce montant a par la suite été augmenté, en particulier pour les emplois à temps partiels (il peut atteindre 800 euros en 2006 pour quelqu'un gagnant un demi SMIC). En parallèle, une prime de 1000 euros a été instaurée pour favoriser la reprise d'emploi des chômeurs de longue durée et des bénéficiaires de minima sociaux.
On peut cependant remarquer que la PPE en France n'est peut être pas assez ciblée (trop grand nombre de bénéficiaires) et qu'elle n'est donc que faiblement incitative pour les bénéficiaires.

5.5 Conclusion

Le chômage est le fléau des économies européennes. Si un certain nombre de mesures on été récemment mises en œuvre pour tenter d'abaisser le niveau du chômage structurel, une partie de la solution proviendra d'un renforcement de la croissance potentielle en Europe. Nous verrons au chapitre 7 (les moteurs de la croissance : Education et Innovation) et chapitre 6 (ouverture internationale et globalisation), que c'est la capacité de l'Europe à développer son potentiel de croissance et à s'insérer dans l'économie mondiale qui pourra permettre de capitaliser les réformes entreprises sur le marché de l'emploi ces dernières années.

Mais il serait vain d'attendre de la croissance et de l'ouverture une solution au chômage si les réformes visant à faire tomber certaines des rigidités existantes sur le marché du travail ne sont pas entreprises.

En effet, si les forts mouvements de réallocation de main d'œuvre qui devront accompagner les futures mutations sectorielles et technologiques ou qui résulteront de la concurrence toujours plus forte des économies émergentes sont freinés, alors les ajustements nécessaires aux entreprises ne seront pas réalisés et le chômage demeurera à un niveau élevé en Europe.

Le chômage ne saurait être vaincu que par une économie innovante, moins frileuse par rapport au risque, et plus sûre des atouts de ses entreprises sur le marché mondial. Cependant, il convient également d'accompagner ces réformes par la mise en place de politiques de formation et de reconversion afin d'aider les victimes des mutations économiques nécessaires à retrouver un emploi, afin que la solution au problème du chômage n'occasionne pas un renforcement des inégalités et de la précarité.

5.6 Annexes

5.6.1 Le modèle néo-classique du marché du travail

Le modèle néo-classique de base s'intéresse à l'échange de travail entre une firme et un travailleur. L'agent offre sa force de travail et consomme un bien grâce au revenu issu de son travail (salaire). La satisfaction de l'agent s'accroît avec la quantité C de bien consommé mais diminue avec la quantité de travail qu'il fournit (H) puisque le travail engendre une désutilité en réduisant le temps que l'agent peut consacrer aux loisirs (L). L'arbitrage que réalise cet agent est le suivant (***arbitrage « consommation – loisir »***). Il dispose d'un nombre d'heures disponibles T qu'il partage entre travail et loisir (L=T-H). Travailler une heure de plus diminue sa satisfaction (le loisir diminue) mais accroît sa consommation en accroissant le revenu salarial. La quantité de travail que l'agent décide d'offrir à l'équilibre dépend donc de la forme de ses préférences, reflétée par une fonction d'utilité U(C,T-H), du prix du bien de consommation (p) et du taux de salaire (w).

La fonction U(.,.) est croissante vis-à-vis de ses deux arguments et donc décroissante vis-à-vis de T.

On est en situation de concurrence pure et parfaite et aucun des agents ne peut influencer la formation des prix. La contrainte budgétaire de l'agent indique que son revenu (salarial) doit couvrir ses dépenses de consommation : wH=pC. Le programme du consommateur est de déterminer le couple (C,H) qui maximise son utilité sous la contrainte budgétaire : $\underset{C,H}{Max}\, U(C, T-H)$ s.c. $pC = wH$. En remplaçant dans U(.,.) la consommation C par w/pH, le programme se réécrit : $\underset{H}{Max}\, U\left(\frac{w}{p}H, T-H\right)$. A l'équilibre la quantité optimale de travail offert par l'agent va provenir de l'égalisation du TMS au rapport des prix (cf. l'annexe sur **le programme du consommateur** en fin de polycopié) :

$$\frac{w}{p} = \frac{U_1\left(\frac{w}{p}H, T-H\right)}{U_2\left(\frac{w}{p}H, T-H\right)} = \frac{\text{utilité marginale de la consommation}}{\text{utilité marginale du loisir}}$$

En se donnant une forme explicite pour la fonction d'utilité, on tire de l'égalité précédente la fonction d'offre de travail $H^O(w/p)$. Sous des hypothèses raisonnables[86] portant sur la fonction d'utilité, la fonction d'offre est croissante vis-à-vis du salaire réel.

L'entreprise qui formule la demande de travail a une technologie de production à rendements décroissants Y=f(H). Son profit est la différence entre ses recettes et ses coûts. $\pi(H) = pf(H) - wH$. Le profit de l'entreprise est donc maximal lorsque la productivité marginale du travail est égale au salaire réelle[87] : $f'(H) = \frac{w}{p}$. La fonction de demande de travail qui résulte de cette égalité est alors : $H^D(w/p)=f'^{-1}(w/p)$. Elle est décroissante vis-à-vis du salaire réel.

L'intersection de l'offre et de la demande de travail définit le niveau du salaire réel d'équilibre (w/p)*. Dans cette vision néo-classique, l'équilibre du marché du travail correspond au plein emploi. Au point (H*, (w/p)*), il y a égalité entre productivité marginale du travail, taux marginal de substitution consommation / loisir et salaire réel. Cette double égalité définit ***l'efficience productive***. Le plein emploi est assuré par flexibilité totale du salaire réel. Le modèle néo-classique du marché du travail ne peut pas expliquer l'existence de ***chômage involontaire*** (situation dans laquelle des agents désireux de travailler pour un salaire plus faible que celui d'équilibre ne parviennent pas à trouver un emploi).

En revanche, il peut exister du ***chômage volontaire*** (situation où le travailleur refuse d'offrir son travail en dessous d'un certain taux de salaire $\overline{w}$/p, notamment parce que des indemnités chômage lui assure une utilité U($\overline{w}$/p,H)). Si $\overline{w}$>w*, le salaire w* n'est pas praticable car il ne compense pas la

[86] La hausse du salaire réel engendre en fait deux effets en sens contraire sur le l'offre de travail : (i) un *effet de substitution :* l'individu remplace du loisir par du temps de travail (car la hausse du salaire réel augmente le coût d'opportunité du loisir (constitué par le renoncement à une heure de paye)), (ii) un *effet de revenu* : l'individu est incité à travailler moins longtemps tout en conservant le même salaire (voir en annexes générales en fin de volume l'annexe sur **effet revenu / effet substitution**). Pour les néo-classiques, le lien croissant entre offre de travail et salaire réel repose donc sur l'hypothèse « raisonnable » que l'effet de substitution l'emporte sur l'effet revenu chez les individus.

[87] Voir en annexes générales l'annexe sur le **programme du producteur** ainsi que l'annexe sur **les rendements de la fonction de production** à la fin du chapitre Marché et régulation.

désutilité du travail. L'offre de travail est alors nulle en dessous de $\bar{w}$/p. Le chômage se mesure alors par l'écart entre $H^D(\bar{w}/p)$ et $H^S(\bar{w}/p)$.

5.6.2 Le modèle de salaire d'efficience

Shapiro et Stiglitz (1984) ont lié la théorie du salaire d'efficience à l'existence d'une asymétrie d'information (voir l'annexe sur l'**aléa moral** dans les annexes générales). Lorsque l'entreprise embauche un salarié, elle ne peut pas savoir si celui-ci va fournir l'effort requis ou bien si c'est un tire-au-flanc (qui va faire semblant de travailler mais ne fournira pas l'effort). Pour encourager le salarié à fournir un effort productif élevé , l'entreprise va être contrainte de fixer un salaire supérieur à celui que le salarié pourra retrouver sur le marché du travail en cas de licenciement. L'attitude du tire-au-flanc devient alors irrationnelle car les risques d'être détectés et licenciés existent avec la certitude de ne pas retrouver un emploi aussi bien rémunéré ailleurs.

Le modèle de salaire d'efficience explique donc un mécanisme de fixation du salaire au dessus du salaire d'équilibre car le salarié possède une rente informationnelle. Ce salaire est rigide à la baisse car il n'est pas intéressant pour l'entreprise de proposer un salaire qui ne soit pas incitatif. Ce salaire supérieur au salaire d'équilibre explique la présence de chômage à l'équilibre.

Le modèle est le suivant.

Soit une économie dans laquelle de nombreuses entreprises produisent un bien identique. Le bien est produit à partir du travail et est consommé. Les individus ont une durée de vie infinie et ont une préférence pour le présent déterminée par un facteur d'actualisation ρ. Chaque travailleur est soit salarié, soit chômeur. Les salariés reçoivent un salaire *w*. L'utilité qu'il retire du salaire est $w - e$ où $e>0$ représente le coût de l'effort demandé en contrepartie du salaire. Les emplois sont détruits à un taux exogène *q*. Les chômeurs perçoivent une allocation chômage w_u. Pour un chômeur, la probabilité de retrouver un emploi est *a*.

L'espérance d'utilité d'un salarié percevant un salaire *w*, notée *Ve* est la suivante :

$$V_e = w - e + \rho\left(qV_u + (1-q)V_e\right)$$

L'espérance d'utilité d'un chômeur, noté *Vu*, est pour sa part :

$$V_u = w_u + \rho\left(aV_e + (1-q)V_u\right)$$

La population active comprend N travailleurs. L'emploi total est donc $N(1-u)$ et le nombre de chômeurs uN. A l'équilibre « stationnaire » du modèle (lorsque le taux de chômage est constant), le flux de sorties de l'emploi vers le chômage est égal au flux de sorties du chômage vers l'emploi, de sorte qu'on a l'égalité suivante :

$$au = q(1-u)$$

Les travailleurs qui occupent un emploi peuvent fournir un effort nul ($e=0$) sans être systématiquement détectés par l'employeur. L'utilité qu'un tire-au-flanc retire de son salaire est w et non $w - e$ puisque $e = 0$. La probabilité qu'un travailleur qui tire au flanc soit détecté est *d*. Lorsqu'un « tire-au-flanc » est détecté, il reçoit le salaire de la période comme solde de tout compte et se retrouve au chômage.

L'espérance d'utilité d'un « tire-au-flanc », notée $\tilde{V}_e$, est la suivante :

$$\tilde{V}_e = w + \rho d V_u + \rho(1-d)\left(qV_u + (1-q)\tilde{V}_e\right)$$

Le problème de l'employeur est de fournir un salaire qui incite le salarié à fournir l'effort *e*.

C'est le cas si $V_e > \tilde{V}_e$. L'employeur doit donc proposer un salaire élevé qui assure au salarié qui fournit l'effort une espérance d'utilité supérieure à celle qu'obtient un chômeur ($V_e>V_u$).

$$V_e = V_u + \frac{e}{\rho d(1-q)}$$

Cette rente qui est accordée au salarié est le seul moyen de l'inciter à fournir l'effort. L'employeur n'a donc aucun intérêt à embaucher des chômeurs prêts à accepter des salaires plus faibles (car ces salaires seraient désincitatifs). On est bien en présence de chômage involontaire. Le salaire d'efficience d'équilibre w* est alors supérieur au salaire d'équilibre (w_u+e) est rigide à la baisse du fait du comportement de l'employeur.

$$w^* = e + w_u + e\frac{\left(1-\rho\left(1-\frac{q}{u}\right)\right)}{\rho d(1-q)}$$

5.6.3 Le modèle de négociation salariale

En France, les contrats de travail résultent à 85% de conventions collectives au niveau de la branche. Les contrats de travail sont donc, dans la plupart des cas, négociés entre les institutions représentatives que sont les syndicats et le patronat. Dans le modèle de « négociations salariales » dit de « droit à gérer » de Nickell et Andrews (1983), on considère que les négociations entre firmes et syndicats ne portent que sur le salaire. L'emploi est ensuite déterminé de manière unilatérale par la firme. Les salariés vont trouver un intérêt à l'action syndicale dès lors que celle-ci permet d'élever le salaire au-dessus du salaire d'équilibre constaté sur le marché du travail. Cela conduit cependant les entreprises à réduire le nombre d'emplois qu'elles proposent ce qui peut expliquer en retour une partie du chômage. Toute hausse du pouvoir de « marchandage » du syndicat entraîne une hausse du salaire et donc une baisse de l'emploi. L'action syndicale est ainsi défavorable à l'emploi. Pour Lindbeck et Snowers (1986), la défense des intérêts des travailleurs en place (les « insiders ») se fait alors au détriment des chômeurs (les « outsiders »).

Les hypothèses du modèle sont les suivantes : (i) les travailleurs participent aux négociations par l'intermédiaire d'institutions collectives (syndicats, comités d'entreprise), (ii) L'objectif de l'institution qui représente les salariés est d'obtenir le niveau de salaire le plus élevé possible de façon à maximiser la somme des utilités des travailleurs présents dans un bassin d'emploi, (iii) l'objectif de l'entreprise est d'obtenir le profit le plus élevé possible.
En résumé, la négociation ne va porter que sur les salaires et l'entreprise conserve la possibilité d'ajuster l'emploi.
Soit L le niveau d'emploi. La somme des utilités des travailleurs situés dans le bassin d'emploi d'une entreprise qui paye un salaire w, notée V_S, est :

$$V_S = LV_e + (N - L)V_u$$

Où V_e et V_u sont respectivement les utilités d'un travailleur et d'un chômeur, N la population active.
La technologie de l'entreprise est représentée par la relation suivante : $Y = \frac{AL^{\alpha}}{\alpha}$ avec $\alpha \in]0,1[$, où Y est le niveau de production. Avec p=1, le profit de l'entreprise est : $\pi = Y - wL$.
La demande de travail et le profit optimal ont pour expression (en fonction du taux de salaire w) :

$$L(w) = \left(\frac{A}{w}\right)^{\frac{1}{1-\alpha}} \qquad \pi(w) = \left(\frac{1-\alpha}{\alpha}\right) A^{\frac{1}{1-\alpha}} w^{-\frac{\alpha}{1-\alpha}}$$

Le salaire est négocié à chaque période lors de la négociation collective. Le résultat de la négociation est obtenu par la maximisation (par rapport au salaire w) du produit pondéré des gains que les « partenaires sociaux » retirent de leur relation. Les gains de l'entreprise correspondent au profit (en cas d'échec des négociations, le profit sera nul). Les gains du syndicat correspondent à la somme des utilités des travailleurs présents dans le bassin d'emploi, diminuée de la somme des utilités qu'ils obtiendraient si la négociation échouait (V_u puisque tous les travailleurs seraient alors au chômage).

$$\underset{w}{Max} \quad (L(w)(V_e - V_u))^{\gamma} (\pi(w))^{1-\gamma}$$

Où γ est le pouvoir de négociation du syndicat. Comme dans la théorie du salaire d'efficience, présentée dans la précédente annexe, les salariés obtiennent une espérance d'utilité supérieure à celle des chômeurs.

$$V_e = w - e + \rho(qV_u + (1-q)V_e)$$

Cette différence entre V_e et V_u augmente avec le pouvoir de négociation des syndicats. Elle décroît en revanche avec la valeur absolue de l'élasticité de la demande de travail par rapport au salaire ($1/(1-\alpha)$) car plus cette élasticité est grande et plus les gains en salaire se traduisent par des diminutions d'emplois importantes.
Le niveau de salaire qui résulte de la négociation est supérieur au niveau de salaire d'équilibre de concurrence :

$$w^* = \frac{e + w_u}{1 - \gamma\left(1 - \rho\left(1 - \frac{q}{u}\right)\right)}$$

5.6.4 Le modèle WS-PS

Le taux de chômage d'équilibre peut être déterminé par le modèle WS-PS, d'inspiration néo-keynésienne. La courbe WS (« ***Wage Setting** relation* » ou «***Wage Schedule*** ») décrit la relation négative qui existe entre le salaire réel et le taux de chômage. C'est la logique de la relation originelle estimée par Phillips (plus le chômage est élevé et plus il existe une pression à la baisse sur les salaires réels, des études récentes indiquent qu'une augmentation du chômage de 10% fait baisser le salaire réel de 1%). La relation WS trouve également des justifications théoriques dans deux autres modèles récents de la pensée néo-keynésienne (le modèle de négociation salariale collective et le modèle de salaire d'efficience, voir les annexes précédentes).

L'équation qui caractérise la détermination des salaires est donc la suivante : $w = F(u, z)$.

Où *w* est le salaire nominal, *p* le niveau des prix, *u* le taux de chômage et *z* une variable prenant en compte l'ensemble des autres facteurs déterminant la fixation du salaire. Ceci est donc une forme réduite des deux équations liant le salaire et le niveau du chômage dans le modèle de salaire d'efficience et le modèle de négociation salariale (voir les deux annexes précédentes).

La relation PS (« ***Price Setting** relation* » ou « ***Price Schedule*** ») décrit la fixation des prix par les entreprises. En partant de la fonction de production simple $Y = \frac{AL^{\alpha}}{\alpha}$ on obtient la relation suivante qui va permettre de définir la courbe PS : $w = \frac{A}{(N(1-u))^{1-\alpha}}$. La demande de travail diminue (et le chômage augmente) lorsque le salaire s'accroît.

L'équilibre sur le marché du travail implique que le salaire réel déterminé lors des négociations salariales (WS) soit égal au salaire découlant du processus de détermination des prix (PS). Le taux de chômage d'équilibre est qualifié de taux de chômage structurel (il dépend de caractéristiques structurelles de l'économie comme la législation ou le montant des allocations chômage, prises en compte à travers le paramètre z).

Lorsque le degré de vérifiabilité de l'effort diminue (voir le modèle de salaire d'efficience) ou que le pouvoir de négociation salariale des syndicats s'accroît (voir le modèle de négociation salariale), alors la courbe WS se déplace vers le haut. Le taux de salaire d'équilibre s'accroît ainsi que le taux de chômage.

Examinons les effets de deux politiques économiques sur le taux de chômage d'équilibre : une politique consistant à accroître les gains des chômeurs (w_u) et une augmentation des prélèvements payés par les employeurs.

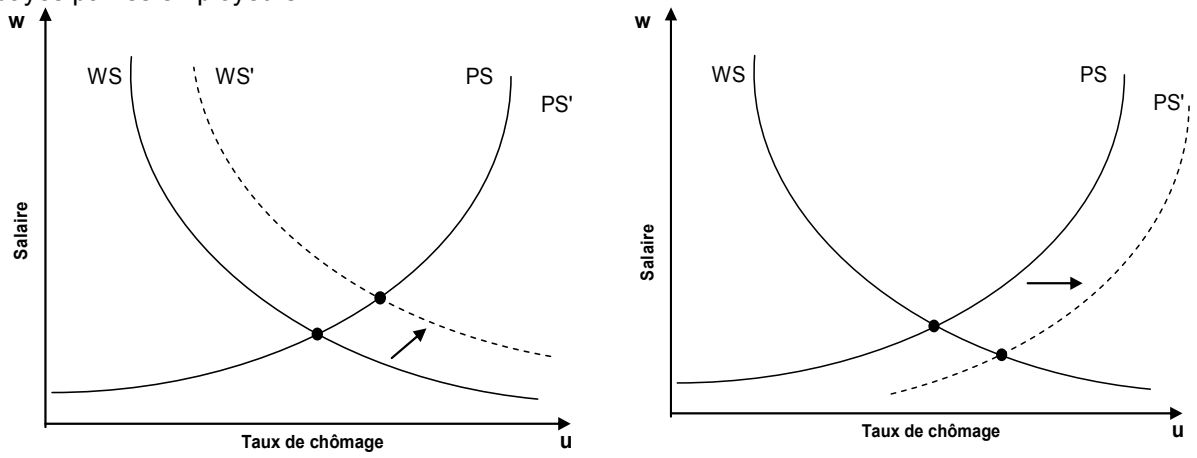


Figure 55 : Le modèle WS-PS

Une hausse des allocations chômage peut être traduite par une hausse du paramètre z puisque cela a pour effet d'augmenter le salaire fixé lors des négociations salariales à taux de chômage donné (il y a augmentation du salaire de réserve). La courbe WS se déplace vers le haut et vers la droite. Le taux de chômage structurel augmente suite à cette mesure (figure de gauche ci-dessus).

Une augmentation du taux des prélèvements payés par les employeurs déplace la courbe PS vers le bas, ce qui accroît le chômage et diminue les salaires (figure de droite ci-dessus).

5.6.5 La courbe de Beveridge

Le nombre d'embauches dans l'économie dépend de la confrontation entre les offres et les demandes d'emplois. Lorsque les travailleurs sont adaptés à n'importe quels emplois proposés et que l'information sur la disponibilité des postes à pourvoir est parfaite, le nombre d'embauches correspond au minimum de l'offre ou de la demande (le marché du travail est efficace). Ce n'est bien sûr pas le cas dans la réalité. L'information n'est pas parfaite et les travailleurs sont hétérogènes, de sorte que certains individus ne trouvent pas d'emplois alors que des entreprises ont des emplois vacants. L'ensemble des obstacles qui s'oppose à l'efficacité du marché du travail peut être résumé à travers une fonction d'appariement indiquant une relation entre embauches, offres et demandes d'emplois.

La fonction d'appariement décrit à chaque date le nombre d'embauches réalisées (M) lorsqu'il y a V emplois vacants et U chômeurs. La fonction M(V,U) est supposée croissante par rapport à chacun de ses arguments. Le nombre d'embauches augmente lorsque les nombres d'emplois vacants ou de demandeurs d'emplois s'accroissent. On va également faire l'hypothèse que cette fonction d'appariement présente des rendements d'échelle constants.

La probabilité de pourvoir un emploi vacant peut être exprimée en fonction du rapport qui existe entre le nombre d'emplois vacants et le nombre de chômeurs θ=V/U (θ est un indicateur de la tension qui règne sur le marché du travail). $\frac{M(V,U)}{V} = M\left(1,\frac{U}{V}\right) = m(\theta)$. La probabilité de pourvoir un emploi vacant diminue avec l'indicateur de tension (puisque m'(θ)<0, pour un nombre donné de chômeurs, il est plus dur pour une firme de pourvoir un poste lorsque le nombre d'emplois vacants augmente). La probabilité de sortie du chômage dépend également de l'indicateur de tension du marché du travail. $\frac{M(V,U)}{U} = \frac{V}{U}\frac{M(V,U)}{V} = \theta m(\theta)$. La probabilité de sortie du chômage est une fonction croissante de la tension sur le marché du travail (pour un nombre donné de chômeurs, chacun d'eux a plus de chance de trouver un emploi lorsque le nombre de postes vacants s'accroît).

La fonction d'appariement et l'équilibre des flux définissent, à l'état stationnaire, une relation entre emplois vacants et chômage. A l'état stationnaire, le nombre de travailleurs qui perdent leur emploi est tout juste égal au nombre de chômeurs qui en retrouvent un : $au = q(1-u)$ avec $a = \theta m(\theta)$.

Cela permet d'exprimer une relation décroissante entre le taux de chômage (u) et le taux d'emplois vacants (v=V/N). Cette relation, qui a pour expression ; $u = \frac{q}{q + \theta m(\theta)}$, est la « courbe de Beveridge). Elle indique l'efficacité du processus d'appariement. Une amélioration de l'efficacité du processus d'appariement se traduit par un déplacement de la courbe de Beveridge vers l'origine.

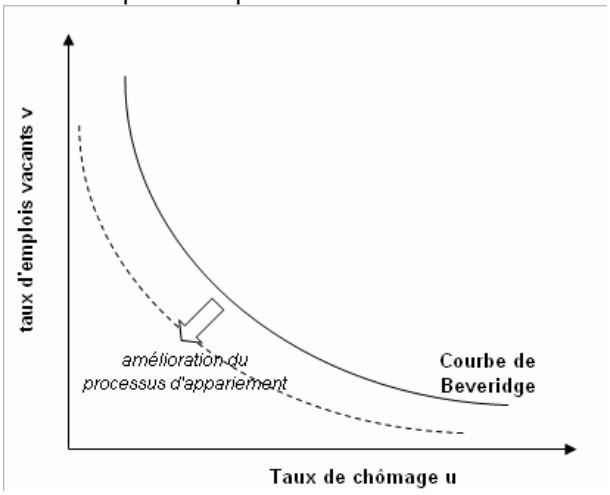


Figure 56 : Courbe de Beveridge

5.6.6 Coût d'appariement et chômage d'équilibre

Les entreprises sont confrontées à des coûts de recrutement de la main d'œuvre. Soit c le coût d'un emploi vacant. Chaque emploi a une probabilité $m(\theta)$ d'être pourvu à chaque période. Le gain marginal pour l'entreprise de pourvoir un emploi est la différence entre la productivité marginale du travail et le coût du travail, sachant que l'emploi peut être détruit avec une probabilité q : $\Pi_m = \frac{1}{1+\rho}\left(AL^{\alpha-1} - w + (1-q)\Pi_m\right)$, c'est-à-dire $\Pi_m = \frac{AL^{\alpha-1} - w}{q+\rho}$ où ρ est le facteur d'actualisation. L'entreprise va donc déclarer des postes vacants jusqu'au point où le coût d'un emploi vacant c sera égal à l'espérance de gain qu'il procure, $m(\theta)\Pi_m$. Cette égalité peut s'exprimer sous la forme $AL^{\alpha-1} = w + \frac{c(q+\rho)}{m(\theta)}$ et définit une relation entre le coût du travail, l'emploi et la variable de tension sur le marché du travail. Elle montre que la productivité du travail doit être supérieure au coût du travail lorsqu'il existe des coûts de recrutement de la main-d'œuvre. Cela va forcément réduire les embauches et créer du chômage par rapport à une situation concurrentielle où le coût du travail est égal à la productivité marginale. En postulant que le salaire croît avec la variable de tension sur le marché du travail, w=w(θ) et en remplcant L par N(1-u), on peut obtenir une courbe **VS** (pour ***Vacancy Schedule***) qui décrit le comportement d'offre d'emplois vacants de l'entreprise lorsque celle-ci fait face à des coûts de recrutement.

$$AN^{\alpha-1}(1-u)^{\alpha-1} = w(\theta) + \frac{c(q+\rho)}{m(\theta)}$$

Cette courbe peut alors être mis en regard de la courbe de Beveridge dans le plan (u,v)

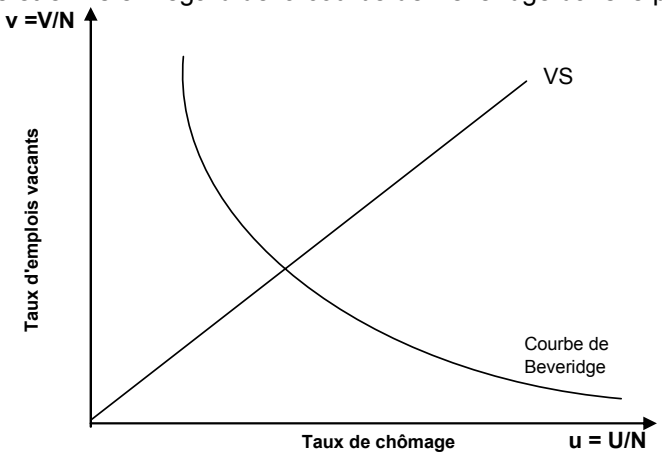


Figure 57 : Courbe de Beveridge et courbe VS

La prise en compte du processus d'appariement permet d'enrichir les conclusions du modèle WS-PS en étudiant l'impact des réallocations d'emplois et de l'efficacité du processus d'appariement sur le chômage. On peut ainsi distinguer le chômage de réallocation (ou chômage frictionnel) du chômage conjoncturel qui résulte des chocs macroéconomiques.

L'ensemble des éléments, qui augmentaient le taux de chômage d'équilibre dans le modèle WS-PS (augmentation du gain des chômeurs, hausse des cotisations sociales employeurs), déplace la courbe VS vers la droite du plan (u,v) sans modifier la courbe de Beveridge. Il y a alors une diminution du nombre d'emplois vacants et une augmentation du taux de chômage d'équilibre. Les chocs de réallocation vont déplacer la courbe de Beveridge vers le haut, ce qui accroît à la fois le taux de chômage et le taux des emplois vacants. Les chocs de réallocation peuvent donc entraîner des variations de même signe des taux de chômage et d'emplois vacants pour peu que le déplacement de la courbe de Beveridge soit suffisamment important par rapport à celui de la courbe VS.

6 Ouverture internationale et globalisation

Depuis la fin de la seconde guerre mondiale, on observe une forte croissance des échanges internationaux de biens et services ainsi qu'une circulation accrue des capitaux. Doit-on s'en réjouir ? L'échange international est souvent présenté comme un facteur d'accroissement des richesses produites, et donc *in fine* une source de bien-être pour le consommateur. Cependant, certains redoutent que la mondialisation des économies aggrave les inégalités ou le chômage. Est-ce que ces peurs nécessitent la mise en place de barrière au commerce mondial ?

Le débat entre partisan du libre-échange et tenant du protectionnisme n'est pas nouveau. Il a structuré le développement de la discipline qu'est l'économie internationale depuis ses débuts. L'économie internationale couvre toute les questions qui ont trait à l'ouverture d'une économie, c'est-à-dire toutes les raisons qui peuvent expliquer que les différentes économies au monde soient reliées par des échanges (de biens, de services, de capitaux, ou encore de personnes) ainsi que l'impact de ces liens sur les équilibres internes des économies.

On distingue habituellement deux champs principaux en économie internationale : le **commerce international** qui analyse les flux réels (causes et implications des gains à l'échange, politiques de restriction et d'ouverture aux échanges lorsqu'on ignore les déséquilibres macroéconomiques (plein emploi, balance commerciale équilibrée)), et la **finance internationale** qui analyse les flux financiers (détermination des taux de change, effets de l'intégration financière, effets des politiques conjoncturelles en économie ouverte, études des déséquilibres budgétaires et commerciaux…). Nous allons mettre à la fois l'accent sur les aspects « commerce international », même si nous ne pourrons couvrir la totalité de ce champ de l'analyse économique (notamment les développements les plus récents dits des nouvelles théories du commerce international), et sur les aspects de finance et de macroéconomie internationales. Notre exposé en matière de finance internationale se limitera à l'étude des composantes de la balance des paiements et à une succincte présentation des mécanismes de change et des enjeux liés aux évolutions du système monétaire et financier international[88].

L'objectif de ce chapitre est donc de vous donner les éléments d'analyse nécessaires pour mieux comprendre les enjeux des négociations au sein de l'OMC, et de vous permettre de compléter votre appréciation du fonctionnement de l'économie par les éléments primordiaux en matière de fonctionnement du système monétaire et financier international.

Nous allons donc bien sûr parler du libre échange et du protectionnisme. Cependant, les théories en commerce international ne se limitent pas à la seule description et l'opposition des bienfaits du libre échange et du protectionnisme. En effet, une question est apparue de plus en plus prégnante à mesure que le poids des firmes internationales dans le commerce mondial prenait de l'importance. Quel est véritablement le rôle joué par les firmes multinationales dans le commerce international ? Sont-elles les vecteurs de la mondialisation ? Concourent-elles au développement ou à la croissance des pays dans lesquels elles s'installent ?

Nous verrons que la prise en compte du rôle des firmes internationales tout comme la description plus fine des échanges de biens similaires entre des économies de même niveau de développement a abouti à un profond renouvellement des théories du commerce international. Il n'en demeure pas moins que l'objectif reste le même : savoir si l'ouverture des frontières et la plus forte mobilité des capitaux est profitable ou non à une économie donnée et à l'économie mondiale prise dans son ensemble.

Les vertus du libre échange se doivent d'être analysées à l'aune de la réduction des inégalités entre pays, et à l'intérieur des pays qu'il permet ou qu'il suscite. On peut sur ce point présenter deux exemples contradictoires : (i) les pays industrialisés d'Asie du sud-est (Tigres et Dragons…) représentent un exemple d'intégration dans l'économie mondiale, de rattrapage, par le commerce international, (ii) au contraire les pays de l'Afrique sub-saharienne s'enfoncent eux dans le sous-développement et les exportations de matières premières ou de biens agricoles qu'ils réalisent ne semblent pas changer cet état de fait.

[88] Nous n'abordons la macroéconomie en économie ouverte qu'à travers un réexamen des politiques économiques dans un modèle IS-LM élargie pour permettre la prise en compte de l'équilibre de la balance des paiements, c'est-à-dire travers une annexe sur **le modèle Mundell-Flemming : IS-LM-BP**.

La mondialisation semble donc s'accompagner d'un creusement des inégalités entre nations puisque l'avance acquise confère des avantages en terme d'effet d'expérience, d'économie d'échelle et de financement de la recherche, qui sont les sources des avantages comparatifs de demain. Les inégalités s'amplifient également à l'intérieur des nations. C'est le cas des pays en développement (où les secteurs traditionnels repliés sur l'intérieur coexistent avec des secteurs modernes tournés vers l'extérieur). C'est également le cas des pays industrialisés où l'on constate une fracture entre les individus qualifiés (qui profitent de la mondialisation) et ceux qui la subissent (moins qualifiés) en raison de la concurrence des pays en développement sur des secteurs traditionnels de main-d'œuvre où le coût de la main d'œuvre joue un rôle primordial. D'où les peurs et les réticences face à cette vague actuelle de mondialisation des échanges qui, dans l'histoire de l'humanité, n'est pourtant pas la première.

Mais les peurs suscitées par cette mondialisation des échanges et la concurrence accrue de certains pays BRIC (Brésil, Russie, Inde, Chine) s'accompagnent aussi d'incertitudes concernant la robustesse et la durabilité du système monétaire et financier international. Les flux financiers journaliers qui transitent sur le marché des changes sont gigantesques (1500 Md $ chaque jour, peut-être plus…, c'est-à-dire l'équivalent du PIB de l'Italie !). La volatilité de ces flux ne cesse d'inquiéter car elle constitue un élément de fragilité du système monétaire et financier international. Le point crucial tient notamment aux évolutions des taux de change qui ont des effets sur la compétitivité des nations et donc sur l'économie réelle. Une des questions en suspens concerne la valeur du dollar, par rapport à l'euro et au yen notamment. Cette question est reliée à ce qu'on appelle les « déficits jumeaux » aux Etats-Unis (déficit commercial et déficit public). Durant l'hiver et le printemps 2007, le dollar a baissé par rapport au yen et à l'euro mais ne s'est pas encore écroulé. Or ce double déficit pourrait bien un jour entraîner une chute plus brutale du dollar. Pourquoi ces déficits jumeaux existent-ils et pourquoi le dollar ne dévisse-t-il pas pour l'instant ? En partie parce que la Chine recycle une partie de ces excédents commerciaux dans le financement de la dette américaine, ce qui soutient le dollar, la croissance américaine et… les exportations chinoises. Il est cependant à redouter que ce mécanisme d'intérêt mutuel ne puisse durer indéfiniment, notamment en cas de récession américaine[89]. La chute du dollar qui pourrait en résulter occasionnerait certainement une crise économique mondiale majeure.

6.1 Quelques faits

6.1.1 D'une mondialisation à l'autre…

Le phénomène de mondialisation n'est pas nouveau. En ce qui concerne les flux commerciaux, l'essor a débuté dès le milieu du XIX^{ème} siècle. La raison du développement des échanges commerciaux fût la diminution des coûts de transport, grâce aux développements du chemin de fer et à l'amélioration des moyens de transports maritimes (entre 1840 et 1913, le commerce mondial a été multiplié par 12 !).

Les guerres mondiales et la crise des années trente ont mis un terme à cette première mondialisation. Ainsi, certains économistes considèrent que le niveau actuel d'ouverture commerciale de l'économie mondiale n'est qu'un juste retour à la situation qui prévalait au début du siècle dernier. Il convient cependant de relever certaines différences entre les deux « mondialisations », notamment le fait que le type de biens échangés et les protagonistes des échanges se sont transformés par rapport au début du XX^{ème} siècle. A ce moment-là, les pays riches (pays du « nord ») exportaient des biens manufacturés et importaient beaucoup de biens agricoles et de matières premières (des pays du « sud » moins développés). Nous allons voir dans la sous-section suivante que la nature des échanges a profondément évolué et que la plus grande partie du commerce mondiale est désormais constituée de commerce intra-industriel entre pays du Nord.

[89] En cas de retournement de conjoncture aux Etats-Unis, les importations américaines baisseraient de manière importante. La croissance chinoise fondée sur les exportations, notamment vers les Etats-Unis, serait elle aussi amoindrie. De plus faibles excédents commerciaux chinois entraîneraient une moindre possibilité de financer le déficit public américain à travers l'achat de titres publics et donc un décrochage du dollar (moins de placements aux Etats-Unis donc moins besoin d'acheter du dollar).

6.1.2 Les aspects essentiels des échanges mondiaux

Nous allons examiner quelques faits stylisés à même d'éclairer le phénomène actuel de globalisation.

Les échanges de biens et services représentent une part croissante de l'activité des pays. L'augmentation est particulièrement importante comme le montrent les tableaux ci-dessous, à la fois pour les pays développés mais également pour certains pays en développement.
Les grands pays commercent moins avec l'étranger que les petits (voir la différence entre les Etats-Unis et la Tunisie par exemple). La diversité des ressources et des productions rend en effet moins important la nécessité du commerce mondial pour les grands pays.

	PIB (milliards de $ courants)			Commerce international (exports + imports, en % du			Flux d'IDE (nets entrants, en % du PIB)		
	1970	1990	2001	1970	1990	2001	1970	1990	2001
Etats-Unis	1026	5751	10065	11	21	18	0,12	0,84	1,43
Japon	206	3052	4141	20	20	20	0,05	0,06	0,15
France	147	1216	1310	30	43	54	0,42	1,28	4,21
Allemagne	-	1689	1846	39	54	68	-	0,18	1,84
Chine	92	355	1159	4	32	49	0	0,98	4,04
Brésil	42	465	509	14	15	27	0,93	0,21	4,41
Mexique	36	263	624	17	38	57	0,91	1,00	4,06
Tunisie	1	12	20	47	94	99	1,11	0,73	2,43

Tableau 8 : L'essor du commerce mondial et des investissements directs dans différents pays développés et en développement

(source : Banque Mondiale (2003), CNUCED (2003))

Ces dernières années, les flux commerciaux et les flux de capitaux ont connu une croissance plus rapide que les revenus au niveau mondial (voir le tableau ci-dessous).
La période récente est notamment marquée par un essor des investissements directs à l'étranger (IDE). Initialement les investissements directs visaient principalement à contourner des barrières tarifaires importantes (droits de douanes élevés) en installant sur place des capacités de production. Ils s'inscrivent désormais davantage dans des stratégies de délocalisations vers les pays à faible coût de main d'œuvre ou bien dans une stratégie de décomposition internationale du processus de production (DIPP) en fonction des avantages comparatifs ou fiscaux des différents pays. La baisse des coûts de transports permet en effet de délocaliser la réalisation de certains sous-ensembles de biens ou services vers des pays à bas coûts de main d'œuvre[90].

	1986-1990	1991-1995	1996-2000
PIB	10,8	5,6	1,3
Exportations de B&S	15,6	5,4	3,4
IDE entrants	23,1	21,1	40,2
IDE sortants	25,7	16,5	35,7

Tableau 9 : Evolutions comparées, du commerce international et des IDE dans le monde (taux de croissance)

(source : World Investment Report (2003))

La période de l'après guerre a été marquée par une croissance mondiale forte accompagnée d'une augmentation des échanges mondiaux encore plus importante. Une des raisons de l'explosion des échanges est la chute régulière des droits de douanes. Ces diminutions ont résulté des multiples négociations internationales dans le cadre du GATT (*General Agreement on Tarif and Trade*). Les négociations multilatérales menées au sein du GATT ont permis notamment de ramener les droits de douane moyens des pays industrialisés sur les produits manufacturés de 40% en 1947 à 3,9% à l'issue du dernier round de négociations.

[90] Des langoustines pêchées en Ecosse peuvent ainsi être congelées puis expédiées en… Thaïlande pour y être décortiquées plus de trente jours après (où les ouvrières y sont payées 19 fois moins qu'en Ecosse) puis réexpédiées en Europe pour y être conditionnées et vendues.

	Tarif douanier moyen						
	1875	1913	1931	1950	1980	1990	2001-2002
Etats-Unis	40-50	44	48	14	7	4,8	3,4
France	12-15	20	30	18	-	-	-
Union Européenne	-	-	-	-	5,7	5,9	4,1

Tableau 10 : Tarifs douaniers sur les biens manufacturés (évolution sur longue période)

(source : Bairoch (1994), OMC (2002))

Le tableau ci-dessous illustre l'augmentation de la part des biens manufacturés dans le commerce des pays développés au cours du siècle dernier. C'est, nous l'avons signalé en introduction, une différence majeure entre la précédente et l'actuelle mondialisation. Les pourcentages d'exportations et d'importations de produits manufacturés dans l'ensemble des exportations et importations des pays développés étant désormais très proches, on peut en conclure que l'on est en présence de commerce de similitude. La plus grosse part du commerce mondial est constitué de commerce intra-industriel Nord-Nord. Les biens ou composants de biens échangés appartiennent aux mêmes industries. Les théories classiques basées sur les avantages absolus ou comparatifs ou sur la différence dans les dotations de facteurs de production ne sont plus alors suffisantes pour comprendre et expliquer ce type de commerce devenu pourtant majoritaire. C'est pour cela que les nouvelles théories du commerce international ont mis en avant la différenciation horizontale et la concurrence imparfaite pour expliquer ces types de flux.

	1910		2000	
	Exportations	Importations	Exportations	Importations
Etats-Unis	47,5	40,7	83,2	79,6
Allemagne	74,5	24,4	83,4	68
France	59,2	25,3	82,9	79
Japon	80,8	38,3	93,8	56
Royaume-Uni	75,4	24,5	80,7	80,8

Tableau 11 : Parts des biens manufacturés dans le commerce de certains pays développés (en %)

(source : Kuznets (1967), OMC (2003))

Une autre caractéristique importante des flux actuels de commerce est leur caractère régional de plus en plus marqué. La part du commerce mondial qui se situe à l'intérieur d'une zone régionale est de plus en plus importante. Cette orientation du commerce résulte de l'adoption de nombreux accords de libre-échange régionaux (ALENA, Union Européenne...).

Ainsi, près d'un tiers du commerce mondial est le fait du seul commerce intra-européen. Plus des deux tiers du commerce des pays européens est constitué de commerce intra-européen[91].

	Parts dans les exportations totales en %				Date d'entrée en vigueur
	1970	1980	1990	2001	
UE	59,5	60,8	65,9	61,2	1957
APEC	57,8	57,9	67,5	71,8	1989
MERCOSUR	9,4	11,6	8,9	20,8	1991
ALENA	36,0	33,6	41,4	54,8	1994
ANASE	22,4	17,4	19,0	22,4	1992

Tableau 12 : Parts des exportations intra-régionales

(en % des exportations totales) (source : OMC (2003))

UE : Union Européenne, APEC : Asia-Pacific Economic Cooperation, MERCOSUR : Marché Commun du sud, ALENA : Association du libre Echange Nord Américain, ANASE : Association des Nations de l'Asie du Sud-Est.

[91] C'est pour cette raison qu'il faut se méfier des explications de la piètre performance du commerce français tenant au niveau jugé trop élevé de l'euro. En effet, les deux tiers des exportations françaises sont à destination de pays de la zone euro. La balance commerciale française est légèrement excédentaire vis-à-vis de la zone hors Europe et le déficit commercial français (30 Md d'€ pour 2006) est en partie causé par notre faible spécialisation dans des produits de haute technologie et de forte valeur ajoutée, qui constituent une bonne part des échanges européens.

L'articulation entre le mouvement d'intégration régionale et la libéralisation commerciale menée de matière multilatérale sous l'égide de l'OMC (l'Organisation Mondiale du Commerce qui a remplacé le GATT depuis 1994) est complexe. Ces accords sont-ils une étape supplémentaire vers une libéralisation plus poussée des échanges ou bien est-ce une manifestation d'un certain repli face aux craintes de la mondialisation ? Répondent-ils à un désir en faveur de plus de libre-échange ou des tentations protectionnistes ? Nous tenterons de répondre à cette interrogation[92], mais il nous faut auparavant examiner les théories normatives en matière de commerce international.

6.2 Le commerce international

Les théories du commerce international peuvent être classées en trois catégories selon le questionnement qu'elles retiennent : (i) la question des effets des échanges internationaux sur les pays échangistes, l'objectif étant alors de savoir si les vertus du libre échange sont avérées ou si ses méfaits doivent être entravés par des politiques protectionnistes, (ii) la question de la détermination des prix des flux internationaux, (iii) la question de la composition des flux internationaux.
Centrons l'analyse sur les éléments importants des grands courants de la théorie économique du commerce international. Nous allons donc examiner tour à tour les avantages du libre échange et les gains à attendre pour une économie de l'ouverture au commerce international. Nous verrons ensuite quelques arguments en faveur du protectionnisme, malgré les pertes de bien être engendrées par la mise en place de barrières tarifaires ou non tarifaires. Nous survolerons ensuite les théories les plus récentes du commerce mondiale, plus à même de décrire les réalités actuelles du commerce mondial. Nous conclurons sur le rôle de l'OMC dans la régulation des échanges mondiaux.

6.2.1 Les arguments pour le libre échange

6.2.1.1 Des avantages absolus (Smith) aux avantages comparatifs (Ricardo)

Nous pouvons tout d'abord donner une définition du libre échange : « *Le libre échange est une doctrine économique qui vise à limiter les obstacles à la circulation des biens, des services et des capitaux entre les économies nationales* ». Pour les auteurs classiques, partisans du libre échange, l'échange international se justifie par la nécessaire division du travail entre les nations qui va permettre des gains de productivité et donc une production à moindre coût.
L'économiste A. Smith (1776) a mis en avant la notion d'**avantage absolu**. Un avantage absolu est obtenu dans l'échange international par une nation qui produit et vend un bien à un prix inférieur à celui des nations concurrentes. Pour Smith, chaque pays a intérêt : (i) à se spécialiser dans les productions pour lesquelles il détient un avantage absolu par rapport aux autres nations, (ii) à se procurer, à moindre coût, les productions pour lesquelles il ne possède aucun avantage par rapport à l'extérieur.
D. Ricardo (1817) a, pour sa part, mis en exergue la notion d'**avantages comparatifs**. Un avantage comparatif est obtenu, dans l'échange international, par une nation, lorsque, comparativement aux autres biens, son désavantage sur un bien, en termes de coût et de prix, est moindre. Pour Ricardo, et sa théorie des avantages comparatifs : (i) les nations sans avantages absolus doivent se spécialiser dans les productions pour lesquelles elles sont le moins désavantagées, c'est-à-dire selon leurs avantages comparatifs, (i) les richesses produites au niveau mondial seront ainsi accrues.
La théorie des avantages comparatifs de Ricardo ne précise pas la répartition des gains à l'échange ou le niveau auquel se fixe les prix relatifs (voir l'annexe en fin chapitre sur **le principe des avantages comparatifs et les gains à l'échange**). Si cette théorie, qui reste encore l'un des arguments majeurs en faveur de l'ouverture aux échanges internationaux, montre la possibilité que les échanges commerciaux soient bénéfiques même pour une économie qui ne présente pas d'avantage absolu, elle ne précise pas vraiment les fondements de la spécialisation internationale. Ce ne sera que plus d'un siècle plus tard, à travers les travaux de Heckscher, Ohlin et Samuelson, que cette lacune sera comblée.

6.2.1.2 L'importance des dotations factorielles (Heckscher, Ohlin, Samuelson)

L'importance de la dotation en facteurs de production dans l'obtention des avantages comparatifs a été mise en avant par les travaux des économistes suédois Heckscher (1919) et Ohlin (1933) puis par les travaux de l'économiste américain Samuelson après la seconde guerre mondiale.
La loi des dotations en facteurs indique que « les nations se spécialisent dans les fabrications qui incorporent le facteur de production qu'elles possèdent en abondance ». Ainsi, les pays en

[92] Voir notamment l'annexe en fin de chapitre sur **les effets de la création d'une union douanière**.

développement vont exporter des produits qui incorporent beaucoup de main d'œuvre, alors que les pays développés exportent des biens nécessitant un capital important pour leur fabrication.
La théorie des dotations en facteur de production conduit donc à une division internationale du travail en termes de complémentarité entre les pays qui échangent, ce qui est conforme aux principes de la théorie des avantages comparatifs de Ricardo.
On considère donc les théories de Ricardo et de Heckscher, Ohlin et Samuelson comme le fondement de la théorie classique des échanges internationaux.
Cependant, on a vu dans l'introduction que les échanges actuels avaient plutôt lieu entre des pays au même niveau de développement (échanges Nord-Nord) et portaient sur des biens similaires (échanges intra-industriels). Le corpus classique en matière d'économie internationale échoue donc à véritablement expliquer la réalité actuelle du commerce mondial. C'est pour cette raison que de nouvelles théories du commerce international se sont imposées à côté de ce corpus théorique principal.

6.2.2 Les nouvelles théories du commerce international

C'est depuis la fin des années soixante-dix que sont appliqués au commerce international les modèles de concurrence imparfaite, principalement issus de l'économie industrielle. L'imperfection de la concurrence est associée dans la plupart des cas à l'existence de rendements croissants au niveau de chaque producteur, qui permettent des comportements non concurrentiels.
Dès lors que les rendements d'échelle sont croissants, il devient intéressant de concentrer la production d'un bien en un seul site, donc dans un seul pays. Lorsqu'il existe, par ailleurs, des barrières à l'entrée dans le secteur, on observe des profits positifs pour les producteurs. Lors de l'ouverture des frontières, de telles rentes peuvent donner lieu à du commerce international car les producteurs souhaitent profiter des rentes disponibles sur les différents marchés.

Les modèles de commerce international en concurrence imparfaite expliquent ainsi des phénomènes empiriques pour lesquels le modèle traditionnel (avantages comparatifs) ne fournissait pas d'explication : (i) l'importance des échanges entre pays d'apparence similaire, (ii) l'importance des flux intra-branches (échanges croisés de produits similaires), (iii) la prise en compte de la stratégie des firmes et le rôle croissant des firmes multinationales[93], (iv) les questions liées aux effets de l'ouverture des frontières. Les effets positifs de l'ouverture, comme l'augmentation du degré de concurrence et les effets de rationalisation, ne peuvent être envisagés que dans un cadre de concurrence imparfaite.

Les **économies d'échelle externes**[94] expliquent pourquoi le commerce **inter-branche** est possible entre deux pays similaires dans un cadre de concurrence parfaite. Les **économies d'échelle internes**, conjuguées avec l'hypothèse de concurrence imparfaite, permettent d'analyser le commerce **intra-branche**[95].

6.2.3 Les arguments en faveur du protectionnisme

Le protectionnisme est une doctrine économique qui a pour but de restreindre l'accès aux marchés nationaux pour les concurrents étrangers. Le protectionnisme repose sur l'idée que le libre-échange n'est pas toujours source d'accroissement des richesses, dans le temps et dans l'espace. L'échange international peut parfois être considéré comme un obstacle au développement ou un facteur d'accroissement des inégalités.
Malgré les effets négatifs[96], en terme de bien-être, de la mise en place de mesures tarifaires (droits de douanes) ou non tarifaires (norme, quota…), certains économistes continuent de défendre l'idée qu'une protection des économies nationales face à la concurrence étrangère par des pratiques protectionnistes est nécessaire ou souhaitable.

---

[93] Les théories traditionnelles ne traitent pas des questions de stratégie des firmes. Le cadre de concurrence parfaite évacue le problème du comportement des firmes et rend difficile l'analyse des questions liées à l'existence et aux actions des firmes multinationales.

[94] On dit qu'il y a des économies d'échelle externes lorsque les rendements sont constants pour chacune des firmes mais que le coût de production baisse avec la production globale du secteur (exemple : le coût de production marginal d'un producteur de vêtement baisse avec l'augmentation du nombre de producteurs dans le pays car chacun apprend des expériences diverses des firmes qui l'entourent dans le secteur). Aucune des firmes n'a investi pour engendrer cette baisse (forme d'externalités).

[95] Les élèves qui souhaitent aller plus loin dans ce domaine peuvent lire le Muchielli, Mayer (2005), Economie Internationale, Dalloz.

[96] Voir l'annexe en fin de chapitre pour une analyse, en termes de surplus, **des effets de la mise en place d'un tarif douanier**.

Les arguments protectionnistes ne datent pas d'aujourd'hui même si la globalisation et l'émergence de concurrents à faibles coûts salariaux tendent à populariser cette doctrine parmi les hommes politiques.
Pour List, économiste allemand de la fin du XIX^{ème} siècle, le surcoût lié au protectionnisme est légitime lorsqu'il permet **l'émergence d'une industrie nationale**. Les industries naissantes sont en effet pénalisées par rapport à la concurrence internationale, parce qu'elles ne bénéficient pas encore d'économies d'échelle suffisantes[97]. Il peut donc être nécessaire de protéger transitoirement les industries naissantes grâce à l'instauration d'un droit de douane.
Certains économistes du développement ont également pointé du doigt l'**inégalité de l'échange international**, ce qui justifierait, à leurs yeux, des mesures protectionnistes de la part des pays qui « subissent » le commerce international.
La spécialisation d'un pays dans les secteurs de haute technologie permet en effet une diffusion des connaissances et de l'expérience pour l'économie nationale (externalités positives) alors que la production de matières premières, de denrées ou de légumes n'assure pas ces effets bénéfiques. L'échange inégal est donc une situation de l'échange international dans laquelle certains pays s'enrichissent grâce à l'échange, au détriment d'autres qui s'appauvrissent[98]. Pour les économistes A. Emmanuel ou S. Amin, le libre échange conduit à une exploitation internationale des pays en développement par les pays développés. Cette exploitation passe notamment par l'intermédiaire des firmes multinationales dont nous allons souligner l'importance dans la prochaine sous-section. Les multinationales obtiennent en effet dans les pays en développement des gains de productivité (faible coût de la main d'œuvre) qu'elles consacrent, en majeure partie, à la baisse des prix des produits qu'elles réexportent vers leur pays d'origine. A l'inverse, les gains de productivité réalisés dans les pays développés sont en général répartis sous la forme de hausses de salaires et des profits (ce qui profite peu aux pays en développement).
Il convient enfin de noter qu'un autre volet du protectionnisme est la promotion des exportations des producteurs nationaux. Cela passe le plus souvent par des subventions allouées aux exportateurs, qui permettent à ceux-ci de vendre leurs produits même si leur coût de production est supérieur au prix mondial[99].

6.2.4 Le rôle des Firmes multinationales

Les théories traditionnelles sur le commerce mondial, qu'elles soient favorables au libre-échange ou au contraire qu'elles prônent le protectionnisme, expliquent insuffisamment la réalité du commerce mondial. Le côté normatif de ces théories a occulté le côté positif. Les principaux acteurs du commerce mondial sont désormais les firmes multinationales qui organisent leurs activités au niveau de la planète.
Une entreprise mondiale est une entreprise qui organise et structure ses activités au niveau mondial. Elle ne considère pas son marché comme une juxtaposition de marchés nationaux cloisonnés mais comme un marché mondial unique.
Le concept de commerce international laisse donc sa place à celui de commerce mondial. Lorsqu'on reste dans le cadre de la nation, on occulte la mobilité internationale des facteurs de production et notamment celle du capital. Or, les investissements des entreprises mondiales à l'étranger montrent qu'une part importante des échanges de biens et services trouve son origine dans les stratégies des entreprises mondiales, sans aucune référence à un quelconque avantage comparatif des pays.
La nationalité d'une firme est celle de son lieu d'enregistrement. Le problème de la nationalité des firmes se pose cependant au consommateur. L'idée sous-jacente est que la nationalité des firmes

[97] Il convient en effet de remarquer que l'ensemble des théories « libérales » est bâti autour de l'hypothèse de rendements d'échelle constants, de sorte que l'augmentation de la production ne conduit pas à une diminution des coûts unitaire de production. On se doute que dans la réalité les rendements d'échelle peuvent être croissants. L'antériorité d'une nation dans la fabrication d'un certain type de produits (effet d'expérience) ou bien encore la taille de son marché intérieur, jouent un aspect non négligeable dans l'obtention d'un avantage comparatif. La protection transitoire permet de laisser les industries naissantes bénéficier d'économies d'échelle et d'effets d'expérience.
Les nouvelles théories du commerce international ont pris en compte des rendements non constants et parviennent à des résultats sensiblement différents de ceux des théories classiques en matière de commerce international.
[98] On dit qu'il y a une dégradation des termes de l'échange. Les termes de l'échange représentent le rapport entre l'indice des prix des biens exportés et l'indice des prix des biens importés. La dégradation des termes de l'échange des pays en développement conduit les pays en développement à travailler toujours davantage pour pouvoir importer une même quantité de biens.
[99] Voir par exemple les subventions aux produits agricoles en Europe et aux Etats-Unis.

suppose une certaine allégeance aux intérêts du pays d'origine. Les entreprises mondiales se caractérisent par le caractère multinational de leur capital. Des entreprises peuvent passer sous le contrôle étranger. Ce qui va compter en premier lieu, c'est l'intérêt des actionnaires (dividendes reçus, valeur du cours de l'action) plus que l'intérêt de la nation.
L'importance des firmes multinationales tient également aux effets qu'elles occasionnent sur la balance commerciale de leur pays d'origine du fait de leur politique de décomposition internationale du processus de production. Lorsqu'elles réimportent des sous-ensembles, fabriqués à l'étranger, d'un produit qu'elles assemblent finalement sur le sol national, cela est comptabilisé comme des importations (des flux physiques passent effectivement la frontière) et cela entraîne une détérioration de la balance commerciale[100].

6.3 La régulation du commerce mondial

Le GATT (*General Agreement on Tarifs and Trade*) est le résultat de négociations entre 23 pays à Genève en 1947. Les craintes de voir les guerres commerciales des années d'avant guerre se reproduire ont largement contribué à la conclusion de cet accord (l'élévation des barrières douanières au cours des années trente avait aggravé la dépression en propageant la crise à travers la réduction des flux de commerce international). Le projet initial comportait la création d'une organisation internationale du commerce (*International Trade. Organization*), placée sous l'égide de l'ONU (accord au cours d'une conférence à la Havane en 1948). Mais le refus américain de ratifier cette création a abouti à ce que le GATT devienne la seule institution de régulation des politiques commerciales internationales jusqu'à sa transformation en une réelle organisation internationale (l'Organisation Mondiale du Commerce (OMC)) en 1995.

6.3.1 Le GATT

Les principaux partenaires du commerce mondial ont mis en place un code de bonne conduite dès 1947 dans le cadre du GATT. Ce code de « bonne conduite » reposait sur deux grands principes : le libéralisme et le multilatéralisme. Le but à atteindre était un libre-échange sans obstacle et, il fallait pour cela procéder au démantèlement des différentes barrières douanières existantes.
L'objectif principal du GATT était d'amener ses membres à s'accorder des baisses multilatérales de droits de douane. Le GATT peut être vu comme un forum de discussion permanent. Les négociations multilatérales (*round*) ont lieu à intervalles réguliers et jouent un rôle important dans le processus.

Un certain nombre d'obligations contraignaient les pays membres : (i) le **principe de non-discrimination** ou clause de la nation la plus favorisée (art. I) stipule que tout avantage tarifaire accordé à un membre doit être étendu à l'ensemble des membres, (ii) le **principe de réciprocité** (art. XXVIII) s'impose en matière d'avantage tarifaire : un pays membre ne peut bénéficier des concessions de ses partenaires sans en accorder lui-même, (iii) le **principe de transparence** (art. XI) est mis en avant : les barrières commerciales non tarifaires (quotas…) doivent être converties en droits de douane afin que leur impact réel devienne transparent.
Deux interdictions viennent se rajouter à ces principes: (i) le **dumping** est interdit. En réponse à cette pratique, le recours par les pays membres à des droits anti-dumping est considéré comme légitime (art. VI), (ii) les **subventions** qui maintiennent des prix artificiellement faibles sont également interdites et peuvent donner lieu à des droits de douane compensatoires (art. XXVIII).
Cependant, les exceptions nombreuses sont de trois types : (i) exceptions aux principes : Autorisation des accords de libre échange et des unions douanières en violation du principe de la nation la plus favorisée, (ii) exceptions pour les pays en développement : certaines restrictions quantitatives sont autorisées en cas de problèmes grave de leur balance des paiements, (iii) exceptions en pratique : Deux industries « sensibles »: l'agriculture et le textile.

Les Etats-Unis, le Japon et l'Union Européenne protègent leur agriculture (puissants groupes de pression) malgré les plaintes répétées de certains pays en développement qui pâtissent des cours faibles de certaines denrées ou productions agricoles sur les marchés mondiaux. Les secteurs liés au textile/habillement ont bénéficié des Accords Multifibres (Accords entre les pays développés (les grands importateurs) et les pays en développement (grands exportateurs de ce secteur)) qui visaient à réguler et essentiellement à contingenter les quantités échangées. Ces accords avaient été négociés sous les auspices du GATT comme une exception aux règles générales mais ont pris fin le

[100] Cet effet n'est pas totalement négligeable, par exemple pour les Etats-Unis.

1er janvier 2005 conformément à ce qui avait été signé lors de l'Uruguay round (accords dits de Marrakech). On se souvient de l'arrivée massive de biens d'habillement en provenance de Chine et les politiques protectionnistes qui n'ont pas tardé aux Etats-Unis et en Europe…
Les négociations du GATT ont abouti à une baisse considérable des droits de douanes entre les pays membres (voir tableaux ci dessous) qui s'est étalée tout au long de la seconde partie du XXème siècle.

Cycles de négociation	Nombre de participants	Résultats
Genève (1947)	23	Baisse moyenne de 15% sur 50% du commerce mondial (produits manufacturés).
Annecy (1949)	13	Baisse moyenne de 25% des tarifs douaniers.
Torquay (1950-51)	38	Réductions tarifaires sur 8700 nouveaux produits
Genève (1955-56)	26	Peu de réductions tarifaires
Dillon Round (1961-62)	26	Réduction tarifaire de 10% environ.

Cycles de négociation	Nb. de participants	Résultats
Kennedy Round (1964-67)	62	Réduction tarifaire de 35% sur les produits. Manufacturés et de 20% sur les produits agricoles. Code antidumping.
Tokyo Round (1973-79)	99	Réduction tarifaire de 34% sur les produits manufacturés. Accords sur les barrières non tarifaires.
Uruguay Round (1986-1994)	115	Création de l'OMC. Accords sur l'agriculture, le textile, les droits de propriété intellectuelle.
Development Round (Doha, 2001)	135	Echec des négociations (Seattle 1999, Cancun 2003, Hong kong 2005)

Tableau 13 : Les négociations commerciales multilatérales

Nous pouvons nous intéresser plus particulièrement à la dernière négociation du GATT ayant aboutie, celle dite de **l'Uruguay Round**. Les négociations avaient débuté en septembre 1986 à Punta del este (en Uruguay). L'acte final a été signé à Marrakech le 15 avril 1994 (vous pouvez ainsi vous rendre compte de la longueur et de la complexité de ces négociations).

On peut distinguer quatre grands thèmes dans l'accord final de Marrakech : (i) la réductions des barrières aux échanges : les tarifs des pays développés baissent sur les biens manufacturés de 38% (taux moyen de 6,3 à 3,9%), (ii) la prise en compte du changement du rôle des pays en développement : Jusqu'à l'Uruguay Round les pays en développement ne prenaient pas part aux négociations et en conséquence, le commerce de produits constituants des avantages comparatifs pour ces pays n'avait pas été libéralisé de manière significative (exemple des accords multifibres qui est largement à l'avantage de ces pays), (iii) de nouveaux sujets apparaissent dont l'importance ne pourra qu'être croissante dans le futur (GATS : accord sur les services, TRIPS : accord sur les droits de propriété concernant le commerce), (iv) création de l'OMC
L'Uruguay Round aboutit donc à la création de l'Organisation Mondiale du Commerce, c'est-à-dire d'une véritable organisation dotée de pouvoir de sanction en cas de manquement aux accords ou de pratiques protectionnistes, en lieu et place du GATT qui n'était qu'un simple club de négociations.

6.3.2 La mise en place de l'OMC et son agenda

L'OMC (Organisation Mondiale du Commerce) est un organisme, issu des accords de l'Uruguay Round qui est chargé d'établir et de faire respecter, par la concertation entre les pays membres, les règles du commerce international. Tous les membres de l'OMC, y compris les pays en développement, doivent signer tous les accords qui font partie de l'accord multilatéral sur le commerce des biens, y compris les accords GATS sur les services et TRIPS sur les droits de propriété intellectuelle.
L'OMC s'est dotée d'un organe de règlement des différends (ORD). Les Etats qui considèrent que les règles du commerce international ne sont pas observées, peuvent saisir l'ORD. De nombreux conflits existent entre les Etats-Unis et l'Europe (notamment sur les questions agricoles ou dans le secteur aéronautique).
Les principaux points du « programme de travail » de l'OMC pour le round du développement (débuté à Doha en 2001) sont les suivants : (i) **la mise en œuvre** : De nombreux points censés être réglés par l'Uruguay Round ne l'ont toujours pas été (accès aux marchés du Nord pour les exportateurs du Sud). Il existe donc une demande forte de la part des pays en développement qu'avant le démarrage de négociations sur de nouveaux sujets, soient mises en œuvre les décisions prises il y a plus de dix ans, (ii) **l'agriculture** où de grandes difficultés sont rencontrées pour parvenir à un accord (réticences des Etats-Unis et de l'Europe à faire des concessions sur ce point), (iii) l'interprétration de l'**accord TRIPs** d'une manière favorable à la santé publique (problème des médicaments génériques et de la mise à disposition des pays du Sud de molécules dont les droits de propriété industrielle sont au Nord, en cas de pandémie), (iv) la mise en œuvre de l'accord GATS sur les services, (v) l'accès aux marchés des pays développés pour les produits non agricoles en provenance des pays en développement.

Les négociations devaient être conclues au plus tard le 1er janvier 2005. La raison du retard et de l'échec de ces négociations est l'opposition entre pays riches et pays en développement (regroupés en groupes de négociations comprenant le Brésil, la Chine, l'Inde) sur les dossiers les plus épineux. Les pays du Nord veulent introduire de nouveaux sujets de négociations ou approfondir certains (IDE, marchés publics, Droits de Propriété Intellectuelle) et améliorer l'accès de leurs produits industriels dans les pays en développement tandis que les pays du Sud sont essentiellement intéressés par la réalisation de progrès dans le domaine du protectionnisme et des subventions agricoles pratiquées par les pays riches.

6.4 Finance internationale

L'ensemble des transactions d'un pays avec le reste du monde, échanges de biens et services et flux financiers, est enregistré dans un document comptable : la balance des paiements. La balance des paiements fait apparaître un certain nombre de soldes caractéristiques et constitue un instrument utile pour juger de l'ampleur et des déséquilibres des échanges au niveau national mais également au niveau mondial. La résorption des déséquilibres en matière de balance des paiements est en effet affectée par les caractéristiques du système monétaire et financier international (monnaie de référence au niveau international, types de régime de change) qui ont profondément évolué depuis la fin de la seconde guerre mondiale.
Dans un premier temps, nous donnerons un aperçu du fonctionnement du marché des changes avant de présenter la balance des paiements. Nous exposerons ensuite les différentes architectures du système monétaire et financier international avant de conclure sur les grands enjeux concernant le fonctionnement et la pérennité du système monétaire et financier international.

6.4.1 Le taux de change

6.4.1.1 Quelques définitions

Le passage d'une monnaie à une autre nécessite l'intermédiaire d'un taux de change. Le taux de change est obtenu par confrontation de l'offre et de la demande d'une monnaie sur le marché des changes (ce n'est rien d'autre qu'un prix de marché). Le change est l'opération qui permet d'obtenir une monnaie étrangère à partir d'une monnaie nationale (ou inversement).
On peut donc s'arrêter sur la définition suivante du **taux de change nominal** : « *Le taux de change est le prix d'une monnaie exprimé en autre monnaie. C'est le nombre d'unités monétaires que l'on peut obtenir en échange d'une unité de l'autre monnaie* ».
Il existe deux types de cotation pour un taux de change : **la cotation à l'incertain** (la monnaie étrangère est exprimée en monnaie nationale, exemple 1$=0,77€) et **la cotation au certain** (la monnaie nationale est exprimée en unité de monnaie étrangère, 1€=1,3$ avec le même exemple). Dans cette définition, il n'est fait aucune référence aux différences de pouvoir d'achat qui peuvent exister entre les monnaies. Pour cela, il faut calculer un taux de change réel, c'est-à-dire établir une comparaison du prix des biens entre deux pays. Soit p^e l'indice des prix à l'étranger et p l'indice des prix pour la zone euro. Le **taux de change réel** (à l'incertain) est $\mathbf{e_r=e_n \cdot p^e / p}$ avec e_n, le taux de change nominal.
Au cours d'une même période, une monnaie peut s'apprécier par rapport à certaines devises et se déprécier par rapport à d'autres. Cependant, pour un pays, les différentes devises n'ont pas la même importance. Le **taux de change effectif** est le taux de change moyen d'une monnaie avec un ensemble de devises (c'est une moyenne pondérée des différents taux de change bilatéraux, dans laquelle les pondérations sont constituées par les poids respectifs de chacun des pays dans les flux de commerce extérieur du pays concerné).
Dernier point, les taux de change s'inscrivent dans un système monétaire international bien déterminé. Ainsi, le **taux de change** peut être **fixe**. Le taux de change fixe est un taux de change fixé à l'avance dans le cadre d'accords internationaux, par référence à un étalon (l'étalon peut être l'or, une autre devise ou un panier de devises). Le **taux de change** peut au contraire être **flexible** (ou flottant). C'est alors un taux de change qui varie librement sur le marché des changes en fonction de l'offre et de la demande pour cette monnaie. La différence entre les deux systèmes est que dans le cadre d'un système de change flottant, il n'y a pas d'engagement des banques centrales à défendre la parité de leurs monnaies pour conserver un taux de change relativement constant par rapport aux

autres monnaies (cours de référence)[101]. Lorsqu'on est dans un système de taux de change fixe est qu'une monnaie est forcée d'adopter un nouveau cours de référence plus faible, on dit qu'il y a **dévaluation**. Lorsqu'on est dans un système de taux de change flottant et qu'une monnaie évolue à la baisse, on dit qu'il y a **dépréciation**[102].

Selon la **théorie de la parité des pouvoirs d'achat (PPA)**[103], le taux de change bilatéral est déterminé par le rapport des pouvoirs d'achat interne de deux monnaies. Le taux de change tend à assurer l'égalité des pouvoirs d'achat entre les deux monnaies. Il existe deux versions de cette théorie : la version absolue et la version relative. Dans la version absolue, un agent économique doit pouvoir se procurer un même panier de biens dans la zone euro et hors de la zone euro (une fois les euros convertis) par exemple. Dans la version relative, ce sont les variations plus que les niveaux des taux de change qui sont étudiés. Le pouvoir d'achat d'une monnaie se dégrade avec l'inflation. Lorsqu'il existe un différentiel d'inflation entre deux zones monétaires, la valeur de la monnaie de la zone dont le pouvoir d'achat se dégrade se déprécie afin que la parité des pouvoirs d'achat soit maintenue.

6.4.1.2 Le marché des changes

Le marché des changes est le lieu de rencontre entre offreurs et demandeurs pour un échange de devises ce qui aboutit à une fixation des prix des devises exprimées entre elles (taux de change). On peut considérer qu'il n'existe qu'un seul marché des changes dans le monde puisque les transactions se réalisent en même temps en plusieurs points du globe. C'est un marché en continu puisque les cours sont côtés 24h/24 en raison des décalages horaires (il y a toujours une place financière ouverte où négocier une devise). Il existe plusieurs types d'intervenants sur le marché des changes. Les banques et les courtiers interviennent sur le marché des changes pour le compte des entreprises et des particuliers (rôle d'intermédiaires). Certaines grandes entreprises ont notamment développé leur propre salle des marchés. Les banques centrales interviennent également pour réguler les fluctuations du cours de leur monnaie. Le marché des changes est donc avant tout un marché interbancaire.

Il existe deux types de marché : le **marché au comptant** et le **marché à terme**. Sur le marché au comptant, la livraison de devises se fait immédiatement (dans les deux jours ouvrables après la négociation), par transfert de compte à compte. Sur le marché à terme, la livraison est reportée à une échéance et à un cours convenus à l'avance (le cours est appelé cours à terme). Le marché des changes à terme permet aux agents économiques de se couvrir contre le risque de change et de spéculer (les deux sont liés puisque pour se couvrir contre un risque, il faut trouver une contrepartie qui accepte de prendre ce risque. Le spéculateur constitue cette contrepartie).

Nous pouvons à présent décrire le fonctionnement de la balance des paiements.

6.4.2 La Balance des paiements

6.4.2.1 Définition et principes

La balance des paiements est un état statistique qui retrace sous forme comptable l'ensemble des flux d'actifs réels, financiers et monétaires entre les résidents d'un pays et les non résidents au cours d'une période donnée. Un résident est une personne physique qui a son domicile principal dans le pays considéré.[104]

La balance des paiements est un compte tenu en crédit (+) et débit (-) conformément aux principes de la comptabilité en partie double. Une opération entre le pays déclarant et le reste du monde donne ainsi toujours lieu à deux inscriptions de même montant de signe contraire. Une inscription traduit la nature économique de la transaction (exportations de marchandises par exemple) tandis qu'une autre traduit le mode de règlement de la transaction. Quel que soit la nature du flux (biens et services, monétaires, financiers), le flux d'un résident vers un non-résident s'inscrit en crédit, le flux d'un non-résident vers un résident s'inscrit en débit.

---

[101] D'où l'absence de la réaction de la BCE à l'appréciation de l'euro face au dollar au printemps 2007. Vous en aurez déduit que nous sommes donc, de nos jours, dans un système de change flexible (voir plus loin pour l'historique…).

[102] Les effets d'une dévaluation (ou d'une dépréciation) sur le solde commercial sont étudiés en annexes (**le théorème des élasticités critiques et la courbe en J**).

[103] C'est l'économiste suédois Cassel qui est l'auteur de la théorie de la parité des pouvoirs d'achat en 1916.

[104] Il résulte de la définition du terme de résident que les échanges entre les sociétés mères et filiales des firmes multinationales sont comptabilisés comme transactions internationales et apparaissent dans la balance des paiements.

Ainsi, une valeur inscrite en crédit traduit une diminution des avoirs sur l'extérieur (augmentation des engagements vis-à-vis de l'extérieur) tandis qu'une valeur inscrite en débit se traduit par une augmentation des avoirs sur l'extérieur (diminution des engagements vis-à-vis de l'extérieur).
Comme les enregistrements comportent quelques erreurs, une ligne erreurs et omissions permet d'équilibrer les masses figurant en crédit et débit.
La balance des paiements est un compte de flux et non de stocks. On y trouve ainsi la variation des avoirs et des dettes d'un pays vis-à-vis du reste du monde au cours d'une période donnée, par le montant de ces avoirs et dettes. La balance des paiements est de ce fait un document équilibré par construction comptable.

6.4.2.2 Balance des paiements et ajustements sur le marché des changes

Les différentes transactions qui sont prises en compte dans la balance des paiements sont regroupées par types d'opérations appelés comptes. Chaque groupe d'opérations donne lieu à un solde (présentant un déficit ou un excédent). La balance des paiements globale étant forcément équilibrée, les soldes se compensent entre eux.
Lorsque la balance est équilibrée, cela signifie que le déséquilibre du solde courant est exactement compensé par les mouvements de capitaux entre résidents et non résidents (la hausse de l'endettement du pays permet par exemple de payer son surplus d'importations par rapport aux exportations).
Lorsque la balance n'est pas équilibrée, cela signifie que le pays doit avoir recours à un financement monétaire de ce déséquilibre. Si un pays est en excédent à la fois au niveau de sa balance des transactions courantes et au niveau des flux de capitaux, le pays accumule des devises ce qui fait augmenter sa masse monétaire (cf. le chapitre 3 sur la monnaie et le financement).
L'inverse peut se produire en cas de déficit de la balance globale, ce qui entraîne une réduction de la masse monétaire du pays à la suite de la baisse de réserves utilisées pour couvrir le déficit.
Les raisonnements ci-dessus font cependant l'hypothèse que l'ajustement sur le marché des changes se réalise par les quantités. Cela peut être le cas lorsque le régime de change est un régime de taux de change fixe. L'ajustement peut néanmoins également se réaliser par les prix, grâce à la flexibilité du taux de change.
Ainsi, en régime de taux de changes fixes, si les échanges sont libellés en dollars et que la zone euro a un déficit de sa balance global, il va en résulter une demande excédentaire de dollars pour payer les importations excédentaires. Cela engendre une pression à la baisse sur l'euro. Pour assurer la fixité du taux de change, la Banque Centrale Européenne devrait dans ce cas intervenir sur le marché pour acheter sa propre monnaie (elle peut réaliser ses interventions soit en puisant dans les réserves de change ou en empruntant des dollars).
En régime de taux de changes flexibles (ou flottants), ce qui est le cas actuellement, en reprenant le même exemple d'une balance globale européenne déficitaire, la demande excessive de dollars entraîne par conséquent une chute de l'euro. La réciproque est vraie : un excédent de la balance globale se traduirait par une appréciation de l'euro.

Crédit	Débit	Solde
Compte des transactions courantes		
=>Exportations de marchandises	=>Importations de marchandises	Solde commercial
=>Exportations de services	=>Importations de services	Solde des biens et services
=>Revenus des salariés et des investissements reçus	=>Revenus des salariés et des investissements versés	
=>Transferts courants reçus	=>Transferts courants versés	Soldes des opérations courantes
Compte de capital		
=>Transferts de capital reçus	=>Transferts de capital versés	
=>Vente de brevets	=>Achat de brevets	Solde des opérations courantes et du compte de capital
Compte financier		
=>IDE entrants	=>IDE sortants	Solde à financer
=>Investissements de portefeuille entrants	=>Investissements de portefeuille sortants	
=>Autres investissements entrants	=>Autres investissements sortants	
=>Désinvestissement des résidents	=>Désinvestissement des non résidents	Solde de la balance globale
=>Réduction des avoirs de réserves des autorités monétaires	=>Augmentation des avoirs de réserves des autorités monétaires	Variation de la position monétaire extérieure
Erreurs et omissions nettes		

Tableau 14 : Les différents comptes et soldes de la balance des paiements

Nous illustrons cet exposé par la présentation des colonnes crédit et débit ainsi que du solde de la balance des paiements pour la France (année 2001)

milliards d'euros	**Crédit**	**Débit**	**Solde**
Comptes des transactions courantes	**521,8**	**498,0**	**23,8**
Biens	324,6	320,6	4,0
Services	89,5	69,7	19,8
Revenus	88,6	72,1	16,5
Transferts courants	19,1	36,6	-17,5
Compte de capital	**1,2**	**1,5**	**-0,3**
Compte financier	**4117,1**	**4144,6**	**-27,5**
Investissements directs	78,1	111,8	-33,7
Investissements de portefeuille	3984,0	3963,0	21,0
Autres investissements	45,4	68,7	-23,3
Produits financiers dérivés	2,8	-	-
Avoirs de réserve	6,8	1,1	5,7
Erreurs et omissions nettes	**4,0**	-	**4,0**
Total général	*4644,1*	*4644,1*	*0,0*

Tableau 15 : La balance des paiements pour la France (2001)

L'un des intérêts de la balance des paiements est de faire apparaître les différents déséquilibres associés à certains types de transactions et les opérations nécessaires à la compensation de ces déséquilibres.

6.4.3 Le système monétaire et financier international

En matière de système monétaire et financier international, on fait sempiternellement référence au système dit de Bretton Woods, mis en place au sortir de la seconde guerre mondiale. Pourquoi plus de soixante ans après sa mise sur pied (1944) et trente ans après sa disparition (1971 puis 1976) cette nostalgie pour le système monétaire et financier de Bretton Woods. Fait-on crédit à celui-ci d'une stabilité qui aurait permis les décennies de croissance forte d'après guerre ? Ce serait oublier que ce système n'était pas parfait et qu'il y avait également des crises (notamment dues aux fortes rigidités et aux contraintes introduites par ce système de changes fixes !).

Cette nostalgie provient peut-être des vertiges que donne l'actuel système monétaire et financier (ou plutôt sa quasi-absence). Désormais, la liberté de circulation des capitaux est totale et les changes sont flottants. Le nouveau système apparaît ainsi moins contraignant pour le rééquilibrage de la balance des paiements qui est automatique à travers les fluctuations des taux de change, mais il est également facteur d'instabilité pour de nombreux pays (pour preuve les nombreuses crises monétaires et financières traversées par des pays en développement ou émergents au cours de la dernière décennie : Thaïlande, Russie, Mexique, Brésil, Argentine…). Est-ce que le système de change flottant est responsable de ces crises ou est-ce plutôt un manque de régulation du système qui engendre des dysfonctionnements ?

Donnons-nous d'abord une définition d'un système monétaire international : « *Le système monétaire international est l'ensemble des règles et des institutions visant à organiser les échanges monétaires internationaux* ».

Dans un premier temps, nous retracerons la mise sur pied puis l'abandon du système de Bretton Woods, avant de discuter des problèmes et des enjeux du système actuel.

6.4.4 Vie et mort du système de Bretton Woods

Les accords de Bretton Woods sont signés en 1944, pour ne pas que se reproduisent les dérives monétaires de l'entre-deux guerres. Ils sont donc destinés à établir un cadre monétaire stable à l'échelle internationale. Les accords de Bretton Woods prévoient la mise en place d'un système de change fixe et la création du Fond Monétaire Internationale (FMI)[105]. Le FMI est l'institution garante de la stabilité du système monétaire et financier international.

Chaque pays voit la parité de sa monnaie définie soit en $ soit en or. La parité entre les deux étalons est fixé à 35$ l'once d'or (1once=28,349 grammes). On a appelé ce régime, qui a duré de 1944 à 1971, le régime d'**étalon de change or** (ou *Gold Exchange Standard*) car il consacrait un étalon (*standard*) construit autour d'une devise (*exchange*), elle-même convertible en or (*gold*). Le système reposait donc autour du pivot constitué par le dollar et sa pérennité dépendait des interventions des

[105] En 1944, la Banque Mondiale est également créée. Sa mission est de lutter contre la pauvreté à l'aide des ressources apportées par les pays les plus riches. Elle intervient dans les pays en développement en prodiguant des conseils et apporte des aides matérielles et financières.

banques centrales des pays adhérents, lorsque la parité de leurs devises s'éloignait de plus ou moins 1% par rapport à la parité fixe vis-à-vis du dollar.
Le Fond Monétaire International (FMI) était le garant de la stabilité des changes. Les Etats devaient demander son autorisation pour des changements de parité dépassant 10%. Le FMI dispose par ailleurs de ressources émanant des adhérents (les quotes-parts). Il peut ainsi intervenir pour soutenir les pays qui éprouvent des difficultés à équilibrer leur balance des paiements (ce qu'il continue de faire par exemple lors des crises des pays émergents).
Le système de Bretton Woods souffrait de trois limitations techniques : (i) le privilège donné au pays émetteur de la monnaie internationale (les Etats-Unis donc) de pouvoir financer leurs déficits puisque les Etats-Unis étaient exonérés de sanction en cas de tendance au déséquilibre de leur balance des paiements, (ii) la soumission des autres pays participants au système à la politique monétaire des Etats-Unis, (iii) le principe des changes fixes s'accomode mal des pressions exercées par l'intégration financière et l'essor des mouvements de capitaux (triangle des incompatibilités de Mundell).
L'augmentation de la quantité de dollars en circulation dans l'économie mondiale a progressivement causé la perte du système de Bretton Woods[106]. Le rapport entre la quantité de dollars en circulation et le stock d'or mondial s'était profondément modifié au cours du temps.
Devant le risque de demandes de conversion des dollars en or, le Président Nixon décida unilatéralement la suspension de la convertibilité des deux étalons, le 15 août 1971. En 1976, la conférence de la Jamaïque entérine la fin des taux de change fixe. Le système de taux de changes flottants succéda alors naturellement au système d'étalon de change-or.

6.4.5 Le régime actuel de changes flottants

Le système actuel de taux de changes flottants provoque des variations de grande envergure des cours des devises, notamment à court terme. Ces fluctuations entraînent des risques de change et donc des coûts supplémentaires pour les agents et les entreprises qui souhaitent se couvrir contre ce risque. Ainsi, depuis trois décennies le dollar a connu des fluctuations très importantes. Or, même si il n'est plus l'étalon du système de changes fixes de Bretton Woods, le dollar occupe une place prépondérante (c'est une monnaie de réserve pour les Banques centrales, c'est également la monnaie valorisant le plus souvent les échanges internationaux). Aussi, ces variations poussent les membres du G7[107] à intervenir lorsque les variations atteignent des seuils jugés critiques (les interventions du G7 en faveur d'une revalorisation du dollar ou d'une baisse de celui-ci sont signalées par des flèches sur le graphique ci-dessous).

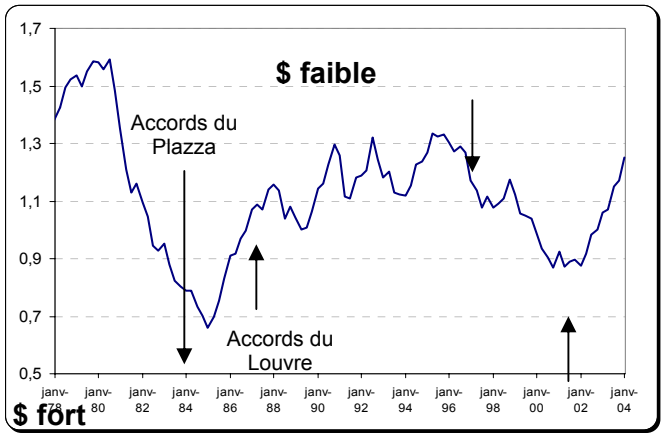


Figure 58 : Evolution du $ euros (1€ = x$)

[106] Le phénomène débute avec l'aide des Etats-Unis aux pays européens pour leur reconstruction (plan Marshall) mais également avec les investissements de entreprises américaines en Europe. Le deuxième flux de dollars en dehors des Etats-Unis est lié à la mutation des échanges commerciaux. A partir de la fin des années cinquante, les excédents commerciaux américains se réduisent. Les européens et le Japon commencent à exporter vers les Etats-Unis. Dans les années soixante apparaissent les **euro-dollars**. Il s'agit de dollar déposés sur des comptes bancaires en dehors des Etats-Unis ce qui permet aux banques qui les ont en dépôts d'accorder des crédits libellés en dollar. Cela provoque donc une création de devises américaines en dehors de Etats-Unis (voir le chapitre 3 sur la monnaie et le financement pour les mécanismes de création monétaire).
[107] Le G7 regroupe périodiquement les dirigeants des principales économies : Etats-Unis, Japon, Allemagne, France, Royaume-Uni, Italie, Canada. Quand la Russie se joint à eux, on appelle alors cela le G8. Lors des derniers sommets, les dirigeants du G8 ont souvent convié les représentants des grands pays émergents (Chine, Brésil, Inde…), signe que les décisions en matières économiques, financières et monétaires peuvent désormais difficilement être prises sans eux.

Si la décennie soixante-dix a été celle de la transformation du système de changes, les années quatre-vingts ont vu l'apparition d'un véritable système financier international (les fameux 3D ; déréglementation, décloisonnement, désintermédiation, ont permis l'amplification des flux de capitaux).
Les mouvements de capitaux ont des effets positifs et négatifs. Ils permettent le financement des déficits des Etats sans crédits et donc sans création monétaire, organisent le rééquilibrage mondial en matière d'épargne et contribuent ainsi au financement des pays en voie de développement[108]. Cependant, la puissance des fonds de pension constitue une contrainte et une menace pour de nombreux Etats ou entreprises. On constate l'apparition d'une spéculation à l'échelle mondiale qui peut aboutir à la déstabilisation complète de certaines économies et donc à des crises financières et monétaires lorsque les appréciations des investisseurs sur le potentiel de certaines économies émergentes évoluent (de manière fondées ou non).

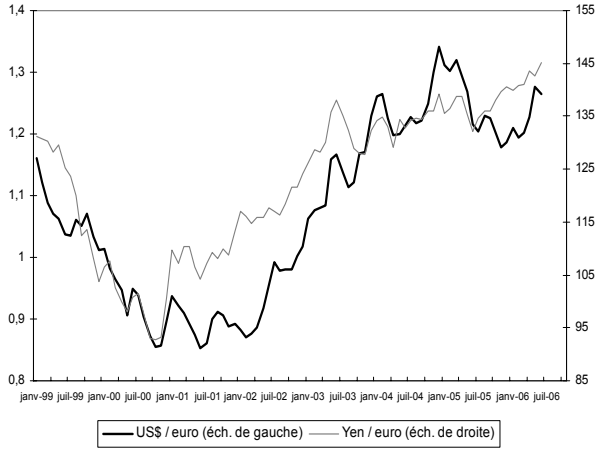


Figure 59 : Evolution du $ et du Yen en euros

Le système monétaire et financier actuel est donc fortement instable, notamment du fait des déficits jumeaux aux Etats-Unis (déficit public et déficit commercial importants) qui pèsent sur la valeur du dollar (voir le graphique ci-dessus). Une accélération de la baisse de celui-ci aurait des conséquences dramatiques pour l'économie mondiale (voir l'introduction).
La gestion alternative (les fameux *hedge funds*) qui vise a décorréler les performances des portefeuilles et l'évolution générale des bourses a abouti à une augmentation du risque systémique du fait du recours de plus en plus important aux produits dérivés, de l'utilisation de l'effet levier (endettement pour augmenter la taille des portefeuilles et donc démultiplier les gains), des placements vers les marchés émergents (qui procurent une rentabilité plus forte au prix de risques plus importants)… Ainsi, les crises de change dans certains pays amènent des arbitrages des investisseurs en faveur de la liquidité et de la sécurité, ce qui occasionne des mouvements de capitaux rapides et de grande ampleur (car ces capitaux sont à court terme). Ces mouvements de capitaux déstabilisent les places financières, y compris celles de pays qui n'étaient initialement pas concernés par cette crise[109].
A cette instabilité de la finance mondiale, s'ajoute par ailleurs une incertitude sur l'aboutissement actuel des négociations à l'OMC (présidée par Pascal Lamy) qui jette un trouble sur la poursuite de la croissance du commerce mondial[110].

L'Europe se doit de parvenir à relancer les négociations du cycle de Doha, sans pour cela accorder trop de concessions, afin de pouvoir continuer à bénéficier de la croissance mondiale plus forte que la croissance européenne. La France doit pour sa part cesser de regarder la mondialisation avec crainte et tenter de gagner une place de choix dans le commerce mondial. Cela ne pourra se faire que par un recentrage sur des productions à plus forte valeur ajoutée et de haute technologie, pour lesquelles la

[108] On a constaté entre 1990 et 1997 une multiplication par 6 des flux d'origine privée vers les pays en développement. Cependant la répartition de ces mouvements de capitaux internationaux est très inéquitable (seuls 2% des capitaux privés se dirigent vers l'Afrique sub-saharienne).
[109] Suite à la crise mexicaine en décembre 1994, d'autres pays alentours subirent les effets de cette crise monétaire et financière : « effet tequila ». La crise brésilienne en 1999 occasionna également une déstabilisation des économies environnantes : « effet samba ».
[110] Etant donnés les désaccords persistants entre les Etats, notamment au sujet de l'agriculture, les négociations du cycle de Doha, ont été suspendues en juillet 2006, même si des discussions se poursuivent.

pression des pays émergents n'est pas encore aussi forte que sur celles des biens incorporant massivement du travail non qualifié. Cette tentative de parvenir à un regain de compétitivité nécessite au préalable le développer les capacités d'offre de l'économie française. Cela passe par une politique économique visant à accroître le capital humain et le potentiel d'innovation des entreprises. Seule une telle politique permettrait de stopper les pertes de parts de marché à l'exportation (et le déficit de la balance commerciale qui va avec), stimulerait la croissance et créerait de l'emploi (trois objectifs du « carré magique » de Kaldor… quant à l'inflation laissons faire la BCE !).

Un tel choix est possible et s'inscrit dans la stratégie de Lisbonne décidée par l'Union européenne en 2000 pour améliorer la croissance et l'emploi en Europe en faisant le pari de l'éducation et de l'innovation. C'est ce que nous allons voir dans le prochain et dernier chapitre.

6.5 Annexes

6.5.1 Le principe des avantages comparatifs et les gains à l'échange

Le principe des avantages comparatifs est le suivant : « *L'échange international est générateur de gains lorsque les pays se spécialisent dans la production des biens pour lesquels leurs coûts relatifs de production sont les plus faibles et importent les biens qui supportent les coûts relatifs de production les plus élevés* ».

L'avantage comparatif provient de la différence de coût relatif en travail. Chaque pays a intérêt à se spécialiser dans le bien pour lequel son coût relatif du travail est le plus faible (bien dont il dispose en abondance). Le prix de l'échange international sera compris entre les prix prévalant en autarcie dans chacun des pays. Sa position dépendra des conditions exactes des demandes de biens dans chacun des pays. L'importance du gain à l'échange va dépendre de la différence entre le prix d'échange et le prix d'autarcie.

Equilibre d'un pays en autarcie

Un pays produit deux biens (1 et 2) en quantité q_1 et q_2 en utilisant comme seul facteur le travail (disponible en quantité L). Les quantités consommées des deux biens sont d_1 et d_2. Les quantités de travail nécessaires pour produire une unité de chaque bien sont c_1 et c_2 (les rendements d'échelle sont constants). Le bien 1 et le numéraire. Q est le revenu national et w est le taux de salaire en unités de bien 1. $p=p_2/p_1$ est le prix relatif du bien 2 en unités de bien 1. L'intégralité du revenu national est consommée. Les préférences du consommateur impliquent qu'une part b du revenu national est affectée à la consommation du bien 1 (0<b<1) et (1-b) à la consommation de bien 2.

Les équations du modèle sont les suivantes :

$$c_1 q_1 + c_2 q_2 \le L$$
$$Q = q_1 + p q_2$$
$$Q = d_1 + p d_2$$
$$Q = w\left(c_1 q_1 + c_2 q_2\right)$$
$$d_1 = bQ$$
$$q_1 = d_1$$

La solution du modèle d'autarcie est :

$$p = \frac{c_2}{c_1}$$
$$Q = \frac{L}{c_1}$$
$$q_1 = d_1 = b\frac{L}{c_1}$$
$$q_2 = (1-b)\frac{L}{c_2}$$
$$w = \frac{1}{c_1}$$

Le prix d'autarcie est égal au coût relatif du bien 2 ($p=c_2/c_1$), c'est-à-dire la valeur absolue de la pente de la droite qui décrit la frontière des possibilités de production.

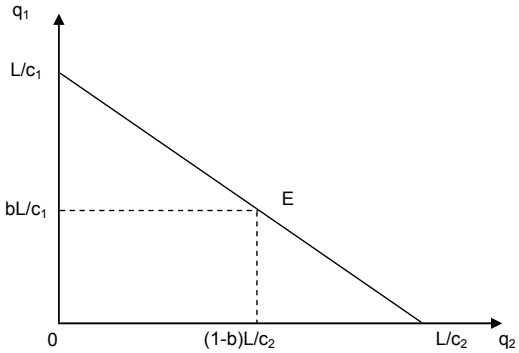


Figure 60 : Equilibre d'un pays en autarcie

Equilibre d'un pays en économie ouverte

Le prix p devient une donnée exogène émanant du marché international. Les productions (q_i) et les consommations (d_i) peuvent désormais différer grâce à l'ouverture sur l'extérieur (la seule contrainte

imposée étant l'équilibre de la balance commerciale). Soit $e_i = q_i - d_i$ la quantité de bien i échangée avec l'extérieur (e_i>0 correspond à des exportations de bien i, e_i<0 à des importations).

Le modèle est désormais :

$c_1 q_1 + c_2 q_2 \le L$

$Q = q_1 + p q_2$

$Q = d_1 + p d_2$

$Q = w(c_1 q_1 + c_2 q_2)$

$d_1 = bQ$

$q_1 = d_1 + e_1$

$e_1 + p e_2 = 0$

La solution du modèle est alors :

$q_1 = \frac{L}{c_1} \qquad q_2 = 0$

$Q = \frac{L}{c_1} \qquad w = \frac{1}{c_1}$

$d_1 = b\frac{L}{c_1} \qquad d_2 = (1-b)\frac{L}{c_1}\frac{1}{p}$

$e_1 = (1-b)\frac{L}{c_1} \qquad e_2 = -(1-b)\frac{L}{c_1}$

Le pays a intérêt à l'échange international si le prix p du marché mondial est différent du prix d'autarcie ($p^a=c_2/c_1$). Le gain à l'échange se traduit par une augmentation de la consommation du bien importé.

Bien entendu trois cas sont possibles : (i) $p<p^a=c_2/c_1$, (ii) $p>p^a=c_2/c_1$, (iii) $p=p^a=c_2/c_1$ (i).

Dans le cas (i), le pays dispose d'un avantage comparatif pour le bien 1 pour lequel le coût relatif de production est inférieur au prix mondial 1/p. Le pays a donc intérêt à se spécialiser entièrement dans cette branche et à importer du bien 2. Le point de production est donc en A. Le commerce international permet au pays en exportant le bien 1 et en important le bien 2 d'accéder au point de consommation C, extérieur à l'ensemble des possibilités de production en autarcie. Le gain collectif à l'échange, par rapport à l'autarcie, se caractérise alors par une augmentation de la consommation du bien 2 représentée par le segment EC. Ce gain est d'autant plus important que le prix p, le prix mondial, est faible (la pente est faible en valeur absolue).

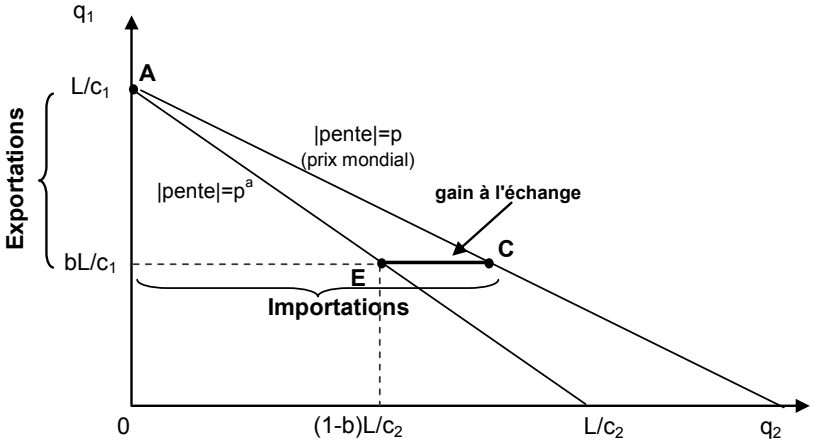


Figure 61 : Equilibre d'un pays en économie ouverte
cas de l'avantage comparatif en bien 1

Bien entendu, le cas (ii) correspond à une spécialisation totale en bien 2 (exportation du bien 2 et importation du bien 1). Le cas (iii) aboutit à une spécialisation et des flux d'échanges indéterminés. Dans ce cas particulier où le prix mondial est égal au prix d'autarcie, le pays n'a aucun intérêt à l'échange.

6.5.2 Effets de la mise en place d'un tarif douanier

Considérons le secteur d'un bien quelconque d'une économie donnée. Les coûts de production de ce secteur induisent que le produit n'est vendu qu'à un prix largement supérieur au prix mondial (p_M) si le pays reste en autarcie (p_A). On considère que cette économie est relativement petite de sorte que (i)

l'offre mondiale au prix p_M pourrait largement combler la demande nationale, (ii) la demande nationale n'a aucun effet sur le prix mondial. Le pays a tout intérêt à ouvrir ses frontières pour bénéficier d'un meilleur prix pour ce bien, ce qu'il fait. Les consommateurs nationaux demandent dans ce cas une quantité Q_O qu'ils achètent au prix mondial puisque les entreprises nationales qui le peuvent sont obligées de s'aligner sur ce prix (rappelez-vous que l'offre mondiale pourrait combler toute la demande intérieure si nécessaire).
Par rapport à la situation d'autarcie, on constate que le surplus du consommateur a largement augmenté. Cette ouverture au commerce mondial est donc particulièrement favorable pour les consommateurs. Mais les entreprises nationales ont vu leur situation se détériorer au profit d'entreprises étrangères (baisse du prix de vente de p_A à p_M, baisse des quantités vendues de Q_A à QM, puisque les entreprises étrangères fournissent la quantité (Q_O-Q_N)). Pour améliorer leur situation, elles peuvent faire du lobbying auprès du gouvernement afin que celui-ci mette en place un tarif douanier dans ce secteur. Disons que ce gouvernement a la faiblesse de céder à cette opération de lobbying et met en place une taxe *ad valorem* τ sur les produits de ce secteur en provenance de l'étranger. Cela va bien entendu renchérir le prix des importations. Le nouveau prix auquel les consommateurs nationaux peuvent acquérir les biens de ce secteur en provenance de l'étranger est donc $p_M(1+\tau)$. C'est sur ce prix que vont s'aligner les entreprises nationales.

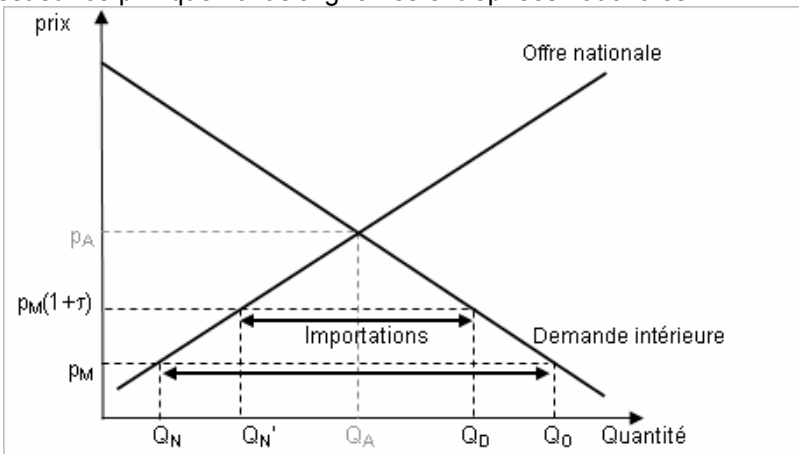


Figure 62 : La mise en place d'un tarif douanier

La mise en place de ce tarif douanier a l'effet escompté par les producteurs nationaux qui en étaient les promoteurs. Les importations diminuent (de (Q_O-Q_N) à (Q_D-Q_N')). La demande adressée aux producteurs nationaux augmente (de Q_N à Q_N') et ceux-ci vendent à un prix supérieur à celui qui prévalait avant la mise en place de ce tarif douanier (augmentation de p_M à $p_M(1+\tau)$). La situation des producteurs nationaux s'est améliorée mais…pas celle des consommateurs nationaux qui subissent une baisse de surplus. Le gouvernement bénéficie pour sa part d'une recette douanière.
Analysons en termes de surplus les différents effets qui suivent la mise en place d'un tarif douanier pour conclure quant à son effet favorable ou non sur l'ensemble de l'économie nationale (tous acteurs confondus).
Le surplus du consommateur a été réduit du fait de la hausse de prix engendrée par la mise en place de ce tarif douanier. Cette diminution du surplus est prise en compte par le trapèze, hachuré, pointillé ou quadrillé de la figure suivante. Cependant, ce trapèze peut être décomposé en quatre aires : (i) le trapèze à gauche de la courbe d'offre, (ii) un triangle rectangle en dessous de la courbe d'offre, (iii) un rectangle et (iv) un triangle rectangle en dessous de la courbe de demande.
Le rectangle (iii) représente un gain puisqu'il s'agit des recettes douanières consécutives à la mise en place du tarif. Cette surface est donc retirée du surplus des consommateurs mais ajouter au surplus global et donc cette baisse de surplus s'annule (si on considère que le gouvernement redistribue intégralement les recettes douanières sous forme de crédit d'impôts, les consommateurs sont indifférents entre la situation initiale et la situation présente où ils payent plus cher le bien pour un niveau d'impôts moindre).

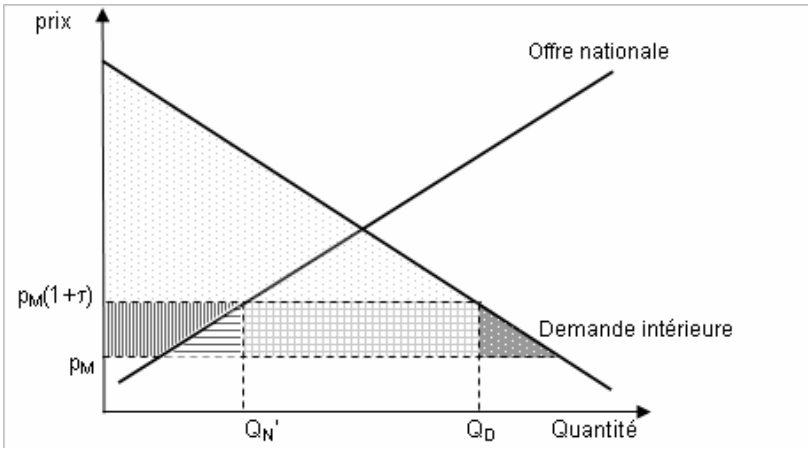


Figure 63 : Mise en place d'un tarif douanier (analyse des surplus)

Le trapèze à gauche de la courbe d'offre représente également un gain pour les producteurs nationaux. Ceux qui étaient déjà capables de produire sans le tarif douanier (donc avec un prix moins élevé) bénéficient d'une rente de situation et réalisent plus de profits. Ce gain en surplus pour les producteurs équilibre donc la perte de surplus pour les consommateurs représentée sur la même aire.
Reste les deux triangles rectangles représentant les pertes non compensées du surplus du consommateur. C'est assez évident pour le triangle rectangle le plus à droite. En ce qui concerne le triangle rectangle de gauche, celui-ci traduit la perte en bien-être engendrée par la production des entreprises nationales non efficaces (celles-ci ne produisaient pas quand il n'y avait pas encore le tarif douanier…). Ces entreprises ne peuvent produire que parce qu'existe ce tarif douanier alors que, « derrière la frontière », des entreprises plus efficace et capables de vendre moins cher sont privées de l'accès au marché national.
L'existence de ces deux triangles rectangle, c'est-à-dire d'une perte de surplus global suite à la mise en place du tarif douanier, montre bien l'inefficacité de ce type de mesure pour l'économie dans son ensemble. Si les producteurs nationaux bénéficient effectivement de cette politique protectionnisme, les consommateurs sont les grands perdants de la mise en place du tarif douanier.
Le libre échange apparaît être plus profitable en terme de surplus global.

N.B. : la mise en place qu'un quota (Q_D-Q_N*) (mesure non tarifaire) en lieu et place du tarif douanier τ (mesure tarifaire) aboutirait à des conclusions encore plus négatives puisque le rectangle représente alors un supplément de profit pour les firmes étrangères plutôt que des recettes douanières pour le gouvernement national. Le tarif douanier est donc moins néfaste que le quota.

6.5.3 Effets de la création d'une union douanière

L'économiste J. Viner s'est interrogé sur les conséquences d'une union douanière. La question à laquelle il entendait répondre est : L'union douanière est-elle meilleure ou pire que la situation avec des droits de douane semblables à l'égard de tous les pays (protectionnisme indifférencié). Viner a ainsi distingué deux cas :

- La **création de trafic** (effet bénéfique en termes de commerce de la mise en place de l'union douanière). Un courant commercial qui n'existait pas auparavant apparaît lorsque le pays passe de la protection indifférenciée à l'union douanière.
- Le **détournement de trafic**. La mise en place de l'union douanière se traduit pour le pays par le remplacement d'un courant d'importation en provenance d'un pays tiers par un courant d'importation depuis un pays membre.

Le détournement de trafic a des conséquences néfastes en matière d'utilisation optimale des facteurs de production au niveau mondial puisqu'un pays qui n'est pas le plus compétitif est avantagé par l'appartenance à cette union douanière, au détriment d'un pays tiers hors de l'union douanière mais possédant un avantage dans cette production.
Appelons A le pays importateur membre de l'union douanière, B le pays membre nouvellement exportateur vers le pays A et C le pays tiers non membre de l'union anciennement exportateur vers le pays A. On suppose pour simplifier que le pays A ne produit pas le bien en question de sorte que la demande d'importations est identique à la demande des consommateurs. Le pays B vend le bien à un

prix supérieur à celui du pays C. Soit τ le taux du droit de douane qui n'est désormais prélevé que sur les importations en provenance de C (non membre de l'union douanière)
Les figures ci-dessous présentent la situation en termes de bien-être du pays A, avant et après la mise en place de l'union douanière. Avant l'Union douanière, le pays A commerce avec le pays C car $p_C<p_B$ et les importations des deux pays sont taxées au même taux τ de sorte que la pays C présente un prix plus intéressant et sert toute la demande des consommateurs du pays A (figure de gauche). Après l'union douanière (figure de droite), les importations en provenance du pays B, membre de l'union, ne subissent plus les tarifs douaniers et donc les importations en provenance du pays B présentent désormais un prix plus avantageux. Le surplus des consommateurs augmente mais le pays A ne bénéficie plus des recettes douanières.

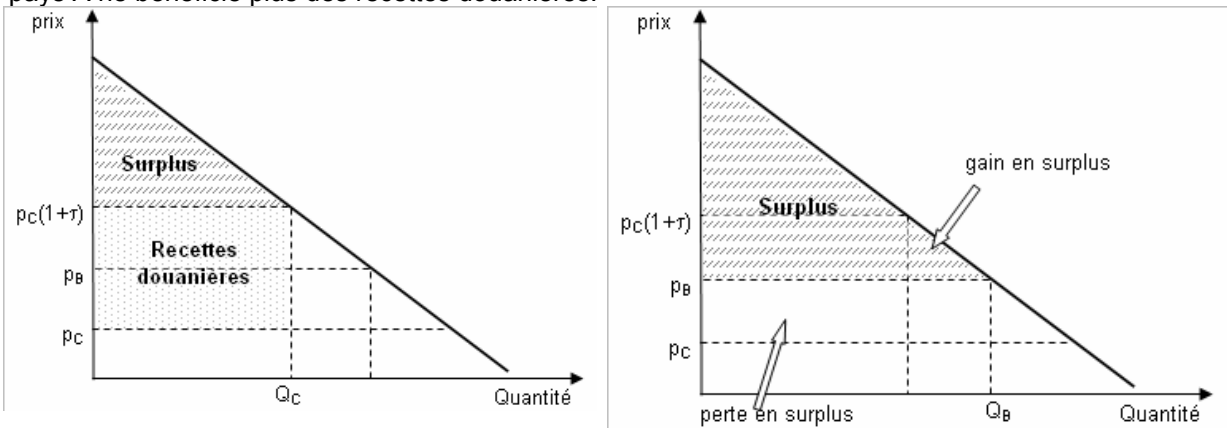


Figure 64 : Effet de la mise en place d'une union douanière

Ce détournement de trafic et la mise en place de cette union douanière ont-ils bénéficié au pays A ? On peut répondre à cette question en procédant à une comparaison de situations en termes de bien-être. Dans notre cas la diminution de recette douanière (rectangle pointillé) est moins que compensée par leur surcroît de surplus du consommateur du fait de la diminution des taxes à l'importation (petit triangle rectangle). Le pays A est perdant lorsqu'il forme une union douanière avec le pays B.

6.5.4 Le modèle Mundell - Flemming : IS-LM-BP

Dans les années soixante deux économistes ont proposé une extension du modèle IS-LM (voir l'annexe sur **le modèle IS-LM** à la fin du chapitre Conjoncture et politiques économiques) à l'économie ouverte. Ce modèle permet de déterminer l'équilibre conjoint des marchés des biens et de la monnaie, et de celui des changes en tenant compte des échanges extérieurs.

Les politiques budgétaires et monétaire que nous avons étudiées au chapitre 4 en économie fermée (conjoncture et politiques économiques) ont des répercussions sur la balance commerciale et les mouvements de capitaux.
Le choix entre la politique budgétaire et monétaire peut ainsi être influencé par des considérations spécifiques et des raisons d'opportunités qui tiennent à la mobilité des capitaux et au type de système de change (change fixe ou change flottant). En économie fermée, on se souvient que les effets de la politique monétaire sont à priori tous positifs alors que la politique budgétaire engendre certains effets négatifs. Nous allons voir ce qu'il advient de ce résultat en économie ouverte.
Nous allons en premier lieu introduire une nouvelle droite dans le système (Y,r) du modèle IS-LM : **la droite BP**.
Nous étudierons ensuite, le cas des politiques économiques en change fixe puis en change flottant. L'efficacité des politiques économiques sur l'extérieur diffère selon que l'économie est à prix fixes ou à prix flexibles. Une économie est dite à prix fixes lorsque le rapport prix intérieurs / prix mondiaux ne varie pas. Nous allons considérer dans ce modèle IS-LM-BP que les prix sont fixes.

La droite d'équilibre de la balance des paiements (voir l'annexe sur **la balance des paiements**) peut être représentée dans le système d'axes (Y,r) (PIB , taux d'intérêt) que nous avons déjà utilisé pour le modèle IS-LM (voir l'annexe sur **le modèle IS-LM** du chapitre 4). La droite d'équilibre de la balance des paiements (BP=0) est le lieu des couples du PIB et du taux d'intérêt pour lesquels la balance des paiements est en équilibre.

La balance des paiements de ce schéma IS-LM-BP correspond à la somme de la balance commerciale et de la balance des capitaux.

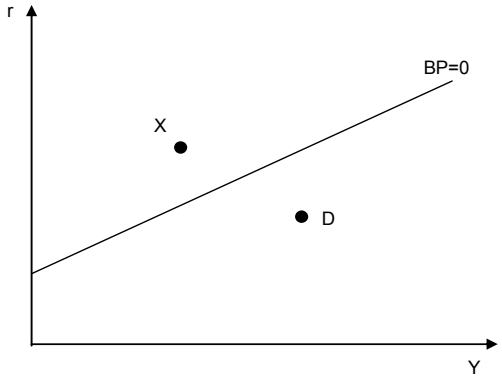


Figure 65 : La droite BP

Dans la zone située au-dessus de la droite BP (point X par exemple sur la figure ci-dessus), la balance des paiements est en excédent parce que la hausse du taux d'intérêt (au-delà du niveau qui permet l'équilibre de la balance des paiements) va induire un afflux de capitaux dans l'économie nationale. Au contraire, en dessous de la droite BP (point D par exemple), la balance des paiements est en déficit parce que le taux d'intérêt est à un niveau inférieur à celui qui permet d'équilibrer la balance des paiements (il y a une diminution des entrées de capitaux). La droite BP est donc le lieu des couples (Y,r) qui permettent un équilibre de la balance des paiements.

Lorsque le PIB est relativement bas, les importations sont faibles ce qui se traduit par une balance des transactions courantes excédentaires. Pour obtenir un équilibre de la balance des paiements, le compte financier doit alors connaître un déficit, ce qui est obtenu par un taux d'intérêt relativement faible (par rapport au taux d'intérêt étranger) induisant une sortie de capitaux. Au contraire, lorsque le PIB est relativement élevé, cela accroît les importations et entraîne du déficit courant. Pour rétablir l'équilibre de la balance des paiements, le compte financier doit être excédentaire ce qui ne peut être obtenu que par un taux d'intérêt élevé permettant d'attirer les capitaux étrangers.

6.5.4.1 Régime de taux de change fixes

En régime de taux de change fixes, la Banque centrale s'engage à conserver une parité fixe avec la (les) autre(s) monnaie(s). Elle ne peut ainsi plus choisir à sa guise la quantité de monnaie en circulation et laisser s'ajuster le taux de change selon l'équilibre du marché des changes.

Mobilité imparfaite des capitaux

Lorsque les taux de change sont fixes, il n'y a pas de déplacement de la courbe BP suite à des mesures de politiques budgétaires ou monétaires. La pente de la droite BP est positive ce qui traduit une mobilité imparfaite des capitaux. Une hausse des taux d'intérêt améliore la balance des capitaux. Cette amélioration requiert un PIB plus élevé ce qui détériore la balance des transactions courantes d'un montant suffisant pour permettre un nouvel équilibre de la balance des paiements. L'équilibre doit toujours se situer sur la balance des paiements initiale afin de conserver la valeur du taux de change fixe.

Prenons un exemple. Sur la figure ci-dessous, au point E, la balance des paiements est excédentaire car le taux d'intérêt est plus élevé que celui qui permettrait un équilibre de la balance des paiements, étant donné le niveau du PIB. Des entrées de devises vont se produire ce qui fait augmenter la quantité de monnaie offerte et la droite LM se déplace jusqu'en LM' permettant la réalisation de l'équilibre extérieur en E' (cela aurait bien entendu fonctionné en sens inverse si le point E avait été initialement en dessous de la droite BP). Les mécanismes monétaires rétablissent donc automatiquement l'équilibre de la balance des paiements.

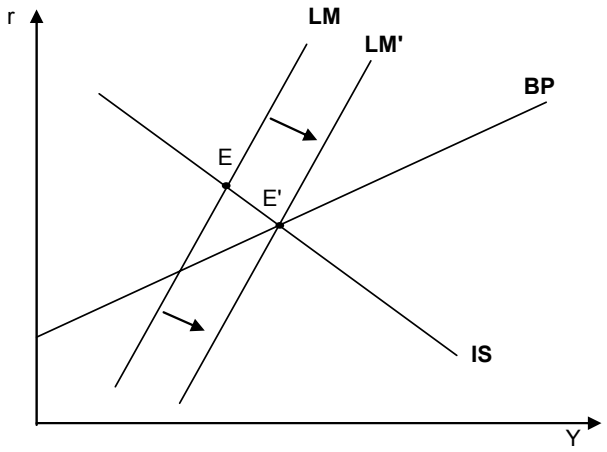


Figure 66 : Le rééquilibrage par entrée de capitaux

En régime de taux de change fixes, la Banque centrale abandonne donc la possibilité d'utiliser la politique monétaire comme un instrument puisque celle-ci est dédiée au maintien de la parité des taux de change.
Si les pouvoirs publics utilisent la seule politique monétaire expansionniste, la courbe LM se déplace vers la droite et un nouvel équilibre IS-LM s'établit. Mais en ce point, les comptes extérieurs sont déficitaires. LM va donc revenir vers l'équilibre initial du fait de la sortie des capitaux (la politique monétaire initiale ayant au pour conséquence la baisse du taux d'intérêt car sortie de capitaux). La politique monétaire utilisée seule n'est donc pas efficace.

Examinons l'effet de la politique budgétaire que les pouvoirs publics peuvent être tentées d'utiliser si on considère qu'au point de l'équilibre IS-LM-BP, le PIB est inférieur à son niveau qui permet le plein-emploi.

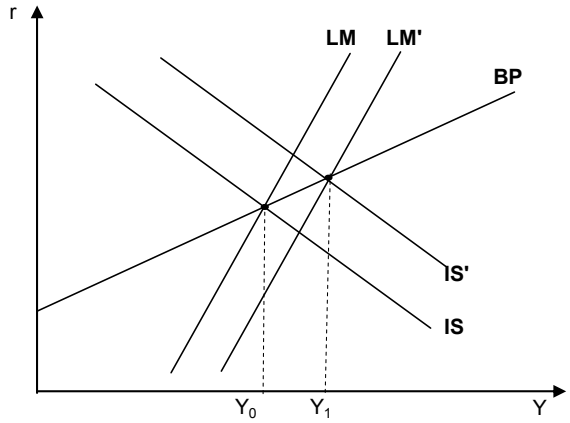


Figure 67 : Efficacité de la politique budgétaire en régime de change fixe
(mobilité imparfaite des capitaux)

Si une politique budgétaire expansionniste est utilisée, alors la courbe IS se déplace vers le haut et la droite et un nouvel équilibre IS-LM apparaît. Toutefois, en ce point, la balance des paiements est excédentaire (puisque la politique budgétaire expansionniste a eu pour conséquence l'élévation du taux d'intérêt). Il y aura donc un rééquilibrage par entrée de capitaux. En régime de changes fixes, la banque centrale ne peut tolérer l'appréciation de la monnaie. Comme l'appréciation de la monnaie mène à une augmentation de la demande de monnaie, la banque centrale accompagne cette augmentation de la demande par une augmentation de l'offre de monnaie. La politique budgétaire déclenche donc une accommodation monétaire. Pour la banque centrale, la solution est donc de trouver le bon dosage entre la politique budgétaire et la politique monétaire qu'elle met en œuvre pour stériliser les effets de la politique budgétaire sur le taux de change.

Mobilité parfaite des capitaux

Lorsque la liberté des mouvements de capitaux est totale, la droite BP est horizontale (au niveau du taux d'intérêt mondial). En effet à la moindre différence de taux d'intérêt, les capitaux affluent dans le territoire qui présente le différentiel de taux positif. Les entrées de capitaux persistent tant que le différentiel de taux existe. La seule possibilité est donc que le taux d'intérêt national soit égal au taux d'intérêt « mondial ».

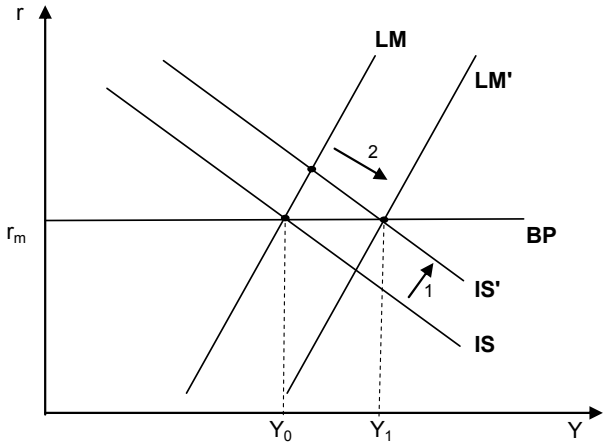


Figure 68 : Efficacité de la politique budgétaire en régime de change fixe
(mobilité parfaite des capitaux)

Examinons alors dans cette configuration, les effets de la politique budgétaire. Lorsque la relance budgétaire est effectuée, la courbe IS se déplace vers le haut et la droite. Au point d'équilibre de IS' et LM, la balance des paiements est excédentaire, ce qui induit des entrées de capitaux et donc une appréciation de la monnaie. La Banque centrale va donc augmenter son offre de monnaie de façon à neutraliser l'effet de l'appréciation du taux de change. En conséquence, l'effet d'éviction ne se produit pas et le multiplicateur exerce pleinement ses fonctions. Dans une économie où les mouvements de capitaux internationaux sont parfaitement libres et les taux de changes fixes, la politique budgétaire est pleinement efficace (plus encore que lorsque la mobilité des capitaux n'est pas parfaite).

6.5.4.2 Régime de taux de change flottants avec mobilité parfaite des capitaux

En régime de taux de change flottants, les autorités n'ont pas à se soucier de maintenir la valeur de leur devise vis-à-vis des autres devises. La fluctuation des devises va au contraire constituer un mécanisme de rééquilibrage de la balance des paiements. Si la balance des paiements est « déficitaire », alors le taux de change va diminuer de sorte que le pays va regagner en compétitivité. Il va réduire ses importations (devenues plus chères) et augmenter ses exportations, de sorte que la balance des paiements va retrouver l'équilibre.

Examinons rapidement l'efficacité des politiques monétaires et budgétaires dans cette configuration de taux de changes flottants.

Lorsque les autorités d'un pays pratiquent une relance budgétaire, nous avons vu au chapitre 4 (annexes sur le modèle IS-LM) que cela avait pour conséquence une hausse du taux d'intérêt. Si les taux d'intérêt de l'économie nationale qui pratique cette relance s'élèvent, cela va engendrer une entrée de capitaux étrangers qui vont être attirés par une plus forte rémunération du fait de cette hausse. L'entrée massive de capitaux aboutit à une élévation du taux de change (puisque pour venir placer leur argent dans cette économie, les agents étrangers doivent auparavant convertir leurs avoirs dans la devise nationale, ce qui provoque une hausse de la demande de la devise nationale sur le marché des changes).

Du fait de l'appréciation de la devise nationale, les entreprises perdent en compétitivité et leurs exportations diminuent, de sorte que la courbe IS va revenir sur elle-même (cf. schéma).

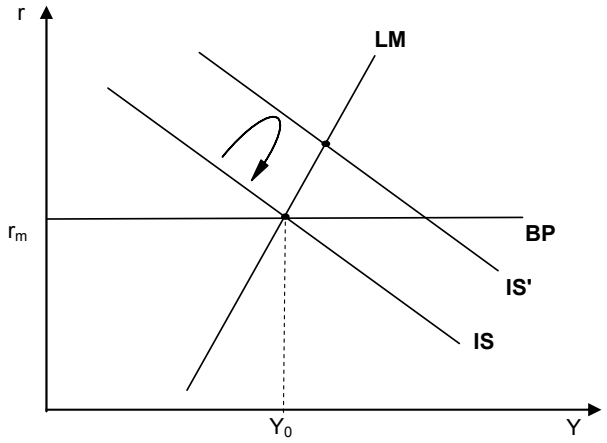


Figure 69 : Inefficacité de la politique budgétaire en régime de changes flottants
(mobilité parfaite des capitaux)

En régime changes flottants, on constate donc l'inefficacité de la politique budgétaire.

Qu'en est-il de la politique monétaire ? Considérons que les autorités d'une nation décident de pratiquer une politique monétaire accommodante. L'offre de monnaie s'accroît (la courbe LM se déplace en LM' sur le schéma ci-dessous (1)). Le nouvel équilibre se situe à l'intersection entre la droite IS et la droite LM'. Le taux d'intérêt national est alors inférieur au taux d'intérêt mondial, ce qui provoque une sortie de capitaux.
Cette sortie de capitaux s'accompagne donc de vente massive de devise nationale sur le marché des changes, ce qui provoque la baisse du taux de change et accroît donc la compétitivité des entreprises nationales.
La plus forte compétitivité des entreprises nationales aboutit à une augmentation des exportations, ce qui provoque le déplacement de la courbe IS en IS'.

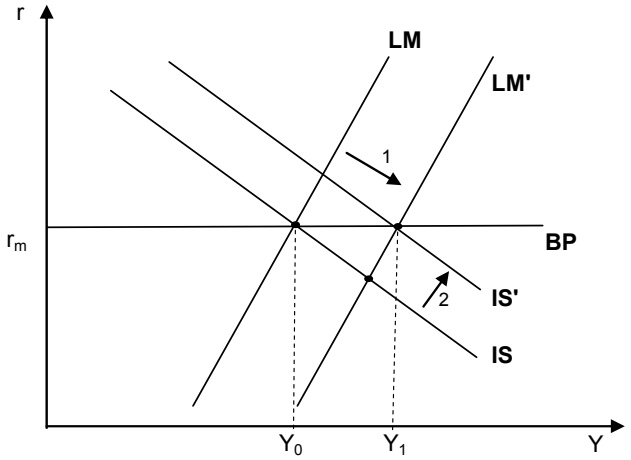


Figure 70 : Efficacité de la politique monétaire en régime de changes flottants
(mobilité parfaite des capitaux)

Au final, le nouvel équilibre se situe à l'intersection des droites IS', LM' et BP. Le taux d'intérêt est toujours égal au taux d'intérêt mondial et le revenu s'est fortement accru. La politique monétaire est donc efficace en régimes de change flottants.

6.5.5 Le théorème des élasticités critiques et la courbe en J

On pourrait penser qu'une dépréciation de la monnaie nationale, puisqu'elle réduit le prix relatif des produits nationaux, améliore systématiquement la balance commerciale (c'est l'hypothèse forte que l'on a fait avec le modèle IS-LM-BP). Or, les effets d'une variation du taux de change sur le solde commercial sont plus complexes que cela. Cette question était bien sûr très importante en régime de changes fixes puisqu'il s'agissait alors de valider la pertinence d'un « instrument » de politique économique : la dévaluation.
Prenons l'expression de la balance commerciale avec e, le taux de change nominal à l'incertain (le taux de change réel est dans ce cas e_R=e.p*/p). Une hausse de e correspond donc à une dépréciation de la monnaie nationale.

$$BC = pX(e) - ep^* M(e)$$

L'effet global de la dépréciation d'une monnaie nationale est en fait la résultante de deux effets de sens contraire : (i) la baisse du taux de change améliore la compétitivité prix des produits nationaux, ce qui augmente le volume des exportations (X) et réduit le volume des importations (M) (effet favorable), (ii) la réduction du taux de change, en élevant le prix en monnaie nationale des produits étrangers (ep*) augmente la valeur des importations (effet défavorable).
En différenciant l'équation de la balance commerciale, en supposant fixe toutes les variables autres que le taux de change nominal et en désignant respectivement par ε_X et ε_M les élasticités prix des exportations et des importations, on obtient l'expression suivante :

$$\frac{\partial BC}{\partial e} = p^* M\left(\frac{pX}{ep^* M}\varepsilon_X - \varepsilon_M - 1\right) \qquad \varepsilon_X = \frac{\partial X}{\partial e}\frac{e}{X} > 0 \text{ et } \varepsilon_M = \frac{\partial M}{\partial e}\frac{e}{M} < 0$$

Une baisse du taux de change nominal ($\Delta e>0$) améliore la balance commerciale ($\Delta BC>0$) si : $\frac{pX}{ep*M}\varepsilon_X - \varepsilon_M > 1$. Si on considère que la balance commerciale était initialement équilibrée ($pX = ep*M$) la condition devient :

$$|\varepsilon_X| + |\varepsilon_M| > 1$$

Cette condition, établie dans les années trente, est connue sous le nom de théorème des élasticités critiques, ou condition de Marshall-Lerner. Elle signifie que l'impact sur les quantités échangées d'une dévaluation (ou d'une dépréciation) doit être suffisamment important pour compenser l'impact négatif sur les prix. Plusieurs remarques : (i) si la balance commerciale est initialement déficitaire, la condition Marshall-Lerner ne garantit pas son amélioration puisque l'effet prix négatif s'applique à des volumes d'importations supérieurs à ceux des exportations, (ii) en régime de change fixe, les « dévaluations compétitives » peuvent être mal perçues par les partenaires, qui peuvent répondre par une dévaluation de leur propre monnaie…, (iii) les raisonnements précédents supposent que les prix nationaux demeurent fixes. Cette hypothèse de rigidité des prix ne vaut qu'à court terme. A long terme, la dépréciation (ou la dévaluation) avive l'inflation (cf. théorie relative de la parité des pouvoirs d'achat)

Si les modifications du taux de change se répercutaient rapidement sur le solde commercial, on devrait constater dans les mois qui suivent la dépréciation une amélioration du solde. Or, on a observé, depuis la fin des années soixante, que les dépréciations ou dévaluations étaient suivies d'une dégradation du solde, puis, après un délai variable (pouvant dépasser une année), d'un redressement. Le profil d'évolution évoque un J majuscule. L'explication repose sur le retard d'adaptation des quantités exportées et importées à la variation des prix relatifs induites par la baisse du taux de change. Le premier effet est donc défavorable (augmentation des prix des importations). La réaction favorable des volumes ne s'effectue qu'avec un délai (réduction du déficit puis enfin, apparition d'un excédent)

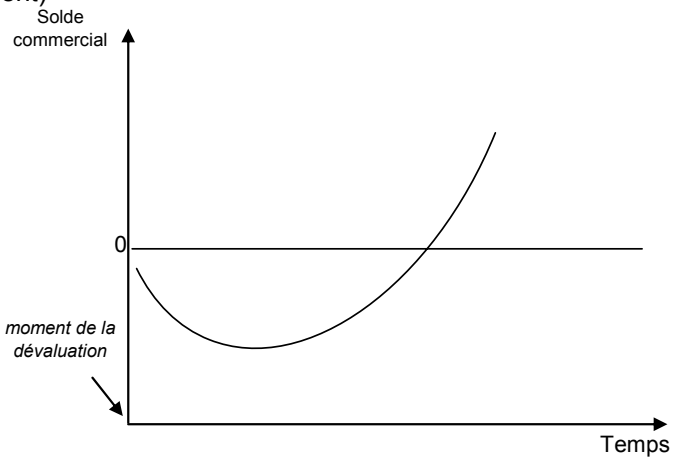


Figure 71 : La courbe en J

7 Les moteurs de la croissance : Education et innovation

On mesure habituellement la richesse d'une nation par le PIB par habitant (PIB/tête), nonobstant les phénomènes d'inégalités qui impliquent que tous les agents d'une économie ne disposent pas des mêmes ressources. C'est cependant au niveau global et dans le long terme un bon indicateur du niveau de développement d'un pays car si les inégalités ne se creusent pas, la croissance bénéficie à tous. Le PIB/tête peut connaître des fluctuations cycliques du fait des variations conjoncturelles du PIB (voir le chapitre sur **les politiques économiques et la conjoncture**) mais ce qui nous intéresse dans ce chapitre, c'est la tendance de long terme de la croissance du PIB/tête (la croissance potentielle). Quels sont les facteurs qui structurent celle-ci ? Pourquoi certains pays ont-ils connu un décollage économique et pas d'autres (que cela soit il y a deux siècles ou plus récemment) ? Pourquoi durant les dernières décennies a-t-on assisté à un rattrapage spectaculaire du niveau du PIB/habitant des économies développées par certaines économies dites émergentes (tigres et dragons asiatiques, désormais Chine, demain l'Inde…) alors que d'autres pays voyaient le niveau de leur PIB/tête stagner ou régresser (Afrique sub-saharienne par exemple) ? Existe-il un moyen de croître plus vite ? Quelles sont les politiques économiques de promotion de la croissance ?
L'acuité de cette question ne vaut pas simplement pour la perspective de développement des pays les plus pauvres. Cette question est en effet primordiale pour certains pays européens qui voient leur potentiel de croissance diminuer (moins de 2% alors que la croissance au niveau mondial s'affiche aux alentours de 5% cette année).

Empiriquement plusieurs variables sembler susceptibles d'expliquer les différences de taux de croissance entre les pays : le taux d'épargne, le taux de croissance de la population, le niveau d'éducation, le fonctionnement des marchés, la stabilité macroéconomique, la stabilité politique…
Dans ce chapitre nous allons plus particulièrement centrer notre analyse sur les ressorts technologiques de la croissance, ainsi que sur l'effet de l'éducation sur la croissance. Nous examinerons également la complémentarité de ces deux moteurs de la croissance en fonction du niveau de développement des économies.

Il semble important que vous ayez à l'esprit les modèles théoriques de croissance reposant sur l'éducation et l'innovation car ils sont sous-jacents à ce qu'on appelle la **stratégie de Lisbonne**. Au conseil européen de Lisbonne en mars 2000, les responsables européens ont en effet entériné « *l'objectif ambitieux de faire de l'Union Européenne l'économie fondée sur la connaissance la plus compétitive et la plus dynamique au monde d'ici la fin de la décennie* ». Au Conseil européen de Barcelone, en mars 2002, les européens ont précisé cette stratégie en se fixant des objectifs chiffrés. En matière d'innovation, ils entendaient «*Renforcer l'espace européen de la recherche et de l'innovation en fixant comme objectif pour la fin de la décennie un montant total des dépenses publiques et privées de* ***R&D égal à 3% du PIB***» (on peut ajouter que les deux tiers de ces 3% devront alors être financés par les entreprises privées, ce qui n'est pas encore le cas partout en Europe). Pour atteindre l'objectif de Barcelone, cela implique l'embauche de 700 000 chercheurs supplémentaires ce qui amènerait le nombre de chercheurs dans la population à 8 pour 1 000 avant la fin de la décennie. C'est un enjeu bien entendu très important tant on a vu que la croissance serait primordiale pour résorber le chômage et réduire les déficits et la dette publique. C'est un enjeu important également au niveau microéconomique pour les entreprises privées qui doivent être le vecteur de cette politique et investir massivement en R&D pour s'assurer le leadership sur les marchés des produits de demain. Cette stratégie européenne passe par l'obtention d'un leadership technologique dans les sciences de la vie et la biotechnologie, un rattrapage du retard européen en matière de nouvelles technologies de l'information et de la communication.
Les technologies d'avant-garde apparaissent comme un facteur-clé de la croissance future et la nécessité d'une politique structurelle forte visant à promouvoir l'innovation et l'éducation en Europe semble communément admise. Cependant, la bonne volonté politique doit désormais se transformer en action et ne efforts car la fin de décennie dont parlaient les européens en 2000, c'est demain.
Après avoir brièvement rappelé certains faits empiriques de la croissance (convergence et divergence de certains pays par rapport à la trajectoire de croissance des pays développés durant les quatre

dernières décennies, « paradoxe de la productivité » dans les décennies soixante-dix et quatre-vingts, « nouvelle économie » et redémarrage de la croissance de la productivité aux Etats-Unis durant la décennie quatre-vingt-dix) et noté le retard européen en matière d'éducation et de recherche, nous allons donc examiner les modèles économiques qui expliquent la croissance de long terme. Nous verrons lesquels sont à même de décrire les effets d'une intensification des efforts de recherche et d'éducation et les prédictions de ces modèles théoriques. Nous pourrons alors conclure sur la plausibilité de réussite de la stratégie européenne de Lisbonne.

7.1 *Quelques* faits stylisés

7.1.1 β-convergence, σ-convergence et rattrapage

Pour assister à un processus de **rattrapage** du niveau de richesse des pays développés par les pays en développement, il faut que le taux de croissance du PIB/tête soit durablement supérieur dans les pays en développement à ceux des pays développés.
On dit qu'il y a **convergence** lorsqu'il existe une relation décroissante entre les PIB/tête initiaux et les taux de croissance du PIB/tête annuels moyens sur longue période (**β-convergence**, les pays en retard connaissent une plus forte croissance). On dit qu'il y a **σ-convergence** lorsque la variance du PIB/tête diminue pour un ensemble de pays. La β-convergence permet la convergence des niveaux de revenu tandis que la σ-convergence n'aboutit qu'à une réduction des inégalités de revenu.
Jetons à présent un regard sur le développement économique des différentes économies.

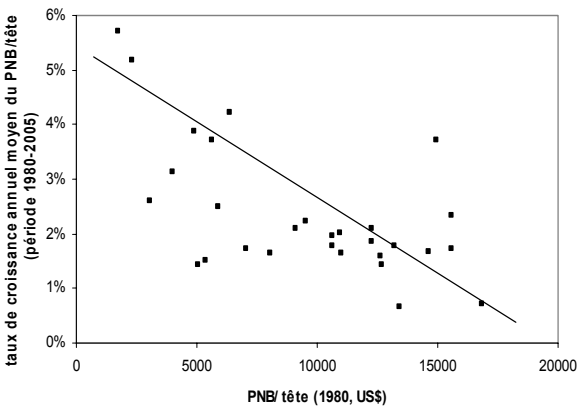
7.1.2 Convergence et développement

On peut distinguer quatre grandes périodes. Jusqu'à la fin du Moyen-âge, on considère que la croissance du PIB/habitant n'était que de 0,2% par an en moyenne au niveau mondial. Le niveau du PIB/tête mondial peut être estimé aux environs de 100$ à la fin du Moyen-âge. La croissance mondiale connaît une accélération avec la révolution industrielle. De 200$ en 1800 le PIB/tête passe à 1 000$ en 1914. Puis c'est la croissance forte au XX^{ème} siècle, néanmoins interrompue par deux guerres. En 2000, le PIB/tête mondial est de 6 500$.
On peut relever un certain nombre de faits en matière de croissance et de développement au cours de la seconde moitié du XX^{ème} siècle :

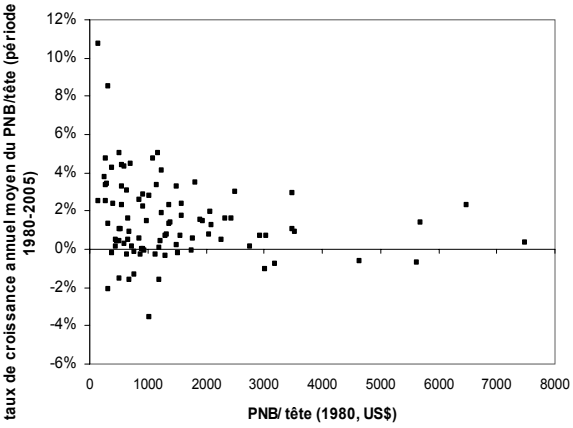
- Le Japon et l'Europe ont convergé vers les Etats-Unis au cours de la seconde moitié du XX^{ème} siècle mais plafonnent depuis le choc pétrolier au niveau de 80% du revenu américain.
- Certains pays émergents ont décollé les uns après les autres (Asie du Sud-Est, Chine…).
- D'autres pays stagnent ou régressent en termes relatifs (Afrique sub-saharienne…)

La hiérarchie en ce début de XXI^{ème} siècle est donc différente de celle qui prévalait début du XX^{ème} siècle[111].
Les figures suivantes illustrent ces phénomènes. Les différents pays ont été scindés en deux groupes : les pays « développés » (pays dont le PIB/habitant est supérieur à 14 000$ en 2005) et les pays « en développement » (117 pays dont le PIB/habitant est inférieur à 14 000$ en 2005). Que constate-t-on ? Pour les pays développés, on est en présence d'un phénomène de rattrapage de la part des pays qui avaient initialement le PIB/tête le plus faible. Il y a donc convergence. En revanche pour le pays « en développement », on ne constate pas un tel processus. D'une part, on constate des phénomènes de stagnation et de régression. D'autre part, il n'y a pas de véritable relation décroissante entre le PIB/tête initial et le taux de croissance (pas de β-convergence donc).
C'est le premier fait stylisé marquant en ce qui concerne la croissance et le développement. Le modèle classique de croissance introduit par Solow (1956) (voir l'annexe en fin de chapitre présentant le **modèle de Solow**), qui prévoit que le taux de croissance sera d'autant plus élevé que le stock de capital par tête est initialement faible, est invalidé par l'expérience des pays que nous avons qualifié de pays en développement.

[111] Le PIB/habitant de l'Argentine dépassait celui de l'Espagne de 70% en 1913. Il était de 50% inférieur en 2000…



"Pays développés" : Australie, Autriche, Belgique, Canada, Suisse, Chypre, Allemagne, Danemark, Espagne, Finlande, France, Royaume-Uni, Grèce, Hong Kong, Irlande, Islande, Israël, Italie, Japon, Corée, Luxembourg, Pays-Bas, Norvège, Nouvelle-Zélande, Portugal, Singapour, Suède, Taïwan, Etats-Unis



"Pays en développement" : Afrique, Europe centrale, anciens pays du Commonwealth, pays d'Asie en développement, Moyen-Orient (sauf pays exportateurs de pétrole), Amérique du Sud.. (117 pays pour lesquels données disponibles)

(Source : World Economic Outlook Database, FMI)

Figure 72 : Rattrapage et stagnation

C'est pourquoi nous étudierons également d'autres modèles de croissance qui parviennent à « endogénéiser » la croissance, c'est-à-dire à décrire le lien entre le taux de croissance du PIB/tête et certains déterminants propres à chaque économie. Les modèles de croissance schumpétérienne ont ainsi mis en exergue la relation existante entre le rythme du progrès technique qui résulte de l'investissement en recherche et développement des firmes privées, et le taux de croissance d'une économie.

7.1.3 Résidu de Solow et paradoxe de la productivité

L'économiste Robert Solow apparaît omniprésent lorsqu'on présente les théories et les faits concernant la croissance. Le **modèle de Solow (1956)**, même si il échoue à décrire des trajectoires de croissance divergente, reste l'outil néo-classique de base pour étudier la croissance et le long terme. Solow a également initié la méthodologie de comptabilité de la croissance, qui permet d'isoler le **résidu de Solow**, c'est-à-dire la part de la croissance de long terme qui n'est pas expliquée par les facteurs traditionnels et qu'on considère donc résulter de l'évolution du progrès technique, considéré comme exogène. C'est ainsi qu'apparaît un troisième élément Solowien en matière de croissance : **le paradoxe de Solow**. Solow a en effet fait remarquer en 1987 que la croissance de la productivité globale des facteurs n'avait jamais été aussi faible alors qu'on s'attendait à la voir renforcée par l'introduction dans l'économie de la technologie informatique censée améliorer la productivité. Solow a ainsi déclaré : « *On voit des ordinateurs partout sauf dans les statistiques de comptabilité nationale* ».

Notre deuxième série d'illustrations empiriques a donc trait à ce qu'on appelle le « paradoxe de la productivité » ou « paradoxe de Solow ». Les tableaux ci-dessous décrivent les contributions à la croissance des différents facteurs de production (travail, capital) et du progrès technique. La méthodologie utilisée est celle dite de la comptabilité de la croissance (voir annexe en fin de chapitre sur **l'exercice de comptabilité de la croissance**). L'idée est que la part de la croissance inexpliquée par les facteurs traditionnels, ce qu'on appelle le « résidu de Solow » ou encore la Productivité Globale de Facteurs (PGF), est une mesure par défaut du progrès technique. Celui-ci est donc considéré comme exogène comme dans le modèle de croissance de Solow, sous-jacent à cet exercice de comptabilité de la croissance.

Taux de croissance annuels moyens	1913-1950	1950-1973	1973-1987
Travail	0,0	0,4	0,0
+durée du travail	-0,8	-0,4	-1,0
=Quantité de travail	-0,8	0,0	-1,0
+qualité du travail	0,5	0,2	0,6
=Facteur Travail Augmenté (1)	**-0,3**	**0,2**	**-0,4**
Capital (hors logement)	1,2	5,1	4,5
+ qualité du capital	1,5	2,1	1,3
=Facteur capital augmenté (2)	**2,7**	**7,2**	**5,8**
Capital logement (3)	**0,4**	**2,8**	**2,3**
Ensemble (4)=0,7*(1)+0,23*(2)+0,07*(3)	**0,4**	**2,0**	**1,2**
PIB (5)	**1,2**	**5,0**	**2,2**
PGF (6) = (5) – (4)	**0,8**	**3,0**	**1,0**

Tableau 16 : Comptabilité de la croissance (1913-1987), France – source : Maddison (1991)

Taux de croissance annuels moyens	1913-1950	1950-1973	1973-1987
Travail	1,3	1,5	1,9
+durée du travail	-0,9	-0,4	-0,5
=Quantité de travail	0,4	1,1	1,4
+qualité du travail	0,5	0,5	0,4
=Facteur Travail Augmenté (1)	**0,9**	**1,6**	**1,8**
Capital (hors logement)	2,1	3,2	3,3
+ qualité du capital	1,4	1,7	1,4
=Facteur capital augmenté (2)	**3,5**	**4,9**	**4,7**
Capital logement (3)	**1,8**	**3,3**	**2,5**
Ensemble (4)=0,7*(1)+0,23*(2)+0,07*(3)	**1,6**	**2,5**	**2,5**
PIB (5)	**2,8**	**3,7**	**2,5**
PGF (6) = (5) – (4)	**1,2**	**1,2**	**0,0**

Tableau 17 : Comptabilité de la croissance (1913-1987), Etats-Unis – source : Maddison (1991)

Les tableaux ci-dessus sont issus des travaux de comptabilité de la croissance de Maddison (1991)[112]. Ils illustrent bien, notamment pour le cas américain, le paradoxe de la productivité. Pour la **France**, on constate que, sur la première partie du vingtième siècle, 1913-1950 (perturbée par deux guerres…), le taux de croissance du PIB s'est élevé à 1,2% par an, en moyenne. Les baisses de la durée du travail font que la contribution du facteur travail à cette croissance est faiblement négative. En revanche, l'accumulation du capital permet d'expliquer une croissance annuelle moyenne de 0,4%. Aussi, sur cette première période, la productivité globale des facteurs (le progrès technique) explique 0,8% de croissance annuelle. La période 1950-1973 (période des « trente glorieuses ») est une période de forte croissance pour la France. Le taux de croissance annuel moyen est de 5%. L'accumulation du capital est très forte sur cette période de rattrapage mais les facteurs traditionnels n'expliquent finalement que 2% de croissance annuelle. Aussi la croissance de la productivité globale des facteurs est elle considérée comme importante durant cette période (3%). Les décennies qui suivent (1973-1987) sont moins propices à la croissance en France. Le taux de croissance annuel moyen est de l'ordre de 2,2%. Les facteurs traditionnels expliquent 1,2% de croissance annuelle (c'est encore l'accumulation du capital qui semble le principal vecteur de la croissance, car la contribution du facteur travail à la croissance est de nouveau négative du fait de la baisse de la durée du travail). La croissance de la PGF est de 1% par an. Il n'y a donc pas de véritable paradoxe de la productivité en France. On doit plutôt considérer que la période 1950-1973 fut exceptionnelle et que la période 1973-1987 signe un certain retour à la croissance de la productivité d'avant la seconde guerre mondiale.
Pour les **Etats-Unis**, l'étude de Maddison aboutit à des conclusions différentes. D'une part la croissance de la période d'après guerre est beaucoup moins marquée que celle de l'économie française. Le taux de croissance annuel moyen passe de 2,8% à 3,7%. Mais comme la contribution des facteurs travail et capital est plus importante, la croissance de la PGF est tout à fait stable sur la période 1913-1973, de l'ordre de 1,2% par an. La période 1973-1987 est également aux Etats-Unis celle d'un ralentissement de la croissance par rapport à la période précédente. Mais ce qui est surtout marquant, c'est la disparition totale de la croissance de la productivité. Le tableau ci-dessus pour les Etats-Unis illustre donc parfaitement le paradoxe de Solow. Alors que la mise en place des technologies informatiques dans les années soixante-dix et quatre-vingts laissait attendre un renforcement de la productivité globale des facteurs, celle-ci devient nulle. Il semblait ne plus y avoir de progrès technique aux Etats-Unis sur cette période !

7.1.4 « Nouvelle économie » et fin du paradoxe de Solow ?

Durant la seconde moitié de la décennie quatre-vingt-dix, l'émergence des nouvelles technologies de l'information et de la communication, la « nouvelle économie » a semblé mettre fin au paradoxe de la productivité (Solow a lui-même déclaré en 2000 que son paradoxe semblait enfin résolu). Le tableau ci-dessus présente les résultats d'une étude de Oliner et Sichel (2000)[113] sur les Etats-Unis. La première période est légèrement différente de celle retenue par Maddison (1987) mais les deux études restent comparables de sorte que l'étude d'Oliner et Sichel permet de raconter la fin de

[112] **Maddison A. (1991)**, Dynamic Forces in Capitalist Development, Oxford University Press.
[113] **Oliner S. et Sichel D (2000)**, « The Resurgence of Growth in the late 1990s : is Information Technology the Story ? », *Journal of Economic Perspectives*, vol.14.

l'histoire débutée par Maddison. La méthodologie est légèrement différente puisqu'elle montre la contribution à la croissance du PIB/tête et non du PIB (la variable dont on décrit les facteurs de croissance est donc la productivité du travail). Ainsi deux facteurs peuvent accroître la croissance de la productivité du travail : (i) la croissance de l'intensité capitalistique (K/L), la croissance de la PGF (progrès technique). Pour la contribution de l'augmentation du capital/tête à la croissance du PIB/tête, Oliner et Sichel (2000) ont par ailleurs mis en exergue la contribution de la croissance des TIC (investissements en logiciels et en matériels d'information et de communication). On constate qu'une part importante de l'intensification du capital repose sur une diffusion des TIC au cours de la décennie quatre-vingt-dix aux Etats-Unis. Par ailleurs, on relève un redémarrage progressif sur les deux périodes (1991-1995 et 1996-1999) de la croissance de la PGF (qui sur la dernière période semble avoir retrouvé un niveau similaire à sa croissance régulière aux Etats-Unis avant 1973). Il semble donc bien que la diffusion des TIC ait eu un effet significatif sur la productivité globale des facteurs et permis de renforcer le potentiel de croissance aux Etats-Unis.

	1974-1990	1991-1995	1996-1999
Taux de croissance de Y/L	1.37	1.53	2.57
Contribution de la croissance de K/L	0.81	0.62	1.10
Capital TIC	0.44	0.51	0.96
Matériel Informatique	0.25	0.23	0.59
Logiciels	0.09	0.23	0.27
Équipements de communication	0.09	0.05	0.10
Autre capital	0.37	0.11	0.14
Qualité du travail	0.22	0.44	0.31
PGF	0.33	0.48	1.16

Tableau 18 : Effet des TIC sur la PGF, Etats-Unis – source : Oliner et Sichel (2000)

Reste cependant à comprendre pourquoi il a fallu attendre vingt ans pour voir poindre les premiers effets de la technologie informatique sur la croissance. Plusieurs explications ont été fournies pour expliquer ce retard dans la contribution à la croissance des technologies informatiques. L'économiste Paul David a fait une analogie avec la diffusion de l'électricité qui avait également mis une vingtaine d'années avant d'avoir une contribution significative à la croissance au XIX^{ème} siècle. Des réorganisations dans les méthodes et dans l'organisation du travail avaient été nécessaires pour pouvoir profiter pleinement des gains de productivité permis par ces technologies (les entreprises avaient été forcées de repenser complètement leurs processus de production adaptés à la machine à vapeur). Ces réorganisations prennent du temps ce qui explique le retard. Il en est de même pour l'application et la diffusion des TIC (nécessité de formation et d'adaptation des utilisateurs).

On peut maintenant s'interroger sur la situation européenne (et donc française) par rapport à la situation américaine. Puisque la Stratégie de Lisbonne met l'accent sur l'importance de l'économie de la connaissance, on peut effectuer une comparaison entre la diffusion des TIC en France et aux Etats-Unis et ses effets sur la croissance du PIB/tête et de la productivité globale des facteurs. Cela est rendu possible par l'étude de Van Ark et Inklaar (2005)[114].

UE15

	1987-1995	1995-2000	2000-2004
Taux de croissance de Y/L	2,3	1,8	1,1
contribution de la croissance de K/L	1,2	1,0	0,8
Capital TIC	0,4	0,6	0,3
Autre Capital	0,8	0,4	0,5
PGF (*secteurs prod. TIC*)	0,2	0,4	0,2
PGF	0,9	0,4	0,0

Etats-Unis

	1987-1995	1995-2000	2000-2004
Taux de croissance de Y/L	1,2	2,3	2,8
contribution de la croissance de K/L	0,6	1,2	1,1
Capital TIC	0,5	1,0	0,6
Autre Capital	0,1	0,2	0,5
PGF (*secteurs prod. TIC*)	0,4	0,7	0,3
PGF	0,2	0,4	1,4

Tableau 19 : Source de la croissance de la productivité du travail dans l'UE15 et aux Etats-Unis – source : van Ark et Inklaar (2005)

La méthodologie de l'étude de van Ark et Inklaar est similaire à celle d'Oliner et Sichel (seul le découpage des sous périodes est légèrement différent). On constate d'une part un ralentissement de la croissance de la productivité du travail (du PIB/tête) pour l'UE15 au cours des trois sous périodes. La croissance de la productivité du travail passe de 2,3% de croissance annuelle moyenne, durant la

[114] **Van Ark B. et Inklaar R.C. (2005)**, « Catching up or getting stuck ? Europe's troubles to exploit ICT's productivity potential », *Research Memorandum GD-79, Groningen Growth and Development Centre*.

période 1987-1995, à 1,1% sur la période 2000-2004. Au contraire, la croissance de la productivité du travail aux Etats-Unis accélère à chacune des sous périodes (la croissance annuelle moyenne passant de 1,2% à 2,8%). Sur la période 1995-2004, la contribution de l'intensification du capital à la croissance de la productivité du travail est plus importante aux Etats-Unis que dans l'UE15, notamment du fait d'un plus fort investissement dans les TIC. Cette diffusion s'accompagne de gains de productivité dans les secteurs producteurs de TIC, gains de productivité qui se diffusent avec un peu de retard à l'ensemble de l'économie. Comme relevé par Oliner et Sichel, les Etats-Unis en fin de période retrouvent une croissance importante de la productivité globale des facteurs. L'Europe voit au contraire un ralentissement de la croissance de la PGF (celle-ci est nulle sur la période 2000-2004), ce qui explique la faible croissance du PIB/tête.
Cherchons maintenant à déterminer les raisons de cette chute de la productivité globale des facteurs au niveau européen.

7.1.5 Le retard européen en matière d'éducation et d'innovation

Les quelques tableaux ci-dessous permettent de comprendre les raisons de la plus faible croissance du PIB/tête et de la moindre productivité en Europe. Les effort consentis par les européens (il s'agit ici de la zone euro) sont plus faibles que ceux de leurs principaux concurrents en matière de leadership technologique, à savoir les Etats-Unis et le Japon. Les deux tableaux ci-dessous illustrent par exemple les évolutions récentes du fameux ratio dépenses de R&D/PIB (celui que l'Europe aimerait porter à 3% pour compléter sa stratégie de Lisbonne/Barcelone). On voit que la zone euro est encore bien en deçà des efforts consentis par ses concurrents et que l'effort des entreprises privées n'est pas suffisant.

	Etats-Unis	Zone euro	Japon	Suède
1997	2,58	1,82	2,83	3,55
1998	2,6	1,83	2,94	3,62
1999	2,65	1,87	2,94	3,65
2000	2,72	1,91	2,98	3,55
2001	2,74	1,94	3,09	4,27
2002	2,67	1,92	3,12	4,12
2003	2,62	1,95	3,15	3,98

Tableau 20 : Dépenses intérieures brutes de R&D (en % du PIB) – source : OCDE, Eurostat

	Etats-Unis	Zone euro	Japon	Suède
1997	1,91	1,14	2,05	2,35
1998	1,95	1,15	2,1	2,42
1999	1,98	1,2	2,1	2,47
2000	2,04	1,23	2,12	2,35
2001	2	1,24	2,26	3,07
2002	1,87	1,25	2,32	3,35
2003	1,79	1,23	2,38	3,14

Tableau 21 : Dépenses intérieures brutes de R&D (en % du PIB) des entreprises– source : OCDE, Eurostat

La Suède, qui ne fait pas partie de la zone euro mais appartient à l'Union européenne, a fait le pari d'une politique éducative et de recherche ambitieuse. La Suède semble l'exemple à suivre pour les européens. L'enjeu est donc de reproduire à l'échelle européenne ce semble fonctionner chez l'un de nos partenaires. Cela implique des changements importants dans de nombreux domaines : (i) un cadre efficace de protection des brevets à l'échelle communautaire, (ii) une amélioration de la coopération entre les entreprises et les autorités, (iii) une accélération des réformes réglementaires visant à améliorer la concurrence sur un certain nombre de marché puisque celle-ci apparaît comme un aiguillon de l'innovation, (iv) une meilleure coordination entre éducation et innovation.
Le retard européen peut également être souligné par les deux tableaux suivants. Le premier présente le nombre de chercheurs en entreprises pour 1000 emplois et le second le montant des dépenses en matière d'enseignements rapportées au nombre d'élèves et d'étudiants.

	Etats-Unis	Zone euro	Japon	Suède
1998	7,41	2,64	6,59	-
1999	7,61	2,81	6,71	5,61
2000	7,58	2,88	6,54	6,09
2001	7,2	2,97	6,72	6,58
2002	7,1	3,04	6,81	6,64
2003	7,37	3,09	7,26	6,71

Tableau 22 : Nombre de chercheurs en entreprises (pour 1000 emplois) – source : OCDE, STI

	Etats-Unis	Zone euro	Japon	Suède
1995	20 207	7 800	9 691	12 361
1999	19 220	8 524	10 278	14 222
2000	20 358	9 085	10 914	15 097
2001	22 234	9 214	11 164	15 188
2002	20 545	9 657	11 716	15 715

Tableau 23 : Dépenses annuelles au titre des établissements d'enseignements par élève/étudiant (en équivalent $ convertis sur base PPA) (4 grands pays pour la zone euro) **–** source : OCDE

Ces tableaux soulignent bien le décrochage de la zone euro en matières d'innovation et d'éducation. Alors que la croissance potentielle européenne demeure faible et que la volonté politique avait semblé définir une politique structurelle permettant d'améliorer cette situation, le rythme d'amélioration des

efforts consentis en matière d'innovation et d'éducation semble encore top faible pour espérer un rattrapage rapide des concurrents américains et japonais.
Cela est d'autant plus dommageable que l'Europe se veut un exemple dans le domaine de la protection de l'environnement. Le progrès technique dans le domaine de l'énergie par exemple, doit accompagner la croissance pour que celle-ci soit « soutenable » dans le long terme.

7.1.6 Croissance et environnement

Au début des années 1990, plusieurs études empiriques ont constaté l'existence d'une courbe en U-inversé entre divers indices de pollution et le niveau des revenus par tête. Ces auteurs lui attribuèrent le nom de "courbe environnementale de Kuznets"[115] à cause de la similitude avec les résultats des travaux de Kuznets (1955) sur les liens entre la croissance du revenu et les inégalités sociales.
Plus précisément, pour Grossman et Krueger (1993) (1995), le point de retournement pour la pollution par le SO2 intervient pour un revenu moyen par habitant de l'ordre de 4 000 à 5 000 dollars US. Il est souvent plus élevé pour les autres polluants mais, en général, se trouve aux alentours de 8 000 dollars US par habitant. Le message est le suivant : au-delà d'un certain niveau de richesse, la croissance économique s'accompagnerait d'une amélioration de l'état de l'environnement. Le graphique ci-dessous illustre ce phénomène :

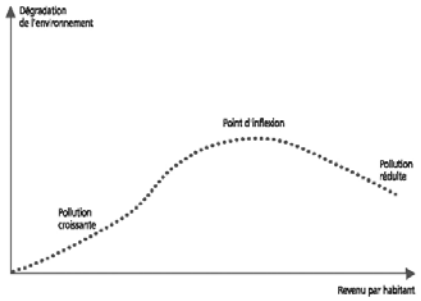


Figure 73 : Courbe de Kuznets « environnementale »

Grossman et Krueger (1993) (1995) ont cherché à établir une corrélation générale entre la croissance économique et les évolutions environnementales contemporaines, en rapprochant le revenu par habitant et les mesures concernant un certain nombre d'émissions polluantes. Les résultats de l'étude économétrique laissent à penser que les émissions polluantes augmentent en fonction des revenus jusqu'à une certaine limite, puis décroissent. L'explication fournie par les auteurs est qu'initialement il y a peu d'émissions polluantes du fait du faible niveau de production. Puis les débuts mal maîtrisés de l'industrialisation provoquent un surcroît de pollution. Enfin, les moyens financiers dégagés par l'augmentation de la richesse, le poids croissant des services et les changements des préférences des individus (plus portés vers la qualité de la vie à mesure que leur revenu individuel augmente) permettent de réduire les émissions.
Force est de préciser ici qu'il existe de nombreuses incertitudes sur l'évaluation de l'existence de cette courbe. Des résultats probants existent cependant pour certaines ressources (forêts) ou polluants (pollution de l'eau et certains gaz atmosphériques). Mais il est clair que ces résultats concernent un petit nombre de polluants et les régions les plus riches. De plus, il est important de souligner le fait que pour de nombreux économistes l'existence de la courbe environnementale de Kuznets est conditionnée par la mise en œuvre de politiques environnementales et de l'innovation. La décroissance de la pollution n'a rien d'automatique, et lorsque la courbe environnementale de Kuznets est vérifiée pour certains pays et certains polluants, elle est souvent liée à des objectifs de protection de l'environnement plus élevés que se fixent les pays riches, objectifs qui sont généralement accompagnés d'un changement technique délibéré. Le rôle du progrès technique dans la réduction des pollutions est donc central et nous amène à étudier le concept de développement durable (voir l'encadré sur **L'origine du concept de Développement Durable** page suivante).

Après avoir relevé l'importance de la technologie pour parvenir au développement durable, nous pourrons nous intéresser aux sources de la croissance (le capital humain et la formation, la R&D et l'innovation).

[115] Kuznets a reçu le prix Nobel 1971.

L'origine du concept de Développement Durable

Cet encadré pourrait également s'intituler « Existe-t-il des limites écologiques à la croissance économique ? », tant la question de la croissance économique et les enjeux environnementaux actuels sont interconnectés. Le changement climatique anthropique et la perte de la biodiversité sont les problèmes principaux de l'environnement : ils ont abouti au concept de développement durable qui s'est peu à peu imposé sur la scène internationale, comme au niveau local. Bien qu'il apparaisse dès 1980 dans un rapport de l'International Union for the Conservation of Nature, il faut attendre le rapport Bruntland du World Commission of Environment and Development (WCED) (1987) pour que le concept trouve la définition qui assurera ce succès (tout le monde a entendu parler de développement durable) :"***Le développement durable est celui qui répond aux besoins du présent sans compromettre la capacité des générations futures à répondre à leurs propres besoins***".
L'idée est donc de concilier le bien-être intertemporel (dont la qualité de l'environnement fait partie) et la poursuite du développement économique. Nous allons voir pourquoi ; après quelques rappels historiques relatifs à l'approche économique du développement durable.
Le premier débat contradictoire entre économistes au sujet de l'environnement oppose, dès la fin du XVIII^{ème} siècle, les tenants de la pensée malthusienne[116] et les adeptes de J.S. Mill[117]. Le débat se centre autour de l'épuisement des ressources naturelles et de son effet négatif sur le développement.
Cette controverse resurgit après plus d'un siècle avec le rapport du club de Rome intitulé Halte à la croissance (Meadows, Meadows, Randers et Behrens (1972)) qui prévoyait, pour le début du XXI^{ème} siècle, une décroissance forte de l'industrialisation et de la taille de la population mondiale à cause des pollutions excessives, de la production agricole limitée, des ressources minérales épuisées (les stocks d'or, de mercure, de pétrole, de zinc, d'argent...)...[118] Force est de constater après coup que cette prévision apocalyptique n'a pas eu lieu. Pourquoi ?
D'abord, les hypothèses du modèle reposaient sur une vision aprioriste des liens qui unissent ses différentes variables. En particulier, les prévisions en matière démographique ne reflétaient pas les tendances estimées à l'époque et reconnues par la communauté scientifique (le consensus). Ensuite, du point de vue de la Science économique, le modèle du club de Rome n'avait aucun système de prix, ni pour les ressources ni pour la production, ce faisant le retour du niveau des prix sur les quantités demandées n'était pas pris en compte. À la demande de l'Organisation des Nations Unies (ONU), Leontieff[119] réalise l'exercice de prévision suivant en 1977 : il reprend les hypothèses les plus pessimistes du rapport du club de Rome (même les moins consensuelles) et y ajoute les effets macroéconomiques des variations des prix sur les quantités demandées. Au final, seules deux ressources minérales mineures sont en danger d'épuisement à l'horizon de la fin du XX^{ème} siècle, les autres conclusions alarmistes du club de Rome (selon lesquelles la poursuite de la croissance économique est impossible) sont rejetées intégralement.

De ces différentes phases historiques, on tire l'enseignement que l'utilisation des ressources naturelles amène à trois types de limites potentielles à la croissance. Premièrement, l'épuisement des ressources non-renouvelables (nécessaires à la production des biens et des services) rend impossible cette production en l'absence de substituts. En effet, les ressources non-renouvelables peuvent être assimilées à des biens privés dont le stock s'épuise avec les prélèvements de façon inexorable. Deuxièmement, l'utilisation trop intensive des services rendus par les ressources renouvelables, empêchant leur régénération, imposerait un taux maximal d'utilisation de ces ressources qui viendrait limiter, en aval, la production des biens utilisant ces services rendus par l'environnement. Les ressources renouvelables (air pur, eau potable, etc.) ont une capacité naturelle d'assimilation qui les distinguent donc des ressources non-renouvelables (pétrole, gaz, etc.) dont le stock est fini. Troisièmement, si l'on intègre la qualité de l'environnement dans la mesure du bien-être économique, alors l'utilisation de l'environnement viendrait en elle-même réduire la possibilité d'accéder à un bien-être supérieur pour les générations futures. On considère généralement que la qualité de l'environnement se dégrade avec les pollutions émises à mesure que le système économique exploite les ressources naturelles. De plus, l'environnement a une capacité naturelle à assimiler une partie de ces pollutions ce qui permet de l'identifier à une ressource renouvelable (capacité de regénération).

Les principales conditions de mise en oeuvre du développement durable qui ont été mises en évidence dans les travaux récents des économistes sont au nombre de quatre : (i) le capital (physique et intellectuel) est un bon

[116] Malthus (1798) écrivait dans son Essai sur les principes de la population que la capacité de la race humaine à se multiplier allait, de façon tout à fait inexorable, buter sur la disparition des ressources naturelles ; l'équilibre social et économique se caractérisant alors par la misère et la privation. Pour Malthus, le rôle du progrès technique consiste à faire croître la production et le bien-être de façon transitoire avant qu'il ne s'évanouisse sous l'effet de l'épuisement et de la destruction des ressources. La solution pour éviter ce terrible équilibre de long terme passe par le contrôle démographique de la population, laquelle ne doit plus croître au-dessus de son niveau de subsistance.
[117] John Stuart Mill (1862) voit, au contraire, le développement de la connaissance dans le domaine agricole comme un élément clé qui permettra de répondre au défi posé par les ressources naturelles limitées. En plus de considérer, comme ses contemporains, l'environnement comme un facteur de production, J. S. Mill reconnaît à l'environnement une valeur en tant que source de bien-être (aménités). Ce dernier peut donc être considéré comme un économiste précurseur dans le domaine de l'économie de l'environnement.
[118] "*Au cas où aucun changement n'interviendrait dans notre système actuel, l'expansion démographique et l'expansion économique s'arrêteraient au plus tard au cours du siècle prochain*", Meadows et al. (1972).
[119] Leontieff reçoit le Prix Nobel d'économie 1973.

substitut de la ressource non-renouvelable qui s'épuise ou de la ressource renouvelable qui est surexploitée, (ii) l'épargne des ménages qui doit permettre l'investissement en capital doit être suffisamment élevée, (iii) la population ne doit pas croître trop rapidement, (iv) le rôle du progrès technique est primordial. Développons ensemble ce dernier point.

Pour certains modèles issus des nouvelles théories de la croissance[120], la croissance économique repose en fait sur une augmentation de la qualité des biens permettant à la production de rester constante en terme d'unité matérielle. La technologie peut donc aider à découpler le développement économique et la dégradation de l'environnement, mais ce découplage n'a en fait rien d'automatique. Comme le changement technologique apparaît de plus en plus souvent en réponse à une incitation économique, les besoins qui ne font pas l'objet d'une demande clairement identifiée sur un marché risquent de ne pas être pris en compte. Il n'est donc pas acquis que les progrès techniques actuels répondent aux besoins liés à l'environnement. Mais le rythme et la direction du progrès technique peuvent être influencés par certaines interventions : par exemple, la mise en place de marchés des permis d'émissions négociables (cf. l'encadré du chapitre 2 sur ce thème) ou de taxes sur les émissions polluantes.[121] La technologie est donc nécessaire au développement durable… mais pas suffisante, dans la mesure où la condition centrale des modèles réside dans la volonté des agents à substituer fortement de la consommation présente en faveur de la consommation future, afin de financer les efforts de court et moyen terme nécessaires à la découverte des technologies. Au final, l'efficacité des politiques repose donc sur le comportement des ménages. De plus, et c'est un élément qui s'impose à nous, les mécanismes naturels de régénération de l'écosystème doivent permettre d'assimiler la pollution à un rythme adéquat (faute d'un paramètre de régénération assez fort, le développement durable n'est techniquement pas possible).

7.2 Les effets économiques de l'éducation

Pour Becker (1964) (prix Nobel d'économie en 1992) et sa théorie du capital humain, les agents arbitrent dans leur décision d'études entre le supplément de salaire qu'ils retireront de celles-ci et la perte de revenu (perte de salaires potentiels (coût d'opportunité) et frais des études) occasionnée par la poursuite d'études.

Il existe toute une littérature micro-économétrique qui a cherché à étudier les rendements de l'éducation. En régressant le logarithme du salaire des individus sur leurs nombres d'années d'études, on peut ainsi mesurer le rendement d'une année supplémentaire d'étude par son effet marginal sur le salaire (sensé refléter la productivité marginale du travail des individus).

La mesure par le salaire représente bien le rendement d'une année supplémentaire d'études si les deux conditions suivantes sont respectées : (i) la productivité marginale et le salaire réel sont proportionnels, (ii) le surcroît de productivité lié à une année d'études supplémentaire peut directement être imputé au système éducatif.

C'est la deuxième condition qui semble la moins susceptible d'être satisfaite. L'éducation remplit en effet deux fonctions : (i) la fonction de transmission de connaissances (dont on espère qu'elle accroît la productivité), (ii) la fonction de signal. La deuxième fonction, tel que présentée par Spence (1973)[122], indique que le système éducatif sert également de filtre. L'obtention du diplôme sert autant à valider les compétences acquises qu'à attester d'un talent inné. Les entreprises sont en effet incapables d'observer les capacités et le talent inné des individus (asymétrie d'information). Le signal procuré par le diplôme permet de résoudre ce problème.

Le rendement de l'éducation, mesuré par l'impact d'une année supplémentaire sur le salaire peut donc être surestimé car la capacité à apprendre des individus est corrélée avec les compétences requises sur le marché du travail, les individus les plus doués sont aussi ceux qui ont le plus d'intérêt à poursuivre leurs études. Le supplément de salaire d'une année supplémentaire rémunère donc tout autant le supplément intrinsèque de qualité des individus que le bénéfice de l'éducation. Il convient cependant pour mesurer pleinement le rendement de l'éducation de prendre en compte un certain nombre d'externalités (dont certaines opérant à travers le progrès technique) ce qui ne peut être possible qu'au niveau macroéconomique.

Au niveau de l'impact macroéconomique de l'éducation, deux approches apportent des résultats différents. En considérant que le salaire réel est égal à la productivité marginale du travail et en utilisant comme base le modèle néo-classique de croissance de Solow (1956), des auteurs tels que Mankiw, Romer et Weil (1992) ont tenté de mesurer l'impact du capital humain, auquel ils font jouer un rôle identique à celui du capital physique (augmentation de l'efficacité productive du travail à technologie donnée). Accumuler du capital humain permet d'améliorer la productivité du travail et retarde donc le temps où l'économie vient buter sur les rendements décroissants du capital physique.

---

[120] Voir par exemple le Chapitre 5 de Aghion P., Howitt P. (1998), Endogenous Growth Theory, MIT Press.

[121] Pour les détails de ces résultats voir le modèle fondateur à croissance exogène de Stockey, N. (1998) "Are There Limits To Growth*?", International Economic Review*, vol.39, pp.1-31.

[122] Spence M. (1973), « Job market signaling », *Quaterly Journal of Economics*, vol.87, pp.355-374.

Le taux de croissance du PIB/tête pendant une période donnée est donc proportionnel au taux de croissance du niveau d'éducation pendant la même période[123] (le coefficient de proportionnalité peut être interprété comme le rendement macroéconomique de l'éducation). Le progrès technique étant exogène, l'impact de l'éducation sur la croissance se limite à ce seul effet sur l'efficacité du travail. Le maintien d'une croissance à long terme implique qu'il faudrait continuellement augmenter le niveau d'éducation de la population…

L'une des critiques qui a été adressée à l'approche de Mankiw, Romer et Weil est l'absence de prise en compte de l'effet de l'éducation sur le progrès technique. Or, dans une économie avec progrès technique, le niveau d'éducation affecte la croissance de long terme à travers un effet sur la vitesse d'adaptation au changement technologique (idée formulée par Nelson et Phelps (1966) et remise au goût du jour par Benhabib et Spiegel (2005)[124]). Cependant pour explorer cette piste, il convient de disposer d'un modèle de croissance dans lequel le progrès technique n'est plus exogène (comme chez Solow (1956)) mais endogène (comme dans les nouvelles théories de la croissance dites aussi théorie de la croissance endogène fondée sur l'innovation).

Voyons donc un peu plus précisément les enjeux de l'endogénéisation du progrès technique pour comprendre les faits en matière de croissance.

7.3 Les théories de la croissance

Malinvaud (1982)[125] énonce plusieurs objectifs que toute bonne théorie de la croissance économique doit atteindre. Malinvaud avait retenu les trois suivants : (i) expliquer les différents profils de croissance à long terme d'un pays à l'autre (d'une région à l'autre) et formuler des relations vérifiables et stables entre la croissance à long terme et tout une gamme de facteurs économiques sous-jacents, (ii) rendre compte des évolutions historiques à long terme (en particulier la convergence ou la divergence entre pays) et ainsi réconcilier la temporalité abstraite des modèles dynamiques avec le déroulement des changements économiques réels, (iii) éclairer conjointement les transformations institutionnelles et les changements technologiques.

Comme le note Aghion (2002), les théories de l'époque, notamment le modèle de Solow (1956), ne pouvaient prétendre atteindre aucun de ces objectifs[126]. L'apparition depuis le début de la décennie quatre-vingt-dix, d'un corpus théorique dit de la « croissance endogène » a permis des avancées très importantes dans la réalisation de ces objectifs et une meilleure appréhension des phénomènes de croissance.

7.3.1 Le modèle néoclassique : l'accumulation du capital

Le modèle néoclassique de croissance de Solow (1956) souligne que la croissance fondée sur l'accumulation du capital physique peut se tarir en raison de la décroissance de la productivité du capital. A long terme, seule la dynamique exogène du progrès technique assure une certaine croissance du PIB/tête[127].

Comme nous l'avons vu dans les faits stylisés présentés en début de chapitre, une implication essentielle du modèle de Solow est que les pays en retard dans l'accumulation du capital physique doivent finir par rattraper les pays initialement en avance (notion de convergence conditionnelle à défaut d'être absolue), ce qui s'avère empiriquement discutable. Le deuxième « défaut » du modèle

---

[123] Mankiw, Romer et Weil (1992) ont ainsi mis en évidence un effet significatif du taux de scolarisation des 12-17 ans (mesuré en moyenne sur la période 1960-1985) sur le niveau du PIB/tête en 1985.

[124] Nelson R., Phelps E. (1966), « Investment in Humans, Technological Diffusion, and Economic Growth », *American Economic Review*

Benhabib J., Spiegel M. (2005), « Human Capital and Technology Diffusion », in Handbook of Economic Growth, Aghion Ph. Durlauf S. (eds), Elsevier.

[125] Malinvaud E. (1982), Théorie économique, Dunod.

[126] Malinvaud (1994) note à propos de la théorie néo-classique de la croissance : « (…) *la nature du régime limite n'était pas l'objet d'un intérêt prioritaire. On était évidemment conscient de ce que la modélisation avait des effets à ce sujet. Mais comme on s'intéressait à autre chose, on se contentait de vérifier la compatibilité des évolution limites avec des idées intuitives sur ce qui devait caractériser le taux de croissance de long terme de nos économies, la signification concrète du long terme étant alors laissée délibérément floue. Nous étions loin d'imaginer que, trente ans plus tard, le critère principal de l'intérêt porté à un modèle de croissance puisse devenir le fait de savoir si le taux de croissance limite y apparaissait comme « endogène » ou « exogène »* ».

Malivaud E. (1994), « Regard d'un ancien sur les nouvelles théories de la croissance », Revue Economique, vol.44, pp.171-188.

[127] Voir l'annexe en fin de chapitre sur **le modèle de Solow**.

de Solow est le fait que la croissance repose sur un progrès technique exogène (« une manne tombée du ciel »…) : le seul moteur de la croissance de long terme du produit par tête n'est pas expliqué. C'est l'incapacité du modèle néo-classique à intégrer le lien entre le progrès technique et les efforts de recherche et développement, à endogénéiser le progrès technique, qui a abouti à la mise en place d'un nouveau corpus théorique en matière de croissance : la croissance endogène fondée sur l'innovation.

7.3.2 Les modèles de croissance endogène fondée sur l'innovation

7.3.2.1 Les nouvelles théories de la croissance

Trois grandes idées sont à la base des nouvelles théories de la croissance : (i) la croissance est principalement stimulée par le taux d'innovations technologiques, (ii) la plupart des innovations sont le fait d'entrepreneurs privés qui prennent des risques et investissent en R&D pour mettre à jour de nouveaux produits ou de nouvelles méthodes de production, (iii) la motivation à investir en R&D est influencée par l'environnement économique (protection des droits de propriété intellectuelle, niveau des taux d'intérêt…).

Les nouvelles théories de la croissance impliquent que les différences observées à la fois en niveaux de PIB/tête et en taux de croissance de la productivité (à court et moyen termes) d'un pays à l'autre, sont largement dues à des différences dans les systèmes et politiques de R&D et également aux différences entre les systèmes éducatifs, dans la mesure où ces systèmes conditionnent l'offre de travail qualifié, indispensable à l'activité de R&D à la base du progrès technique.

Deux types principaux de modèles de croissance endogène fondée sur l'innovation ont été introduits. L'un fait reposer la croissance sur la multiplication du nombre de produits disponibles (la variété). C'est le modèle de croissance fondée sur **l'innovation horizontale**. La multiplication des biens induit une plus forte spécialisation ce qui accroît soit l'utilité des ménages soit la productivité des entreprises (voir l'annexe en fin de chapitre sur **le modèle de croissance endogène fondée sur l'innovation : Romer (1990)**). Il en ressort une croissance fondée sur l'innovation qui peut être entretenue ou stimulée (d'où le vocable de croissance endogène) grâce à des politiques économiques appropriées. Une subvention publique à la recherche et développement réalisée par les entreprises privées permet d'accroître la quantité de ressources privées allouée à cette activité et renforce ainsi le niveau de l'innovation et le taux de croissance de long terme[128].

Le second modèle fait reposer la croissance sur l'amélioration de la qualité des biens disponibles ou l'amélioration de la productivité des biens intermédiaires. C'est le modèle dit de croissance fondée sur l'**innovation verticale**. Ce modèle introduit par Aghion et Howitt (1992) permet par ailleurs de décrire un phénomène que l'économiste J. Schumpeter avait mis en lumière en 1942 : le phénomène de « destruction créatrice ». Lorsqu'un bien de meilleure qualité apparaît, la consommation se reporte sur le nouveau bien qui présente un meilleur rapport qualité-prix et donc les rentes de l'ancien monopole disparaissent. L'innovation créé une nouvelle rente pour celui qui innove mais détruit les profits de l'innovateur anciennement en place. La dynamique de l'innovation est un incessant remplacement des situations de leadership (voir la présentation détaillée de ce modèle dans l'annexe de fin de chapitre **le modèle de croissance schumpétérienne : Aghion et Howitt (1992)**).

Cependant, la clarté de ce message initial s'est un peu perdue dans les débats sur la plausibilité empirique de ces modèles.

7.3.2.2 Investissement en R&D et « effet d'échelle »

Le développement des nouvelles théories de la croissance n'a pas été sans heurt ou débat entre théoriciens. Le principal débat qui a agité les théoriciens de la croissance fondée sur l'innovation au milieu de la décennie quatre-vingt-dix fût celui soulevé par Jones (1995). Celui-ci fît remarquer que les modèles de croissance endogène fondée sur l'innovation de première génération (Romer (1990) donc, mais également Aghion et Howitt (1992) ou Grossman et Helpman (1991)) avaient tous le travers de présenter un effet d'échelle, c'est-à-dire que l'un des déterminants du taux de croissance de long terme du PIB/tête était la taille de l'économie ou sa population ce qui implique qu'une croissance de la population entraînerait une accélération de la croissance (constatation bien entendu invalidé par les faits). Les modèles qui parvenaient à supprimer l'effet d'échelle, comme celui de Jones (1995), présentaient la particularité d'avoir un taux de croissance de long terme dépendant uniquement de paramètres exogènes (retour à Solow…). C'est ce qu'on a appelé la croissance semi-endogène. Sans rentrer trop dans les détails de cette querelle de théoriciens, nous relèverons

[128] La politique économique retrouve une capacité à influencer le niveau de croissance de long terme, ce qui n'est pas envisagé chez Solow où une politique économique ne peut avoir que des effets transitoires.

néanmoins que le corpus des nouvelles théories de la croissance fondée sur l'innovation a convergé à la fin de la dernière décennie sur un modèle de croissance endogène sans effet d'échelle (voir par exemple Howitt (1999)[129] mais de nombreux autres auteurs ont introduit des modèles théoriques similaires). L'idée que la croissance reposant sur des innovations, réalisées par des entreprises privées, peut être renforcée par une politique de subvention à la R&D fait désormais l'objet d'un large consensus parmi les économistes.

7.3.2.3 Les implications de ces modèles en matière de politique économique

En résumé, le message de ces nouveaux modèles de croissance est encourageant : le niveau du taux de croissance de long terme dépend de facteurs propres à chaque économie, en particulier des politiques de R&D qui sont menées. C'est pourquoi, je vous précisais dans l'introduction que ces modèles apparaissent en filigrane derrière les imprécations de la stratégie de Lisbonne. L'idée qu'une augmentation des ressources affectées aux activités d'innovation (l'objectif de 3% de dépenses de R&D par rapport au PIB) serait bénéfique à la croissance européenne n'est rien d'autre que le message délivré par les modèles de croissance fondée sur l'innovation de la décennie précédente.
Cependant, nous l'avons vu dans la sous-section sur les effets économiques de l'éducation, la pertinence d'une politique de promotion de la recherche ne peut être que renforcée par le choix d'une politique d'éducation visant à accroître le niveau de capital humain (améliorer les qualifications de la population et notamment être capable de « fournir » la population qualifiée nécessaire à la recherche et développement).
Pour comprendre cela, nous pouvons jeter un regard sur la figure ci-dessous qui décrit le marché du travail des travailleurs qualifiés (ce graphique est utilisé par Romer (2000)).
Les subventions à la recherche et développement privée (dont on a vu l'efficacité pour accroître le taux de croissance de long terme) aboutissent donc à une plus forte demande de travail d'individus qualifiés (ingénieurs, chercheurs…). On peut considérer que l'offre de travail qualifié est à court terme relativement inélastique (constante) car dépendante des structures de formation existante. Aussi lorsqu'une subvention à la recherche et développement est introduite, cela risque d'avoir pour seul effet d'accroître les salaires de la population qualifiée, si jamais l'offre de travail qualifié n'est pas capable de s'ajuster.

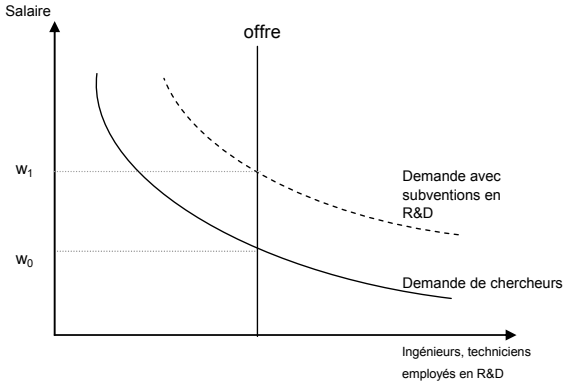


Figure 74 : Effets de subventions à la R&D quand l'offre de travail qualifié est inélastique

Pour que l'offre de travail qualifié s'ajuste, c'est-à-dire que le niveau de l'offre se déplace vers la droite, il convient que le système de formation et d'éducation soit capable d'améliorer sa performance. C'est à cette seule condition que les efforts des politiques d'innovation seront pleinement couronnés de succès.
Il convient de préciser qu'en France, cela pourrait être obtenu relativement aisément par un redéploiement partiel des ressources affectées à l'éducation primaire et secondaire vers l'éducation supérieure. En effet, on peut considérer que la France est relativement proche de la frontière technologique. Elle a donc besoin de développer ses capacités d'innovation (c'est-à-dire former des chercheurs et des ingénieurs…) plus que ses capacités d'imitation (techniciens, ouvriers qualifiés…). C'est sur ce point que nous conclurons.

[129] **Howitt P. (1999)**, « Steady Endogenous Growth with Population and R&D Inputs Growing », *Journal of Political Economy*, vol.107, pp.715-730, voir aussi **Jones C.I. (1999)**, « Growth : With or Without Scale Effects », *American Economic Review*, vol.89, pp.139-144.

7.4 Conclusion : Quelle politique de croissance ?

Le développement et la croissance ne sont pas toujours influencés par les mêmes déterminants, comme le montre le modèle de Aghion, Acemoglu et Zilibotti (2002) (celui-ci est présenté succinctement en annexe, **Education, innovation et distance à la frontière**).

L'idée de Aghion, Acemoglu et Zilibotti est que les deux modes d'acquisition du progrès technique, l'imitation et l'innovation, ne requièrent pas les mêmes institutions « appropriées ». L'imitation est plutôt favorisée par des investissements de long terme dans des entreprises de grande taille bénéficiant de crédits bancaires. L'innovation peut au contraire requérir plus d'initiative, davantage de prise de risques. Un pays qui veut favoriser l'innovation doit par conséquent favoriser les institutions basées sur le marché et la flexibilité, augmenter le rôle des marchés financiers par rapport aux banques (cf. chapitre 3), favoriser la concurrence et l'entrée de nouvelles firmes (cf. chapitre 2), permettre une plus grande flexibilité sur le marché du travail (cf. chapitre 5).

Les institutions les plus favorables à la croissance changent à mesure qu'un pays se rapproche de la frontière technologique. Loin de la frontière, un pays croît plus vite lorsqu'il adopte ce qu'Aghion, Acemoglu et Zilibotti (2002) nomment les **institutions ou politiques d'investissement** (voir la politique française des trente glorieuses ou l'expérience de la Corée du Sud après guerre). Un pays proche de la frontière technologique devra pour améliorer sa croissance mettrent en place des **institutions ou politique d'innovation**.

L'un des facteurs clés de la politique d'innovation réside dans le développement ou l'amélioration de l'enseignement supérieur qui apparaît complémentaire avec la recherche et l'innovation. A mesure qu'une économie s'approche de la frontière technologique mondiale, c'est l'ambition de l'économie européenne, une priorité doit être donnée à ce type d'éducation pour que les efforts déployés par les entreprises en matière d'innovation ne soient pas vains. Les différents paramètres « institutionnels » à même de placer l'Europe sur une trajectoire de croissance vertueuse sont connus (plus de concurrence sur le marché des biens et services, plus de flexibilité sur le marché du travail, un marché financier réactif et un système éducatif performant…). Nous avons cependant signale différentes difficultés tout au long de ce polycopié (en matière de manque de concurrence, en matière de financement, en matière de manque de flexibilité…).

Enfin, à la lumière du modèle de Aghion, Acemoglu et Zilibotti présenté en annexe, on comprend mieux la réussite de la France (de l'Europe) à rattraper son retard de productivité par rapport aux Etats-Unis sur la période 1945-1970. Eloignée de la frontière technologique, elle a pu bénéficier d'externalités importantes en matière d'imitation et son système scolaire (primaire et secondaire) de qualité a accru ses capacités à assimiler les techniques, notamment venues des Etats-Unis.

A mesure que la France s'est rapprochée des Etats-Unis, les possibilités offertes par l'imitation, c'est-à-dire la mise en œuvre de solutions techniques inventées ailleurs, se sont amoindries et l'importance de l'innovation comme moteur de croissance est apparue primordiale. La nouvelle vague technologique des années quatre-vingt-dix a encore renforcé l'importance de l'innovation et le besoin de transférer des ressources au sein même de son système de formation et d'éducation[130].

Le retour du rythme de la croissance de la productivité en France à son niveau des années quatre-vingts passe par la capacité de l'économie française à assimiler et exploiter les nouvelles opportunités technologiques, celles offertes par les TIC et celles que constituent les technologies du futur (biotechnologies, énergies propres…).

C'est avec une politique ambitieuse de recherche et d'éducation que l'Europe pourra réussir son pari (la stratégie de Lisbonne), renforcer sa croissance et donc améliorer un certain nombre de problèmes structurels et s'atteler à définir les modalités d'une trajectoire de croissance compatible avec un développement durable[131].

---

[130] Ces chiffres sont connus mais il convient ici de les rappeler. Alors qu'un élève du secondaire coûte 36% de plus en France que dans la moyenne des pays de l'OCDE, un élève de l'enseignement supérieur (toutes formations confondues) coûte 11% de moins en France que dans la moyenne des pays de l'OCDE. L'enseignement supérieur est le parent pauvre de l'éducation nationale. En 1999, la France investissait 1,1% de son PIB en matière d'enseignement supérieur, les Etats-Unis 2,3%...

[131] En matière de politique liée au dérèglement climatique, il est en effet impératif de souligner que l'enjeu majeur réside dans les négociations internationales qui ont lieu actuellement et qui prépare l'après Protocole de Kyoto. En effet, fin 2012, ce premier protocole prendra fin, il s'agira alors de réaliser un nouvel accord qui cette fois engagera les principaux pollueurs (les Etats-Unis à l'heure actuelle, la Chine et l'Inde demain : ces pays n'étant pas contraints par le Protocole actuel).

7.5 Annexes

7.5.1 Le modèle de Solow

Le propos de l'analyse néo-classique de la croissance (modèle de Solow [1956][132]) n'était pas d'expliquer les sources de la croissance, puisque celles-ci sont considérées comme exogènes, mais de montrer sous quelle condition l'économie pouvait converger vers son sentier de croissance régulier potentiel déterminé par les facteurs démographique (taux de la croissance de la population) et technologique (productivité du travail).
En cela, le modèle néo-classique de croissance entendait s'inscrire en faux contre le modèle de croissance instable de l'économiste Harrod, d'obédience keynésienne.

Le modèle de Solow avec épargne exogène peut être résumé sous la forme suivante.
On considère que le taux de croissance de la population est n. La force de travail au temps t est $L(t)=L(0)e^{nt}$. On suppose que la productivité du travail croît régulièrement à un taux exogène μ. La productivité du travail au temps t est donc $A(t)=A(0).e^{\mu t}$. Soit $Y(t)=F(K(t),A(t).L(t))$ la fonction de production qui caractérise la technologie de production dans cette économie. On fait l'hypothèse qu'il existe des rendements d'échelle constants dans les facteurs K(t) et (A(t).L(t)).
On peut exprimer les variables sous forme de variable par tête ajustée de la productivité. Le produit par tête efficace est $y(t)=\frac{Y(t)}{A(t)L(t)}$ et le capital par tête efficace est $k(t)=\frac{K(t)}{A(t)L(t)}$. La fonction de production s'écrit alors $y(t)=f(k(t))$. Soit s la part du produit consacré à l'épargne. L'épargne sert à l'investissement I(t). Le taux d'épargne est donc : $s=\frac{Y(t)-C(t)}{Y(t)}=\frac{I(t)}{Y(t)}$. Soit δ le taux de dépréciation du capital physique. L'accroissement du stock de capital est la différence entre l'investissement brut et la dépréciation du capital : $\dot{K}(t)=I(t)-\delta.K(t)$.
L'évolution du capital par tête efficace est l'équation fondamentale du modèle de Solow. Le taux de croissance du capital par tête efficace est le suivant :

$$\frac{\dot{k}(t)}{k(t)}=s\frac{y(t)}{k(t)}-(\delta+\mu+n)$$

L'équation d'évolution du capital est quant à elle : $\dot{k}(t)=sy(t)-(\delta+\mu+n)k(t)$.
Le long du sentier de croissance régulière, le stock de capital va croître à un taux égal à la somme du taux de croissance de la population et du taux de croissance de la productivité du travail, de sorte que le capital par tête efficace va demeurer constant ($\dot{k}=0$).
Le produit par tête efficace d'équilibre est : $y^*=f(k^*)=\frac{(\delta+\mu+n)}{s}k^*$. Connaissant la forme de la fonction f(.), on peut alors déterminer le niveau d'équilibre du stock de capital par tête efficace. Pour ce niveau du stock de capital par tête efficace, l'épargne égalise l'investissement et permet de maintenir constant le ratio capital-travail ajusté de la productivité, malgré la croissance démographique, le progrès technique et la dépréciation du capital.
Les propriétés du modèle sont les suivantes : (i) le taux de croissance de l'économie est égal à la somme du taux de croissance de la population active et de la productivité du travail, (ii) l'intensité capitalistique et la productivité du travail croissent au taux μ, (iii) l'équilibre est stable et le retour à l'équilibre est spontané en cas de déviation initiale ou transitoire.

[132] **Solow Robert [1956]**, « A contribution to the Theory of Economic Growth », *Quaterly Journal of Economics*, pp.65-94.

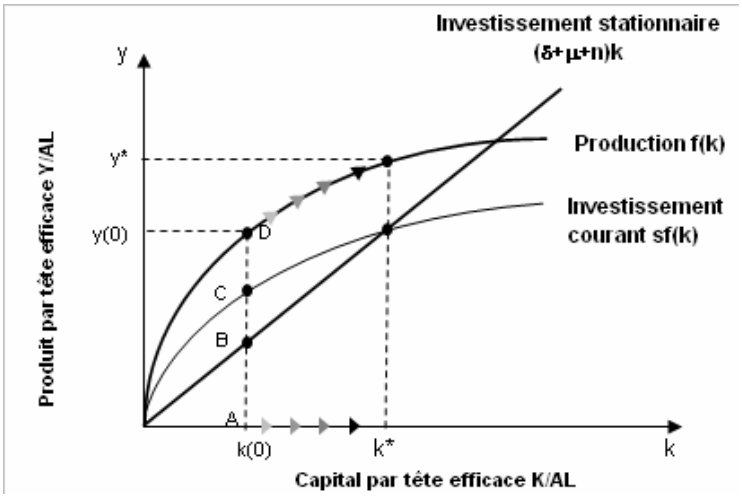


Figure 75 : Dynamique du modèle de Solow

On peut donner une interprétation graphique de la dynamique du capital et de la production par tête efficace. Considérons, sur la figure ci-dessus, que l'économie est initialement dotée d'un stock de capital par tête efficace k(0). A ce niveau, le produit par tête efficace est donné par AD. L'investissement par tête efficace est donné par AC. Or la quantité d'investissement par tête efficace nécessaire au maintien de ce niveau de capital est donnée par AB. L'investissement courant excède l'investissement stationnaire, strictement nécessaire au maintien d'un capital par tête efficace. Le capital par travailleur efficace k va donc augmenter, ainsi que la production par tête efficace. Cet ajustement va perdurer jusqu'à ce que l'investissement soit tout juste suffisant pour compenser la dépréciation du capital, la croissance démographique et l'élévation de la productivité des travailleurs, c'est-à-dire au niveau k*. L'état stationnaire de l'économie est caractérisé par le fait que le capital et la production par tête efficace sont constants respectivement aux niveaux k* et y*.

7.5.2 L'exercice de comptabilité de la croissance

La comptabilité de la croissance est une méthode qui permet de séparer ce qui, dans la croissance du PIB, est imputable à la croissance des facteurs de production (capital et travail) de ce qui relève du progrès technique. Cette méthodologie s'est développée à partir de la fin des années cinquante, à la suite des travaux de Solow. Le modèle sous-jacent à cette décomposition comptable est le modèle de Solow, dont les hypothèses sont néo-classiques (concurrence parfaite, rendements d'échelle constants).

On adopte pour cette présentation de l'exercice de comptabilité de la croissance une fonction de production de type Cobb-douglas : $Y(t) = A(t).K(t)^{\alpha} L(t)^{1-\alpha}$ qui satisfait à l'hypothèse de rendements d'échelle constants.

Le niveau du produit Y(t) est relié à deux types d'arguments : les volumes des facteurs de production (capital K(t) et travail L(t)), et un indice du niveau technologique A(t) de l'économie. Le terme A(t) est appelé la productivité globale des facteurs ou PGF. En passant en taux de croissance la fonction de production (en log-différenciant par rapport au temps), on obtient :

$$\frac{\dot{Y}(t)}{Y(t)} = \frac{\dot{A}(t)}{A(t)} + \alpha \frac{\dot{K}(t)}{K(t)} + (1-\alpha)\frac{\dot{L}(t)}{L(t)}$$

Le premier terme ($\dot{A}(t)/A(t)$) représente l'évolution de la PGF, c'est-à-dire le taux de progrès technique exogène. Les autres termes les contributions respectives du capital et du travail à la croissance.

En concurrence parfaite, les facteurs sont rémunérés à leur productivité marginale (voir l'annexe sur le programme du producteur). Les hypothèses de rendements constants de la fonction Cobb-Douglas et de concurrence parfaite permettent d'exprimer les parts des facteurs de la façon suivante :

$$\alpha = \frac{r(t).K(t)}{p(t).Y(t)} \quad \text{et} \quad 1-\alpha = \frac{w(t).L(t)}{p(t).Y(t)}$$

Où r(t) est le coût d'usage nominal du capital et w(t), taux de salaire nominal et p(t) le niveau général des prix associé à la production. Toutes les variables de l'équation de croissance du produit sont observables (taux de croissance du PIB, du capital et du travail, parts de la rémunération du capital

dans le produit supposées le plus souvent constantes) à l'exception du taux de croissance de la PGF. Cette équation va donc permettre de calculer ce terme inconnu, que l'on nomme **résidu de Solow**, en référence à la façon dont il est obtenu :

$$\frac{\dot{A}(t)}{A(t)} = \frac{\dot{Y}(t)}{Y(t)} - \alpha \frac{\dot{K}(t)}{K(t)} - (1-\alpha)\frac{\dot{L}(t)}{L(t)}$$

Une autre manière de présenter la contribution des différents facteurs à la croissance est de raisonner par tête. Soit y=Y/L et k=K/L respectivement le produit par tête et le capital par tête (ou intensité capitalistique). On peut exprimer le taux de croissance du produit par tête de la façon suivante :

$$\frac{\dot{y}(t)}{y(t)} = \frac{\dot{A}(t)}{A(t)} + \alpha \frac{\dot{k}(t)}{k(t)}$$

La croissance de la productivité du travail (Y/L) dépend à la fois de la croissance de la PGF et de l'élévation de l'intensité capitalistique. Lorsque K et L croissent au même taux, la productivité du travail ne dépend plus que du progrès technique.

Quelques remarques sur la méthodologie de comptabilité de la croissance:

- Cette méthodologie repose sur les conditions fortes de concurrence parfaite et de constance des rendements d'échelle, ceci afin que la structure des rémunérations des facteurs de production reflète fidèlement leur contribution à la formation du produit agrégé.
- Seuls deux facteurs de production rémunérés sont identifiés tandis que le progrès technique, l'innovation technologique, est ici considéré comme un bien libre auquel chacun à accès sans coût.

Sous ces hypothèses, le résidu mesure la contribution de la productivité globale des facteurs à la croissance. Si les hypothèses sont inadaptées, il y a fort à parier que le résidu est contaminé par d'autres facteurs que le progrès technique (concurrence imparfaite, facteurs non pris en compte).

7.5.3 Le modèle de croissance endogène fondée sur l'innovation : Romer (1990)

Une partie importante du progrès technique résulte du comportement d'entrepreneurs/innovateurs privés qui recherche le profit. Dans le modèle de Romer (1990)[133] que l'on va exposer, le progrès technique résulte de la production de connaissances techniques, d'idées permettant de produire des biens nouveaux, par des entrepreneurs privés. Ce modèle est dit de croissance endogène car l'activité d'innovation y est précisément décrite et impulse la dynamique de l'économie. L'innovation est protégée par un système de brevets de sorte que l'innovateur obtient un pouvoir de marché et donc réalise des profits. Le progrès technique est assimilé dans ce modèle à une augmentation du nombre de variété de biens intermédiaires. La variété est source de richesse car elle permet d'améliorer la productivité des facteurs de production dans la fonction de bien final (Idée de Young [1928] que la spécialisation sectorielle améliore les rendements).

Le modèle est constitué de trois secteurs :

- Un secteur final qui produit, en condition de concurrence parfaite, le bien final à partir de travail et des N biens intermédiaires existants.
- Un secteur intermédiaire où N entreprises en situation de monopole produise l'un des biens intermédiaires, à partir du bien final.
- Un secteur de R&D, en concurrence, ou de nouvelles technologies sont mises à jour, à partir de travail et du stock de technologies déjà existants.

Le nombre de sous-secteurs dans le secteur de biens intermédiaires (nombre de biens intermédiaires différenciés) se confond donc avec le nombre de technologies découvertes (et donc de brevets).

La fonction de production de bien final est la suivante : $Y = L_Y^{1-\alpha} \int_0^N x_i^{\alpha} di$. Une quantité Y(t) de bien final est produite à partir d'une quantité de travail L_Y et d'un continuum de N biens intermédiaires, chacun en quantité x. Le paramètre α est compris entre 0 et 1. Les rendements d'échelle sont donc constants dans les quantités physiques L_Y et x mais deviennent croissants lorsqu'on assimile N à un input. Soit $X = Nx$, en considérant que $x_i = x \quad \forall i$, alors $Y = N^{1-\alpha} L_Y^{1-\alpha} X^{\alpha}$. Les firmes qui produisent le bien final sont en situation de concurrence. Elles considèrent les prix comme donnés sur tous les marchés. Soit w le taux de salaire et p_x le prix des biens intermédiaires disponibles. La firme représentative du secteur de bien final choisit ses demandes de facteur X et L_Y en égalisant la

[133] **Romer Paul (1990)**, « Endogenous Technological Change », *Journal of Political Economy*, vol.98, pp.71-102.

productivité marginale de ces facteurs et leurs coûts réels. Les demandes de facteurs sont donc les suivantes :

$$L_Y = w^{-\frac{1}{\alpha}} X N^{\frac{1-\alpha}{\alpha}} (1-\alpha)^{\frac{1}{\alpha}}$$

$$x = \frac{X}{N} = p_x^{-\frac{1}{1-\alpha}} L_Y \alpha^{\frac{1}{1-\alpha}}$$

Chaque variété de bien intermédiaire est produite par une firme unique en position de monopole. Ce pouvoir de marché résulte de l'existence d'un système de brevets garanti par l'Etat. Chaque unité de bien intermédiaire est produite à partir d'une unité de bien final. Le programme d'une entreprise produisant un bien intermédiaire est donc $Max\ \pi(x) = p_x x - x \quad s.c.\ x = p_x^{-\frac{1}{1-\alpha}} L_Y \alpha^{\frac{1}{1-\alpha}}$ et aboutit à la fixation du prix $p_x=1/\alpha$ (comme α<1, on constate bien que l'entreprise tarifie au-dessus de son coût marginal, le prix du bien final étant considéré égal à l'unité). En réintroduisant ce prix p_x dans la fonction de demande, on obtient la quantité de chacun des biens intermédiaires spécialisés produite à l'équilibre ainsi que les rentes de monopole qui vont servir à rémunérer le détenteur du brevet :

$$x^* = \alpha^{\frac{2}{1-\alpha}} L_Y$$

$$\pi^* = (1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}} L_Y$$

La création de nouvelles variétés de biens intermédiaires est le résultat des investissements effectués dans l'activité de R&D. Le nombre de variétés peut être perçu comme un stock de connaissances techniques dans lequel chacun peut puiser pour innover. Chaque nouvelle connaissance technique donne lieu à un dépôt de brevet de sorte que l'innovateur va pouvoir s'approprier une partie du revenu généré par l'innovation (versement de royalties à partir des profits réalisés par les entreprises de biens intermédiaires).

La modélisation du flux de nouvelles connaissances techniques se fait sous la forme d'une fonction de production de connaissances de la forme $\dot{N}(t) = F(N(t), L_N)$ où L_N est la quantité de travail mobilisée dans les tâches de recherche et développement, N(t) le stock de connaissances mises à jour et $\dot{N}(t)$ le flux courant de nouvelles connaissances. La fonction F(.,.) est une fonction croissante dans son deuxième argument (et c'est tout…). Il convient en effet de bien distinguer ce qui relève de la décision individuelle des firmes de R&D, de ce qui provient des externalités. Dans ce modèle, deux types d'externalité sont présents dans la production de connaissances :

- L'influence positive (mais possiblement négative) du stock de connaissances techniques antérieurement mises à jour, N(t), sur l'efficacité des ressources privées déployées, L_N (externalités intertemporelles des connaissances du fait du caractère non exclusif[134] et non rival[135] de la connaissance[136]). Lorsque F'(N)>0 (cas le plus fréquemment admis) les chercheurs sont aidés par la possibilité de puiser dans les connaissances antérieures pour en élaborer de nouvelles : on parle de « *Standing on Giant Shoulders' Externality* ». Lorsque F'(N)<0, les connaissances les plus faciles à obtenir sont sensées être apparues les premières de sorte qu'à mesure de l'apparition de nouvelles connaissances techniques, il devient de plus en plus difficile d'innover. On qualifie ce cas où le phénomène d'externalités intertemporelles de connaissances est négatif de « *fishing out* » (l'image est celle d'une mare

[134] Un bien est non exclusif lorsqu'il est impossible d'exclure les autres de sa consommation. Il est difficilement appropriable. A la différence d'une machine-outil qui est installée dans une usine, la connaissance se diffuse et peut être captée par d'autres que ceux qui l'ont produite.

[135] Un bien est non rival lorsque son utilisation ou sa consommation par un individu n'empêche pas l'utilisation ou la consommation par d'autres individus. Tous les élèves de 2ème année peuvent utiliser en même temps le théorème de Pythagore sans se gêner puisque c'est une connaissance (bien non rival). En revanche, ils ne peuvent pas tous dormir dans la même chambre ou boire le même café (biens rivaux).

[136] Un bien à la fois non exclusif et non rival possède les propriétés de ce qu'on appelle un bien public. La connaissance a donc les caractéristiques d'un bien public. Pour cela il existe une trop faible incitation à la produire car les producteurs de connaissances ne peuvent s'accaparer l'ensemble des profits de celle-ci du fait de sa diffusion (externalités). C'est pour cela qu'ont été mis en place les systèmes de propriété intellectuelle (brevets) afin de permettre au producteur d'une connaissance technique d'en tirer les bénéfices (position de monopole dans la production) tout en permettant la diffusion à tous de cette connaissance (obligation de divulgation).

où l'on pêcherait. Plus on retire de poissons et moins il en reste dans la mare d'où la baisse de la fréquence de pêche fructueuse…).

- La duplication stérile des activités de R&D lorsque plusieurs entreprises mobilisent des ressources pour entreprendre des recherches sur le même domaine technique. On qualifie cette externalité de « *stepping on toes externality* ».

Pour une firme privée spécialisée dans la R&D, la fonction d'innovation qu'elle perçoit est la suivante : $\dot{N}(t) = \mu L_N$ où μ est la productivité de l'input privé de recherche L_N (l'entreprise n'a pas conscience des externalités qu'elle subit ou dont elle bénéficie, elle ne les intègre pas dans ses calculs). Au niveau global, la fonction de production de connaissances est en fait la suivante :

$$\dot{N}(t) = \beta N(t)^{\phi} L_N{}^{\lambda}$$

Le paramètre ϕ traduit l'influence de l'externalité intertemporelle de connaissances ($\phi>0$ ou $\phi<0$ selon le cas retenu, cf. ci-dessus) et le paramètre λ tient compte de l'effet de duplication de la recherche ($0< \lambda \leq 1$).

La forme de la fonction de production de connaissances est très importante car elle conditionne la réalisation d'un sentier de croissance endogène auto-entretenue. La réalisation d'une croissance régulière de long terme nécessite que N s'accroisse à un taux constant. Romer [1990] impose donc les restrictions suivantes : (i) pas de croissance démographique (L_N constant), (ii) rendements constants en N ($\phi=1$, condition dite « fil du rasoir ») et (iii) pas de duplication de l'effort de recherche ($\lambda=1$, mais ce n'est ni nécessaire ni même important…).

Les brevets qui protègent les innovations sont supposés de durée de vie infinie et constituent les actifs dont disposent les ménages pour transférer leur consommation vers le futur (épargner).

Soit $V_N(t)$ la valeur d'un brevet couvrant une nouvelle variété. C'est la somme actualisée des profits futurs rendus possibles par cette innovation : $V_N(t) = \int_t^{\infty} e^{-r(\tau - t)} \pi(\tau) d\tau$.

Lorsqu'on dérive cette équation par rapport au temps, on obtient la condition de non-arbitrage (équivalence entre les rendements de la détention de l'actif brevet et un placement d'un montant équivalent sur l'actif sans risque : le taux d'intérêt) : $\frac{\pi(t)}{V_N(t)} + \frac{\dot{V}_N(t)}{V_N(t)} = r(t)$. La rémunération de la détention d'un brevet permet l'obtention de dividendes et potentiellement d'un gain en valeur. L'équation de non-arbitrage indique donc que le taux de dividende π/V plus le taux de croissance g_V de la valeur des brevets égalise le taux d'intérêt sans risque r(t).

Le secteur de la R&D étant concurrentiel, il y a libre entrée dans ce secteur de sorte que la valeur d'un brevet égalise son coût d'obtention. Le programme $\max_{L_N} \dot{N}(t) V_N(t) - wL_N \quad s.c. \dot{N}(t) = \mu L_N$ aboutit a l'égalité $V_N(t) = \frac{w}{\beta N(t)}$.

Les deux équations dynamiques qui régissent l'évolution des variables du modèle sont la fonction de production de connaissances (le cœur du réacteur de ce modèle…) et la règle de Keynes-Ramsey (voir l'annexe dans les Annexes générales en fin de polycopié sur **le programme dynamique du consommateur** et l'obtention de cette règle) :

$$\dot{N} = \beta N L_N$$

$$\dot{C} = (r - \rho) C$$

On dispose également pour le bouclage du modèle de deux conditions : la condition de plein emploi (CPE) et la condition de non arbitrage (CNA) :

$$L_Y + L_N = L \qquad \text{(CPE)}$$

$$\frac{\pi}{V_N} + \frac{\dot{V}_N}{V_N} = r \qquad \text{(CNA)}$$

La valeur d'un brevet est $V_N = \frac{\alpha^{\frac{2\alpha}{1-\alpha}}(1-\alpha)}{\beta}$ de sorte que son taux de croissance est nul et la CNA devient $\alpha\beta L_y = r$. La loi des débouchés (tout ce qui est produit est consommé) permet décrire : C=Y-X (le bien final qui n'est pas utilisé pour produire des biens intermédiaires est consommé) et

donc $C = NL_N \alpha^{\frac{2\alpha}{1-\alpha}}\left(1-\alpha^2\right)$. Le modèle peut être résumé par les deux équations différentielles suivantes :

$$\dot{N} = \beta N\left(L - \frac{C}{N\alpha^{\frac{2\alpha}{1-\alpha}}\left(1-\alpha^2\right)}\right)$$

$$\dot{C} = C\left(\alpha\beta\frac{C}{N\alpha^{\frac{2\alpha}{1-\alpha}}\left(1-\alpha^2\right)} - \rho\right)$$

Le long du sentier de croissance équilibrée, C et N croissent au même taux, donc : $C/N = \frac{(\beta L+\rho)\left(\alpha^{\frac{2\alpha}{1-\alpha}}\left(1-\alpha^2\right)\right)}{\beta(1+\alpha)}$. En reportant ce ratio dans l'une des équations du système, on obtient le taux de croissance endogène des variables de consommation et du nombre de biens intermédiaires (de connaissances techniques) le long du sentier de croissance équilibrée :

$$\boxed{g = \frac{\alpha\beta L - \rho}{1+\alpha}}$$

Le taux de croissance de l'économie est d'autant plus élevé que : (i) l'échelle de l'économie L est importante, (ii) les agents sont patients (ρ faible), (iii) la productivité de la recherche est forte (β élevé). La forme de la fonction de production du bien final influence le taux de croissance à travers le paramètre α.

Le modèle de Romer (1990) a permis d'endogénéiser le processus d'innovation et de progrès technique qui n'est donc plus « une manne tombée du ciel » comme dans le modèle de Solow. Le courant de la croissance endogène fondée sur l'innovation (dont Romer (1990) est l'article pionnier avec ceux de Grossman et Helpman (1991) et Aghion et Howitt (1992) que nous décrirons dans la prochaine section) a remis en avant le rôle de la politique économique puisque toute mesure qui permet d'augmenter l'effort de recherche et développement (une subvention à la R&D privée par exemple) augmente mécaniquement le taux de croissance de l'économie.

Le modèle de Romer (1990) souffre cependant d'un défaut majeur, mis en lumière par Jones (1995) : la présence de l'effet d'échelle. Le taux de croissance de l'économie est une fonction linéaire de la taille de l'économie (approximée par la taille de la population active L). Cette propriété découle de la présence des rendements croissants dans le secteur de la R&D et de la production finale : *A nombre de connaissances donné N, multiplier par deux le nombre de chercheurs L_N multiplie par deux le nombre de connaissances supplémentaires, ce qui multiplie par deux la production de bien final et donc le taux de croissance*.
On a donc g=f(L). Cela oblige a supposer que le taux de croissance de la population est nul ($\dot{L}/L = 0$) car dans le cas contraire ($\dot{L}/L = n$ où n est le taux de croissance de la population donc), il y a alors accélération du rythme de croissance. Or, dans la réalité, il y effectivement croissance de la population mais les taux de croissance sont relativement stables. Présenté autrement, l'effet d'échelle n'a aucune plausibilité empirique et il est un défaut majeur de ces modèles car les dépenses en R&D n'ont cessé d'augmenter depuis la fin la seconde guerre mondiale alors que le taux de croissance est resté stable, voire a décru légèrement.

Jones [1995] a donc intégré la croissance démographique dans le modèle de Romer [1990] et proposé une formulation de la fonction de production de connaissances dans laquelle la condition fil du rasoir ϕ=1 est levée. La fonction de production de connaissances est alors $\dot{N}(t) = \beta N(t)^{\phi} L_N(t)^{\lambda}$ avec ϕ<1. Le taux de croissance de l'économie devient $g = \frac{\lambda n}{1-\phi}$. Ce régime de croissance est qualifié de semi-endogène car si le processus d'innovation est bien endogénéisé, les ressorts de la croissance sont de nouveaux exogènes (λ, ϕ, n) comme dans le modèle de Solow. La conclusion de la théorie de la croissance semi-endogène est donc que les politiques de promotion de l'innovation sont sans effet à long terme sur le taux de croissance et donc inutiles.

7.5.4 Le modèle schumpétérien de croissance : Aghion et Howitt (1992)[137]

On considère une économie fermée qui fabrique un seul produit qui peut être utilisé à la fois pour la consommation et pour la production (comme input). Ce produit final est fabriqué selon la technologie de production suivante : $Y = L_Y^{1-\alpha} \int_0^1 A_i x_i^{\alpha} di$. L est le flux de travail employé dans le secteur de production de bien final. x_i correspond à la quantité de bien intermédiaire i utilisé pour fabriqué le produit final. A_i est un paramètre de productivité qui mesure la qualité de la dernière version du bien intermédiaire i. Il existe un continuum de biens intermédiaires ($i \in [0,1]$).

Les biens intermédiaires sont tous produits en employant le capital selon la fonction de production $x_i = \frac{K_i}{A_i}$. La présence du paramètre de productivité sectoriel au dénominateur indique que les générations successives d'un bien intermédiaire sont fabriquées par des technologies de plus en plus capitalistiqueq.

Une innovation technologique porte sur l'un des biens intermédiaires. Une innovation dans le secteur i va permettre une amélioration de ce bien intermédiaire et le remplacement du monopoleur en place par le nouvel innovateur[138] : c'est le processus de destruction créatrice mis à jour par Schumpeter.

Dans chaque secteur de bien intermédiaire, le monopole en place connaît la fonction de demande inverse de son bien par le secteur de bien final ($p_i = \alpha A_i x_i^{\alpha-1} L^{1-\alpha}$), cette dernière étant obtenu en égalisant la productivité marginale du bien intermédiaire i et son prix. La production d'une quantité x_i de bien intermédiaire engendre un coût : $C(x_i) = (r+\delta)K_i$ où (r+δ) est le coût du capital (taux d'intérêt plus taux de dépréciation). Le monopole en place choisit xi de façon à maximiser $\pi_i = \alpha A_i x_i^{\alpha} L^{1-\alpha} - (r+\delta) A_i x_i$. La solution x* de ce programme de maximisation est indépendante de i. A l'équilibre, toutes les entreprises de biens intermédiaires fournissent donc la même quantité. En définissant $K = \int_0^1 K_i di$, le stock total de capital, $A = \int_0^1 A_i di$, la productivité moyenne de l'économie, et k=K/(AL), le stock de capital par travailleur productif, on peut écrire la condition du premier ordre de ce programme de maximisation qui va constituer notre première condition d'équilibre (***équation de capital***)

$$\alpha^2 k^{\alpha-1} = (r+\delta)$$

Les innovations résultent des investissements en R&D. Les secteurs de R&D utilisent du bien final comme facteur de production. Une innovation intervenant dans le secteur i à la date t permet au paramètre de productivité A_i de ce secteur d'atteindre le niveau de la productivité de pointe dans l'économie : $A_{max} = \max_{i \in [0,1]} A_j$. On suppose que les innovations interviennent dans n'importe quel secteur de bien intermédiaire, suivant un processus de Poisson avec un taux d'occurrence: λn où λ est le paramètre qui mesure la productivité du secteur de la recherche et n la quantité de biens finals consacrée à la R&D, ajustée par A_{max} de façon à prendre en compte la difficulté croissante à innover à mesure que le niveau de productivité de pointe dans l'économie progresse (effet de *fishing out*).

La condition de libre entrée dans la recherche stipule que le coût marginal de la R&D (1-τ), où τ est le taux de subvention à la R&D, est égal à la valeur de l'innovation : $V = \frac{(1-\tau)}{\lambda}$

---

[137] **Aghion Phillipe et Howitt Peter (1992)**, « A Model of Growth through Creative Destruction », *Econometrica*, vol.60, pp.323-351.

[138] Dans le modèle d'Aghion et Howitt [1992], aucune innovation n'est le fait des monopoleurs en place. C'est « l'effet de remplacement à la Arrow ». L'accroissement de profit du fait de l'innovation est moindre pour l'entreprise en place que pour une nouvelle entreprise car la première jouit déjà d'une rente de monopole. La moindre récompense pour l'innovation réalisée par le monopole aboutit à ce que l'innovation soit toujours le fait d'un nouvel entrant. Le caractère trop irréaliste de cette hypothèse (*cf.* Microsoft ou Intel…) a amené les chercheurs à proposer des modélisations plus complexes, soit pour examiner le possibilité de duopole et examiner l'incidence de degré de concurrence sur l'incitation à l'innovation (**Aghion, Harris, Howitt, Vickers (2002)** "Competition and Innovation: An inverted-U Relationship", *Quaterly Journal of Economics*, Vol.120, pp.701-728), soit pour permettre de prendre en compte l'innovation par des entreprises déjà leader (**Segerstrom (2006)**, "Intel Economics", *International Economic Review*, forthcoming).

La valeur de l'innovation est la somme actualisée (actualisation prenant en compte le risque de création destructrice) des flux de profits réalisés grâce à cette innovation ce qui permet d'exprimer la seconde condition du modèle (***équation d'arbitrage***) :

$$\lambda\left(\frac{\pi(k)}{r+\lambda n}\right)=(1-\tau)$$

Les équations de capital et d'arbitrage définissent conjointement l'équilibre stationnaire de la R&D en fonction des paramètres de l'économie. Le niveau d'équilibre de la R&D sera d'autant plus important que les taux d'intérêt sont faibles ou que le taux de subvention à la R&D est important.
Pour passer du niveau de la R&D au taux de croissance de l'économie, il suffit de faire l'hypothèse d'une diffusion de connaissances d'un secteur à l'autre (rappelez-vous que la connaissance est un bien non-exclusif et non-rival) de sorte que la productivité de pointe va croître à un rythme proportionnel au flux des innovations dans l'économie.

$$g=\frac{\dot{A}_{\max}}{A_{\max}}=\mu\lambda n$$

Où m>0 mesure l'importance du phénomène de diffusion d'un secteur à l'autre. A l'équilibre le taux de croissance de la productivité moyenne de l'économie est égale au taux de croissance de la productivité de pointe car la distribution des ratios de productivité $A_i / A_{\max}$ est stationnaire.

Toute politique qui permet d'élever le niveau d'équilibre de la R&D permet donc d'augmenter le taux de croissance de la productivité de pointe, celui de la productivité moyenne et celui de l'économie.

7.5.5 Education, innovation et distance à la frontière

Aghion, Acemoglu et Zilibotti (2002) ont intégré la prise en compte d'institutions « appropriées » (au sens de Gerschenkron (1962)[139]) dans un modèle de croissance. Cette prise en compte des fondements institutionnels de la croissance a permis à ces auteurs de montrer que les différents types d'arrangements institutionnels ou de politiques structurelles (éducative par exemple…) affectaient la croissance différemment selon la distance du pays considéré à la frontière technologique.
Voici la présentation de ce modèle[140] :
Soit une économie composée de plusieurs pays. Dans chaque pays un bien final unique est produit en utilisant un continuum de biens intermédiaires selon la fonction de production suivante :

$$y_t=\int_0^1 A_t(i)^{1-\alpha}\,x_t(i)^{\alpha}\,di$$

Où $A_t(i)$ est le niveau de productivité du bien intermédiaire produit par le secteur i, et $x_t(i)$ le flux de bien intermédiaire i rentrant dans la production du bien final. Chaque bien intermédiaire est produit par un monopole local, qui utilise une unité de bien final pour produire une unité de bien intermédiaire. Le secteur de production de bien final est en situation de concurrence, de sorte que chaque producteur de bien intermédiaire vend son bien intermédiaire i au secteur du bien final à un prix égal à la productivité marginale de celui-ci. Le profit réalisé par chaque monopole de bien intermédiaire est alors :

$$\pi_t(i)=\delta A_t(i)$$

Où δ est une constante.

Soit $A_t \equiv \int_0^1 A_t(i)di$ la productivité moyenne dans le pays à la date t. Soit $A_t^{\max}$ la productivité à la frontière technologique mondiale à cette période (on va faire l'hypothèse que cette productivité mondiale de pointe croît au taux constant *g* d'une période à l'autre). Définissons $a_t=\frac{A_t}{A_t^{\max}}$ comme la proximité du pays à la frontière technologique mondiale à la date t. A chaque période, un producteur de bien intermédiaire peut améliorer sa productivité de deux façons : (i) soit en « imitant »

[139] Dans *Economics Backwardness in Historical Perspective*, Gershenkron A. (1962) montre que les économies « attardées » peuvent rattraper relativement rapidement les pays les plus avancés lorsqu'elles se dotent d'institutions « appropriées ».
[140] Je m'appuie sur la présentation qu'en font Aghion Ph. et Cohen E. dans leur rapport « Education et Croissance » au Conseil d'Analyse Economique.

la technologie à la frontière de la période précédente $A_{t-1}^{\max}$, (ii) soit en innovant, c'est-à-dire en améliorant la technologie locale existante A_{t-1}.

La productivité de la firme i en fin de période t peut donc s'exprimer sous la forme suivante :

$$A_t(i) = \eta A_{t-1}^{\max} + \gamma A_{t-1}$$

Où η et γ sont respectivement le poids de l'imitation et de l'innovation dans le progrès technique.

En intégrant sur l'ensemble des secteurs et en utilisant le fait que la productivité de pointe (la frontière technologique) croit au taux g, on obtient la relation suivante entre les niveaux de la proximité à la frontière aux date *t* et *t-1* : $a_t = \frac{(\eta + \gamma a_{t-1})}{1+g}$.

L'importance relative de l'innovation pour la croissance de la productivité augmente à mesure que le pays s'approche de la frontière technologique ou sous l'effet d'une révolution technologique (augmentation de γ).

Supposons que l'innovation repose sur le travail qualifié (éducation supérieure) tandis que les tâches d'imitation peuvent être accomplies par des individus moins qualifiés (éducation secondaire).

L'approche de Aghion, Acemoglu et Zilibotti (AAZ) prédit alors que plus le pays se rapproche de la frontière technologique et plus l'investissement dans l'éducation supérieure est rentable en termes de croissance et de productivité.

Pour présenter le modèle AAZ adapté à la politique éducative, nous conservons la même structure productive mais modifions légèrement l'équation décrivant la dynamique du progrès technique :

$$A_t = A_{t-1} + \lambda\left((1-f)^{\beta} s_{imi}^{\,1-\beta}\left(A_{t-1}^{\max} - A_{t-1}\right) + \gamma s_{ino} A_{t-1}\right)$$

Où $0<\beta<1$ mesure le poids relatif du travail non qualifié dans l'imitation et γ représente le poids de l'innovation dans le progrès technique. Le paramètre λ est un facteur d'échelle. *1-f* est la fraction de la population qui n'a pas reçu d'éducation supérieure (qui s'est arrêtée disons au secondaire), s_{imi} est la fraction de la population ayant reçu une formation supérieure qui est employée dans des tâches d'imitation et s_{ino} est la fraction de la population ayant reçu une éducation supérieure qui œuvre à l'innovation. Bien entendu, le plein emploi nécessite l'égalité suivante : $s_{imi} + s_{ino} = f$.

Le taux de croissance de la productivité s'écrit alors :

$$\boxed{g = \gamma f + \gamma\left(\left(\frac{(1-\beta)(1-a)}{a\gamma}\right)^{1/\beta}\left(\frac{\beta}{1-\beta}\right)\right)(1-f)}$$

L'impact marginal d'une augmentation de la fraction f d'individus ayant reçu une éducation supérieure sera positif si et seulement si : $a \geq \frac{\beta^{\beta}(1-\beta)^{1-\beta}}{\gamma + \beta^{\beta}(1-\beta)^{1-\beta}}$ (effet seuil). On peut noter que le seuil critique de a, au-delà duquel davantage d'éducation supérieure favorise la croissance, diminue lorsque g augmente. Lorsqu'une révolution technologique rend l'innovation plus efficace et plus importante (cas des TIC) par rapport à l'imitation, il est intéressant pour un plus grand nombre de pays d'investir dans l'enseignement supérieur. La contribution du travail qualifié à la croissance augmente à mesure qu'on se rapproche de la frontière technologique.

8 Annexes générales

8.1 Le cadre de comptabilité nationale

8.1.1 Le Produit Intérieur Brut

Nous allons présenter les grandes lignes du cadre de comptabilité nationale afin d'être à même de savoir comment est défini le concept de base de produit agrégé, c'est-à-dire la richesse produite au sein de la nation au cours d'une période donnée.
Il existe trois manières de mesurer cet indicateur synthétique : une optique basée sur la production, une autre basée sur la dépense et une dernière sur le revenu.

L'optique de la production

Dans le cadre comptable actuel, on retient pour évaluer la production tous les biens et tous les services socialement organisés (que ces services soient vendus sur un marché ou non, la condition est qu'ils soient obtenus à partir de facteurs de production s'échangeant sur un marché). Les services rendus par les administrations ou les services à la personne (ménages,…) sont ainsi pris en compte car il y a bien achat des services du travail sur le marché.
Le système comptable distingue la **production marchande** et la **production non-marchande**. Les biens et services marchands sont évalués au prix de base. Le prix de base est la recette effective du producteur par unité produite (montant reçu de l'acheteur pour chaque unité de produit moins les impôts sur les produits plus les subventions sur les produits). Les services non-marchands sont évalués à leurs coûts de production.

Les opérations de production consistent à fabriquer, à partir de travail, de matières premières et d'immobilisations diverses (machines, bâtiments) des produits plus élaborés.
Les **consommations intermédiaires** représentent les biens consommés au cours du processus courant de production. Il y a disparition complète de ces biens, soit par incorporation dans les produits plus élaborés (les matières premières transformées au cours du processus de production) soit par destruction (fuel utilisé pour faire tourner les machines).
On doit bien distinguer les consommations intermédiaires des biens qui constituent le **capital fixe** (bâtiments, machines). Ces derniers sont également utilisés lors du processus de production mais ne disparaissent pas au cours du processus courant de production (ils servent durant plusieurs périodes).
On définit la **valeur ajoutée brute** comme la différence entre la valeur de la production et la valeur des consommations intermédiaires.
La mesure la plus usuelle de la production agrégée est le **Produit Intérieur Brut** (en abrégé PIB). Il est défini comme la somme des valeurs ajoutées (plus les subventions sur les produits moins les impôts sur les produits) :

PIB = ΣVA + IMPÔTS s/ Produits – SUBVENTIONS s/ Produits

Le PIB mesure la production réalisée sur le territoire de l'économie considérée par les unités de production résidentes (un entreprise résidente est une entreprise installée sur le territoire national depuis plus d'un an).
Le **Revenu National Brut** (RNB, anciennement Produit National Brut, PNB) mesure la production réalisée par des facteurs de production fournis par des résidents (sur ou en dehors du territoire national) :

RNB = PIB
+ revenus de facteurs reçus du Reste Du Monde
- revenus de facteurs versés au Reste Du Monde

Les deux évaluations (RNB et PIB) ne diffèrent que faiblement dans les pays développés (moins de 5% de différence). On pourra donc les assimiler en première approximation. La différence peut en revanche être importante pour certains pays en développement (Exemples : Koweït (RNB>PIB de l'ordre de 35%) et Brésil (PIB>RNB de l'ordre de 14%). Le Koweït, exportateur de pétrole a placé les revenus de ses exportations de pétrole à l'étranger et reçoit donc des revenus du reste du monde, alors que le Brésil est un pays vers lequel se dirige beaucoup d'investissements directs à l'étranger, et

les multinationales rapatrient une partie des profits réalisés, d'où les versements importants de revenus au reste du monde).

L'optique de la dépense

Le PIB peut également être mesuré à partir des dépenses réalisées au cours d'une période dans l'économie considérée. On distingue traditionnellement les dépenses suivantes :

- La consommation finale des ménages de biens et services (C),
- L'investissement (ou Formation Brute de Capital Fixe) destiné à accroître le stock de capital de l'économie (I),
- La variation de stocks (biens autres que les biens de capital fixe détenus à un instant donné par les unités de production) (ΔS),
- Les dépenses publiques (Etat, Administrations publiques centrales et locales) qu'il s'agisse de dépenses de fonctionnement (salaires des fonctionnaires..) ou d'investissement (construction d'une école ou d'une route…) (G),
- Les exportations (X) nettes des importations (M)

L'égalité suivante représente cette optique de la dépense :

$$\mathbf{PIB = C + G + I + X - M + \Delta S}$$

Cette égalité mise sous la forme équivalente suivante est appelée **équilibre Ressources-Emplois** :

$$\mathbf{PIB + M = C + G + I + X + \Delta S}$$

Ressources Emplois

L'optique du revenu

De l'opération de production naît la valeur ajoutée qui engendre un montant de revenus strictement équivalent. Pour produire, il faut rémunérer les facteurs de production (travail, intérêts sur les emprunts, dividendes versés aux détenteurs des capitaux, versement de loyer).
Il convient néanmoins de constater que tous ces revenus ne font pas l'objet de distribution effective (une partie des profits réalisés est conservée pour être ultérieurement investie).

Approche production	Approche demande	Approche revenus
Valeur ajoutée au prix de base 1395,3	Dépenses de consommation finale 1243,2	Rémunération des salariés 821,2
+ Impôts sur les produits 180,8	+ FBC (FBCF + DS) 294,8	+ Excédent Brut d'Exploitation et revenu mixte brut 294,8
- subventions sur produits -18,8	+ Exportations de b&s 401,9	+ Impôts sur la production et les importations 242,2
	- Importations de b&s -382,6	- Subventions -30,4
Total 1557,3	Total 1557,3	Total 1557,3

PIB 2003 à prix courants (en milliards d'euros)

8.2 Le programme du consommateur

8.2.1 Le programme statique : détermination des demandes de biens

Le programme statique du consommateur (statique car le temps n'est pas pris en compte : on se focalise sur une période donnée) consiste à déterminer les demandes de biens de ce consommateur en fonction de ses préférences et des données exogènes (prix des biens, budget du consommateur).
Soit un ménage représentatif qui dispose d'un revenu R pour consommer deux biens 1 et 2 dont les prix sont respectivement p_1 et p_2. Ce consommateur a des préférences prises en compte à travers une fonction d'utilité $U(C_1,C_2)$ où C_1 et C_2 sont les quantités de biens 1 et 2 respectivement consommés par l'agent. La dérivée première de la fonction d'utilité par rapport à l'un des ces arguments (appelée utilité marginale) est bien entendu positive. La dérivée seconde est négative (effet de satiété) : l'utilité marginale est décroissante. Un accroissement de la quantité consommée d'un des deux biens accroît toujours l'utilité de l'agent représentatif, mais à mesure qu'il a déjà consommé une quantité importante d'un bien, l'apport d'une unité supplémentaire de ce bien à son utilité est de plus en plus faible.

Soit $x = (C_1,C_2)$ le vecteur de consommation de chacun des biens. On appelle ce vecteur un « panier de consommation ». On suppose que le consommateur est capable de classer deux paniers de biens x_1 et x_2 en fonction de leur attrait respectif, c'est-à-dire des services qu'ils vont lui rendre et de la satisfaction qu'ils vont lui procurer. Comment le consommateur est-il capable d'opérer cette hiérarchisation dans ses préférences ? Des **hypothèses sur les préférences** du consommateur sont nécessaires. Voici le système de préférences du consommateur (avec $\succ$ signifiant « préfère strictement », ~ « est indifférent » et $\underline{\succ}$ « préfère faiblement » (peut être indifférent)) :

$x_1 \succ x_2 \Leftrightarrow$ le consommateur préfère strictement le panier 1 au panier 2.
$x_1 \sim x_2 \Leftrightarrow$ le consommateur est indifférent entre le panier 1 et le panier 2.
$x_1 \underline{\succ} x_2 \Leftrightarrow$ le consommateur préfère faiblement le panier 1 au panier 2.

Les préférences du consommateur sont rationnelles (c'est-à-dire cohérentes) ce qui conduit à poser trois axiomes : (i) la relation de préférence, $\underline{\succ}$, est une relation complète (on peut comparer toute paire quelconque de paniers de biens), (ii) la relation est réflexive ($x_1 \underline{\succ} x_1$), (iii) la relation est transitive : $x_1 \underline{\succ} x_2$ et $x_2 \underline{\succ} x_3 \Rightarrow x_1 \underline{\succ} x_3$.
Par rapport à un panier de consommation donné x*, le consommateur est capable de définir l'ensemble des paniers faiblement préféré (zone hachurée sur la figure ci-dessous). On appelle **courbe d'indifférence** (lorsqu'on se limite à deux biens comme c'est le cas ici), l'ensemble des paniers qui sont strictement équivalents pour notre consommateur au panier x*. La courbe d'indifférence passant par le panier x* est la frontière de l'ensemble faiblement préféré à x*.

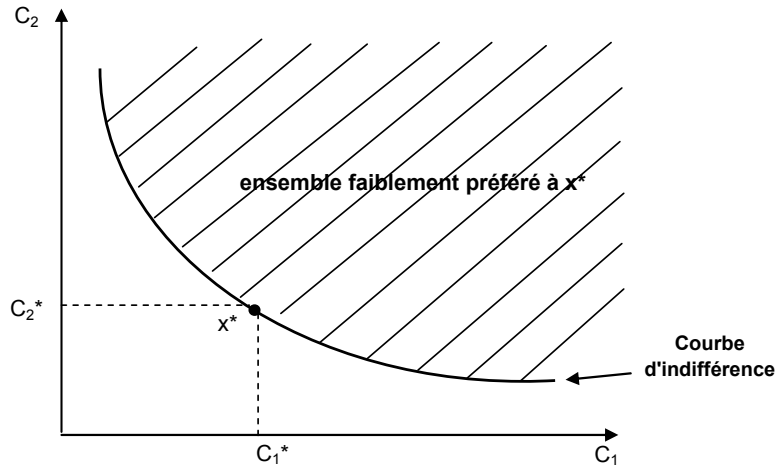


Figure 76 : Courbe d'indifférence

Il existe une infinité de courbes d'indifférence puisque tout point du plan représente un panier de deux biens considérés par lequel passe une courbe d'indifférence. Deux courbes d'indifférence distinctes ne peuvent pas se couper (du fait de l'axiome de transitivité).

On appelle donc **courbe d'indifférence** la courbe dans le plan (C_1,C_2) qui décrit l'ensemble des couples de consommation aboutissant au même niveau d'utilité $\overline{U}$. Le long d'une courbe d'indifférence la variation d'utilité consécutive à des variations des quantités consommées des deux biens est nulle puisque le niveau d'utilité est constant : $\Delta U = \Delta C_1 \frac{dU}{dC_1} + \Delta C_2 \frac{dU}{dC_2} = 0$. On peut alors exprimer l'égalité suivante :

$$\frac{\frac{dU}{dC_1}}{\frac{dU}{dC_2}} = -\frac{\Delta C_2}{\Delta C_1}.$$

On appelle **Taux Marginal de Substitution** (TMS) le rapport des utilités marginales. Ce n'est rien d'autre que la pente (en valeur absolue) de la courbe d'indifférence en un point donné.

La **contrainte budgétaire** de cet agent représentatif est la suivante : $R \geq p_1C_1 + p_2C_2$

L'objectif de cet agent représentatif, de ce consommateur, est de maximiser son utilité sous la contrainte budgétaire. Pour cela il a la possibilité de choisir les quantités consommées des biens 1 et 2 qui vont rendre son niveau d'utilité le plus élevé possible. Il est cependant bien entendu contraint par son revenu (R).

$$\underset{C_1,C_2}{Max}\, U(C_1,C_2)$$

$$\text{s.c. } R \geq p_1C_1 + p_2C_2$$

On peut former le Lagrangien suivant pour résoudre ce programme d'optimisation :

$L = U(C_1,C_2) + \lambda\left[R - p_1C_1 - p_2C_2\right]$ où λ est le multiplicateur de Lagrange associé à la contrainte.

Les deux conditions du premier ordre sont les suivantes :

$$\frac{dU}{dC_1} = -\lambda p_1$$
$$\frac{dU}{dC_2} = -\lambda p_2$$

auxquelles on doit ajouter la saturation de la contrainte : $R = p_1C_1 + p_2C_2$.

En égalisant les deux CPO, on obtient l'égalité suivante : $\frac{\frac{dU}{dC_1}}{\frac{dU}{dC_2}} = \frac{p_1}{p_{21}}$, c'est-à-dire l'égalité entre le taux marginal de substitution et le rapport des prix. En utilisant cette égalité et la contrainte saturée on peut obtenir les demandes de biens 1 et 2 ; $C_1(p1,p_2,R)$ et $C_2(p1,p_2,R)$ en fonctions des exogènes (prix et revenu).

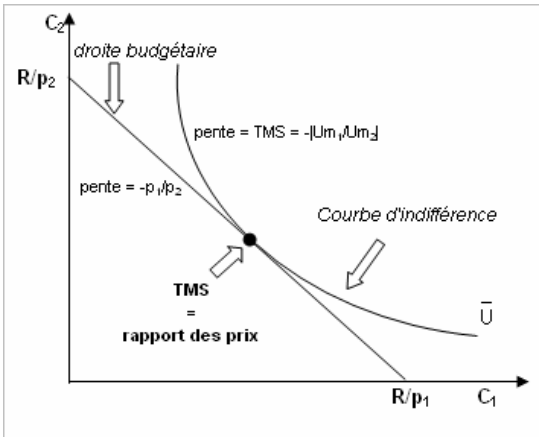


Figure 77 : Programme du consommateur

Graphiquement, le programme revient à déterminer le point de tangence entre la droite qui traduit la contrainte budgétaire et la tangente à la courbe d'indifférence (**rapport des prix=TMS**).

8.2.2 Effet Revenu / Effet Substitution

Considérons un consommateur représentatif dont les préférences sont décrites par une fonction d'utilité $U(C_1,C_2)$, dont le revenu est R et qui fait face au système de prix p_1 et p_2. L'effet de l'évolution du prix de l'un des deux biens sur les demandes des deux biens peut être appréhendé à travers deux effets : l'effet de revenu et l'effet de substitution. Prenons l'exemple d'une augmentation du prix du bien 1 ($p_2 \uparrow$).

L'effet de revenu mesure la baisse de consommation des deux biens provenant de la perte de pouvoir d'achat du consommateur du fait d'un système de prix (p_1,p_2) moins favorable ($C_1 \downarrow$ et $C_2 \downarrow$).

L'effet de substitution mesure le report de consommation du bien 1 vers le bien 2 du fait de la baisse du prix relatif de ce bien ($C_1 \downarrow$ et $C_2 \uparrow$).

L'effet de la hausse du prix du bien 1 sur la consommation de ce bien est donc sans ambiguïté. En revanche, la consommation du bien 2 augmente à la suite de l'augmentation du prix du bien 1 si l'effet de substitution l'emporte sur l'effet revenu, et diminue dans le cas contraire.

Calculer l'effet revenu revient à répondre à la question suivante : De quel revenu le consommateur devrait-il disposer pour conserver son niveau d'utilité initial avec le nouveau système de prix (c'est-à-dire demeurer sur la courbe d'indifférence U).

Calculer l'effet substitution revient à résoudre le problème suivant sous l'hypothèse que le consommateur dispose du revenu fictif précédemment décrit : Quelles sont alors ses consommations des deux biens et son niveau d'utilité avec le nouveau système de prix ?

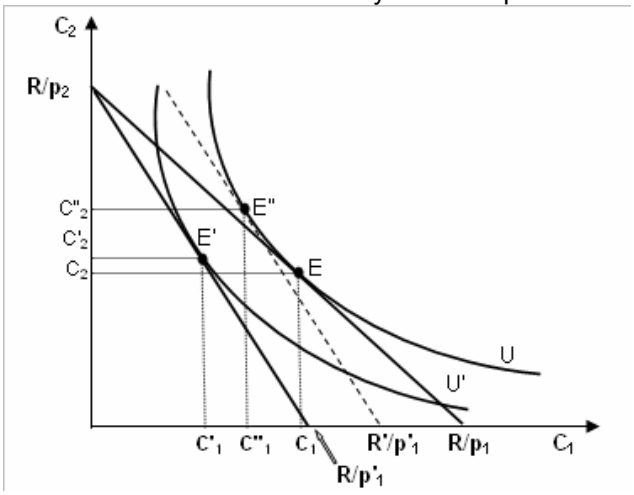


Figure 78 : Effet revenu et effet substitution

La figure ci-dessus montre l'équilibre de la situation initiale ($E(C_1,C_2)$ lorsque le prix était (p_1,p_2)) et finale ($E'(C'_1,C'_2)$ lorsque le système de prix est (p'_1,p_2) suite à l'augmentation du bien 1). Elle présente également la situation fictive E'' permettant de scinder l'effet total (passage de E à E') en deux effet : l'effet revenu (passage de E'' à E') et l'effet de substitution (passage de E à E'').

Pour déterminer le point E'' correspondant à la situation fictive, il, suffit de trouver une droite budgétaire parallèle à la nouvelle droite budgétaire (c'est-à-dire prenant en compte le nouveau système de prix) et tangente à l'ancienne courbe d'indifférence (représentant le niveau d'utilité initial).

Pour déterminer la situation finale E' il faut trouver le point de tangence entre la nouvelle droite budgétaire et la courbe d'indifférence la plus intéressante pour le consommateur.

Dans l'exemple retenu, l'effet substitution l'emporte sur l'effet revenu et le consommateur augmente sa consommation de bien 2 suite à la hausse du prix du bien 1.

8.2.3 Le programme dynamique du consommateur : Modèle de Ramsey.

Le modèle de Ramsey est un modèle de croissance reposant sur la description du choix du consommateur en matière d'épargne en fonction de ses préférences. Le ménage dont la durée de vie est infinie maximise son utilité intertemporelle U dont la forme est

$$U = \int_0^{\infty} e^{-\rho t} u(c(t))dt$$

où ρ est le taux de préférence pour le présent et u(.) la fonction d'utilité instantanée qui vérifie les conditions suivantes : u'(.)>0 et u''(.)<0. La consommation par tête est notée c. L'utilité instantanée satisfait également les conditions d'Inada, à savoir : $\lim_{c\to 0} u'(c) \to \infty$ et $\lim_{c\to \infty} u'(c) \to 0$.

Donc *u(c)* est croissante et concave. Le postulat de concavité traduit le désir d'étaler la consommation dans le temps : les consommateurs préfèrent un profil relativement uniforme à celui où la consommation serait très basse à certaines périodes et très élevée à d'autres. Enfin, on considère que ρ est positif : les unités de consommation courantes sont préférées aux unités de consommation futures.

Les ménages détiennent des actifs sous forme de droits de propriété sur le capital ou bien sous forme de prêts qu'ils ont accordés. Le capital et les prêts sont considérés comme parfaitement substituables en tant que réserve de valeur. Ils rapportent le même taux de rendement réel r(t). Les prêts négatifs sont des dettes.

Soit a(t) les actifs d'un ménage mesurés en termes d'unité de consommation. La contrainte de budget du ménage s'exprime, en termes de flux, sous la forme suivante :

$$\dot{a}(t) = w(t) = r(t)a(t) - c(t)$$

Les actifs du ménage augmentent avec le revenu (w(t)+r(t)a(t)) et baissent avec la consommation.

Le problème d'optimisation du ménage est de maximiser U sous la contrainte de budget, étant donné le stock de ses actifs initiaux a(0). On peut former le Hamiltonien suivant pour résoudre ce programme d'optimisation dynamqie :

$$H = u(c(t))e^{-\rho t} + \mu(t)(w(t) + r(t)a(t) - c(t))$$

La variable μ représente la valeur actualisée du prix implicite du revenu. C'est la valeur, mesurée en unités d'utilité disponible au temps 0, accordée à une augmentation de revenu reçue au temps t.

Les conditions du premier ordre pour un maximum de U sont :

$$\frac{dH}{dc(t)} = 0 \Rightarrow \mu(t) = u'(c(t))e^{-\rho t}$$

$$\dot{\mu}(t) = -\frac{dH}{da(t)} \Rightarrow \dot{\mu}(t) = -r(t)\mu(t)$$

En dérivant μ(t) par rapport au temps (dans la CPO 1) , $\frac{\dot{\mu}(t)}{\mu(t)} = \left[\frac{u''(c(t)).c(t)}{u'(c(t))}\right]\frac{\dot{c}(t)}{c(t)} - \rho$,et en remplaçant le résultat dans la CPO 2, on obtient le critère de base du choix de la consommation au cours du temps (***condition d'Euler***):

$$r(t) = \rho - \left[\frac{u''(c(t)).c(t)}{u'(c(t))}\right]\frac{\dot{c}(t)}{c(t)}$$

Cette équation indique que les ménages choisissent leur consommation de façon à égaliser le taux de rendement r(t) de leur épargne au taux de préférence pour le présent ρ augmenté du taux de diminution de l'utilité marginale de la consommation u' due à l'augmentation de la consommation par tête.

Le membre de droite de l'égalité précédente peut être considéré comme le taux de rendement de la consommation. Les agents souhaitant optimiser leur consommation doivent égaliser les deux taux de rendement de façon à être indifférents, à la marge, entre consommer et épargner.

Les agents préfèrent la consommation présente à la consommation future pour deux raisons :

- Parce que les ménages déprécient l'utilité future au taux ρ.
- Parce que si $\dot{c}/c > 0$ cela veut dire que *c* est plus faible aujourd'hui que demain. Or, comme les agents préfèrent une consommation régulière dans le temps (car u''(c)<0), ils vont niveler le flux en transférant une partie de leur consommation du futur vers le présent (second terme du m.d.d.).

Les ménages choisissent un profil de consommation parfaitement uniforme ($\dot{c}/c = 0$) si ρ=r(t). Les ménages acceptent de dévier de ce profil de consommation uniforme si ils reçoivent en compensation une rémunération r(t) de leur épargne suffisamment supérieure à ρ ($\dot{c}/c > 0$ si $r(t) > \rho$). Le montant de cette compensation est indiqué par le terme $\left[-\frac{u''(c(t)).c(t)}{u'(c(t))}\right]\frac{\dot{c}(t)}{c(t)}$ avec, entre crochets, l'élasticité de l'utilité marginale par rapport à la consommation. Cette élasticité mesure la concavité de la fonction d'utilité instantanée u(c) et détermine le niveau duquel r(t) doit dépasser ρ. L'écart est d'autant plus élevé que l'élasticité est forte.

Exemple : En spécifiant la forme fonctionnelle suivante pour l'utilité instantanée (cette fonction est appelée fonction d'utilité à *élasticité de substitution inter-temporelle constant*e (CIES)) : $u(c(t)) = \frac{c(t)^{1-\theta} - 1}{1-\theta}$ où θ>0, de sorte que l'utilité marginale est égale est constante à θ, la condition d'optimalité se simplifie de la façon suivante : $\frac{\dot{c}(t)}{c(t)} = \frac{1}{\theta}(r(t) - \rho)$. Plus θ est élevée, plus la diminution proportionnelle de u'(c) est rapide en réaction à l'augmentation de la consommation et de ce fait, moins les ménages sont enclins à consentir des déviations par rapport à un profil uniforme dans le temps.

8.3 Le programme du producteur

8.3.1 Programme statique : détermination des demandes de facteurs

Le programme statique du producteur consiste à déterminer les demandes de facteurs de production (capital, travail) en fonction de la technologie dont il dispose et des prix exogènes de ces facteurs (coût du capital, taux de salaire).

Il existe deux types de facteurs de production : les facteurs fixes et les facteurs variables. A court terme, seuls les facteurs variables sont ajustables (embauche de personnels) tandis qu'à moyen terme on peut également ajuster les facteurs fixes (construction d'une nouvelle usine).
Soit r le coût du capital et w le taux de salaire.
Les économistes utilisent une **fonction de production** pour décrire le résultat en terme de production de l'utilisation de quantités données de facteurs de production. Il existe plusieurs formes fonctionnelles possibles pour les fonctions de production (Cobb-Douglas, CES…), les différences résidant dans le degré de substitution possible entre les facteurs de production.
La fonction de production **Cobb-Douglas** est de la forme suivante : $Q = F(K,L) = K^{\alpha} L^{1-\alpha}$ où Q est le niveau de production (l'*output*), K le niveau du stock de capital (exprimé en quantité physique) et L la quantité de main d'œuvre (nombre de travailleurs x heures travaillées). Le paramètre α (compris entre 0 et 1) est la part de la rémunération du capital dans la valeur ajoutée (α=rK /pQ).
La fonction de production **C.E.S.** (pour *Constant Elasticity of Substitution*) est de la forme suivante : $Q = \left[\beta_K K^{\rho} + \beta_L L^{\rho}\right]^{\frac{1}{\rho}}$ où ρ est l'élasticité de substitution entre capital et travail.
L'entreprise représentative va déterminer ses demandes de facteur à travers le programme de maximisation du profit. Le programme est le suivant : $\underset{K,L}{Max}\, \pi(K,L) = p.\mathrm{F(K,L)} - \mathrm{wL} - \mathrm{rK}$

Les conditions du premier ordre sont les suivantes :

$$\frac{d\pi}{dL} = 0 \Leftrightarrow \frac{dF(K,L)}{dL} = \frac{w}{p}$$

$$\frac{d\pi}{dK} = 0 \Leftrightarrow \frac{dF(K,L)}{dK} = \frac{r}{p}$$

La productivité marginale du travail égalise le salaire réel. La productivité marginale du capital égalise le coût réel du capital. On détermine aisément les demandes de facteurs à partir des deux égalités précédentes en fonction des exogènes et des prix des facteurs. On peut illustrer ceci avec la forme fonctionnelle CES où le programme aboutit aux demandes suivantes :

$$L^* = \frac{Q}{\frac{w}{p} + \frac{r}{p}\left(\frac{w}{r}\right)^{\frac{1}{1-\rho}} \left(\frac{\beta_K}{\beta_L}\right)^{\frac{1}{1-\rho}}} \qquad K^* = \frac{Q}{\frac{r}{p} + \frac{w}{p}\left(\frac{r}{w}\right)^{\frac{1}{1-\rho}} \left(\frac{\beta_L}{\beta_K}\right)^{\frac{1}{1-\rho}}}$$

On peut ensuite réintroduire dans la fonction de production ces demandes pour déterminer le niveau de production en fonction des exogènes.

8.4 La boîte d'Edgeworth

La représentation de la boîte d'Edgeworth permet de sélectionner, parmi les répartitions possibles de deux biens entre deux individus, celles qui sont optimales. L'optimalité se définit au sens de Pareto : une répartition est dite optimale au sens de Pareto si, à partir de celle-ci, il n'est pas possible d'augmenter l'utilité d'un agent sans diminuer celle d'au moins un autre.
On introduit les courbes d'indifférences correspondant à divers niveaux d'utilité des deux consommateurs dans le même rectangle (celles de l'individu A en gras, celles de l'individu B en trait fin). Par un point I, correspondant aux dotations initiales des individus en biens 1 et 2, passe nécessairement une courbe d'indifférence et une seule dans chaque réseau de courbes. Ce point n'est cependant pas I car ce n'est pas une situation optimale. Il est en effet possible d'atteindre un niveau d'utilité plus grand pour les deux individus avec toutes les répartitions représentées par des points situés dans la zone grisée.
Lorsque les deux courbes d'indifférence sont tangentes, les individus ne peuvent plus améliorer leur situation par des échanges. A l'optimum de Pareto, la tangente à une courbe d'indifférence d'un des deux individus est également tangente à la courbe d'indifférence de l'autre individu, de sorte que le Taux Marginal de Substitution (TMS) entre les deux biens est le même pour les deux consommateurs.
L'optimum de Pareto n'est pas unique. Les différents optima de Pareto peuvent être reliés par une courbe (cf. figure) contenue dans la boîte d'Edgeworth et passant pas OA et OB (la courbe des contrats que nous allons définir ci-dessous).

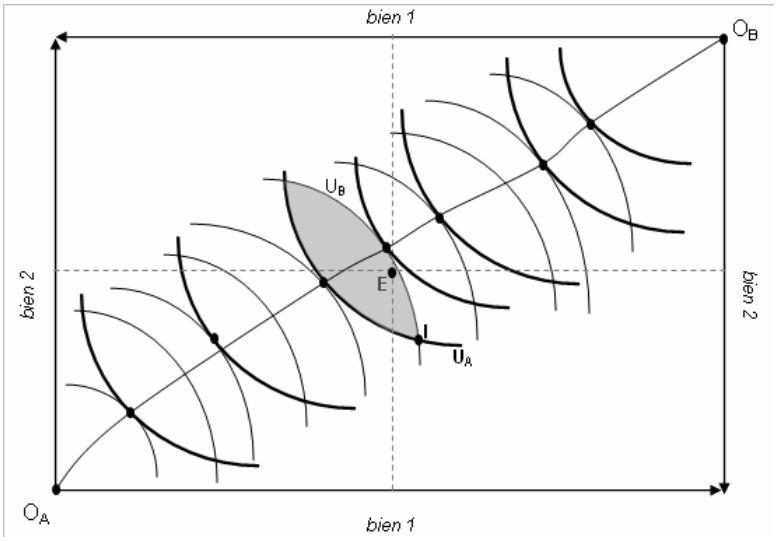


Figure 79 : Diagramme d'Edgeworth

On note qu'un optimum de Pareto n'implique pas nécessairement que la répartition soit équitable (c'est-à-dire socialement égalitaire). Inversement, une situation équitable (juste socialement) n'est pas forcément optimale (cf. le point E sur la figure).
Parmi toutes les répartitions représentées dans la boîte d'Edgeworth, voyons quelles sont celles qui peuvent être considérées comme stables. Repartons de la dotation initiale des biens 1 et 2 représentée par le point I. Le point I n'est pas un optimum de Pareto (nous l'avons déjà vu) : les consommateurs ont intérêt à l'échange afin d'améliorer leur satisfaction simultanément. I est donc instable et les seules répartitions stables sont les points de la courbe de contrats compris dans la zone grisée (cette partie de la courbe des contrats est appelée le cœur ou noyau de l'économie).
Une répartition des ressources entre les deux individus A et B est dite stable, si elle est, pour chacun d'eux, au moins aussi favorable que leur dotation initiale, et s'il n'est pas possible que les agents accroissent simultanément leur satisfaction en échangeant à partir de cette répartition.

8.5 Brève introduction à la théorie des jeux

La théorie des jeux est un outil d'analyse, permettant de prendre en compte les interactions stratégiques entre les agents économiques, qui s'est développé à partir des travaux de Von Neumann et Morgenstern (1944)[141]. Les individus rationnels interagissent dans des situations conflictuelles puisque le comportement d'un individu peut affecter directement la satisfaction d'autres individus. Il est parfois indispensable pour formaliser certains problèmes économiques d'intégrer ces interactions stratégiques. Des individus dont les décisions entraînent des conséquences pour les autres peuvent faire leur choix économique de deux manières : soit (i) en coopérant (décision d'un commun accord, en fonction d'un objectif commun), soit (ii) en se comportant de manière individualiste.
Ces deux types de comportement débouchent en théorie des jeux sur deux concepts d'équilibre : (i) l'**équilibre coopératif** (correspondant à l'optimum social), et (ii) l'**équilibre non-coopératif** (équilibre de Nash).
Les éléments de la théorie des jeux ont permis d'étudier l'interaction des prises de décisions des entreprises. Un jeu stratégique représente une situation dans laquelle la satisfaction d'un joueur est affectée par les actions qu'il peut entreprendre mais aussi par les actions des autres.

8.5.1 Représentation d'un jeu

La théorie des jeux concerne l'étude des situations dans lesquelles les individus interagissent dans un environnement d'interdépendance stratégique (c'est-à-dire que les gains d'un joueur dépendent aussi des actions choisies par les autres agents). Un jeu est une représentation formelle de cette situation. Deux modèles de représentation des jeux existent : la représentation stratégique et la représentation extensive. Nous étudions d'abord les hypothèses puis décrivons les deux types de représentation.

Hypothèses
L'analyse économique des comportements par l'intermédiaire d'un jeu requiert de spécifier :

- *l'ensemble des joueurs*, noté $I=\{1, ... , N\}$ qui sont les décideurs.
- *les stratégies possibles*, qui décrivent de manière exhaustive le comportement des joueurs et leurs choix possibles, face à différentes contingences. Chaque joueur i choisit une stratégie a^i parmi un ensemble de stratégies A^i. Par conséquent, $A^i=\{a_1^i, a_2^i , ..., a_k^i\}$ si k^i stratégies sont disponibles pour le joueur *i.*
- *les règles du jeu*, qui vont contraindre le comportement des joueurs sur la base des actions entreprises. Ces règles spécifient à tout moment les actions que les joueurs peuvent décider d'entreprendre.
- *les paiements*, qui sont les résultats du jeu. Pour chaque joueur *i*, une fonction de gain $\prod^i$ donne la valeur du paiement pour chaque résultat du jeu. Le jeu peut être à somme nulle (ce qui est perdu par l'un et gagné par l'autre) ou à somme non nulle (les joueurs peuvent gagner ou perdre, l'un peut perdre plus que ce que l'autre gagne ou inversement).
- *l'information*, que possèdent les joueurs. Il est important de distinguer l'information que possède un joueur sur la structure même du jeu (information complète ou incomplète) et l'information qu'il possède sur la décision prise par les autres joueurs (information parfaite ou imparfaite). Ainsi chaque participant qui a une information complète connaît son ensemble de choix, l'ensemble des choix des autres joueurs et toute la gamme des issues possibles du jeu et des gain associés. Dans le cas contraire, le jeu est à information incomplète. Par ailleurs, chaque joueur peut connaître le choix des autres joueurs au moment où il prend sa décision : il a une information parfaite. Lorsque le choix des autres joueurs est ignoré 'en partie ou en totalité), l'information est imparfaite.

La représentation extensive (sous la forme d'un arbre de décision)
La forme extensive est représentée sous la forme d'un arbre de décision, qui clarifie la séquence des actions des joueurs. Cette forme développée décrit de façon séquentielle toutes les stratégies des joueurs. Les joueurs doivent déterminer à chaque nœud de décision (nœud de l'arbre) une action qui est décrite par une flèche (branche de l'arbre), compte tenu de l'information dont il dispose.

Exemple : Sony contre Philips et le choix de standardisation.

[141] J. Von Neuman et O. Morgenstern (1944), Theory of Games and Economic Behavior, *Princeton University Press.*

Supposons que seulement deux entreprises fabriquent des walkmans numériques : Sony et Philips. Au préalable, ces deux entreprises se doivent de rendre compatible leur système d'exploitation en adoptant ou non la même cassette magnétique (mini-disc ou DCC). La compatibilité bénéficierait aux deux entreprises car elle permettrait à Sonny de vendre ses cassettes aux possesseurs d'un lecteur Philips et inversement. Considérons, par exemple, que Sony possède une avance technologique sur Philips dans la conception des lecteurs ce qui lui permet d'annoncer le premier le type de cassette retenu. Les profits de chaque firme dépendent des choix décidés par chaque acteur du jeu.

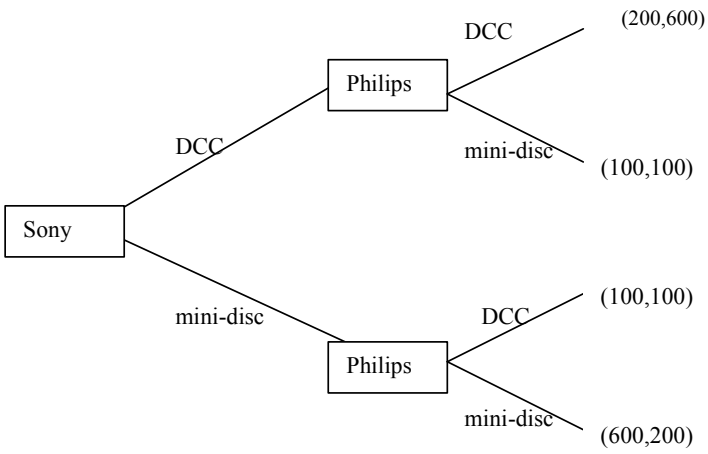


Figure 80 : Représentation d'un jeu sous forme d'un arbre de décision

La représentation stratégique (matricielle)

La forme stratégique offre une version condensée du jeu, représentée par une matrice. C'est une version simultanée, correspondant au cas où les joueurs déterminent leurs actions au même moment ou lorsque les joueurs ne peuvent pas observer les coups de leurs adversaires.

Exemple : Sony contre Philips et le choix de standardisation.

Le jeu entre Sony et Philips sur le choix de standardisation du produit peut être représenté sous forme matricielle suivante :

		Philips	
		mini-disc	DCC
Sony	mini-disc	(600,200)	(100,100)
	DCC	(100, 100)	(200,600)

Figure 81 : Représentation d'un jeu sous forme matricielle

Si Sony adopte le mini-Disc, sachant que Philips a aussi adopté le mini-disc, le gain de Sony est 600, celui de Philips est de 200 Le premier chiffre correspond au gain du joueur qui est en ligne, le deuxième est celui du joueur en colonne.

$\prod^{sony}$*(mini-disc, DCC)=600* et $\prod^{Philips}$*(mini-disc, DCC)=100.*

Détermination des équilibres

L'équilibre est le résultat de l'application d'une axiomatique qui détermine les stratégies pouvant être choisies par les joueurs. La solution d'un jeu est l'ensemble des stratégies choisies par chaque joueur.

L'équilibre en stratégie dominante

Une stratégie particulière ($a^{i*} \in A^i$) est une stratégie dominante du joueur *i* si, quelles que soient les stratégies choisies par les autres joueurs, a^{i*} maximise le gain du joueur *i*.

L'équilibre de Nash

Le choix optimal d'un joueur va-t-il dépendre de la stratégie adoptée par l'autre joueur ? Parfois, il n'y aura pas de stratégie dominante ou de stratégie dominée. La résolution du jeu peut alors nécessiter

de recourir à l'équilibre de Nash[142], correspondant au choix de stratégies tel que tous les joueurs choisissent la meilleure réponse au comportement des autres joueurs. A l'équilibre de Nash, aucun joueur n'a intérêt à changer unilatéralement sa stratégie. Tout équilibre en stratégies dominantes est aussi un équilibre de Nash mais chaque équilibre de Nash n'est pas nécessairement un équilibre en stratégies dominantes.

Nous pouvons illustrer cette présentation par deux exemples fameux de théorie des jeux : le « dilemme du prisonnier » et la « guerre des sexes » afin de percevoir la méthode, les conséquences de ses résultats et ses limites.

8.5.2 Le « dilemme du prisonnier »

Le dilemme du prisonnier est une parabole habituellement utilisée pour présenter l'équilibre non-coopératif et l'équilibre coopératif en théorie des jeux. L'histoire est la suivante : deux hommes soupçonnés d'un larcin sont arrêtés et interrogés par la police. Cette parabole montre que, bien choisies, les menaces de peine et les propositions de remises de peine faites à chacun des deux séparément, par les policiers qui les interrogent amènent chacun des deux accusés à avouer ce vol, alors que cela aurait été leur intérêt de le nier. Nous présenterons cet exemple de manière plus triviale avec deux élèves soupçonnés par leur professeur d'avoir triché à un examen…

Soit deux élèves A et B, ayant obtenu tous les deux la note de 14 à un examen, mais soupçonnés d'avoir triché par leur professeur. Les deux étudiants sont interrogés séparément par celui-ci, sans avoir eu la possibilité de se parler auparavant. Aucun des deux ne sait donc ce que l'autre va dire au professeur (on verra que cette hypothèse n'est en fait pas cruciale…). Le Professeur annonce à chacun le marché suivant :

1. « *si l'autre élève avoue et que vous niez, vous aurez la note de 0* »
2. « *si l'autre élève avoue et que vous avouez également, vous aurez 5 points en moins sur votre note*»
3. « *si vous niez tous les deux, vous aurez 2 points en moins sur votre note* »
4. « *si l'autre suspect nie et que vous avouez, vous conservez votre note* »

Le point 1. punit celui qui continue de nier alors que les preuves du forfait ont été obtenues par l'aveu de l'autre. Le point 2. cherche à rendre l'aveu profitable par les deux élèves tricheurs puisque le professeur fait preuve d'une relative indulgence en cas d'aveu (« faute avouée, à moitié pardonnée »…). Le point 3. décrit qui se passera si le professeur n'arrive pas à obtenir un aveu d'au moins un des étudiants. Comme le Professeur n'a que des soupçons et pas de preuves tangibles de la tricherie, il ne sanctionnera que très légèrement les deux élèves si ceux-ci nient farouchement (ses soupçons sont néanmoins suffisamment fondés pour qu'il leur enlève quelques points…). Le point 4. est une incitation à la dénonciation puisqu'il instaure une règle d'indulgence pour l'élève qui permet la mise à jour de la tricherie alors que l'autre tente de la dissimuler (*cf.* la règle d'indulgence instaurée par les autorités de régulation de la concurrence pour les entreprises qui permettent la dénonciation de pratiques anti-concurrentielles).

A l'aide de la théorie des jeux on va montrer que les deux étudiants ont intérêt à avouer, c'est-à-dire qu'ils choisissent une situation non optimale d'un point de vue collectif.
Pour cela on va mettre les résultats du marché proposé par le professeur aux deux élèves sous la forme d'une **matrice de gains**. Celle-ci présente la note de chacun des élèves en fonction de sa décision (avouer ou nier la tricherie) et de celle de son comparse.
On peut à partir de cette matrice de gain, déduire la décision de chacun des élèves prise de manière rationnelle. Mettons nous à la place de l'élève A. Si il pense que B va avouer, il préfère avoir 9 que 0, donc il avoue. Si il pense que B va nier, comme il préfère la note de 14 à celle de 12, il a également intérêt à avouer. L'hypothèse que l'individu est rationnel et maximise son utilité (ici sa note à l'examen à l'issue de la décision du professeur), aboutit donc à que l'élève A avoue quelle que soit la décision de B. Bien entendu, B fait le même raisonnement en conjecturant l'une ou l'autre décision de A. Finalement, les deux élèves avouent la tricherie et se retrouvent chacun avec une note de 9. L'équilibre que l'on vient de déterminer, lorsque chaque élève maximise son utilité (sa note) de manière individuelle, est l'**équilibre non-coopératif** (9,9). Bien entendu, un équilibre plus profitable pour les élèves apparaît dans la matrice de gain : celui où les deux élèves s'entêtent à nier et où le professeur n'enlève que quelques points. C'est l'équilibre coopératif où les deux élèves adoptent un comportement d'entre-aide ou de coopération (disons qu'ils ont réussi à se parler juste avant d'être

[142] J.F. Nash (1954), « Non-Cooperative Games », *Annals of Mathematics*, 54, p. 286-295.

interrogés par le Professeur et qu'ils se sont promis de nier tous les deux) afin d'avoir la situation la plus profitable au niveau des deux élèves (la somme des deux notes, 24., est la plus élevée de toutes les possibilités).

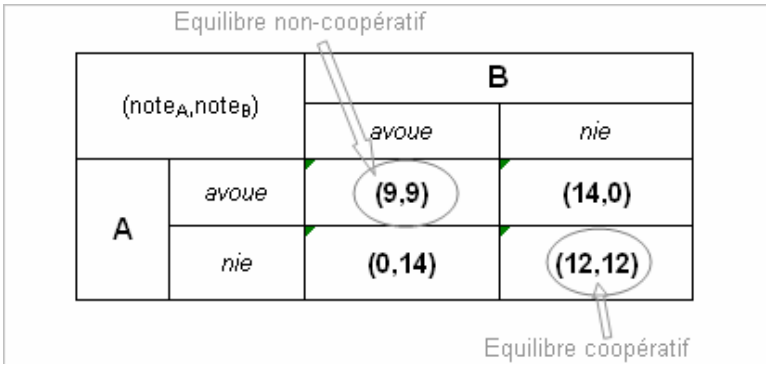


Figure 82 : Dilemme du prisonnier

Mais alors pourquoi, alors que cette solution coopérative existe, l'équilibre qui survient est celui où la situation des deux élèves est moins profitable (tous les deux en dessous de la moyenne) ? Tout simplement parce que l'équilibre coopératif n'est pas un équilibre stable.
Reprenons notre exemple. Disons qu'effectivement les deux élèves s'étaient promis de nier tous les deux. Mettons nous à la place de A. Il sait que B va nier puisqu'il lui a juré qu'il agirait de la sorte. Si A est sûr que B va nier alors A à tout intérêt à avouer et ainsi conserver sa note de 14 (c'est toujours deux points de pris et tant pis pour son malheureux comparse qu'il a trahi…). Cependant, B lui aussi a tout intérêt à agir de la sorte et à avouer pour profiter, pense-t-il, de la négation de A et conserver sa note initiale. Ainsi, les deux élèves vont avouer en pensant que l'autre nie et on retombe sur l'équilibre non-coopératif qui, lui, est stable. L'équilibre coopératif est instable, car la recherche du gain individuel pousse les individus à en dévier pour finalement se retrouver dans le cas de l'équilibre non-coopératif moins profitable.

La théorie des jeux est donc un outil puissant pour modéliser les interactions stratégiques occultées par la théorie néo-classique quand celle-ci se focalise sur l'équilibre de concurrence et l'hypothèse d'atomicité.
C'est également un « pavé dans la mare » de l'économie néo-classique puisque le sacro-saint principe de la « main invisible » est ici mis en défaut : La recherche par chacun de son intérêt individuel ne débouche pas sur l'intérêt collectif.
Bien entendu, des recherches nombreuses ont tenté d'examiner les moyens pour les individus d'adopter la solution coopérative (jeux répétés, modélisation des habitudes…). Le champ de recherche en théorie des jeux est fécond et les avancées nombreuses comme le prouve l'obtention en 2005 du « prix Nobel d'économie » par deux théoriciens des jeux (Aumann et Schelling).
L'outil de la théorie des jeux est donc plutôt séduisant mais son plus gros défaut est de ne pas toujours être concluant. Dans de nombreuses situations, il n'y pas un équilibre unique qui émerge mais plusieurs. C'est le problème de la multiplicité des équilibres que l'on peut mettre en lumière par l'exemple suivant.

8.5.3 La « guerre des sexes »

Le jeu et le suivant : Un couple discute pour déterminer où sortir pour la soirée. Monsieur souhaite aller voir un film d'action et madame aller au restaurant (il est déjà tard et seule une des deux possibilités est réalisable…). La satisfaction de chacun des deux individus qui constitue ce couple va dépendre de la décision de l'autre puisqu'ils passeront, quoi qu'il en soit, la soirée ensemble.
Les satisfactions de monsieur et madame en fonction de la décision de chacun sont exprimées dans la matrice de gain suivante. Si monsieur est intransigeant (« nous irons au cinéma et rien d'autre ! ») et que madame est intransigeante (« c'est le restaurant ou bien c'est fini entre nous ! »), le couple ne parvient pas à se mettre d'accord. La soirée est gâchée et chacun rumine de son côté dans l'appartement (les deux ont une insatisfaction (-1,-1)). Si monsieur est intransigeant et que madame est accommodante (« L'essentiel est d'être à tes côtés mon chéri… »), le couple va au cinéma. Monsieur est très satisfait (satisfaction de niveau 3) et madame légèrement satisfaite (satisfaction de niveau 1) puisqu'elle est en compagnie de monsieur… Idem si c'est madame qui est intransigeante et monsieur accommodant (« va pour le restaurant, mais je choisis le vin ! »). Madame est alors très satisfaite (3) et monsieur se console en buvant une agréable bouteille en compagnie de madame (satisfaction de niveau 1). Si les deux sont accommodants (« Oh non, chéri, choisis toi… » « mais non

mon amour vas-y, toi d'abord »), ils ne parviennent pas à décider et restent à l'appartement. Mais à la différence de la situation où les deux ont été intransigeants, ils ne se font pas la tête, et donc il n'y a pas d'insatisfaction (situation (0,0)). Voilà l'histoire…

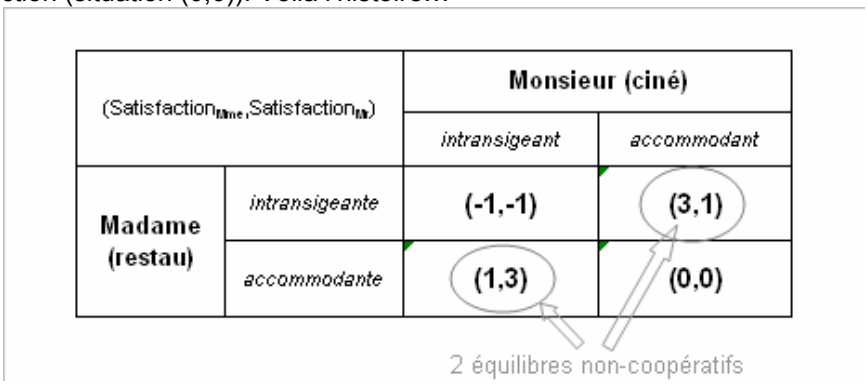


Figure 83 : La « guerre des sexes »

Sachant que monsieur et madame sont rationnels et qu'ils connaissent les différents résultats en termes de satisfaction en fonction de leur décision et de celle de l'autre, que vont-ils faire ?
Mettons-nous à la place de madame. Elle pense : « Si il est intransigeant, j'ai intérêt à être accommodante (car 3>1) ; Si il est accommodant, j'ai en revanche tout intérêt à être intransigeante (car 1>0) ». Sa décision dépend donc de celle de monsieur. Monsieur produit un raisonnement analogue : « Si elle est accommodante, j'ai tout intérêt à rester intransigeant, mais si elle est intransigeante mieux vaut que je sois accommodant ».
Ainsi, contrairement au dilemme du prisonnier, il y a ici non pas un mais deux équilibres non-coopératifs et on ne peut pas conclure définitivement à l'aide de la théorie des jeux sur cette situation. On ne peut que produire un résultat du type « si A fait cela, il est optimal pour B de faire ceci » mais cela est grandement insatisfaisant. La théorie des jeux n'a pas réponse à tout. La prise en compte des interactions stratégiques fait que l'on est souvent incapable de dire quelle décision sera prise dans les faits.
Bien entendu, l'introduction d'hypothèses supplémentaires (c'est un vieux couple qui se connaît bien, les caractères sont bien établis ; madame domine monsieur et impose sa décision…) ou le raffinement du jeu (jeu répété) permettrait de parvenir à la détermination d'un équilibre unique.

8.6 Les asymétries d'information et la théorie des contrats

Le modèle concurrence pure et parfaire fait l'hypothèse que l'information dont disposent les agents est parfaite. Cette hypothèse a été remise en cause à travers les travaux des économistes sur l'asymétrie d'information, dont le pionnier fut G. Akerlof (1970)[143].
L'**économie de l'information** étudie les comportements d'agents confrontés à des problèmes d'acquisition de l'information. L'économie de l'information utilise les résultats de l'économie de l'incertain puisqu'un agent, confronté à un problème d'acquisition de l'information, se retrouve dans le cadre d'une situation risquée. L'objectif de l'économie de l'information est de définir les structures d'information (quelles sont les formes du risque) et d'analyser les problèmes qui émergent dans chaque structure.
Une information imparfaite et asymétrique peut amener des différences importantes dans la nature de l'équilibre de marché. Les informations privées sont toutes les informations pertinentes pour la détermination d'une allocation efficace mais qui ne sont connues que d'une partie des intéressés. Lorsque les agents économiques ne disposent pas de la même information sur l'état de rareté ou la qualité d'un produit, l'équilibre entre l'offre et la demande peut ne pas être stable. L'information délivrée par le système de prix sera moins efficace mais s'affinera à long terme par corrections d'erreurs.

L'exemple canonique de la situation des asymétries d'information est celui du problème principal - agent, où un acteur (le principal) souhaite déléguer à un tiers (l'agent) le soin d'exécuter un tâche en étant intéressé à son résultat. En d'autres termes, cette situation désigne un ensemble de problèmes rencontrés lorsque la satisfaction du principal dépend de l'action ou de la nature d'un agent, sur lequel

[143] **Akerlof G. (1970**), « The Market for Lemons : Quality uncertainty and the Market Mechanism » , *The Quaterly Journal of Economics*, vol.84, pp.488-500.

le principal est imparfaitement informé. Le principal fait face à deux catégories de défaut d'information représentées par : -la « sélection adverse » (« anti-sélection ») ou le « hasard moral » (« aléa moral »). Ce modèle principal - agent est au cœur de la théorie de l'agence

8.6.1 La « sélection adverse »

La « sélection adverse » ou « anti-sélection » a trait à des situations où le type des agents n'est pas observable de sorte qu'un côté du marché doit estimer le type ou la qualité d'un produit sur base du comportement de l'autre côté du marché. Dans cette situation, le principal ignore les caractéristiques de l'agent qui ont un effet sur l'issue de l'accord entre l'agent et lui. L'exemple canonique de la situation de section adverse est le problème de l'assurance lorsque le principal (ici l'assureur) propose un contrat aux agents (les clients) sans pouvoir distinguer les bons et les mauvais comportements de ces derniers face au risque. En présence de sélection adverse, le principal P peut améliorer son information sur l'agent A en proposant un contrat en deux parties, qui obligent les agents à s'auto-sélectionner et ainsi à révéler leur qualité. Chaque agent est caractérisé par un paramètre $\theta \in [0,1]$, inobservable par le principal. Le principal propose un contrat unique au prix *p* sans franchise. L'agent θ retire du contrat une utilité $U(p,\theta)$ croissant avec θ décroissant avec *p* et le principal obtient un profit $\prod = p - c(\theta)$ avec *c* le coût probable d'assurance de l'agent croissant avec θ. Le paramètre θ peut s'interpréter comme une probabilité d'accident pour un agent. L'ensemble des agents est de la forme $\left[\theta_{\inf}, 1\right]$. Le principal ne connaît que la répartition des θ dans l'économie mais ignore le type de chaque agent.

L'agent accepte le contrat lorsque sa contrainte de participation est vérifié : $U(p,\theta) \geq 0$
c'est-à-dire qu'il retire un niveau de satisfaction positif lorsqu'il accepte le contrat.

Le principal détermine le contrat à partir de la maximisation de sa fonction objective sous contrainte de participation de l'agent :

$$\begin{cases} \max \prod(p) = \int_{\theta_{\inf}}^{1} p - c(\theta) d\theta \\ sc \quad U(p,\theta) \geq 0 \end{cases}$$

Si le principal connaissait les types de agents θ, il pourrait leur faire payer un prix $p^*(\theta)$ tel que $U(p^*(\theta),\theta) = 0$. Dans le cas d'une information parfaite sur le type des agents, le principal obtient alors une satisfaction maximale : $\prod^* = \int_{\theta_{\inf}}^{1} p^*(\theta) - c(\theta) d\theta$. Si les agents sont averses au risque, $\forall \theta, p^*(\theta) > c(\theta) d\theta$. Alors le principal supporte un coût de l'aléa moral pour l'assurance mesuré par la différence : $\prod(p) - \prod^*$.

8.6.2 Le « hasard moral »

Le « hasard moral » ou « aléa moral » a trait à une situation où un côté du marché ne peut pas observer les comportements de l'autre côté. Dans cette situation, l'agent s'engage à accomplir une action pour le compte du principal dont le résultat final dépend d'un paramètre inconnu du principal. L'asymétrie d'information dote l'agent de la possibilité d'utiliser à son avantage une information privée tel que le principal ne peut en constater ou en contrôler la nature. Les exemples d'aléa moral font appel aux théories de l'assurance (cacher un risque à l'assureur, prendre plus de risques une fois assuré, déclarer un sinistre non couvert par un contrat, de la protection sociale (fraudes à l'assurance chômage, arrêt maladie).

Un exemple d'aléa moral est celui mobilisé dans la relation contractuelle entre un exploitant de terre (l'agent) et un propriétaire terrien (le principal) qui ne peut contrôler directement ex-ante l'effort de travail de l'exploitant mais seulement la récolte ex-post. Le propriétaire prend toute la récolte R moins

ce que mange l'exploitant S. La récolte $R = R(e,\theta)$ dépend de deux paramètres inconnus et inobservables pour le principal l'effort de travail de l'exploitant *e* et de la situation météorologique. La fonction de récolte est strictement croissante avec le niveau d'effort du paysan $e \in [e_{\inf}, 1]$. Différentes situations contractuelles sont envisageables entre le principal et l'agent : le salariat (le propriétaire garde la récolte et paie à l'exploitant une somme fixée ex-ante), le métayage (les frais et la récolte sont partagés pour moitié entre l'exploitant et le propriétaire) et le fermage (l'exploitant supporte tous les risques et frais en échange de conserver la récolte pour laquelle il paye un loyer fixé au propriétaire).

8.6.2.1 Situation du salariat

Le propriétaire souhaite que l'exploitant fournisse un grand niveau d'effort tel que la récolte soit maximum : $\max_e R(e,\theta)$. La satisfaction du paysan est représentée par son niveau d'utilité $U = R - P - e$ avec P la part de récolte versée au propriétaire. Le principal prend toute la récolte sauf $U - S - e$. L'objectif du paysan (l'agent) est de minimiser son effort sous contrainte d'obtenir une part de la récolte fixée ex-ante (contrainte de participation) :
$\{\min_e R(e,\theta) / sc\ R(e,\theta) \geq S\}$. Le paysan a donc intérêt à faire un effort minimum $e_{\inf}$ et annoncer en retour au propriétaire les mauvaises conditions climatiques. Le principal est alors victime de l'aléa moral, ne peut infirmer le contraire.

8.6.2.2 Situation du métayage

Le principal prélève une part $q \in [0,1]$ de la récolte tel que le revenu qu'il obtient est donné par : *P=qR*. Le métayer (l'agent) détermine son effort pour obtenir un niveau maximum de satisfaction représenté par la différence entre le surplus de la récolte et l'effort engagé : $\max_e U=(1-q)R-e$. La condition du premier ordre est donnée par : $(1-q)\frac{\partial R(e,\theta)}{\partial e}(e_m,\theta) = 1$, c'est-à-dire que l'agent choisit un niveau d'effort e_m tel que la part du surplus de la récolte qui découle d'un accroissement marginal de son effort est égal à la désutilité de cet effort.

8.6.2.3 Situation du fermage

Le propriétaire reçoit une somme fixe F, quelque soit le niveau de la récolte. Le paysan conserve la totalité de la récolte et obtient un niveau de satisfaction $U=R-F-e$. Le programme de l'agent consiste donc à déterminer son niveau d'effort e_f tel que : $\max_e U=R(e,\theta)-F-e$. La condition du premier ordre est déterminée par : $\frac{\partial R(e,\theta)}{\partial e} = 1$.

En comparant la situation du métayage et du fermage :
$\frac{\partial R(e,\theta)}{\partial e}(e_f,\theta) = 1 < \frac{\partial R(e,\theta)}{\partial e}(e_m,\theta) = \frac{1}{(1-q)}$ comme $\frac{\partial R(e,\theta)}{\partial e}$ est décroissante en *e* alors $e_f > e_m$ tel que $R(e_f,\theta) > R(e_m,\theta)$. Le fermage conduit donc le fermier à faire un effort qui maximise la récolte. Ainsi lorsque l'agent supporte tout le risque, il est conduit à faire un effort maximum.

Bibliographie

Abraham-Frois G. (2005), Introduction à la macroéconomie contemporaine, Economica. ***(Pour les chapitres 3, 4 et 5)***

Assouline M. (2001), Mécanismes Économiques Fondamentaux, Polycopié de cours de 1ère année d'études, Ecole Centrale Paris, année 2001-2002. ***(Pour les chapitres 2, 4, 5)***

Begg D., Fischer S., Dornbusch R. (2002), Macroéconomie (2ème édition), Dunod. ***(Pour les ch. 3, 4, 5 et 6)***

Begg D., Fischer S., Dornbusch R. (2002), Microéconomie (2ème édition), Dunod. ***(Pour les ch. 1 et 2)***

Blanchard O., Cohen D. (2002), Macroéconomie, Pearson Education. ***(Pour les ch. 3, 4, 5 et 7)***

Burda M., Wyplosz C. (1998), Macroéconomie Une perspective européenne, Prémisses, De Boeck Université. ***(Pour les chapitres 3, 4, 5 et 6)***

de Montbrial T., Fauchart E. (2001), Introduction à l'économie (2ème édition), Dunod. ***(Pour chap. 1, 2, 4, 5)***

Gaffard J.-L. (2005), Analyse économique, Polycopié de cours de 2ème année d'études, Ecole Centrale Paris, année 2005-2006. ***(Pour tous les chapitres)***

Giraud P.-N. (2001), Initiation à l'économie, Polycopié de cours, Ecole des Mines de Paris, année 2005-2006. ***(Pour tous les chapitres)***

Guillochon B. et Kawecki A. (2003), Économie Internationale Commerce et macroéconomie (4ème édition), Dunod. ***(Pour le chapitre 6)***

Hairault J.-O. (2000) ***(sous la direction de)***, Analyse Macro-Économique, « Manuels Repères », La découverte (Tomes 1 et 2). ***(Pour les chapitres 4, 5, 6 et 7)***

Hairault J.-O. (2004), La croissance Théories et Régularités empiriques, Economica. ***(Pour le chapitre 7)***

Longatte J., Vanhove P. (2001), Économie Générale, Dunod. ***(Pour tous les chapitres sauf 1 et 2)***

Mucchielli J.-L., Mayer Th. (2005), Économie Internationale, HyperCours, Dalloz. ***(Pour le ch. 6)***

Prager J.-C., Villeroy de Galhau F. (2006), 18 leçons sur la politique économique, éditions du seuil. ***(Pour tous les chapitres)***

Jones C.I. (2000), Théorie de la croissance endogène , Prémisses, De Boeck Université. ***(Pour chapitre 7)***

Schubert K. (2000), Macroéconomie Comportements et Croissance (2ème édition), Vuibert. ***(Pour le ch. 7)***

Stiglitz J. (2000), Principes d'économie moderne , De Boeck Université. ***(Pour tous les chapitres)***

Varian H. (1995), Analyse microéconomique (3ème édition), Prémisses, De Boeck Université. ***(Pour le chapitre 2)***

Rappel : Les chapitres entre parenthèses au bout des références sont les chapitres du polycopié qui sont concernés et qui peuvent être approfondis par ces références.