

Introduction aux sciences économiques

École polytechnique - ECO 361

Olivier Gossner, Jean-Baptiste Michau, Vincent Rollet

2023

Table des matières

1	Valeur, demande et surplus	19
1.1	La mesure de l'utilité de travaux publics	19
1.1.1	Fibrage d'une zone rurale	20
1.1.2	Demande d'accès internet rapide et surplus	21
1.1.3	Surplus et propension marginale à payer	23
1.1.4	Surplus et propension totale à payer	25
1.2	Loi de la demande	25
1.3	Élasticité-prix de la demande	26
1.3.1	Déterminants de l'élasticité-prix de la demande	31
1.3.2	Déterminants de la demande	32
2	Coûts, offre et profits	37
2.1	Production de pétrole	38
2.2	Fonction d'offre	40
2.2.1	L'offre individuelle	40
2.2.2	La courbe d'offre	45
2.2.3	Élasticité-prix de l'offre	46
3	Marchés et équilibres de marché	53
3.1	L'équilibre de marché	55
3.2	Chocs d'offre et de demande	56
3.3	Propriétés de l'équilibre de marché	57
3.3.1	Maximisation du surplus	57
3.3.2	Efficience des marchés et premier théorème du bien-être	59
3.3.3	Contrôle des prix et perte sèche	61
3.3.4	Coûts fixes et condition d'ouverture	65
3.3.5	Surplus et efficacité	67
3.4	La libre entrée	69
3.5	Défaillances des marchés	71
3.5.1	Pouvoir de marché	71
3.5.2	Externalités	72
3.5.3	Biens publics	72
3.5.4	Asymétries d'information	72
3.5.5	Autres conditions du premier théorème du bien-être	72

4	Le commerce international	81
4.1	Le modèle de Ricardo	82
4.1.1	Productivités, avantages absolus et avantages comparatifs	83
4.1.2	Spécialisation et échanges	86
4.1.3	Un exemple historique: l'ouverture au commerce du Japon	91
4.1.4	Les enseignements du modèle de Ricardo	92
4.1.5	Remarques méthodologiques	93
4.2	Gagnants et perdants du libre-échange	94
4.2.1	Ouverture des frontières d'une petite économie	95
4.2.2	Imposition d'un tarif douanier	98
4.3	Autres effets du libre-échange	100
4.3.1	Bénéfices du libre-échange	100
4.3.2	Dangers du libre-échange	101
4.4	Quantifier les bénéfices du libre-échange	104
4.5	Déficits commerciaux	105
5	Économie publique	111
5.1	Les ressources et l'action de l'État	111
5.2	L'incidence fiscale	114
5.2.1	Imposer l'offre ou la demande ?	114
5.2.2	Qui paye l'impôt ?	117
5.2.3	Application au marché du travail	119
5.2.4	L'incidence de l'imposition des entreprises	120
5.2.5	L'incidence des tarifs douaniers de Donald Trump	121
5.3	Les distorsions fiscales	121
5.3.1	La perte sèche	121
5.3.2	Les distorsions de comportement	124
5.3.3	L'impact du niveau d'imposition	125
5.3.4	Qu'est ce qu'une bonne fiscalité ?	126
5.4	Les externalités	127
5.4.1	Les externalités négatives	128
5.4.2	Les externalités positives	130
5.4.3	Théorème de Coase	131
5.4.4	Les taxes pigouviennes	132
5.4.5	Application au réchauffement climatique	132
5.4.6	Taxes <i>vs.</i> régulations	136
5.5	Les biens publics	137
5.5.1	Les différents types de biens	137
5.5.2	Le problème du passager clandestin	138
5.5.3	Application au réchauffement climatique	138
5.5.4	La tragédie des biens communs	139
5.5.5	Les biens de clubs	140

5.6	Les interventions de l'État	140
6	Concurrence imparfaite	149
6.1	Le pouvoir de marché	149
6.1.1	Qu'est-ce que le pouvoir de marché ?	149
6.1.2	D'où provient le pouvoir de marché ?	150
6.2	Analyse du monopole	151
6.2.1	Comparaison avec l'équilibre de marché	153
6.2.2	Majoration des prix par le monopole	153
6.2.3	Effet sur le surplus	154
6.3	Analyse du duopole	155
6.3.1	Le modèle de duopole de Cournot	156
6.3.2	Fonction de meilleure réponse	156
6.3.3	Équilibre de Cournot-Nash	157
6.4	Augmentation du nombre de firmes	162
6.5	Compétition en prix : duopole de Bertrand	163
6.6	La régulation du pouvoir de marché	165
6.6.1	Régulation de la concurrence	165
6.6.2	Les incitations à l'innovation	166
6.6.3	Tendances récentes	167
7	Théorie des jeux	175
7.1	Concept de jeu	175
7.1.1	Le dilemme du prisonnier	176
7.1.2	Externalités et dilemme du prisonnier	177
7.1.3	Définition formelle d'un jeu	179
7.2	Équilibre de Nash	180
7.2.1	Les jeux de coordination	181
7.2.2	Stratégies mixtes, extension mixte et théorème de Nash	184
7.2.3	Interprétation de l'équilibre de Nash	190
7.3	Asymétries d'information	193
7.3.1	Le marché aux guimbardes	193
7.3.2	La sélection adverse	195
7.3.3	L'aléa moral	196
8	L'économie du travail	205
8.1	Un marché du travail compétitif	205
8.1.1	L'offre de travail	206
8.1.2	La demande de travail	207
8.1.3	Équilibre du marché du travail	209
8.1.4	Application empirique	211
8.2	Robots et emploi	213
8.3	Inégalités de revenu et redistribution	216

8.3.1	Les sources d'inégalités de salaires	217
8.3.2	Les politiques de redistribution	219
8.3.3	La dynamique des inégalités	220
8.3.4	Les inégalités de salaires entre les hommes et les femmes	221
8.4	Le chômage	222
8.4.1	Chômage et participation au marché du travail	223
8.4.2	Les causes du chômage	226
8.4.3	Le salaire minimum	229
8.4.4	Les allocations chômages	232
8.4.5	La protection de l'emploi	234
8.5	La régulation du marché du travail	236
9	Les marchés financiers	243
9.1	Le marché du crédit	243
9.1.1	Taux d'intérêt réel et nominal	244
9.1.2	L'offre de crédit	244
9.1.3	La demande de crédit	245
9.1.4	Equilibre du marché du crédit	247
9.2	L'actualisation	248
9.3	Les actifs financiers	249
9.3.1	Les obligations	249
9.3.2	Les actions	252
9.3.3	Hypothèse des marchés efficients	253
9.3.4	Actions ou obligations	255
9.4	La crise financière de 2008	257
9.4.1	La bulle immobilière	257
9.4.2	L'effet de levier	258
9.4.3	La crise bancaire	259
10	Macroéconomie	265
10.1	Les agrégats macroéconomiques	265
10.1.1	Le Produit Intérieur Brut	266
10.1.2	L'inflation	271
10.1.3	Les revenus	273
10.1.4	Les dépenses	276
10.1.5	Trois mesures du PIB	278
10.2	L'analyse macroéconomique	278
10.3	Paradoxes de la macroéconomie keynésienne	280
10.3.1	Le paradoxe de l'épargne	281
10.3.2	Le paradoxe de la flexibilité	281
10.4	Le rôle de la monnaie	282
10.4.1	La monnaie	282

10.4.2	Le rôle de la monnaie à long terme : la théorie quantitative de la monnaie	284
10.4.3	Le rôle de la monnaie à court terme : la théorie de la préférence pour la liquidité	287
10.5	Politique monétaire optimale	291
10.6	La Grande Récession de 2008	293

Remerciements

Nous remercions Romain Farthoat et Alex Tordjman pour leur excellente assistance dans la réalisation de ce polycopié, ainsi que Yves LeYaouang pour ses commentaires.

Introduction

Chaque année, les presque huit milliards d'humains qui peuplent la terre produisent et s'échangent entre eux un immense nombre de biens et de services, dont la valeur est estimée à près de 100 000 milliards de dollars. Pour produire tout cela, nos sociétés disposent de ressources limitées. Les ressources naturelles, les terres agricoles, et notre temps de travail ne sont pas en quantités infinies. En conséquence, on ne peut pas produire tout ce dont nous voudrions bénéficier. Les entreprises, les ménages, et les États décident d'allouer les ressources disponibles entre leurs différents usages, et cela mène à l'échelle mondiale à la production de près de 110 millions de tonnes de textile, 80 millions de voitures, des centaines de milliards d'heures d'éducation pour des élèves en primaire, etc.

Cette production ne se fait pas de façon uniforme : la Chine fournit ainsi plus de la moitié de la production textile mondiale, mais une faible part de nos avions. Différentes régions d'un pays ont également tendance à se spécialiser dans différents secteurs. La consommation des biens que l'on produit est également loin d'être uniforme : un Américain moyen consomme près de 23 fois plus qu'un Vietnamien moyen, et au sein de chaque pays il y a de grandes disparités de niveau de vie.

Les sciences économiques s'intéressent à ces phénomènes : pourquoi produisons-nous collectivement telle ou telle quantité de voitures, de blé, et d'heures d'éducation ? Dans chaque cas, produisons-nous en trop ou pas assez ? Pourquoi cette production se fait-elle à tel endroit et pas tel autre ? Et pourquoi ces biens sont-ils consommés par telle personne et pas telle autre ?

De façon générale, les sciences économiques étudient la création et de la répartition des ressources dont dispose la société. Cette question extrêmement large recouvre une longue série de questions qui animent les économistes, telles que les suivantes :

- Baisser la TVA sur les produits de première nécessité est-il un bon moyen d'améliorer les conditions de vie des plus pauvres ?
- Quel est l'impact de la robotisation des chaînes de production sur l'emploi ?
- Y a-t-il aujourd'hui plus ou moins de mobilité sociale qu'il y a cinquante ans ? Pourquoi ?
- Quel est le rôle des brevets dans le développement de nouvelles technologies ?
- Quel est l'impact du commerce international sur les inégalités ?
- Quels sont les moyens les plus efficaces de lutter contre le réchauffement climatique ?
- Comment répartir les élèves entre les différents collèges, lycées, ou universités d'une ville ou d'un pays ?

- Quel soutien une banque centrale doit-elle apporter à l'économie ?

Approches théorique et empirique

Pour répondre à une question en sciences économiques, deux approches sont utilisées. La première est une approche théorique, où l'on simplifie la réalité dans un modèle pour comprendre différents mécanismes à l'œuvre autour de nous. La seconde est une approche empirique où l'on analyse rigoureusement des données pour quantifier ces mécanismes.

Un exemple de démarche théorique

Une grande partie de la théorie économique a été élaborée à partir du champ des mathématiques. Émile Borel, John von Neumann, John Nash sont des mathématiciens qui ont apporté des contributions fondamentales à la théorie des jeux, méthodologie centrale en économie théorique.

Prenons l'exemple des travaux de David Gale et Lloyd Shapley en 1962 sur l'allocation des places en université pour les étudiants. Dans leur modèle, nous avons d'un côté des étudiants, chacun ayant des préférences sur les universités qu'ils ou elles sont susceptibles de rejoindre, et de l'autre des universités, chacune ayant un classement des étudiants. Comment effectuer un appariement entre places d'un côté, et étudiants de l'autre ? La question est de nature théorique, et Gale et Shapley mettent en avant une propriété désirable (on dit qu'elle est *normative*) : la stabilité.

Si un appariement associe un étudiant A à une université α , et B à β , mais A préfère β à α tandis que β classe A au dessus de B , l'appariement est dit *instable*. Un appariement qui n'est pas instable est dit *stable*.

Pour comprendre ce qu'est un appariement instable, imaginons la situation de la définition dans laquelle A préfère β à α tandis que β classe A au dessus de B . Dans ce cas, l'université β aurait intérêt à rejeter le candidat B et à faire une offre au candidat A , tandis que le candidat A aurait intérêt à rejeter l'offre de α pour accepter celle de β .

Cette notion de stabilité fait intervenir deux notions fondamentales en économie : celles de préférences d'un côté et de choix de l'autre. Les étudiants et les universités sont des exemples d'agents économiques ayant des préférences sur leur environnement. Ces agents sont amenés à faire des choix tels que le choix par les universités des candidats à qui faire une offre et le choix par les étudiants d'accepter ou refuser ces offres. Il s'agit là de l'expression des préférences des candidats et des universités. Plus généralement, ces choix peuvent être des choix de consommation, de production, de vote, d'allocation du temps entre travail et loisir. . . Les choix peuvent être plus ou moins contraints par des règles, mais les économistes supposent généralement que les préférences des agents guident leurs choix : c'est ce à quoi nous faisons référence en parlant d'hypothèse de *rationalité des agents*. Un agent rationnel choisit l'allocation qu'il préfère parmi toutes celles qui lui sont disponibles.

Gale et Shapley se sont posés la question de l'existence d'un appariement stable entre étudiants et universités, ainsi que celle de la manière de générer un tel appariement. Pour y

répondre, ils ont mis en avant un algorithme qui génère un appariement stable pour n'importe quel ensemble de préférences. Cet algorithme procède par offres successives : chaque université commence par faire des offres aux étudiants qu'elle préfère le plus. Chaque étudiant accepte temporairement la meilleure offre reçue et rejette les autres offres, libérant ainsi des places dans des universités. Ces places sont ensuite offertes par les universités à des étudiants de plus en plus bas dans leur classement, et chaque étudiant qui reçoit une nouvelle offre choisit entre cette offre et celle préalablement acceptée, libérant plus de places et ainsi de suite. . . Gale et Shapley ont démontré rigoureusement qu'un tel algorithme aboutit systématiquement à un appariement stable et montrent ainsi l'existence d'un tel appariement.

D'expérience, nous savons que les participants à un tel système d'appariement se posent la question de leur intérêt à suivre les règles de l'algorithme. Par exemple, est-il souhaitable pour un étudiant de rejeter une de ses premières offres, en espérant générer de meilleures offres ensuite ? L'économie s'intéresse beaucoup à de tels comportements dits *stratégiques* dans lesquels les agents ont intérêt à ne pas montrer leurs véritables préférences pour obtenir un résultat qu'ils préfèrent. Ce genre de comportement est particulièrement visible lors de scrutins, lorsque des électeurs ont intérêt à ne pas voter pour leur candidat préféré s'il a peu de chances de gagner, et à reporter leur voix sur un candidat mieux placé dans la course. Pour le cas de l'allocation des places à l'université, il a été démontré que chaque étudiant a intérêt à suivre l'algorithme en acceptant toujours la meilleure offre possible¹.

Lors de leur publication en 1962, les travaux de Gale et Shapley étaient surtout vus comme une curiosité mathématique. Depuis, les utilisations de la théorie de l'appariement se sont vastement généralisées, que ce soit pour les postes d'internes dans les hôpitaux, pour les lycées, ou encore pour les crèches. De manière plus surprenante, cette théorie a permis de sauver des vies en allouant de manière plus efficace les dons d'organes aux personnes en attente de dons. Et désormais, l'ingénierie de marché, c'est à dire la construction de mécanismes d'appariement tels que ceux utilisés par Uber ou Google (pour la publicité en ligne) sont des domaines en forte expansion. En 2012, à l'âge de 89 ans, Lloyd Shapley a obtenu le prix Nobel d'économie pour ces travaux réalisés 50 ans auparavant.

Un exemple de démarche empirique : la prime au gagnant en politique

Les chercheurs en économie réfléchissent souvent à des questions théoriques comme celle posée par Gale et Shapley, mais la majorité de la recherche actuelle est empirique, c'est à dire qu'elle utilise des données pour répondre à une question précise. Par exemple, une question importante en économie politique (la branche des sciences économiques qui s'intéresse aux interactions entre l'économie et les institutions politiques) est celle de la prime au sortant : lorsqu'un parti politique gagne une élection, cela lui donne-t-il un avantage à l'élection suivante ? Plusieurs mécanismes théoriques prédisent un tel avantage : les partis qui gagnent une élection ont plus de visibilité et ont un accès plus simple à des financements qui peuvent favoriser leur réélection.

1. En revanche, dans certaines situations, les universités peuvent avoir intérêt à faire des offres à des étudiants moins bien classés avant d'en faire à d'autres pourtant mieux classés pour s'assurer de leur venue. C'est un défaut du système, mais en pratique les contraintes légales interdisent aux universités de le faire.

Pour estimer cette prime au gagnant avec des données, on pourrait naïvement prendre un échantillon d'élections, regarder quel parti gagne chaque élection et regarder si les partis qui gagnent une élection au moment t ont une probabilité plus élevée de gagner l'élection suivante à $t + 1$. On trouverait sans doute que c'est le cas, mais pourrait-on conclure à l'existence d'une prime au gagnant ? Non, car ce résultat refléterait probablement des caractéristiques très différentes entre les partis qui gagnent des élections et ceux qui les perdent : les partis avec de meilleurs candidats et mieux financés ont une probabilité plus élevée de remporter les élections à la fois à t et à $t + 1$, donc il est logique de trouver que les partis qui gagnent une élection ont une probabilité plus élevée de gagner la suivante. La difficulté consiste donc à séparer l'effet de remporter l'élection de l'effet d'autres caractéristiques elles-mêmes corrélées au fait de gagner une élection.

Si vous aviez pour projet de mesurer la prime au sortant lors des élections parlementaires américaines, vous pourriez faire l'expérience suivante : prenez 400 circonscriptions, et divisez-les en deux groupes tirés aléatoirement. Ensuite, manipulez (sans que personne ne puisse le savoir) les résultats des élections de façon à ce que le candidat Démocrate gagne dans toutes les circonscriptions du premier groupe, et que le candidat Républicain gagne dans toutes les circonscriptions du second groupe. Ensuite, attendez deux ans, et mesurez la part des élections gagnées par des Démocrates dans chacun des deux groupes. Si cette part est la même dans les deux groupes, on pourra conclure qu'il n'y a pas de prime au sortant dans ce contexte. Au contraire, si la part d'élections gagnées par des Démocrates est plus importante dans les circonscriptions, qui, pour une raison aléatoire, avaient été assignées à un représentant Démocrate deux ans plus tôt, on pourra conclure à la présence d'une prime au sortant.

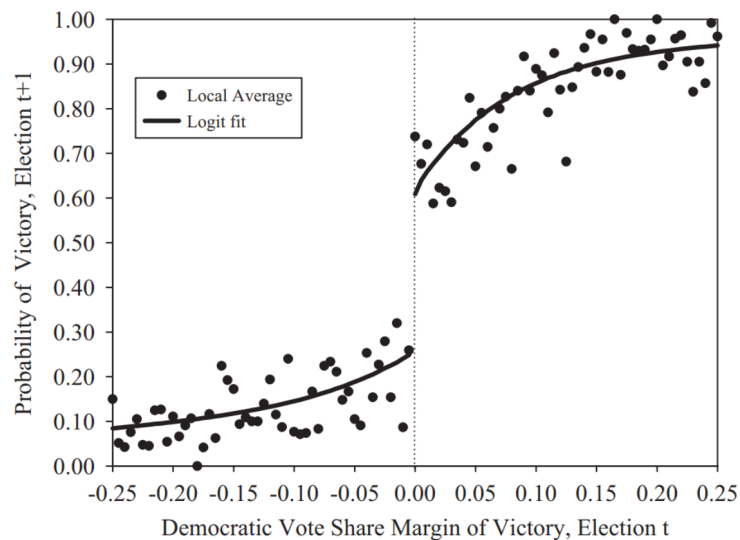
En faisant une telle expérience, vous feriez ce que l'on appelle un essai randomisé contrôlé. Cette méthode est extrêmement répandue en médecine pour évaluer l'effet d'un médicament sur une maladie. Dans ce contexte, il s'agit de prendre une population touchée par la maladie, et de donner le médicament à un échantillon aléatoire de celle-ci. Le groupe qui reçoit le médicament est appelé le groupe de traitement, alors que les autres patients (auxquels on aura donné un placebo – un produit qui a l'apparence du médicament mais ne contient pas de principe actif) constituent le groupe de contrôle. Si les patients du groupe de traitement se portent mieux que ceux du groupe de contrôle après quelque temps, on pourra conclure à l'effet bénéfique du médicament. Une alternative aux essais randomisés contrôlés sont les études observationnelles. Vous feriez une telle étude si, pour tester l'efficacité d'un médicament, vous donniez celui-ci à des patients et observiez par la suite si l'état de ceux-ci s'améliore ou non. Une telle étude vous donnerait des indices sur l'efficacité du médicament, mais comme vous ne savez pas comment la santé des patients aurait évolué si vous ne leur aviez pas donné de médicament, il vous sera difficile de conclure. En vous donnant accès à une comparaison de bonne qualité, les essais randomisés contrôlés fournissent des preuves plus solides que celles données par une étude observationnelle.

S'il est en général raisonnable de mener un essai randomisé contrôlé pour l'analyse d'un nouveau médicament, il est clair (et heureux) qu'il est impossible de mener l'expérience politique décrite plus haut pour mesurer la prime au sortant lors d'élections parlementaires.

Or, comme nous l'avons expliqué, une étude observationnelle est susceptible de nous donner des résultats erronés dans ce cas. Pour pouvoir tout de même apporter des réponses à des questions pour lesquelles l'expérimentation est impossible, les économistes ont développé plusieurs méthodes qui permettent d'exploiter ce que l'on appelle des "expériences naturelles".

Ainsi, David Lee (2008) a eu l'idée de s'intéresser aux élections très serrées pour mesurer la prime au sortant : en prenant un échantillon d'élections législatives aux États-Unis, il a calculé la probabilité pour le parti Démocrate de remporter une circonscription en fonction de sa marge de victoire dans la circonscription deux ans plus tôt. La figure 1 montre son résultat : à droite de la discontinuité (à une marge nulle), on trouve des candidats Démocrates qui ont gagné une élection au moment t , et à gauche de la discontinuité ce sont des candidats Démocrates qui ont perdu une élection au moment t . En ordonnées, on trouve pour chaque marge de victoire x la probabilité $p(x)$ que le parti Démocrate remporte le scrutin suivant (calculée comme la part des candidats Démocrates ayant une marge de victoire de x qui ont gagné l'élection à $t + 1$).

FIGURE 1 – Prime au vainqueur dans l'étude de Lee (2008). La ligne verticale représente la limite à partir de laquelle le candidat démocrate remporte l'élection au temps t .



Sur ce graphique, on peut voir une relation croissante entre x et $p(x)$: plus la marge Démocrate à une élection était importante, plus le parti avait de chances de remporter l'élection suivante. Cela peut refléter le fait que les préférences politiques restent à peu près stables d'une élection à l'autre, et que les circonscriptions Démocrates à l'élection t restent souvent Démocrates à l'élection $t + 1$. Ce qui est frappant est que cette relation est discontinue : lorsque le parti Démocrate perd de peu une élection (juste à gauche de la discontinuité), il a à peine 25% de chances de gagner l'élection suivante. Par contre, lorsque le parti Démocrate remporte de très peu l'élection (juste à droite de la discontinuité), il a un peu plus de 60% de chances de remporter l'élection suivante. Or, au moment t il n'y a pas de différences en moyenne entre les circonscriptions juste à droite de la discontinuité et celles juste à gauche : un peu par hasard, certaines ont élu un Républicain tandis que d'autres ont élu un Démocrate. Très proche de la discontinuité, nous sommes dans une situation proche

de l'expérience « idéale » décrite plus haut, où l'on assignait aléatoirement à certaines circonscriptions un élu Démocrate, il s'agit des circonscriptions du groupe de traitement, et à d'autres un élu Républicain, dans le groupe de contrôle. C'est pour cette raison que l'on parle d'expérience naturelle. Si l'on observe au moment $t + 1$ une différence entre les circonscriptions des deux côtés de la discontinuité, ces différences ne peuvent être liées qu'au résultat de l'élection et pas à des divergences préexistantes. Ici, on observe un saut important de la probabilité de réélection à la discontinuité, mettant en évidence la prime au gagnant : gagner une élection a un impact causal sur la probabilité de victoire à l'élection suivante.

L'étendue des sciences économiques

On pourrait s'étonner de ce dernier exemple : ne relève-t-il pas des sciences politiques plutôt que de l'économie ? En réalité, il reflète la très grande diversité des sujets abordés en sciences économiques. Ainsi, notre discipline s'intéresse aux systèmes politiques, à la famille, à la criminalité, à l'éducation, à la culture ou encore à la psychologie. Il y a deux raisons à cette diversité du champ d'étude.

Premièrement, on ne peut pas isoler l'économie du reste de la société. Les régimes politiques ont un impact considérable sur le développement des pays. La famille est le lieu où sont prises de nombreuses décisions économiques et où sont transmis la culture et le savoir. La psychologie a un impact considérable sur nos comportements, notamment lorsque nous devons effectuer des choix qui portent sur l'avenir et dans un environnement incertain.

Deuxièmement, pour répondre à leurs questions, les économistes ont développé des outils de modélisation théorique et d'estimation empirique qui sont très largement applicables à l'ensemble des problèmes relevant des sciences humaines. Ainsi, ce qui distingue les sciences économiques des autres sciences sociales est notre tendance à reposer sur la formalisation mathématique et notre désir de comprendre les relations de causalité, de les séparer des simples corrélations – ce qui permet notamment de bien comprendre l'impact des politiques publiques et de les rendre plus efficaces. Aujourd'hui, notre discipline se définit davantage par les méthodes qu'elle utilise que par les sujets dont elle traite.

Les sciences économiques sont-elles pour autant véritablement une science au même titre que les mathématiques ou la biologie ? N'est-ce pas trop prétentieux de modéliser les choix humains, n'y a-t-il pas une incompatibilité entre cette ambition et le libre arbitre de chacun ? La très grande diversité des comportements individuels implique qu'on ne peut pas prétendre en économie au même déterminisme qu'en physique. L'économie est assurément un système complexe et les relations de causalité sont à prendre au niveau statistique, et non individuel. Par la mise en relation de théories formalisées et de données, la discipline économique s'appuie indubitablement sur une méthodologie scientifique. La climatologie, l'épidémiologie, et même la médecine (ou, plus précisément, la physiologie) sont d'autres disciplines scientifiques s'intéressant à des systèmes complexes. La réalité du climat, celle d'une épidémie, ou de l'économie seront toujours plus riches et subtiles que la compréhension que nous en aurons. Malgré cela, l'approche scientifique de ces phénomènes permet une accumulation considérable de connaissances, utiles pour éclairer les choix collectifs et

notamment l'action publique.

Soulignons ici une particularité des sciences économiques. La compréhension que nous avons du monde dans lequel nous vivons peut affecter nos comportements. Ainsi, une théorie doit prendre en compte son impact sur la réalité qu'elle cherche à décrire. Ces considérations sont particulièrement importantes dans le domaine de la finance, où des avancées méthodologiques permettent le développement d'outils mis à disposition des ménages et entreprises, lesquels deviennent ensuite objets d'étude.

Le contenu de ce cours

Ce cours d'introduction ne pourra pas faire justice à toute l'étendue de la discipline économique et n'a pas vocation à donner une réponse définitive aux questions évoquées plus haut. Nous avons fait le choix d'insister sur les fondamentaux qui structurent notre discipline ainsi que sur leurs principales applications. Nous introduirons donc un certain nombre d'outils conceptuels utilisés en sciences économiques, comme la notion de demande, d'offre, de marché ou de théorie des jeux. Ces outils seront ensuite appliqués à l'étude de thèmes importants, tels que le commerce international, le rôle des taxes, le marché du travail, ou encore les marchés financiers.

La difficulté des sciences économiques n'est pas mathématique mais conceptuelle. Par conséquent, nous avons fait le choix d'une introduction mathématiquement simple, mais conceptuellement avancée. Au fur et à mesure de l'introduction du formalisme, nous réfléchirons en détail aux comportements individuels des ménages ou des entreprises auxquels ce formalisme correspond. Cette compréhension intuitive de l'économie permet de mettre en évidence des concepts qui sont bien plus robustes et bien plus généraux que le détail des équations utilisées pour les décrire. Ainsi, l'analyse économique vous donnera des outils vous permettant de comprendre une multitude de situations auxquelles vous pourriez être confrontés, sans qu'il ne soit nécessaire d'avoir recours à la modélisation. Bien sûr, lorsqu'un problème est vraiment nouveau ou complexe, une étude plus formelle s'impose, tant sur le plan théorique que empirique. Il faut alors utiliser des outils mathématiques et statistiques plus sophistiqués qui sont enseignés dans des cours plus avancés.

Bibliographie

Gale, D. and Shapley, L. S. (1962), 'College admissions and the stability of marriage', *The American Mathematical Monthly* **69**(1), 9–15.

Lee, D. S. (2008), 'Randomized experiments from non-random selection in US house elections', *Journal of Econometrics* **142**(2), 675–697.

Chapitre 1

Valeur, demande et surplus

Les sciences économiques s'intéressent à la répartition des ressources dans la société en adoptant deux angles qu'il est important de bien distinguer. Le premier est l'angle *positif* qui consiste à décrire la réalité telle qu'on l'observe, sans aucun jugement de valeur. Comprendre pourquoi les salaires sont plus élevés à Paris qu'à Marseille ou pourquoi l'activité économique engendre de la pollution sont des questions positives. Le second angle est dit *normatif* : il consiste à déterminer la répartition des ressources que l'on *devrait*, ou voudrait, avoir. Par exemple, quel devrait être notre niveau de pollution et comment réaliser les efforts pour y parvenir. L'analyse normative doit nécessairement reposer sur une analyse positive rigoureuse. En outre, elle peut faire intervenir des jugements de valeur, sur lesquels nous reviendrons. Un point commun des deux approches est qu'elles nécessitent de comprendre la *valeur* que chacun accorde à telle ou telle ressource.

1.1 La mesure de l'utilité de travaux publics

Lorsqu'un décideur public doit allouer des fonds, par exemple pour financer différents projets, il doit être capable d'évaluer la valeur ainsi créée par ces projets pour la collectivité. Par exemple, vaut-il mieux allouer des ressources à l'entretien des routes existantes ou à la construction de nouvelles routes, à la construction de nouvelles lignes de transport ferroviaire rapide ou à la subvention de lignes régionales ? Ces projets doivent-ils être financés par la collectivité, ou bien par les usagers des services ? Le coût pour l'utilisateur doit-il être égal pour tous, ou bien différencié en fonction de ce que chacun est prêt à payer ?

Élève de l'École polytechnique et ingénieur du corps des Ponts et Chaussées en charge de l'étude des routes et des péages, Jules Dupuit posa dès 1844 les bases de l'économie publique¹ dans son ouvrage « De la mesure de l'utilité des travaux publics » en s'intéressant à de telles questions.

1. Économie publique est le terme moderne utilisé pour désigner le champ consacré à l'étude des politiques publiques. Jules Dupuit utilisait celui d'*économie politique* qui désigne maintenant un champ distinct de l'économie étudiant le lien entre systèmes politiques et le processus d'élaboration des politiques publiques.

1.1.1 Fibrage d'une zone rurale

Prenons l'exemple d'une région qui projette d'améliorer l'accès internet dans une zone rurale par des investissements en fibres optiques. Afin de valider ou non ce projet, nous avons besoin de mesurer la valeur ainsi créée pour la zone touchée par cet investissement. Supposons que la collectivité bénéficiant de l'investissement soit composée de 10 000 foyers. Parmi eux, 500 ont un abonnement à internet par satellite leur coûtant 45€ par mois, et la fibre permettrait de réduire ce coût à 30€ par mois. Supposons par ailleurs pour simplifier que la qualité d'accès à internet est la même pour l'abonnement satellite et l'abonnement fibre. Le gain du passage à la fibre pour chacun des 500 foyers abonnés au satellite est mesurable par l'économie qu'ils font en passant à la fibre : il est de 15€ par mois. Cette méthode – proposée par Jean-Baptiste Say avant les travaux de Jules Dupuit – mesure le gain pour la collectivité dû à l'installation de la fibre à $500 \times 15€ = 7500€$ par mois.

Le calcul précédent mesure bien le gain pour les abonnés au satellite, mais quid des personnes n'ayant pas d'abonnement internet rapide ? Quel est l'effet économique pour elles d'une possibilité d'accéder à la fibre ? Ce que le calcul de Say oublie, nous dit Dupuit, c'est que certains des foyers qui ne sont pas preneurs d'un abonnement à 45€ par mois seront intéressés par un abonnement à 30€ par mois. Par conséquent, le nombre total de foyers accédant à internet augmentera du fait d'une baisse des tarifs d'accès.

Pour pouvoir mesurer le gain économique dû à l'augmentation des abonnements lors de la réduction de tarif de 45€ à 30€, nous avons besoin d'évaluer, pour chacun de ces foyers, la valeur qu'ils accordent à un abonnement internet. Plus généralement, on pose donc la question de la valeur qu'un individu attribue à un bien, et Jules Dupuit propose de mesurer cette valeur par le montant maximal que cet individu est prêt à payer pour jouir de ce bien ; montant qu'on appelle la *propension à payer*.²

Si on imagine une baisse progressive du prix de l'abonnement de 45€ à 30€, un nombre croissant de foyers vont pouvoir profiter d'un accès internet. Pour un foyer dont la propension à payer est de 40€ par mois, une baisse du prix de 45€ à 40€ n'engendre pas de valeur supplémentaire : à 40€, le foyer est indifférent entre accéder à internet pour 40€ par mois et ne pas avoir accès à internet en conservant cette somme. Néanmoins, toute baisse consécutive engendre un gain pour ce foyer, et lorsque le prix passe de 40€ à 30€, le surcroît de valeur obtenu par ce foyer est de 10€.

En généralisant ce raisonnement à l'ensemble des foyers, nous voyons qu'une mesure du gain pour la collectivité engendré par une baisse du prix de l'abonnement peut-être obtenue selon la méthode suivante :

- Estimer, pour chaque foyer, la propension maximale à payer de ce foyer pour un abonnement internet rapide.
- Si cette propension à payer est supérieure à 45€, le foyer bénéficie d'un gain estimé à

2. Contrairement au prix d'échange sur un marché, qui est indépendant de la personne possédant le bien, la valeur est un concept qui dépend intrinsèquement de l'individu. Par exemple, la valeur d'échange d'une automobile d'occasion peut être fixée selon une cote, mais cette automobile sera plus utile à certains individus qu'à d'autres, qui y attribueront par conséquent une valeur plus grande. Cette distinction entre prix et valeur adoptée par Jules Dupuit suit les travaux d'Adam Smith.

15€ par mois en passant du satellite à la fibre.

- Si elle est située entre 30€ et 45€, le gain est la différence entre la propension à payer et le nouveau tarif de 30€.
- Si elle est inférieure à 30€, le gain est nul car le foyer ne prendra toujours pas d'accès au nouveau tarif.

Enfin, on peut mesurer le gain pour l'ensemble de la collectivité comme la somme des gains pour tous les foyers.

1.1.2 Demande d'accès internet rapide et surplus

Supposons qu'à l'aide d'un sondage ou de statistiques nationales, il ait été établi que la propension à payer pour un accès rapide à internet se répartit selon les données du tableau 1.1. Nous calculons, pour chaque propension à payer, le gain par foyer puis le gain pour l'ensemble des foyers dû à la baisse du coût d'accès à internet. Le gain total pour la collectivité obtenu comme leur somme est estimé à 41 000€ par mois.

TABLE 1.1 – Propensions à payer de foyers pour un internet rapide, gain par foyer et gain total en € par mois consécutifs au passage de tarifs de 45€ à 30€.

Prix	Foyers	Gain/Foyer	Gain Total
> 60	0	-	0
60	100	15	1 500
50	500	15	7 500
45	1 000	15	15 000
40	1 100	10	11 000
35	1 200	5	6 000
30	1 300	0	0
20	1 900	0	0
10	1 700	0	0
0	1 200	0	0
Total	10 000		41 000

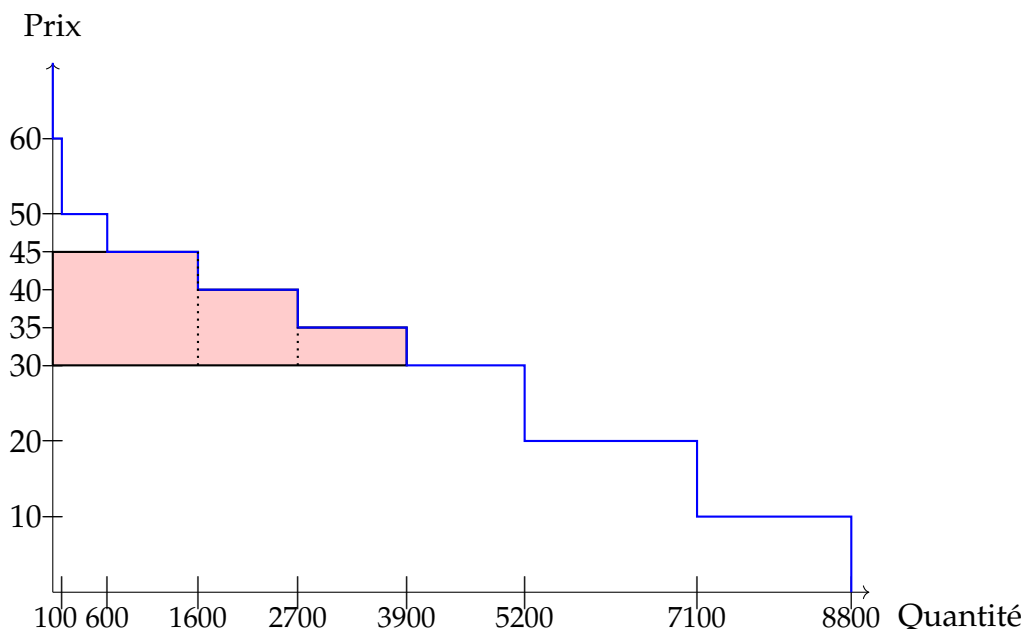
Nous voyons que les gains ou pertes économiques dus à des variations des prix peuvent être calculés dès que nous connaissons la propension à payer pour le bien de tous les foyers. Les économistes ont l'habitude de représenter les propensions à payer à travers ce qu'on appelle la *fonction de demande* qui représente, pour chaque valeur du prix, le nombre de foyers souhaitant acquérir le bien à ce prix. Pour déterminer la demande d'un bien à un prix p , on compte simplement l'ensemble des foyers prêts à payer p ou plus pour le bien. Le tableau 1.2 représente la demande pour un accès internet à haut débit. Aux points d'indifférence tels qu'à 40€, la demande est discontinue car pour un prix très légèrement inférieur, les 1 100 foyers ayant une propension à payer de 40€ consomment le bien, tandis qu'ils le rejettent à un prix très légèrement supérieur.

Le graphe de la fonction de demande est représenté figure 1.1. Notons que, conformément à l'usage en économie, le prix est représenté en ordonnée et les quantités en abscisse.

TABLE 1.2 – Demande pour un accès internet rapide en fonction du prix.

Prix en €	Demande
$p > 60$	0
$60 > p > 50$	100
$50 > p > 45$	600
$45 > p > 40$	1 600
$40 > p > 35$	2 700
$35 > p > 30$	3 900
$30 > p > 20$	5 200
$20 > p > 10$	7 100
$10 > p > 0$	8 800

FIGURE 1.1 – Nombre d'abonnements à internet demandés en fonction du prix (courbe bleue) et gain économique dû à la baisse de prix de 45€ à 30€ (aire rouge).



Visuellement, on constate que :

- L'aire du rectangle situé entre 45€ et 30€ et à une demande de 0 à 1600 correspond au gain des foyers ayant une propension à payer supérieure à 45€.
- L'aire du rectangle situé entre 40€ et 30€ et à une demande entre 1600 et 2700 correspond au gain des foyers ayant une propension à payer de 40€.
- L'aire du rectangle situé entre 35€ et 30€ et à une demande entre 2700 et 3900 correspond au gain des foyers ayant une propension à payer de 35€.

On lit donc sur ce graphique que le gain économique dû à une baisse du prix de l'abonnement internet est mesuré par l'aire située en abscisse sous la fonction de demande, et en ordonnée entre le prix d'origine et le nouveau prix.

En suivant la terminologie d'Allais introduite en 1943, on définit le *surplus* du consommateur au prix p comme l'aire située sous la fonction de demande des consommateurs, et

au dessus du prix p . Il s'en suit que le gain économique pour les consommateurs consécutif à un passage d'un prix p à p' est mesuré par l'augmentation correspondante du surplus du consommateur.

1.1.3 Surplus et propension marginale à payer

Dans le cas d'un abonnement à internet, la décision des agents économiques est binaire : souscrire à un abonnement ou ne pas y souscrire. Pour d'autres biens, les choix sont multiples : prenons le cas de l'eau, qui souvent est fournie par des compagnies publiques. Tout le monde consommera de l'eau, et la question est désormais : en quelle quantité ? Pour comprendre ce choix, il faut à nouveau se poser la question de la valeur du bien : quelle est la valeur de l'eau pour des consommateurs ? L'eau a de nombreux usages, et selon la quantité disponible, ou selon le prix, elle sera utilisée en priorité pour s'hydrater, puis pour les soins de toilette et la cuisine, pour l'arrosage du jardin, enfin d'autres usages moins importants tels qu'une piscine, etc. La valeur d'un litre d'eau dépend donc de la quantité disponible. Le premier litre quotidien est nécessaire à la survie d'un individu, tandis que les quantités suivantes ont progressivement de moins en moins de valeur. Ceci est mesurable par la *propension marginale à payer* de chaque individu, définie, pour une quantité q déjà disponible, comme le montant maximal que cet individu est prêt à payer pour une unité supplémentaire. Cette propension marginale à payer est, selon le raisonnement que nous avons effectué, décroissante avec la quantité q disponible.

Nous avons vu dans l'exemple d'accès internet que la notion de fonction de demande était particulièrement utile. Comment peut-on lier *propension marginale à payer* et *fonction de demande* d'un individu ?

Partons de la propension marginale à payer d'un individu, supposée connue. A prix p fixé, quelle sera la quantité du bien consommée par l'individu ? Si p est supérieur à la propension marginale à payer, chaque unité consommée autour de la quantité q coûte p , mais l'individu n'est prêt qu'à payer moins que ce montant, il choisira donc de réduire sa consommation. Symétriquement, si le prix est inférieur à la propension marginale à payer, l'individu souhaitera augmenter sa consommation. On obtient donc la relation suivante :

La quantité demandée égalise le prix et la propension marginale à payer.

Ce principe nous permet donc de dériver la courbe de la fonction de demande d'un individu à partir de ses propensions marginales à payer. Cependant, ceci correspond à un exercice théorique car les propensions marginales à payer sont rarement observées directement. De manière empirique, les économistes observent des quantités demandées en fonction des prix, et en déduisent les propensions marginales à payer.

Une fois ce principe établi, comment peut-on mesurer la valeur économique pour un individu liée à l'utilisation de q litres d'eau ? Le premier litre a, nous l'avons vu, une valeur plus élevée que le second, lequel a une valeur plus élevée que les unités suivantes. Pour obtenir la valeur associée à q litres d'eau, il faut donc sommer la valeur du premier litre, du

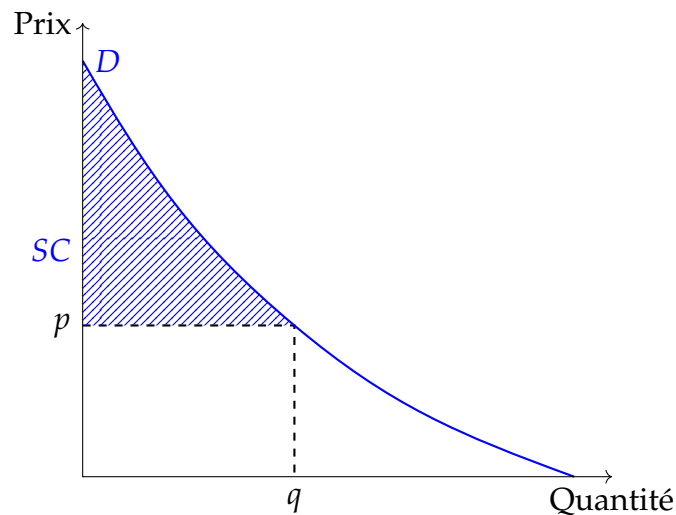


FIGURE 1.2 – Fonction de demande et surplus du consommateur. A prix p donné, la quantité q demandée égalise p et la propension marginale à payer. L'aire située sous la courbe de demande et au dessus de p est le surplus du consommateur, noté SC .

second,... , jusqu'à q ème, c'est à dire effectuer la somme des propensions marginales à payer du premier litre jusqu'à q . Pour obtenir le surplus du consommateur, il faut retrancher à cette quantité le prix pq payé par celui-ci. Graphiquement, le surplus du consommateur est donc mesuré par l'aire située sous la courbe de demande et au dessus du prix p égal à la propension marginale à payer pour un individu consommant q litres.

Le surplus d'un consommateur est défini comme l'aire située sous sa courbe de demande, et au dessus du prix.

Passons maintenant du cas d'un consommateur isolé à un ensemble de consommateurs. Pour une valeur donnée d'un prix p , on obtient, comme dans le cas de l'accès internet, le surplus de l'ensemble des consommateurs comme la somme des surplus individuels. Il s'agit donc de la somme des aires situées au dessus de p et au dessous des courbes de demandes individuelles, ou encore de l'aire située au dessus de p et au dessous de la somme des demandes individuelles.

La demande agrégée est la somme des demandes individuelles. Le surplus de l'ensemble des consommateurs est la somme des surplus de tous les consommateurs, et s'obtient comme l'aire situé sous la courbe de demande agrégée et au dessus du prix.

La valeur associée à l'accès à un bien est donc utilement mesurée par la notion de surplus de l'ensemble des consommateurs. Ce surplus s'obtient directement de la fonction de demande agrégée, et est mesurable à partir de l'observation d'un comportement agrégé des consommateurs.

1.1.4 Surplus et propension totale à payer

Nous avons décrit le surplus du consommateur comme l'aire située sous sa courbe de demande et au dessus du prix. Étant donné que la courbe de demande correspond au graphe de la propension marginale à payer u_i d'un consommateur i , le surplus du consommateur i est donné par :

$$SC_i = \int_0^q (u_i(x) - p) dx$$

En notant $U_i(q) = \int_0^q u_i(x) dx$ une primitive de u_i , on obtient la reformulation suivante du surplus :

$$SC_i = U_i(q) - pq.$$

Un consommateur est prêt à payer à un prix unitaire $u_i(0)$ sa première unité infinitésimale de consommation du bien, puis à payer à un prix unitaire $u_i(x)$ chaque unité infinitésimale supplémentaire lorsqu'il qu'il a déjà acquis x du bien. Le montant total qu'il est prêt à payer pour acquérir q unités du bien est la somme des montants qu'il serait prêt à payer pour toutes les unités infinitésimales, soit $U_i(q) = \int_0^q u_i(x) dx$. Pour cela, on appelle U_i la fonction de *propension totale à payer*. Cela nous conduit à une nouvelle interprétation du surplus du consommateur :

Le surplus d'un consommateur est la différence entre le montant total que ce consommateur aurait été prêt à payer pour q unités du bien, soit $U_i(q)$, et le montant effectivement payé pour acquérir cette quantité, soit pq .

À un prix unitaire p , la quantité q choisie par le consommateur égalise la propension marginale à payer $u_i(q)$ et le prix : $u_i(q) = p$. On en déduit que la dérivée de $U_i(x) - px$ est nulle au point $x = q$. Ceci nous conduit à réinterpréter le choix du consommateur :

La quantité q choisie par le consommateur à prix p donné maximise son surplus
 $U_i(q) - pq$.

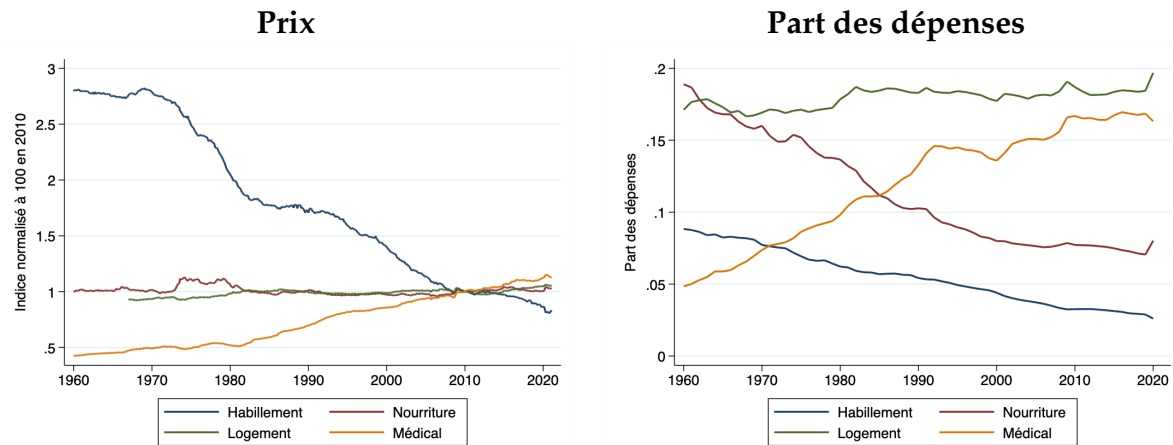
1.2 Loi de la demande

Nous avons introduit les fonctions de demande individuelle et agrégée comme des fonctions décroissantes du prix car nous avons supposé que la propension marginale à payer pour un bien est décroissante avec la quantité détenue de ce bien.

Loi de la demande : règle empirique selon laquelle la demande pour un bien décroît lorsque le prix augmente.

Mais en pratique, cette relation est-elle observée dans les données ? Prenons l'exemple des prix et dépenses d'habillement aux USA. La figure 1.3 montre qu'au cours des dernières décennies, le prix de l'habillement a considérablement chuté mais que sa part dans les dépenses des Américains a également beaucoup diminué.

FIGURE 1.3 – Évolution d'indices de prix et des dépenses pour différents postes de dépenses. Source : FRED Economic Data.



La loi de la demande ne serait-elle pas satisfaite dans cet exemple ? En réalité, deux phénomènes ont eu lieu pendant la période étudiée : le prix des vêtements a baissé, en augmentant la consommation, et les États-Unis sont devenus plus riches, augmentant la consommation de l'ensemble des biens. Même si les Américains dépensent aujourd'hui une part plus faible de leur revenu dans l'habillement que dans les années 1960, la quantité d'habits achetés a explosé pendant la même période (voir la figure 1.4). La loi de la demande prédit que pour un revenu fixé, la demande diminue avec le prix. La figure 1.3 ne permet donc pas de déduire une fonction de demande pour l'habillement car le revenu des Américains n'était pas fixé pendant la période étudiée.³

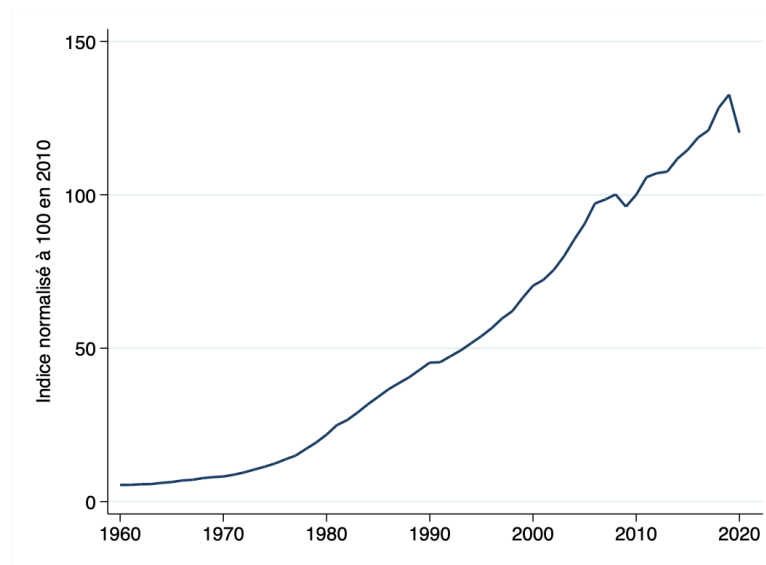
Pour comprendre l'effet d'une variation du prix d'un bien sur les dépenses pour ce bien, nous avons besoin d'un outil qui nous permette de représenter l'intensité avec laquelle une variation de prix génère une variation de la demande. L'élasticité-prix de la demande représente exactement cette notion.

1.3 Élasticité-prix de la demande

Définition de l'élasticité-prix de la demande. Précédemment, nous avons vu que la fonction de demande était une fonction décroissante du prix, et qu'elle était déterminée par les propensions marginales à payer des différents consommateurs. Cette fonction de demande nous permet de comprendre comment les consommateurs adaptent leur comportement lorsque les prix varient, par exemple lorsqu'une nouvelle taxe est mise en place ou

3. Notons que pour certains biens, il peut arriver que la demande soit (localement) croissante avec le prix. On parle alors de biens Giffen.

FIGURE 1.4 – Évolution de la consommation d’habillement par habitant aux États-Unis. Cette quantité a été mesurée comme la dépense d’habillement par habitant (en dollars) divisée par le prix d’un panier représentatif d’habillement. Source : FRED Economic Data.



lorsque des progrès technologiques permettent de produire un bien plus efficacement.

L'élasticité-prix e_D représente la variation de la demande en fonction de la variation du prix : si la demande pour un bien baisse de $x\%$ lorsque son prix augmente de 1% , alors on dit que l'élasticité-prix de la demande est de x :

$$e_D = \frac{\text{variation en \% de la demande}}{\text{variation du prix de 1\%}} = -\frac{\frac{dD}{D}}{\frac{dp}{p}} = -\frac{d \log D}{d \log p}'$$

Il est important de noter que ces variations de quantités et de prix sont exprimées en termes relatifs, et non absolus (on ne s'intéresse pas à la variation de la demande lorsque le prix augmente de 1 €, mais lorsqu'il augmente de 1%). Cela permet de faire de l'élasticité une mesure adimensionnelle, qui ne variera pas selon que l'on mesure les prix en euros, en milliers d'euros ou en dollars (et de même pour les quantités)⁴.

Élasticité et forme de la fonction de demande. Connaître l'élasticité-prix permet de connaître la forme de la courbe de demande. Lorsque celle-ci est faible ($e_0 \ll 1$), alors de grandes variations de prix n'entraînent pas des variations importantes de demande et la courbe de demande est relativement verticale. Plus l'élasticité augmente, plus la courbe de demande est aplatie. Dans ces cas, des variations de prix vont entraîner des variations importantes de la demande (voir figure 1.5).

4. Dans certaines conventions, e_D est défini comme égal à $\frac{d \log D}{d \log p}$ (et l'élasticité-prix de la demande est alors négative).

FIGURE 1.5 – Courbes de demande : très inélastique à gauche, élasticité de 1 au milieu, et très élastique à droite



Application : impact d’une augmentation de prix sur le budget alloué à un bien. Connaître ce paramètre d’élasticité-prix de la demande nous permet de comprendre l’impact des variations de prix sur le comportement des consommateurs. Reprenons l’exemple de l’habillement : lorsque le prix des vêtements diminue, la quantité achetée augmente. La dépense totale pour l’habillement, égale au prix multiplié par la quantité, pourrait donc augmenter ou diminuer a priori.

Le consommateur alloue à l’habillement la somme $B(p) = pD(p)$. Nous avons donc $\log B(p) = \log p + \log D(p)$ et

$$\frac{d \log B(p)}{d \log p} = 1 + \frac{d \log D}{d \log p} = 1 - e_D$$

Où e_D est l’élasticité-prix de la demande.

Ainsi, lorsque $e_D > 1$, le budget alloué à l’habillement diminue lorsque le prix augmente et lorsque $e_D < 1$, le budget alloué à l’habillement augmente avec le prix. Enfin, lorsque $e_D = 1$, le budget alloué à un bien est indépendant de son prix.

Ce résultat est assez intuitif lorsque l’on s’intéresse aux cas extrêmes : lorsque $e_D = 0$ (la demande est totalement inélastique), alors la quantité demandée ne s’ajuste pas à des variations de prix, et $D(p)$ est constant à une certaine valeur D . Le budget alloué par le ménage à ce bien est pD , qui augmente avec le prix. Au contraire, lorsque $e_D \rightarrow \infty$, une hausse infinitésimale du prix entraîne l’arrêt total de la consommation du bien, et le budget qui lui est alloué devient alors nul.

L’élasticité-prix de la demande d’habillement est estimée aux environs de 0,4, ce qui signifie qu’une baisse des prix s’accompagne d’une baisse du poste de dépense, en adéquation avec les faits empiriques observés⁵.

Pourquoi s’intéresser à l’élasticité? Avoir une idée de l’élasticité-prix de la demande d’un bien peut se révéler crucial dans la décision publique.

Ainsi, une entreprise qui vend un produit dont la demande est très inélastique pourra le vendre à des prix élevés sans voir ses ventes décroître fortement. En une décennie, le prix de l’insuline aux États-Unis a été multiplié par trois, atteignant en 2016 un prix moyen de

5. Source : “La nouvelle sensibilité des consommateurs aux prix”, par Pascale Hebel, Nicolas Fauconnier et Morgane David, Cahier du CREDOC, 215, 2005.

450\$ par mois. Vous ne serez pas étonnés de savoir que la quantité d'insuline achetée par les Américains ne s'est pas effondrée pour autant. L'élasticité-prix de la demande détermine donc en partie la capacité des entreprises à faire des profits. Dans certains cas, les institutions publiques ont un rôle de régulation à jouer lorsque les entreprises peuvent placer les prix de leurs produits très au-dessus de leur prix de fabrication, comme c'est le cas pour l'insuline américaine. Nous reviendrons sur ces problématiques plus tard dans ce cours.

Pour un décideur public, avoir une idée des élasticité-prix des produits auxquels il souhaite soit favoriser l'accès (comme dans les transports en commun ou l'accès à internet) ou décourager l'accès (comme pour les carburants polluants) est crucial. Dans notre exemple d'abonnements à internet haut débit, une méthode de sondages peut permettre d'estimer la distribution des propensions à payer dans la population et de tracer la courbe de demande. Cette méthode est néanmoins assez coûteuse et imprécise, car les réponses ne sont pas toujours sincères ou exactes. Une estimation de l'élasticité de la demande pour l'accès internet haut débit ainsi que la demande à prix donné permet de déduire une approximation de la courbe de demande. Cela permet d'estimer le nombre d'abonnés attendus à 30€, ainsi qu'une estimation du surplus qu'une baisse des tarifs de 45€ à 30€ peut engendrer.

La mesure d'une élasticité : le pricing d'Uber. Il existe différentes méthodes pour calculer des élasticités-prix et estimer des courbes de demande. Prenons l'exemple d'une telle mesure pour un service : les courses Uber. Lorsque vous commandez une course, Uber vous propose un prix de vente qui correspond à un prix de base multiplié par un facteur de surcoût, qui augmente avec l'affluence. Par exemple, en cas de forte affluence, Uber pourrait vous proposer un prix qui correspond au tarif de base multiplié par 1,6, 1,7, 1,8... Lorsqu'Uber vous propose un tarif pour une course, vous pouvez l'accepter ou renoncer à la commande. Une approche naïve consisterait à comparer les décisions des consommateurs faisant face à différents niveaux de surcoût. Par exemple, face aux données de la table 1.3, nous pourrions être tentés de faire le raisonnement suivant : lorsque le prix des courses augmente de 50%, la demande de courses Uber décroît de 25%, ce qui nous donne une élasticité de 0,5.

TABLE 1.3 – Exemple de quantités demandées - Uber

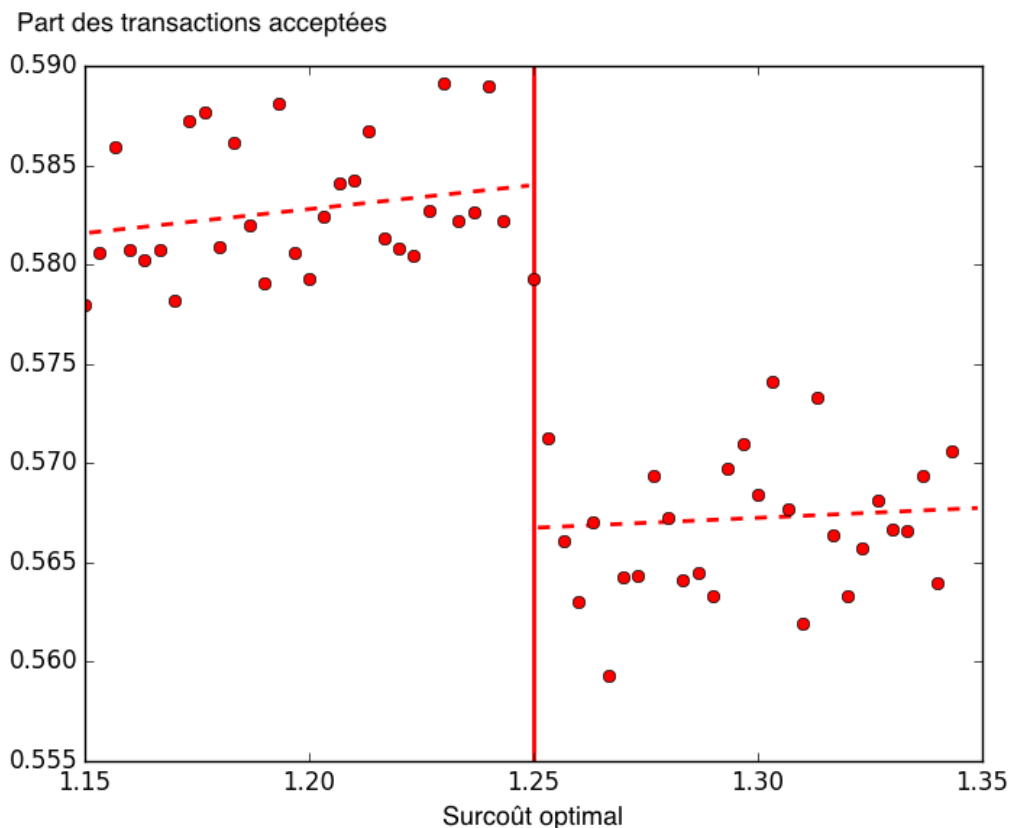
Niveau de surcoût	1,0	1,5
Proportion d'utilisateurs qui acceptent la transaction	0,8	0,6

Cette estimation de l'élasticité serait sans doute biaisée. Les moments où Uber propose un surcoût de 1,5 sont des moments de forte affluence, dans lesquels les utilisateurs ont peut-être moins d'alternatives à Uber que dans les moments sans surcoût (par exemple, le surcoût de 1,5 pourrait être proposé aux moments où les transports en commun ne fonctionnent plus). La proportion d'utilisateurs qui acceptent la transaction n'est donc pas seulement une fonction des prix, mais d'une multitude d'autres facteurs qui peuvent être difficiles à mesurer.

Pour contourner ce problème, un groupe d'économistes (Cohen et al., 2016) a décidé d'utiliser une caractéristique de l'algorithme d'Uber qui détermine les prix des courses. Cet

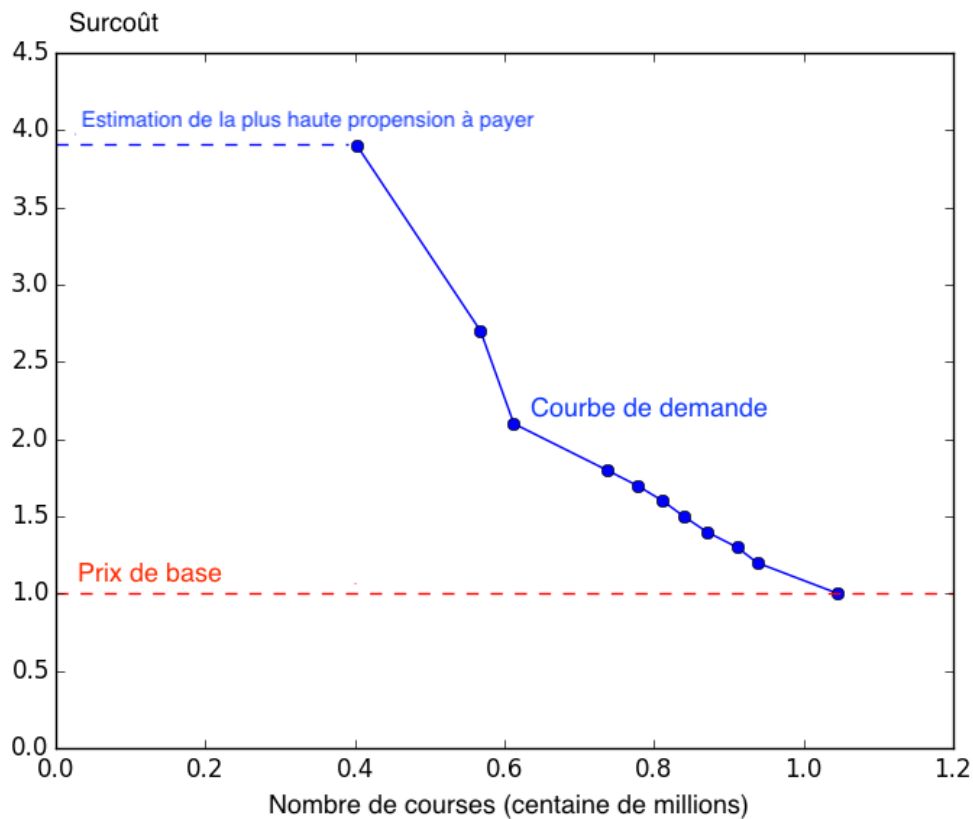
l'algorithme calcule en fonction de l'affluence un surcoût jugé optimal, et propose à l'utilisateur un arrondi. Ainsi, si le surcoût optimal est de 1,249, Uber proposera à l'utilisateur un surcoût de 1,2, mais si le surcoût optimal est de 1,251, l'utilisateur fera face à un surcoût de 1,3. La figure suivante montre la proportion d'utilisateurs qui acceptent la transaction pour différentes valeurs de surcoût optimal. Les points à gauche de la ligne verticale correspondent à des utilisateurs qui font face à un surcoût de 1,2, tandis que ceux qui sont à droite de la ligne font face à un surcoût de 1,3. Les utilisateurs juste à gauche de la ligne verticale sont dans des situations d'affluence très similaires à ceux qui sont juste à droite de la ligne. La seule différence entre ces deux groupes est le prix auquel ils font face. La différence de comportement observés entre les deux groupes correspond donc à l'impact de la variation de prix, et rien d'autre. Le taux de transaction à gauche de la discontinuité est de 58,3%, et celui à droite de la discontinuité est de 56,5%. Lorsqu'on augmente le prix des courses Uber de 8,3%, le taux de transaction baisse de 3,2%. A ce niveau de prix, on peut donc estimer l'élasticité-prix à 0,39.

FIGURE 1.6 – Pour mesurer l'élasticité-prix de la demande de courses Uber en un prix donné, s'intéresse à une discontinuité dans l'algorithme de la plateforme (Cohen et al., 2016)



En répétant cet exercice à toutes les discontinuités présentes dans l'algorithme d'Uber (aux surcoûts jugés optimaux de 1,05, 1,15, 1,25, etc.), on peut estimer l'élasticité-prix de la demande pour différents niveaux de prix, et donc accéder à une approximation de la dérivée de la fonction de demande. Connaissant la quantité demandée au niveau de prix de base, on peut ainsi tracer toute la fonction de demande (voir figure 1.7).

FIGURE 1.7 – Après avoir mesuré l'élasticité-prix de la demande pour différents prix, on peut retracer la courbe de demande des courses Uber



1.3.1 Déterminants de l'élasticité-prix de la demande

Si l'on ne dispose pas de mesures précises de l'élasticité comme dans le cas des courses Uber, un raisonnement économique permet au moins de comprendre si la demande pour un produit est plus ou moins élastique :

- **L'élasticité-prix d'un produit dépend de sa substituabilité.** Par exemple, si le prix des céréales pour petit-déjeuner augmente beaucoup, il est probable que la demande pour ce produit baisse rapidement : celui-ci est en effet facilement substituable à d'autres biens (comme le pain). En revanche, si le prix de la crème solaire augmente, la quantité demandée de ces produits ne devrait pas beaucoup changer dans la mesure où peu de substituts sont disponibles. L'élasticité-prix de la crème solaire est donc inférieure à celle des céréales.
- **Elle peut être différente pour une catégorie de biens et ses sous-catégories.** On peut imaginer que l'élasticité-prix de l'habillement en général est assez faible, dans la mesure où il est difficile de ne pas consommer d'habits. Néanmoins, si l'on s'intéresse à une sous-catégorie d'habillement, comme les jeans Levi's, l'élasticité-prix de la demande sera plus élevée, car ces jeans sont facilement substituables à d'autres pantalons. L'élasticité-prix d'une catégorie de consommation peut être assez faible quand bien même elle serait forte pour chacun des produits qui la composent.
- **Les produits de première nécessité ont une élasticité-prix plus faible.** Ainsi, une aug-

mentation du prix de l'insuline, du savon ou de la nourriture ne s'accompagnera pas d'une grande baisse de la consommation de ces produits. Au contraire, la quantité demandée de croisières ou de billets d'avion est susceptible de varier fortement avec les prix.

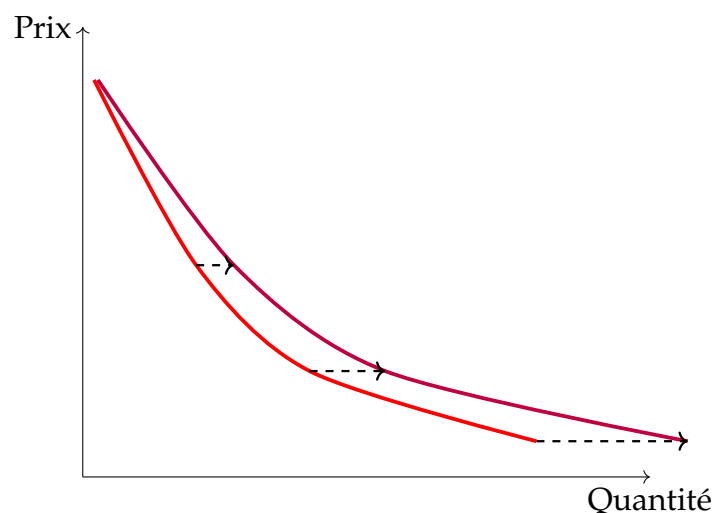
- **L'élasticité-prix d'un produit peut dépendre de l'horizon temporel auquel on s'intéresse.** Ainsi, si le prix du tabac augmente, il est probable que sa consommation ne diminue pas beaucoup à court-terme : l'addiction à la nicotine rend le tabac relativement insubstituable à d'autres produits et diminue son élasticité. En revanche, ces hausses de prix peuvent avoir un important impact de long-terme, dissuadant de potentiels fumeurs de commencer leur consommation. Dans ce cas, il peut être utile de distinguer l'élasticité-prix de court terme de celle de long terme.

1.3.2 Déterminants de la demande

Souvenons-nous que la *demande* pour un bien est une fonction du prix de ce bien. Cette notion doit être distinguée de la *quantité demandée*, qui est la valeur prise par cette fonction à un prix donné. Des modifications du prix modifient donc la quantité demandée, mais ne modifient pas la fonction de demande en elle-même. La fonction de demande n'est par ailleurs pas un objet immuable. Elle dépend de nombreux facteurs économiques. En voici quelques uns parmi les plus importants :

- **Le nombre d'acheteurs** : la *demande* croît avec le nombre d'acheteurs, comme illustré par la figure 1.8.

FIGURE 1.8 – Accroissement de la demande consécutif à une augmentation du nombre de consommateurs



- **Le revenu des ménages** : pour la plupart des biens, la demande croît avec le revenu. On parle dans ce cas de bien *normal*. C'est le cas de voyages, de l'habillement, etc. Certains biens ont une demande décroissante avec le revenu, et on parle alors de biens *inférieurs*. C'est le cas des pommes de terre : lorsque les revenus augmentent, les consommateurs

y substituent la consommation de biens plus onéreux, diminuant alors la demande en pommes de terre.

- **Les prix d'autres biens** sont aussi un facteur influant la demande d'un bien donné.
 - Pour des biens *substituts*, la demande d'un bien croît avec le prix de l'autre. C'est le cas du train et des vols domestiques en avion. Les consommateurs effectuent des arbitrages entre les deux types de biens : plus l'avion est cher, plus la demande de billets de train est forte.
 - Pour des biens *compléments*, la demande d'un bien décroît avec le prix de l'autre. C'est le cas des box internet et des abonnements de video à la demande (VOD). Une baisse du prix des box permet d'augmenter le nombre de box internet, et donc le nombre de consommateurs pouvant s'offrir la VOD. Réciproquement, des prix plus bas de la VOD rendent les box plus attrayantes, et augmentent leur demande.
- **Les anticipations sur le futur** : par exemple, des anticipations d'augmentation de revenus peuvent conduire à une augmentation de la demande par anticipation, des craintes sur des chocs économiques à venir conduisent à de l'épargne de précaution. Ainsi, les achats de voitures chutent de manière importante lors de récessions. Ces aspects dynamiques sont particulièrement importants en finance et en macroéconomie.

Nous verrons dans les chapitres suivants comment des facteurs tels que la concurrence entre producteurs ou les taxes ont un impact sur les prix, et donc sur les quantités demandées.

Bibliographie

Allais, M. (1943), 'A la recherche d'une discipline économique : Traité d'économie pure', *Ateliers Industria, Paris.*(3^e Éd., *Economica* 1994) .

Cohen, P., Hahn, R., Hall, J., Levitt, S. and Metcalfe, R. (2016), Using big data to estimate consumer surplus : The case of uber, Technical report, National Bureau of Economic Research.

Chapitre 2

Coûts, offre et profits

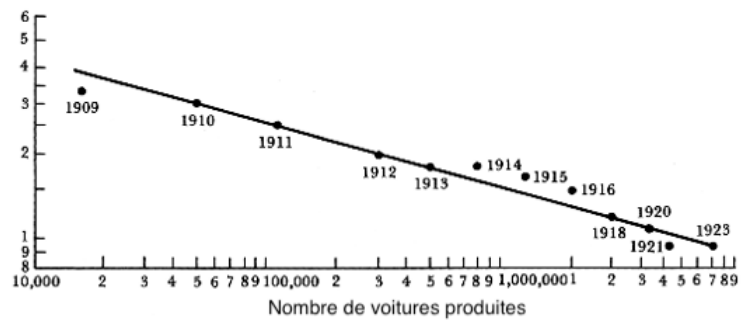
Lorsque nous avons décrit la courbe de demande, nous nous sommes posé la question suivante : pour un prix donné, quelle sera la quantité consommée par les agents économiques? Pour l'instant, nous n'avons pas dit grand chose de l'origine de ce prix : il est néanmoins clair que le prix d'un bien est lié à son coût de production.

Le premier séquençage d'un génome humain coûta près de 500 millions de dollars et fut achevé en 2003. Une fois cette prouesse scientifique réalisée, il était estimé que le séquençage d'un deuxième ADN ne coûterait que 50 millions de dollars. À ce stade, il était difficile de parler du « prix » d'un séquençage ADN, car personne ne vendait vraiment ce service. Les choses ont néanmoins vite changé : en 2006, obtenir un séquençage correct ne coûtait plus que 14 millions de dollars, et en 2015 seulement 1000 dollars. Avec une réduction aussi drastique des coûts, des entreprises commerciales ont commencé à proposer le service de séquençage, et le prix de ce service a baissé petit à petit avec l'amélioration des techniques scientifiques.

Le séquençage de l'ADN est un exemple particulièrement frappant de la descente d'une *courbe d'expérience* : lorsqu'une unité d'un bien est produite, ceux qui le produisent acquièrent de l'expérience, diminuant le coût de production des unités suivantes. La figure 2.1 montre ainsi que le coût de production de l'icône Ford T a été divisé par plus de trois au cours des dix premières années de production. Au fur et à mesure de cette diminution des coûts de fabrication, son prix a également été abaissé, rendant la voiture accessible à de nombreux américains. Plus récemment, la diminution du prix des matériels électroniques ou de l'accès à internet a été fulgurante.

Pour comprendre les échanges de biens et de services, il est donc pertinent de s'intéresser à ce qui se passe du côté des producteurs, qui sont contraints par leurs coûts de fabrication : quand décident-ils de produire ou non, et quelles quantités décident-ils de produire? Nous pouvons commencer par nous poser une question symétrique à celle du chapitre précédent : lorsqu'un producteur fait face à un prix, à quels choix fait-il face et comment ses décisions de production sont-elles guidées?

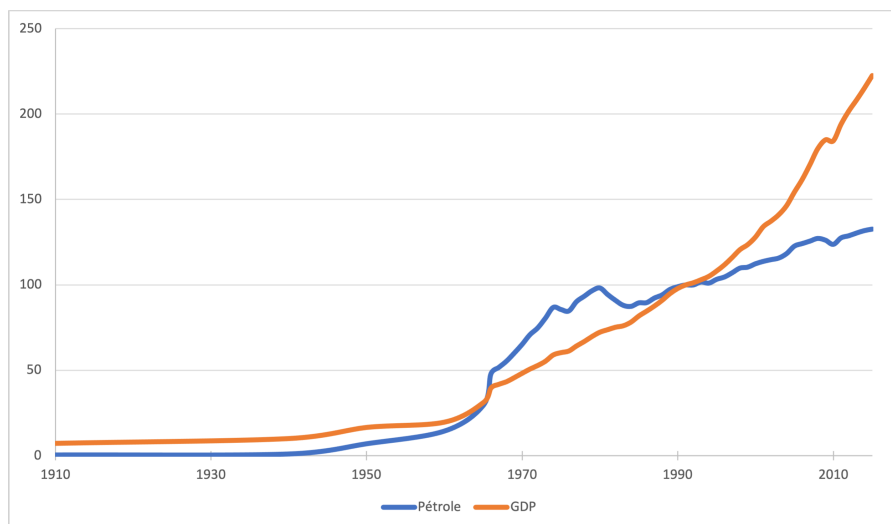
FIGURE 2.1 – Coûts de production de la Ford T (en dollars)



2.1 Production de pétrole

Un exemple intéressant est celui du pétrole : malgré le développement des énergies décarbonées et des gains d'efficacité énergétique, la production de biens et services dans l'économie mondiale en est toujours fortement dépendante. La consommation de pétrole est ainsi en augmentation constante depuis plus d'un siècle. Le graphique 2.2 nous montre une forte corrélation entre l'augmentation de la production mondiale et l'utilisation de pétrole, avec cependant un ralentissement relatif de l'utilisation de pétrole dans la période récente. Identifié comme ressource stratégique dès la fin du XIX^{ème} siècle, l'augmentation de son prix lors des deux chocs pétroliers dans les années 1970 a été suffisante pour mettre un coup de frein à la période de forte croissance d'après guerre communément appelée les 30 glorieuses.

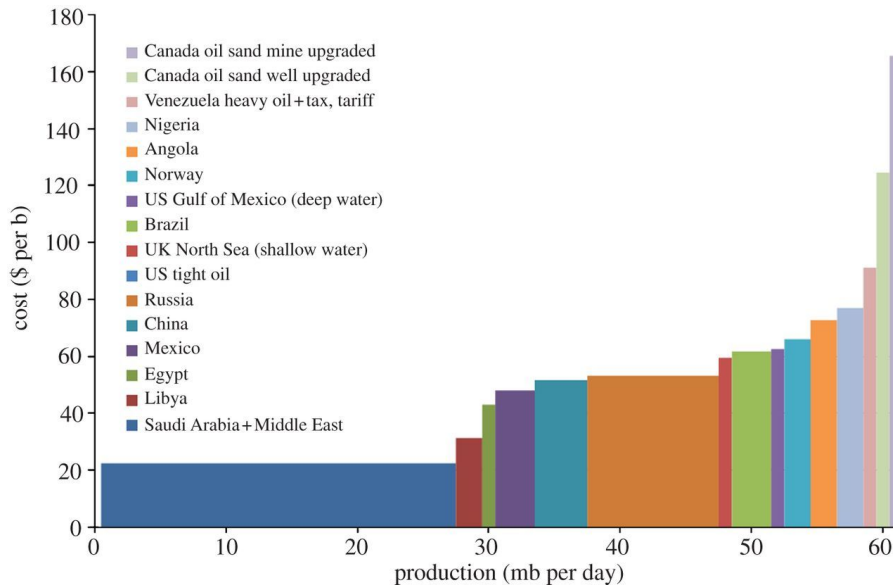
FIGURE 2.2 – Évolution du PIB mondial et de la production de pétrole de 1910 à 2015. Base 100 en 1990. Source : ourworldindata.org



Pour comprendre comment la production mondiale s'adapte aux prix du pétrole, il est important de noter que cette production est en fait la somme de productions de pétrole provenant de différents pays et puits ayant chacun des coûts particuliers. La plupart des pays producteurs sont spécialisés dans une ou quelques technologies d'extraction de pétrole. La figure 2.3 résume les principaux types de production et leurs capacités, ordonnés par type

de production croissante. La largeur de chaque colonne correspond à la capacité de production, et la hauteur au coût. On voit ainsi que le pétrole le moins onéreux et le plus abondant est celui produit par le Moyen-Orient avec un coût de production d'un peu plus de 20\$ par baril, tandis que le pétrole de schiste produit par les USA est plus onéreux, avec un coût de production d'environ 70\$ par baril.

FIGURE 2.3 – Coûts et capacités d'extraction de pétrole par jour



La courbe en escalier ainsi tracée par ces barres est ce qu'en économie nous appelons une *fonction d'offre*. Elle permet de connaître, pour chaque prix, la quantité produite du bien. Par exemple, pour un prix de 35\$ par baril, la production du Moyen-Orient et celle des eaux peu profondes du Royaume-Uni sont rentables, mais les autres ne le sont pas, on peut prédire une production aux alentours de 30 millions de barils par jour. Pour un prix de 60\$ par baril, on lit que la production mondiale peut être estimée à 50 millions de barils par jour, et ainsi de suite...

Bien évidemment, cette courbe n'est elle-même qu'une approximation, car chaque mode de production recouvre des coûts hétérogènes. Par exemple, une partie du pétrole russe est plus aisée à extraire qu'une autre, et donc sera associée à des moindres coûts. Par ailleurs, des pays ou groupes de pays peuvent avoir intérêt, temporairement ou durablement, soit à réduire leur production pour faire augmenter les prix (ce qui est le cas de l'OPEP, phénomène que nous reverrons plus tard dans ce cours), soit à produire à perte (par exemple pour briser les tentatives d'autres pays de développer leur production énergétique). Mentionnons aussi qu'un pays peut avoir intérêt à baisser sa production afin de conserver des réserves futures.

Pour n'importe quel bien ou service, la quantité totale produite correspond à l'agrégation d'un grand nombre de décisions individuelles de producteurs, qui font face à différents coûts de production selon leurs machines, leur localisation géographique, les taxes qu'ils doivent payer, etc. Les économistes formalisent ces phénomènes de la façon suivante : chaque producteur est associé à une *fonction de coûts* qui reflète l'ensemble des contraintes aux-

quelles il fait face. Cette fonction de coûts permet de modéliser les choix du producteur en fonction du prix auquel il fait face, donnant une *fonction d'offre individuelle*. L'agrégation de l'ensemble des fonctions d'offres individuelles nous donne une fonction d'offre globale pour un bien donné, qui sera utile pour comprendre comment s'ajuste son prix.

2.2 Fonction d'offre

2.2.1 L'offre individuelle

Imaginons que nous connaissions les coûts auxquels font face les différents producteurs d'un bien. À partir de quel prix pour ce bien est-il économiquement viable de commencer la production? Pour un prix donné, quelle quantité du bien chaque entreprise décidera-t-elle de produire?

Commençons par nous intéresser à un seul producteur. Typiquement, ses coûts de production se composent de coûts fixes qui sont nécessaires au démarrage de la production, et de coûts variables qui augmentent avec la quantité produite. Ainsi, dans son article de 1844, et concernant l'exemple des routes, Jules Dupuit écrit :

Les frais de production se composent de deux parties : l'une très dispendieuse faite une fois pour toutes ou pour un très grand nombre de fois, l'autre peu dispendieuse faite pour chaque objet.

Notons la quantité produite par l'entreprise q (un nombre qui sera positif ou nul), et notons $C(q)$ le coût de produire q unités du bien, qui est mesurée en montant monétaire (comme ¥, €). Si l'entreprise décide de ne pas produire du tout, ses coûts sont nuls, mais la présence de coûts fixes C_f dès que $q > 0$ est représentée par le fait que $C_f = \lim_{q \rightarrow 0^+} C(q) > 0$. Les coûts augmentent avec les quantités, donc C est croissante.

Une notion importante en sciences économiques est la notion de *coût marginal*. Il s'agit du coût unitaire de la production d'une infime quantité supplémentaire. Si on note $c(q)$ le coût marginal à un niveau de production q , on définit donc :

$$c(q) = \lim_{\epsilon \rightarrow 0} \frac{C(q + \epsilon) - C(q)}{\epsilon}.$$

Quitte à en faire une approximation, on suppose donc la fonction de coûts C dérivable, et sa dérivée $c(q) = C'(q)$ est appelée *fonction de coûts marginaux*. Elle représente le coût de production d'une unité supplémentaire lorsque la production est de q . Elle est mesurée en unité monétaire par unité de bien, comme des €/kg, ou \$/litre. De manière typique, les premières unités du bien sont de moins en moins coûteuses à produire, tandis qu'au delà d'une certaine quantité produite, une augmentation de la production devient de plus en plus coûteuse. On supposera donc les coûts marginaux décroissants jusqu'à une quantité q_0 , puis croissants et tendant vers $+\infty$.

Pourquoi fait-on cette hypothèse de coûts marginaux décroissants puis croissants? Prenez le cas d'une usine qui produit à la chaîne du mobilier. Si l'usine produit peu, une augmentation de la production lui permettra de profiter d'économies d'échelle (plus on produit,

plus on le fait de façon efficace), mais si la production augmente trop, il faudra faire travailler les ouvriers la nuit (ce qui coûte plus cher), les matières premières manqueront et il faudra payer de plus en plus cher pour les acquérir, etc. A l'échelle d'une grande entreprise, des phénomènes similaires se produisent : lorsqu'une entreprise acquiert plus d'usines, elle peut économiser des coûts sur certaines fonctions transversales comme la logistique, mais si l'entreprise grossit trop, elle se heurtera également à une hausse des coûts marginaux : le management devient plus difficile, il devient plus difficile de restructurer l'entreprise lorsqu'elle fait face à des chocs, etc.

Si on fait cette hypothèse de coûts marginaux décroissants puis croissants, que peut-on en déduire sur le comportement de l'entreprise? À prix donné p , le revenu associé à la production d'une quantité q est de pq , et les coûts sont égaux à $C(q)$. La fonction de profit de l'entreprise est donc donnée par $\Pi(p, q) = pq - C(q)$. Le producteur choisira une quantité q^* telle que :

$$q^* = \arg \max_{q \geq 0} \Pi(p, q) = \arg \max_{q \geq 0} [pq - C(q)].$$

Il s'agit là de l'application de l'hypothèse de rationalité au comportement d'une entreprise. Une entreprise rationnelle effectue ses choix de production de manière à maximiser son profit!

On décompose le choix de production de l'entreprise en deux parties. La première est la décision d'ouvrir ou non la production. Ne pas ouvrir revient à produire une quantité nulle, et ouvrir à produire une quantité strictement positive. La seconde décision, si la firme ouvre, est celle de la quantité positive effectivement produite. Afin d'analyser la décision d'ouverture, l'entreprise a besoin d'estimer quel serait son profit en cas d'ouverture. Nous commençons donc par analyser le choix de la quantité produite en supposant que la firme ouvre, puis dans une seconde étape nous porterons attention au choix d'ouvrir ou non.

Analyse de la quantité produite si ouverture : coûts marginaux

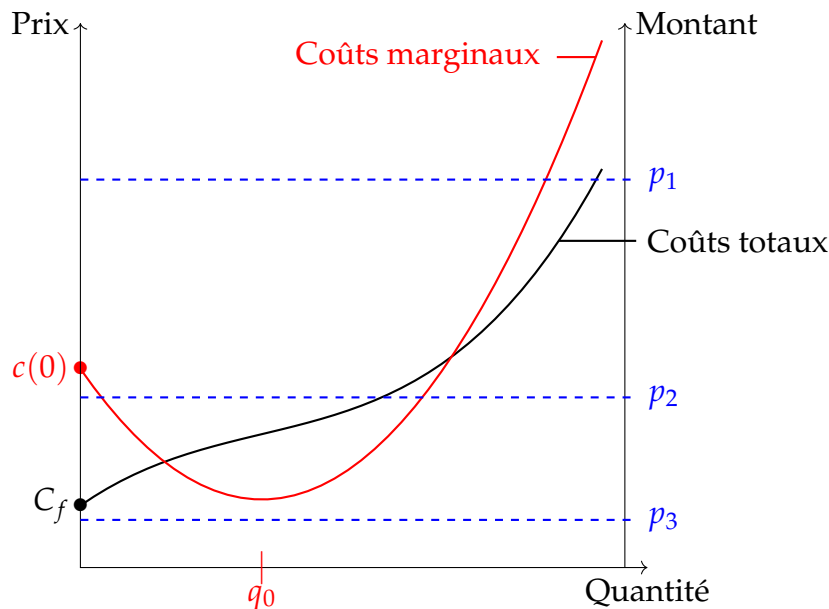
Commençons donc par supposer $q^* > 0$. Comme premier résultat, nous allons voir que la quantité produite choisie q^* égalise les coûts marginaux et le prix : $c(q^*) = p$. Mathématiquement, cela provient du fait qu'à un point q^* qui maximise le profit, la dérivée du profit par rapport à la quantité et qui vaut $p - c(q)$ doit être nulle. Par raisonnement économique, on observe qu'une quantité produite q telle que $c(q) < p$ ne saurait être optimale, car une augmentation infinitésimale de la production rapporterait plus par unité produite (p) qu'elle ne coûterait ($c(q)$). De même, pour une quantité $q > 0$ telle que $c(q) > p$, une diminution infinitésimale de la production permettrait d'économiser plus en coût ($c(q)$) qu'elle ne coûterait en revenus (p).

Ainsi, une entreprise qui maximise son profit raisonne à la marge : est-il profitable de produire un peu plus ou un peu moins? A l'optimum, la quantité choisie est telle qu'il n'est ni souhaitable d'augmenter, ni de diminuer, la production.

Étant donné que les coûts marginaux sont décroissants, puis croissants tendant vers $+\infty$, nous pouvons être dans 3 cas possibles, illustrés figure 2.4.

- **Cas 1** : Pour un prix p supérieur à $c(0)$, il existe une valeur unique q^+ de q pour laquelle $c(q^+) = p$ (voir le prix p_1 sur la figure ci-dessous);
- **Cas 2** : Pour un prix compris entre $c(0)$ et le minimum $c(q_0)$ des coûts marginaux, il existe deux valeurs $q^- \leq q^+$ de q pour lesquelles $c(q^-) = c(q^+) = p$ (voir le prix p_2 sur la figure)¹;
- **Cas 3** : Pour un prix inférieur au minimum $c(q_0)$ des coûts marginaux, il n'existe pas de valeur de q pour laquelle $c(q) = p$ (voir le prix p_3 sur la figure).

FIGURE 2.4 – Exemple de fonction de coûts totaux (en noir, mesurés en montants, échelle de droite) et de coûts marginaux (en rouge, mesuré dans la même unité que les prix, échelle de gauche). Chaque ligne pointillée bleue représente une valeur de p . Selon la valeur de p , il existe soit aucune, soit une, soit deux valeurs de q pour lesquelles $c(q) = p$.



Lorsqu'elles existent, les solutions q^- et q^+ se situent respectivement sur la partie décroissante et sur la partie croissante de la fonction de coûts marginaux. Dans ce cas, l'entreprise va-t-elle choisir de produire q^- ou q^+ ? Considérons une augmentation de la production q de q^- à q^+ : chaque unité supplémentaire est produite à un coût marginal $c(q) < p$ et rapporte p . Par conséquent les profits sont supérieurs en q^+ à ce qu'ils sont en q^- . La production se fera donc à q^+ . Plus généralement, si la quantité optimale à produire n'est pas 0, on a $q^* > q_0$. Intuitivement, dès lors que la courbe de coût marginal est décroissante, chaque unité supplémentaire est plus rentable que la précédente. De manière générale, pour une fonction de coûts quelconque, on a donc la propriété suivante :

Si la quantité optimale produite n'est pas 0, elle se situe nécessairement sur une partie croissante de la courbe de coûts marginaux.

1. Dans le cas où $p = c(q_0)$, les deux solutions sont confondues : $q^- = q^+ = q_0$.

Analyse de la décision d'ouverture : coûts moyens

Maintenant que nous avons déterminé la quantité qu'il est optimal de produire en cas d'ouverture de la firme, nous pouvons, en comparant les profits en ce point et en cas de non ouverture, déterminer s'il est optimal pour la firme d'ouvrir ou non. Étant donné que le profit en cas de non ouverture est nul, il s'agit simplement de déterminer si le profit en cas d'ouverture est positif ou non.

En cas d'ouverture, il est optimal de produire la plus grande solution notée q^+ de $c(q) = p$. On réécrit le profit de l'entreprise en q^+ sous la forme

$$\Pi(p, q^+) = pq^+ - C(q^+) = q^+(c(q^+) - \frac{C(q^+)}{q^+}),$$

où le terme $C(q^+)/q^+$ correspond au *coût moyen par unité produite* jusqu'à q^+ . Par conséquent, en notant $\bar{C}(q) = C(q)/q$ la fonction de coûts moyens :

Il est optimal d'ouvrir la firme si en q^+ , les coûts marginaux excèdent les coûts moyens. La *condition d'ouverture* d'un producteur s'écrit donc :

$$c(q^+) \geq \bar{C}(q^+)$$

Pour comprendre si une firme produit ou non, il faut donc s'intéresser à ses coûts moyens et les comparer aux coûts marginaux. Pour des valeurs très petites de q , les coûts moyens sont très élevés car $C_f > 0$, et ils excèdent les coûts marginaux.

De même, pour toute valeur de q inférieure ou égale à q_0 , les coûts moyens sont obtenus comme une moyenne des coûts fixes et des coûts marginaux entre 0 et q , et sont donc supérieurs aux coûts marginaux.²

On remarque aussi que si les coûts marginaux sont inférieurs aux coûts moyens, augmenter la quantité produite diminue les coûts moyens. Symétriquement, si les coûts marginaux sont supérieurs aux coûts moyens, augmenter la quantité augmente les coûts moyens³ :

$$\bar{C}'(q) > 0 \iff c(q) > \bar{C}(q).$$

On peut montrer⁴ qu'il existe une unique valeur \bar{q} telle que $\bar{C}(\bar{q}) = c(\bar{q})$. Les coûts marginaux sont inférieurs aux coûts moyens pour des quantités inférieures à \bar{q} , et ils sont supérieurs aux coûts moyens pour des quantités supérieures à \bar{q} . On en déduit que la firme fait les choix suivants :

2. Mathématiquement, pour $q \leq q_0$, $\bar{C}(q) = \frac{1}{q}(C_f + \int_0^q c(x)dx) > \frac{1}{q} \int_0^q c(q)dx > c(q)$.

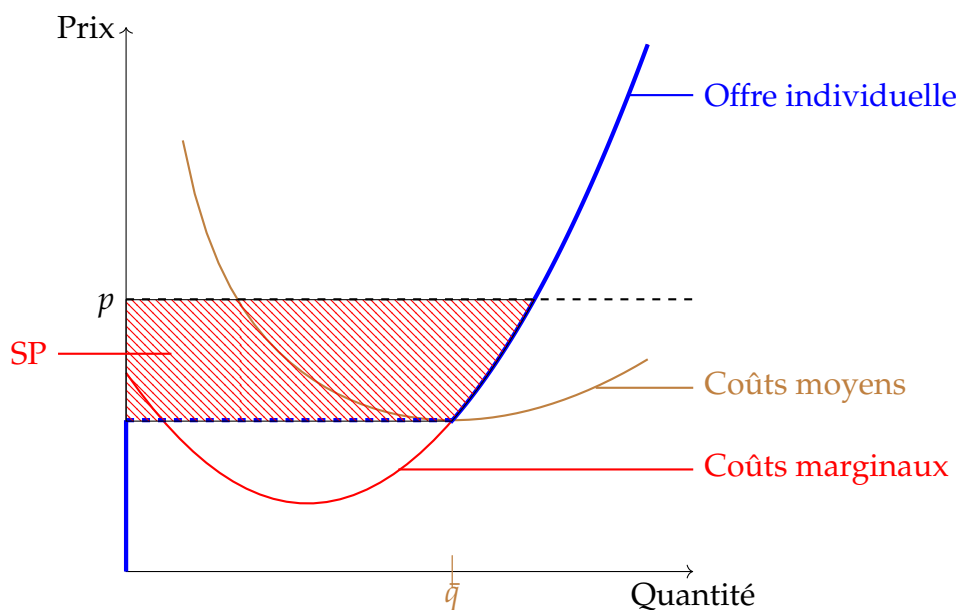
3. On remarque que $(q\bar{C})' = q\bar{C}' + \bar{C}$, et donc \bar{C}' est du signe de $(q\bar{C})' - \bar{C} = c - \bar{C}$.

4. Preuve : Pour des quantités supérieures à q_0 et telles que $\bar{C} > c$, \bar{C} est décroissante tandis que c est croissante. Étant donné que c tend vers $+\infty$ et que les fonctions sont continues, il existe un point d'intersection \bar{q} , solution de $\bar{C}(q) = c(q)$. Peut-on avoir plusieurs solutions à $\bar{C}(q) = c(q)$? Si on note \bar{q}_1 la plus basse des solutions alors $\bar{q}_1 > q_0$ car nous avons vu que $\bar{C}(q) > c(q)$ pour $q \leq q_0$. On a donc $\bar{C}'(\bar{q}_1) = 0$ et $c'(\bar{q}_1) > 0$ ce qui implique que $\bar{C} < c$ dans un voisinage à droite de \bar{q}_1 et $\bar{C} > c$ dans un voisinage à gauche. Supposons qu'il existe une seconde solution plus basse notée \bar{q}_2 . Pour $\bar{q}_1 < q < \bar{q}_2$, nous avons donc $\bar{C}(q) < c(q)$, ce qui est contradictoire avec le fait que $\bar{C}(q) > c(q)$ dans un voisinage à gauche de \bar{q}_2 .

- Lorsque le prix est inférieur à $c(\bar{q})$, le prix est inférieur au coût moyen quelle que soit la quantité produite, la firme ne peut donc jamais faire de profit : elle n'ouvre pas et son offre est donc nulle.
- Lorsque le prix est supérieur à $c(\bar{q})$, la firme produit une quantité qui égalise les coûts marginaux et le prix, et l'offre est donc donnée par la courbe de coûts marginaux.

Ce raisonnement est illustré sur la figure 2.5. En particulier, au point $c(\bar{q})$, l'offre est discontinue et passe de 0 à \bar{q} . Notez également que l'offre est une fonction croissante des prix. En résumé, la courbe d'offre individuelle d'une entreprise correspond à la courbe de coût marginal dès lors qu'il est profitable de produire.

FIGURE 2.5 – La fonction d'offre individuelle



Surplus d'un producteur

Lorsque la firme fait le choix de production optimal, elle dégage un profit aussi appelé *surplus du producteur* :

$$SP(p) = \Pi(p, q^*(p)),$$

où $q^*(p)$ est la quantité optimale produite lorsque le prix est p . Lorsque le prix est inférieur à $c(\bar{q})$, le surplus est nul car il est optimal de ne pas produire. En $p = c(\bar{q})$ le surplus est aussi nul, et la firme est indifférente entre produire ou non. Lorsque le prix augmente au delà de $c(\bar{q})$, chaque augmentation du prix de p à $p + dp$ a deux effets sur le surplus : premièrement, une augmentation du profit sur les $q^*(p)$ unités déjà produites, d'un montant de $q^*(p)dp$; deuxièmement une augmentation de la production de $q^*(p)$ à $q^*(p) + dq^*(p)$, mais étant donné que ces unités sont produites à coût marginal égal à p , cette augmentation de prix n'a pas d'impact net sur le surplus (l'augmentation des bénéfices de $p \times dq^*(p)$ étant compensée exactement par l'augmentation des coûts de $c(q^*) \times dq^*(p)$). Chaque augmentation de dp du prix conduit donc à une augmentation $q^*(p)dp$ du surplus. Cela nous permet d'établir la

relation suivante⁵ :

$$SP(p) = \int_0^p q^*(p) dp.$$

Cette dernière relation a une interprétation graphique simple et utile : le surplus du producteur est représenté par l'aire située au dessus de sa courbe de production, et sous le prix p , voir Figure 2.5.

2.2.2 La courbe d'offre

Offre et surplus agrégés

Considérons désormais qu'il n'y a pas un seul producteur, mais un ensemble J de producteurs. L'offre totale des producteurs est

$$S(p) = \sum_j q_j^*(p),$$

où q_j^* est l'offre du producteur j . Étant donné que chaque demande individuelle est une fonction croissante du prix, on a la "loi de l'offre" :

Loi de l'offre : l'offre agrégée est une fonction croissante du prix.

Par ailleurs, le surplus de l'ensemble des producteurs est :

$$SP(p) = \sum_j SP_j(p) = \sum_j \int_0^p q_j^*(x) dx = \int_0^p S(x) dx.$$

Le surplus des producteurs est représenté par l'aire de la surface située sous le prix et au dessus de la fonction d'offre agrégée. Pour le connaître, il suffit donc de connaître l'offre agrégée, et non chacune des offres individuelles, ce qui est empiriquement beaucoup plus aisé.

Déterminants de l'offre

Nous avons vu comment dériver et analyser l'offre des producteurs en fonction de leurs fonctions de coût. De manière pratique, de nombreux facteurs peuvent modifier l'offre des producteurs :

- Le prix des intrants. Par exemple, si le prix du cuivre augmente, la courbe d'offre de câbles électriques à haute tension est déplacée vers le haut.
- La technologie. Ainsi, l'introduction d'engrais et de pesticides augmente l'offre de blé, déplaçant la courbe d'offre vers le bas. Les effets de courbe d'expérience que nous avons décrits plus haut contribuent également à abaisser la courbe de demande.

5. De manière mathématique, on note que $SP(p) = pq^*(p) - C(q^*(p))$ et donc $SP'(p) = q^*(p) + pq'^*(p) - c(q^*(p))q'^*(p) = q^*(p)$ car $p = c(q^*(p))$, et que par ailleurs $SP(0) = SP(c(\bar{q})) = 0$, ce qui donne la même relation.

- Le nombre de producteurs. Ainsi, la fermeture de puits de pétrole (causée par un conflit ou un boycott par exemple) en diminue l'offre.
- Les anticipations des producteurs, par exemple lorsqu'une firme qui anticipe une croissance du marché peut augmenter sa capacité de production en anticipation.

2.2.3 Élasticité-prix de l'offre

Dans les faits, il est souvent très difficile d'estimer la fonction de coût d'une firme, encore plus de l'ensemble des firmes de l'économie. Néanmoins, il est souvent suffisant d'avoir une approximation locale de cette fonction d'offre : comme nous l'avons vu précédemment avec la demande, cette approximation est donnée par une élasticité-prix. L'élasticité-prix de l'offre se définit naturellement comme la variation relative de l'offre suite à une variation relative du prix :

$$e_S = \frac{\text{variation en \% de l'offre}}{\text{variation du prix de 1\%}} = \frac{\frac{dS}{S}}{\frac{dp}{p}} = \frac{d \log S}{d \log p}$$

Où S est la fonction d'offre. Comme pour la demande, plusieurs paramètres déterminent l'élasticité-prix de l'offre. En voici quelques exemples :

- **L'horizon temporel** : comme l'ajustement des capacités de production prend du temps, l'offre de court-terme peut rester assez stable face à des augmentations de prix. Néanmoins, une hausse durable des prix entraîne à long terme une augmentation des capacités de production : ainsi, l'offre est typiquement plus élastique à long terme qu'à court terme.
- **À court terme, l'utilisation des facteurs de production** : lorsque ceux-ci sont sous-utilisés (par exemple lorsque des machines sont à l'arrêt), l'offre est plus élastique.
- **Les capacités de stockage** : lorsqu'une entreprise peut stocker un produit, elle pourra typiquement en offrir rapidement plus lorsque les prix augmentent, en diminuant ses stocks. Certains biens, comme l'électricité, sont difficilement conservables. Ceci explique que dans certains cas, l'offre ne s'ajuste que peu à une variation des prix, conduisant à des prix plus volatiles. Ainsi, des prix nuls (voire négatifs) des carburants pour avions ont été observés de manière momentanée en 2020, suite à la mise à l'arrêt du trafic aérien mondial consécutive au début de l'épidémie de Covid-19.
- **La mobilité des facteurs de production** : nous avons vu que l'offre était plus élastique lorsque les capacités de production pouvaient s'ajuster rapidement. Cette facilité d'ajustement est fonction de la mobilité des facteurs de production. Par exemple, face à une variation des prix de denrées agricoles, il sera possible pour un agriculteur de changer d'une année sur l'autre les variétés qu'il exploite, rendant l'offre élastique. Au contraire, si l'on s'intéresse au marché du travail, l'élasticité-prix de l'offre sera plus faible. Il est en effet souvent difficile pour des travailleurs de basculer d'un secteur vers un autre (à cause de coûts de formation par exemple), ou de changer de région géographique.

- **Les conditions de concurrence** : la libre entrée des producteurs sur un marché est un facteur d'augmentation de l'élasticité. Lorsqu'une augmentation des prix rend les activités de production profitables, de nouveaux producteurs entrent sur le marché, augmentant ainsi l'offre.

Exercices des chapitres 1 et 2

Exercice 2.1. Propension totale et marginale à payer

Lors du cours, nous avons supposée la propension marginale à payer donnée. Comment peut-on mesurer la propension marginale à payer des consommateurs ?

Supposons qu'il est possible de mettre un consommateur devant des choix binaires du type "acheter X tablette de chocolat à un montant total M ou ne rien acheter du tout". Dans ce cadre, il est possible de mesurer le montant M_X maximal qu'un consommateur est prêt à payer pour X unités du bien.

Prenons l'exemple des tablettes de chocolat, supposons que $M_1 = 4$, $M_2 = 6$, $M_3 = 7$.

1. Quelle est la propension marginale à payer pour passer de 0 à 1 unité, de 1 à 2, de 2 à 3 ?
2. Quelle est la fonction de demande du consommateur ?

Exercice 2.2. Préférences et propension totale à payer

Dans le cours, nous avons vu qu'à prix p , un consommateur dont la fonction de propension marginale à payer est $u_i(\cdot)$ choisit de consommer une quantité q qui maximise

$$U_i(q) - pq.$$

où $U_i(q) = \int_0^q u_i(x)dx$ est la propension totale à payer du consommateur.

- Montrer par un raisonnement que le consommateur est indifférent entre recevoir q unités du bien en payant $U_i(q)$ et ne rien recevoir et ne rien payer
- Montrer que pour des valeurs quelconques de q_1, q_2, t_1, t_2 , le consommateur préfère recevoir q_1 unités et payer t_1 plutôt que recevoir q_2 unités et payer t_2 si et seulement si :

$$U_i(q_1) - t_1 > U_i(q_2) - t_2.$$

Exercice 2.3. Croissance et décroissance de la propension marginale à payer

Lors du cours, on a supposé en cours que la propension marginale à payer pour un bien est décroissante avec la quantité de ce bien.

1. Donner deux exemples de biens pour lesquels cette hypothèse est raisonnable
2. Donner deux exemples de biens pour laquelle elle ne l'est pas forcément
3. Considérons un bien divisible (farine, pétrole, ...) et supposons la propension marginale à payer $m(q)$ dérivable en la quantité q du bien. Montrer que si à prix p , la

demande du consommateur vaut $d = D(p)$, alors m ne peut être croissante au point $d(p)$

Exercice 2.4. Surplus d'un projet public

Considérons un projet de construction portuaire d'un système de grutage. Ce système permettrait d'améliorer le débarquement des marchandises et de baisser le coût de l'ordre de 10 € par tonne de marchandises débarquées. On suppose que ce gain est entièrement reporté vers les consommateurs.

1. Le port débarque actuellement 30 000 tonnes de marchandises par an. En supposant que pour des raisons de contraintes capacités du port cette quantité reste inchangée après construction du système de grutage, quel est le surplus généré par cette construction pour les consommateurs ?
2. Dans le cas où la quantité de marchandises transitant par le port peut s'adapter à la hausse ou à la baisse, que pouvez-vous prédire concernant la quantité de marchandises après construction du système de grutage ?
3. Que pouvez-vous dire du surplus réellement généré en comparaison de l'estimation obtenue en question 1) ?
4. On observe que quantité de marchandises transitant par le port subit une hausse imputable à la baisse des coûts et que la nouvelle quantité est de 36 000 tonnes par an. Pouvez-vous estimer le surplus réellement généré par le nouveau système de grutage ? On pourra faire une approximation linéaire de la demande en fonction du prix si nécessaire.

Exercice 2.5. Création d'un marché unique

Imaginons une situation dans laquelle la France et l'Allemagne ont des marchés séparés pour les pommes de terre, si bien que le prix pour le tubercule est différent dans les deux pays : $p_F = 1,5\text{€}/\text{kg}$ en France et $p_A = 1\text{€}/\text{kg}$ en Allemagne. Supposons que la quantité consommée annuellement est 200 000 tonnes en France et 400 000 tonnes en Allemagne.

L'Union Européenne impose la création du marché unique de la patate par lequel les pommes de terre circulent librement entre les deux pays. Suite à ce marché les pommes de terre s'échangent à un prix du kilo p_U égal dans les deux pays.

1. En supposant que suite au marché unique de la patate, 600 000 tonnes de pommes de terre se répartissent entre France et Allemagne, que pouvez-vous dire de p_U comparé à p_F et à p_A ?
2. Suite au marché unique, quel pays augmente sa consommation et lequel la baisse ? Certains consommateurs sont-ils gagnants et d'autres perdants ?
3. Estimer p_U ainsi que le gain ou la perte de surplus pour les consommateurs allemands, pour les consommateurs français, et pour l'ensemble des consommateurs. On pourra pour cela supposer une élasticité-prix de la demande égale à 0.3 dans chacun des pays.

Si vous deviez conseiller l'Union Européenne, diriez-vous qu'il faut créer le marché unique de la patate ?

Exercice 2.6. Étude de document : élasticité-prix du carburant.

Cet exercice est basé sur le rapport du commissariat général au développement durable que l'on peut trouver sur la page moodle du cours.

On demande de lire la conclusion du rapport en page 29, et de répondre aux questions suivantes :

1. Pourquoi le commissariat général au développement durable s'intéresse-t-il à l'élasticité-prix des carburants automobiles ?
2. Selon le rapport, quelles sont les estimations des élasticités-prix de court terme et de long terme ?
3. Pourquoi ces élasticités diffèrent-elles selon vous ?
4. Comment l'élasticité-prix de long terme varie-t-elle selon les revenus ou la situation géographique des ménages ? Comment l'expliquez-vous ?
5. A la lumière du mouvement social dit des "gilets-jaunes", comment interprétez-vous le dernier paragraphe ?

Exercice 2.7. Élasticités-prix

Ranger par élasticité-prix décroissante de la demande

- Insuline
- Pain
- Places de cinéma
- Tabac

Exercice 2.8. Offre de court versus long terme

Pour une pizzeria, les coûts sont donnés comme suit :

- Location matériel : 40€ /jour
- Matière première : 2€ par pizza.
- Main d'œuvre à la journée : 30€ /jour par employé.

En fonction du nombre d'employés on a la production maximale de pizzas suivante :

Employés	Pizzas
0	0
1	10
2	32
3	47
4	55
5	60
6	63
7	65

1. Décrire la fonction de coûts totaux de la pizzeria.

2. Décrire l'offre de pizzas selon la méthode du cours. Cette offre est aussi appelée offre de long terme.

Nous supposons maintenant que nous avons ouvert la pizzeria, avec un bail de location de matériel d'un an *avant* d'observer le prix de marché. Ces coûts de location sont irrécupérables, et nous sommes obligés de payer cette location pendant un an.

3. Décrire l'offre de pizzas en fonction du prix dans ce cas, aussi appelée offre de court terme.
4. Comparer l'offre de court terme et l'offre de long terme. En pratique, que se passera-t-il si un choc temporaire induit une baisse des prix de vente des pizzas ? Et si cette baisse est durable ?

Exercice 2.9. Coca-Zero, Coca light

1. Karine aime le coca light et le coca zero, et ne fait pas de différence entre ces deux types de boissons. Soit q_l, q_z les quantités de coca light et de coca zero consommées. Représenter dans \mathbb{R}_+^2 un ensemble de paires de quantités (q_l, q_z) entre lesquels Karine est indifférente, aussi appelé courbe d'indifférence.
2. Supposons que Karine a un budget de 3€ et que le prix du coca zero est de 1€. En fonction du prix du coca light, quelle est sa demande de coca light ? de coca zero ?
3. Coca zero et coca light sont-ils substitués ou compléments ?

Exercice 2.10. Gin-Tonic

Jeanne prépare du Gin-Tonic pour ses invités. Pour un bon Gin-Tonic, il faut deux unités de Tonic pour une unité de Gin. Soient (q_G, q_T) les quantités de Gin et de Tonic à disposition.

1. Quelle est la quantité de Gin-Tonic produite à partir de (q_G, q_T) ?
2. Sachant que Jeanne est uniquement intéressée par la quantité de Gin-Tonic produite, représenter sur un graphe dans \mathbb{R}_+^2 un ensemble de paires (q_G, q_T) entre lesquelles Jeanne est indifférente.
3. Si Jeanne a un budget de 20€, que le prix d'une unité de Tonic est 1€ et celui d'une unité de Gin est de 4€, quelle seront les quantités de Gin et de Tonic achetées par Jeanne ? Il est conseillé de s'aider d'une représentation graphique comme dans l'exercice précédent.
4. Fixons maintenant le prix du Tonic à 1€ l'unité. En fonction du prix du Gin, dériver la demande de Gin et la demande de Tonic.
5. Le Gin et le Tonic sont-ils des substitués ou des compléments ?

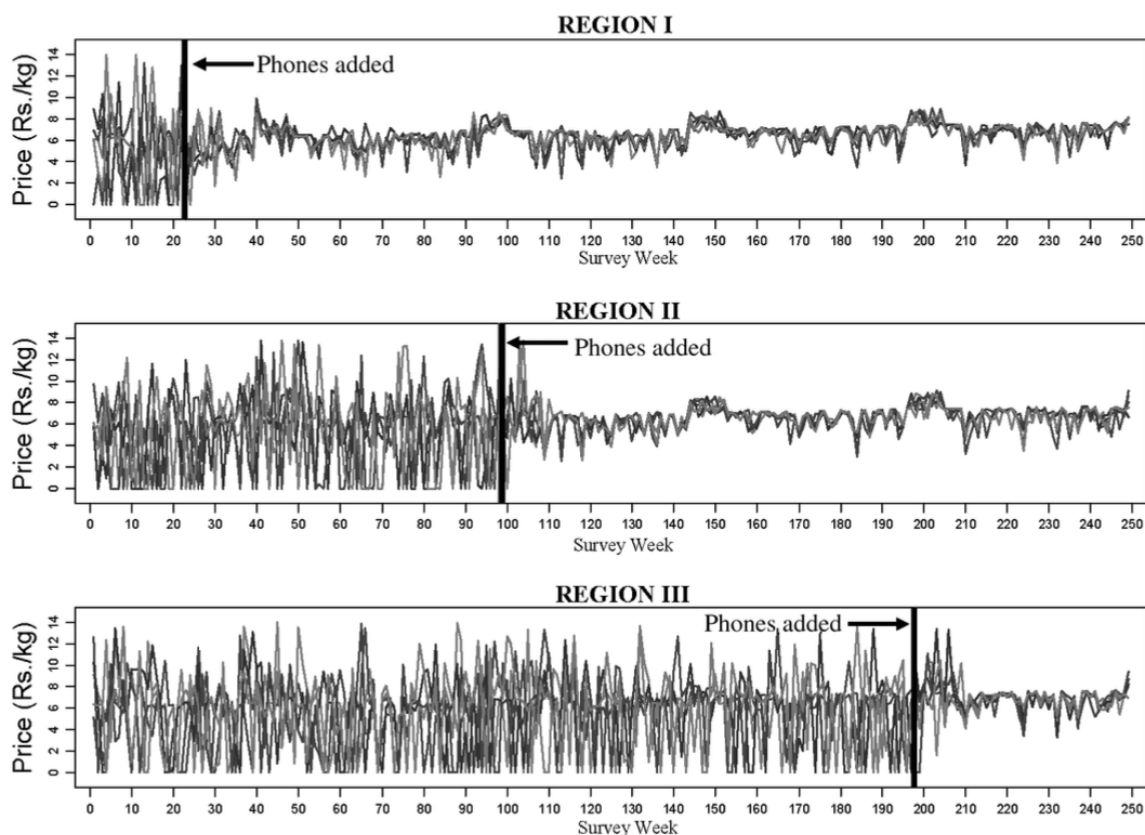
Chapitre 3

Marchés et équilibres de marché

Le Kerala est un État au Sud-Ouest de l'Inde qui compte un grand nombre de pêcheurs. En 1997, des premières antennes de téléphonie mobile commencèrent à y être construites. Les pêcheurs s'équipèrent rapidement de téléphones portables, et cette innovation leur a été particulièrement utile. Chaque matin, entre 5h et 8h, les plages du Kerala deviennent des marchés au poisson dans lesquels les pêcheurs peuvent écouler leurs prises. À la fin de chaque nuit, les pêcheurs doivent décider vers quelle plage aller pour la vente. Cette décision est importante car les plages sont assez éloignées les unes des autres et le marché ne dure pas longtemps : une fois une plage choisie, les pêcheurs doivent s'y tenir. Un pêcheur arrivant à une plage où de nombreux autres pêcheurs sont déjà présents court le risque de ne pas pouvoir vendre l'intégralité de sa prise. Avant l'arrivée du téléphone portable, de 5 à 8% des poissons pêchés étaient jetés faute d'acheteurs. La téléphonie mobile allait changer la donne : en communiquant avec la côte, les pêcheurs avaient accès chaque jour à plus d'information pour sélectionner un marché. Quel a été l'impact économique de ce progrès technologique ? L'économiste Robert Jensen et son équipe de recherche ont réussi à mesurer les prix du poisson proposés sur différentes plages du Kerala chaque semaine entre 1996 et 2001. Les antennes de téléphonie mobile ont été installées à différents endroits à des moments différents : la figure 3.1 montre l'évolution des prix du poisson dans trois régions ainsi que la date d'arrivée de la téléphonie mobile.

Les résultats sont frappants : avant l'arrivée des téléphones mobiles, il y avait une très grande volatilité des prix du poisson. Lorsque les pêcheurs arrivaient trop nombreux sur une plage (tellement nombreux qu'une partie de la production était jetée), les prix étaient très faibles – une bonne affaire pour les clients locaux. Parfois au contraire, très peu de pêcheurs se rendaient sur une plage, et leur prises se vendaient à un prix élevé. Une fois les antennes téléphoniques mises en place, tout changea très vite : en quelques semaines, la majorité des pêcheurs s'est équipée de téléphones mobiles, et pouvait désormais mieux choisir un point de vente. La volatilité des prix a chuté presque du jour au lendemain, le prix du poisson est devenu stable. Qui a profité de ce changement ? Jensen a mesuré qu'après la mise en place du réseau mobile, le gaspillage est devenu presque inexistant : tous les poissons (ou presque) trouvaient un acheteur. Faisant face à moins d'incertitude, les pêcheurs ont pu baisser leurs prix (de 4%), profitant aux habitants des côtes. Malgré cette baisse des

FIGURE 3.1 – Évolution du prix hebdomadaire des sardines sur les plages du Kerala entre 1996 et 2001, dans trois régions ayant installé des antennes téléphoniques. Les barres verticales indiquent le moment où chaque région a eu accès à la téléphonie mobile. Données de Jensen (2007).



prix, les pêcheurs n'ont pas été perdants : leurs profits moyens ont augmenté de 8%. Mais la mise en place du téléphone n'a pas eu un impact que sur les pêcheurs. Dans un autre article de recherche¹, Robert Jensen et Nolan Miller montrent que ce changement a également changé la donne pour les constructeurs de bateaux : lorsque l'échange d'information est devenu plus simple, des constructeurs de bateaux de meilleure qualité ont commencé à embaucher plus d'ouvriers et à construire plus de navires, tandis que les constructeurs de moins bonne qualité ont plus souvent fermé boutique. Dans l'ensemble, le prix des bateaux a également diminué suite à l'introduction des antennes téléphoniques.

En permettant la transmission d'information, la téléphonie mobile a eu un impact économique fort : un prix stable du poisson est apparu, presque identique dans les différentes régions du Kerala. On peut considérer que le téléphone portable a permis de créer un marché du poisson à peu près unifié dans l'ensemble de la région.

Les marchés sont un concept central dans l'étude des sciences économiques. On parle de marché lorsque l'offre et de la demande d'un même bien se rencontrent de façon décentralisée, comme c'est le cas au Kerala. Chaque consommateur et chaque producteur ont la possibilité d'échanger le bien, pour un prix qui leur convient. En pratique, les biens considérés peuvent être aussi divers que les matières premières, les actifs financiers, les biens immobi-

1. Voir Jensen and Miller (2018).

liers, ou encore les heures de travail. Parce qu'ils requièrent une intervention minimale de l'État (qui doit cependant être garant du respect des contrats et des droits de propriété), ils sont un bon point de départ pour comprendre la mise en relation de l'offre d'un bien avec sa demande.

Dans ce chapitre, nous allons décrire la notion de marché, une abstraction qui permet d'analyser de manière simple des interactions dans une économie. Un marché est caractérisé par une demande, une offre, et un équilibre qui s'établit entre les deux. Nous allons utiliser la notion de marché dans plusieurs cadres dans ce cours, mais le plus simple est d'analyser les échanges d'un bien. La fonction de demande indique, en fonction du prix du bien, la quantité demandée par un consommateur ou un groupe de consommateurs. Elle est la résultante des choix effectués par les consommateurs, et donc de leurs préférences. De manière symétrique, la fonction d'offre indique, en fonction du prix, les quantités du bien qui seront mises sur le marché par un producteur, ou par l'ensemble des producteurs. Elle s'obtient à partir des fonctions de coût des producteurs et de leur recherche de la maximisation du profit. A l'équilibre, les échanges du bien se font à un certain prix, les producteurs vendent une certaine quantité du bien qui est achetée par les consommateurs. Une question se pose alors : s'il n'y a pas de régulation du marché, cette quantité est-elle une "bonne" quantité? Nous verrons que selon la situation, l'équilibre de marché est satisfaisant ou au contraire sous-optimal. Dans le cas où le résultat en situation de laissez-faire est insatisfaisant, nous allons étudier à plusieurs reprises dans ce cours comment l'État peut intervenir pour améliorer l'équilibre.

3.1 L'équilibre de marché

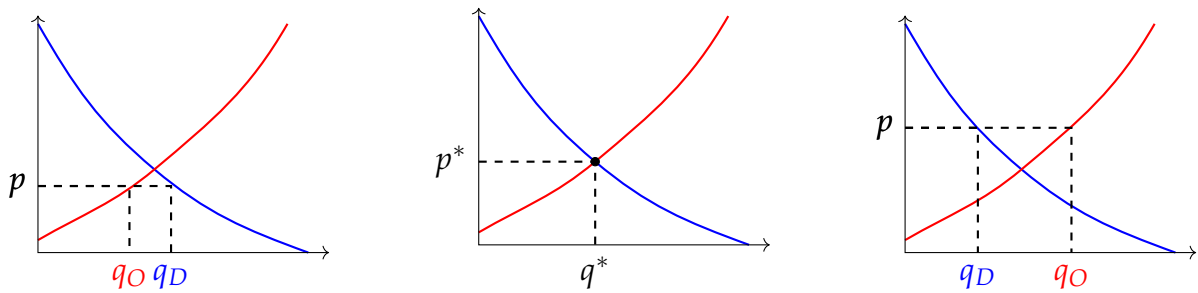
Partons de l'hypothèse de consommateurs et de producteurs librement mis en relation et pouvant s'échanger des quantités du bien contre de l'argent à un prix négocié entre eux. Supposons qu'un consommateur se fournisse auprès d'un producteur à un prix p_1 , tandis qu'un autre consommateur se fournit à un prix p_2 auprès d'un second producteur. Si p_1 est inférieur à p_2 , le second consommateur aurait intérêt à proposer un peu plus que p_1 au premier producteur, ce qui serait bénéficiaire aux deux. Le même raisonnement montre qu'on ne saurait avoir $p_2 < p_1$, et par conséquent $p_1 = p_2$.

Loi du prix unique : Dans un marché sans frictions, tous les échanges d'un même bien se font au même prix.

L'exemple du marché du poisson au Kerala nous fournit une très bonne illustration de la loi du prix unique. Avant l'arrivée du téléphone portable, il y avait plusieurs marchés qui avaient chacun un prix différent. Le téléphone a unifié ces marchés, ce qui a conduit à la convergence des prix entre ces différents marchés.

Étant donnée la loi du prix unique, à quel(s) prix peuvent se faire les échanges? Les fonctions d'offre et de demande nous permettent de répondre à cette question. Lorsque pour un prix d'échange donné, la demande est supérieure à l'offre on parle de *pénurie* du bien.

FIGURE 3.2 – Courbe d’offre en rouge et de demande en bleu. À gauche : situation de pénurie, la quantité demandée q_D est supérieure à la quantité offerte q_O . À droite : situation de surproduction, la quantité demandée q_D est inférieure à la quantité offerte q_O . Au milieu : équilibre de marché, la quantité demandée q_D et la quantité offerte q_O sont égales.



Certains consommateurs sont dans ce cas prêts à proposer un prix plus élevé aux producteurs pour acquérir le bien, ce qui conduit à une augmentation du prix (une inflation). En revanche, lorsque le prix d’échange est tel que l’offre est supérieure à la demande, on dit que nous sommes dans une situation de *surproduction*. Dans ce cas, la trop forte compétition entre producteurs les conduit à baisser leurs prix pour attirer les acheteurs (une déflation).

Le *prix de marché* est déterminé par l’égalisation de l’offre et de la demande. À ce prix, chaque consommateur acquiert la quantité souhaitée du bien, et chaque producteur choisit sa production de manière optimale : il n’y a ni surproduction, ni pénurie. Les trois situations sont illustrées sur le graphe de la figure 3.2.

Au *prix de marché*, la demande et l’offre s’égalisent, et on a une situation d’*équilibre de marché*.

En pratique, le mécanisme de convergence des prix vers le prix d’équilibre dépend de l’organisation exacte du marché. Le mécanisme par lequel une pénurie conduit à une augmentation des prix, et une surproduction à leur diminution, s’appelle le tâtonnement.

3.2 Chocs d’offre et de demande

Dans les pages précédentes, nous avons décrit les deux principaux facteurs qui déterminent le prix d’un bien : la valeur donnée par les consommateurs à ce bien, et le coût de sa production. Ce paradigme d’offre et de demande nous permet de décrire précisément comment ces deux paramètres entrent dans la détermination d’un prix, et comment ceux-ci changent lorsque changent les goûts des consommateurs ou les coûts de production.

Lorsque la valeur donnée à un bien par les consommateurs augmente, la courbe de demande est déplacée vers le haut, ce qui fait augmenter le prix du bien. Si le coût auquel on produit un bien diminue, cela déplace la courbe d’offre vers le bas et fait diminuer le prix.

Les marchés agricoles regorgent d’exemples de tels chocs d’offre ou de demande. Ainsi, le prix de la vanille reflète les événements politiques et climatiques de Madagascar, le pays qui produit le plus de cette épice. En 2004, le pays a été touché par un violent cyclone qui

a détruit une bonne partie des plantations de vanille (ce qu'on peut interpréter comme un choc d'offre négatif), et le prix de la vanille a plus que doublé.

3.3 Propriétés de l'équilibre de marché

Les économistes s'intéressent beaucoup aux propriétés dites d'*efficacité* du marché – c'est-à-dire de savoir si les marchés conduisent à une bonne allocation de la consommation et de la production, ou si au contraire l'issue des marchés est insatisfaisante. Pour faire cela, nous avons besoin de reposer sur des critères dits *normatifs* permettant de juger la désirabilité des allocations de consommation et de production. Nous introduisons deux critères, la maximisation du surplus, et l'efficacité au sens de Pareto.

Le principe des marchés est de laisser acheteurs et vendeurs échanger librement. À l'équilibre de marché, chaque acheteur acquiert autant de biens qu'il l'estime souhaitable au prix de marché. Par conséquent, pour bénéficier à un acheteur, un achat supplémentaire devrait se faire à un prix inférieur au prix de marché. De manière symétrique, pour satisfaire un producteur, une vente supplémentaire devrait se faire à un prix supérieur au prix de marché. Ainsi, à l'équilibre de marché, aucun acheteur et aucun producteur ne souhaitent effectuer de transaction supplémentaire entre eux. Par principe, tous les échanges mutuellement bénéfiques sont réalisés.

Nous formalisons cette intuition dans la section 3.3.1 où nous montrons que, parmi tous les prix possibles, le prix de marché est celui qui maximise le surplus total. Dans la section 3.3.2, nous introduisons le critère d'efficacité de Pareto, selon lequel une allocation des biens est efficace si aucune autre ne pourrait mieux satisfaire chacun des agents économiques. Enfin, dans la section 3.3.5, nous apportons une critique des critères de surplus et de Pareto sous l'angle des inégalités.

3.3.1 Maximisation du surplus

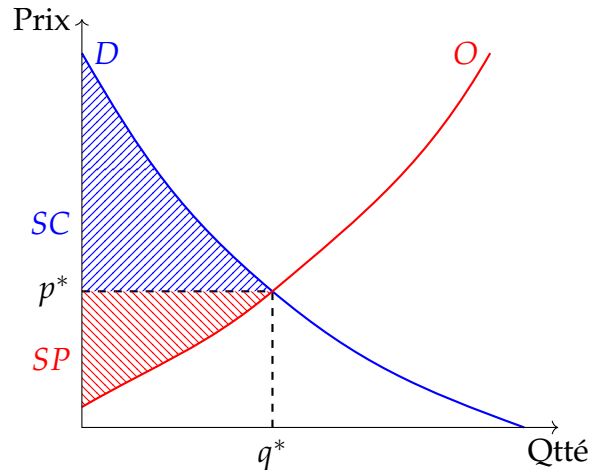
En étudiant les consommateurs et les producteurs, nous avons introduit les notions de surplus des consommateurs SC , et de surplus des producteurs SP . Pour une allocation des biens donnée, le surplus total est défini comme leur somme :

$$ST = SC + SP.$$

Pour montrer que le prix d'équilibre permet de maximiser le surplus total ST , nous allons reposer sur une analyse graphique. À prix d'équilibre p^* et pour une quantité échangée correspondant q^* , le surplus des consommateurs SC est représenté comme l'aire située sous la courbe de demande et au dessus du prix, et le surplus des producteurs SP comme l'aire au dessus de la courbe d'offre et au dessous du prix. Par conséquent, le surplus total est représenté comme l'aire comprise entre la courbe d'offre et la courbe de demande, pour les quantités inférieures à q^* (voir figure 3.3).

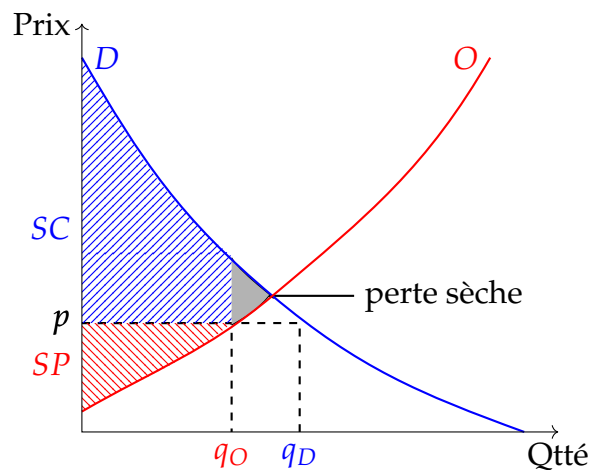
Considérons maintenant une situation de pénurie, c'est à dire un prix p inférieur à p^* .

FIGURE 3.3 – Surplus du consommateur et du producteur au prix d'équilibre de marché



Dans ce cas, la quantité demandée q_D est supérieure à la quantité offerte q_O . La quantité échangée sur le marché est limitée par l'insuffisance de l'offre et vaut q_O . Le surplus des consommateurs est lui aussi limité par la quantité échangée. Si l'on fait l'hypothèse que l'offre satisfait en priorité les consommateurs ayant la propension marginale à payer la plus élevée, on obtient que le surplus des consommateurs est représenté par l'aire située sous la courbe de demande et au-dessus du prix, pour des quantités inférieures à q_O .² Le surplus des producteurs quant à lui est donné par l'aire située au-dessus de la courbe d'offre et sous le prix. Comme représenté sur la figure 3.4, le surplus total est donc donné par l'aire située sous la courbe de demande et sous la courbe d'offre, pour des quantités inférieures à q_O , et est donc inférieur au surplus total à l'équilibre de marché. La différence est appelée *perte sèche*.

FIGURE 3.4 – Surplus et perte sèche (en gris) en situation de pénurie



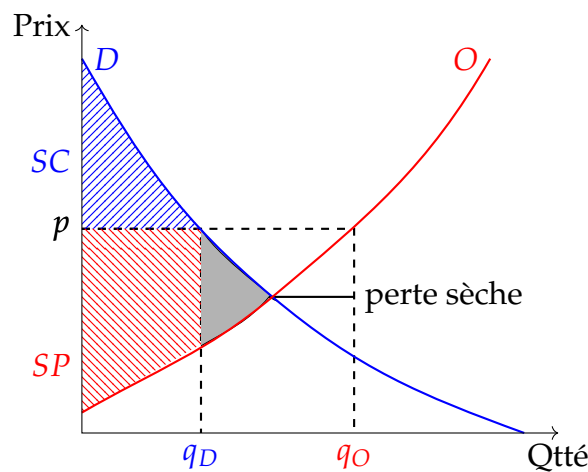
Une situation inverse, dite de surproduction, est représentée en figure 3.5. Pour un prix supérieur au prix d'équilibre de marché, la quantité offerte q_O est inférieure à la quantité demandée q_D . L'insuffisance de la demande limite les échanges, et la quantité effectivement

2. Si cette hypothèse n'est pas vérifiée, le surplus des consommateurs sera inférieur. La méthode proposée conduit donc à une borne supérieure au surplus des consommateurs en situation de pénurie.

échangée est q_D . Le surplus des consommateurs est donné par l'aire située sous la courbe de demande et au dessus du prix, tandis que le surplus des producteurs est représenté par l'aire au dessus de la courbe d'offre et sous le prix, en se limitant aux quantités inférieures à q_D .³ Le surplus total est donc encore une fois inférieur à celui du prix d'équilibre, avec une perte sèche.

Les situations de pénurie et de surproduction ne sont pas économiquement efficaces car des échanges profitables aux producteurs et aux consommateurs n'ont pas lieu. En effet, la partie du graphe située à droite de q_O pour la pénurie ou à droite de q_D pour la surproduction, et au dessus de la quantité d'équilibre, correspond à des consommateurs prêts à payer plus qu'il ne coûte aux producteurs de produire, mais pour laquelle des échanges ne se produisent pas.

FIGURE 3.5 – Surplus et perte sèche (en gris) en situation de surproduction



Nous pouvons donc conclure :

Parmi tous le prix possibles, le prix d'équilibre de marché est celui qui maximise le surplus total.

3.3.2 Efficience des marchés et premier théorème du bien-être

Nous avons vu que parmi tous les prix possibles, celui qui maximise le surplus est celui d'équilibre de marché. Mais peut-être existe-t-il une manière d'allouer les biens d'une autre manière qu'avec la formation d'un prix qui soit plus efficace que l'allocation de marché! Nous pouvons ainsi considérer des subventions à des producteurs qui seraient déficitaires à l'équilibre de marché afin de maintenir leur production, ou encore des tarifs différenciés selon les consommateurs. Ainsi, la SNCF comme les cinémas d'art et d'essai reçoivent des subventions publiques pour leur permettre de rester en activité.

Supposons qu'une économie est formée d'un ensemble I de producteurs et J de consommateurs. La fonction de propension marginale à payer du consommateur $i \in I$ est notée u_i , et

3. On fait ici aussi l'hypothèse implicite que les unités produites le sont à coût moindre. Dans le cas contraire, le surplus obtenu serait encore plus faible.

la fonction de coûts marginaux du producteur $j \in J$ est notée c_j . Imaginons que nous soyons dans le cadre d'une économie planifiée, où un dictateur bienveillant pourrait décider du profil de consommation et de production de l'économie. Une *allocation de biens* $(d_i)_{(i \in I)}, (s_j)_{(j \in J)}$ est donnée par une quantité $d_i \geq 0$ du bien reçu par chaque consommateur i et par une quantité $s_j \geq 0$ du bien produite par chaque producteur j , avec la restriction $\sum_i d_i = \sum_j s_j$ (on ne peut consommer que ce qui est produit). Une *allocation de richesses* $(t_i)_{(i \in I)}, (r_j)_{(j \in J)}$ est donnée par une quantité t_i payée par chaque individu i et une quantité reçue r_j par chaque producteur j , avec la contrainte $\sum_i t_i = \sum_j r_j$.

La définition suivante de l'efficacité d'une allocation de biens et de richesses provient des travaux du sociologue Vilfredo Pareto (1848–1923) :

Une allocation des biens (consommation) et richesses (monnaie) est *inefficace au sens de Pareto* s'il en existe une seconde qui est préférée (au sens strict) par *certaines agents* de l'économie tandis qu'aucun ne préfère (au sens strict) la première à la seconde. Dans le cas contraire elle est efficace au sens de *Pareto*.

Par conséquent, pour une allocation inefficace, il existe un moyen d'augmenter le bonheur de certains individus sans créer de détriment à aucun des autres. Dans ce qui suit, on dira d'un agent qu'il préfère faiblement une première allocation à une seconde lorsqu'il préfère la première ou est indifférent entre les deux. Pour bien distinguer, on dira aussi qu'un agent préfère strictement une allocation à une seconde s'il préfère la première faiblement mais n'est pas indifférent entre les deux.

Autrement dit, pour une allocation Pareto-efficace, le seul moyen d'améliorer la satisfaction de certains agents est de diminuer celle d'autres agents. L'efficacité au sens de Pareto caractérise donc l'absence de gaspillage de ressources.

Dans la section précédente, les pertes sèches représentées sur les figures 3.4 et 3.5 correspondent à des unités pour lesquels la propension à payer des consommateurs, représentée par la courbe de demande, est supérieure au coût de production, représentée par la courbe d'offre. L'absence de production de ces unités implique qu'un échange mutuellement bénéfique n'a pas eu lieu, ce qui est inefficace au sens de Pareto. La perte sèche caractérise donc une inefficacité dans l'allocation des biens.

Supposons un prix de marché p . À ce prix, chaque consommateur i obtient $D_i(p)$ unités et paye $D_i(p)p$, et chaque producteur j produit $S_j(p)$ unités et reçoit $S_j(p)p$. Est-il possible qu'une allocation de biens $(d_i)_i, (s_j)_j$ et de richesses $(t_i)_i, (r_j)_j$ soit préférée par *certaines consommateurs et producteurs* à l'allocation de marché tout en étant préférée faiblement par tous ? Autrement dit, est-il possible que l'allocation de marché soit inefficace ?

Supposons que ce soit le cas. Pour le consommateur i , cela voudrait dire que ce consommateur préfère faiblement recevoir d_i et payer t_i que recevoir $D_i(p)$ et payer $pD_i(p)$. Comme son choix optimal de consommation est $D_i(p)$ à prix p , il préfère faiblement aussi recevoir $D_i(p)$ et payer $pD_i(p)$ plutôt que recevoir d_i et payer pd_i . Par transitivité, il préfère faiblement recevoir d_i et payer t_i , plutôt que recevoir d_i et payer pd_i . On doit donc avoir $t_i \leq pd_i$, avec inégalité stricte si i préfère strictement recevoir d_i et payer t_i que recevoir $D_i(p)$ et payer

$pD_i(p)$. En sommant sur l'ensemble des consommateurs, on a donc $\sum_i t_i \leq p \sum_i d_i$.

Chaque producteur j doit aussi préférer faiblement produire s_j et recevoir un montant r_j que produire $S_j(p)$ et recevoir $pS_j(p)$. Comme son choix de production à prix p est $S_j(p)$, il préfère faiblement produire $S_j(p)$ et recevoir $pS_j(p)$ plutôt que de produire s_j et recevoir ps_j . Par transitivité, on doit donc avoir $r_j \geq ps_j$ avec inégalité stricte si au moins un producteur préfère strictement produire s_j et recevoir un montant r_j que produire $S_j(p)$ et recevoir $pS_j(p)$. Par sommation, on obtient $\sum_j r_j \geq p \sum_j s_j$. Si l'allocation de marché est inefficace, on a $\sum_i t_i < p \sum_i d_i$ et/ou $\sum_j r_j > p \sum_j s_j$.

Or, ceci est incompatible avec les contraintes $\sum_i d_i = \sum_j s_j$ et $\sum_i t_i = \sum_j r_j$. Ceci démontre donc le résultat suivant, connu comme le *premier théorème du bien être* :

L'allocation de biens et richesses issue du marché est efficace au sens de Pareto.

Ce résultat est valide dans un cadre beaucoup plus vaste que celui présenté ici. Le premier théorème du bien-être a été formalisé par Maurice Allais, Kenneth Arrow et Gérard Debreu dans le cadre dit d'*équilibre général* qui traite simultanément de l'ensemble des marchés de l'économie, chaque bien s'échangeant sur un marché qui lui est spécifique. Cette approche permet d'éviter de définir le prix comme la valeur d'un objet de référence, tel que la monnaie. En outre, elle permet des préférences plus générales que celles supposées ici, et prend en compte les aspects d'incertitude et d'horizon temporel de l'économie.

On peut interpréter le premier théorème du bien-être ainsi : *sous certaines conditions*, une société qui laisse une totale liberté de choix aux producteurs et aux consommateurs est assurée de ne pas gaspiller de ressources. Le marché permet alors d'aboutir à une allocation efficace. Pour qu'un centralisateur (tel un dictateur bienveillant) puisse arriver au même résultat, il devrait pouvoir connaître les propensions à payer de chaque consommateur et les coûts de production de chaque producteur, ce qui est impossible en pratique. Le marché, en plus d'être efficace en termes d'allocation réalisée, est aussi très économe en information : il ne repose que sur des décisions décentralisées, et chaque acteur de l'économie ne doit connaître que le prix du bien pour prendre ses décisions de production et de consommation. Adam Smith, qui a été le premier à mettre en avant l'intuition et la logique sous-jacente au premier théorème du bien-être, utilisa l'expression de *main invisible des marchés* pour décrire ce mode de fonctionnement décentralisé.

Lorsque vendeurs et acheteurs peuvent effectuer des échanges librement, sans intervention de l'État ou autre barrière, on parle d'une situation de "laissez-faire".

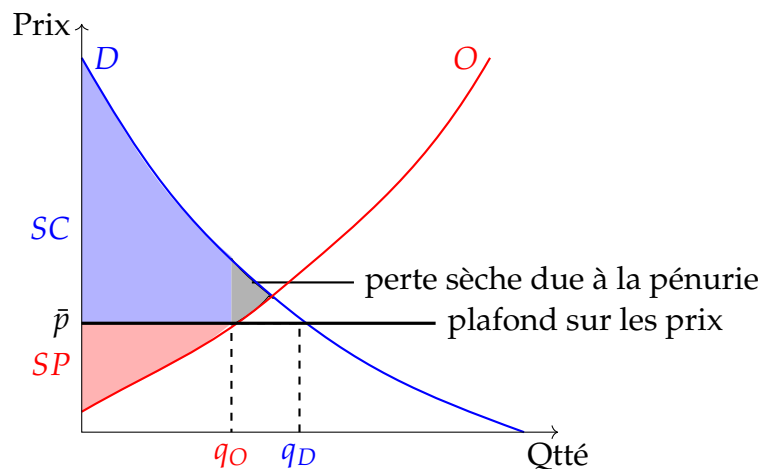
3.3.3 Contrôle des prix et perte sèche

Face à des prix trop élevés, les gouvernements choisissent parfois d'avoir recours au contrôle des prix. Il s'agit d'une législation qui fixe le prix maximum auquel peut être échangé un bien.

Prenez par exemple une politique de contrôle des loyers, qui interdit à un propriétaire de louer un appartement à un prix dépassant un certain plafond. Si ce plafond est en deçà

du prix d'équilibre de marché, cela induira une perte sèche. Supposez ainsi qu'une grande ville décide d'interdire la location d'appartements à un prix dépassant un montant \bar{p} , en deça du prix d'équilibre de marché p^* . Dans cette situation, le plafond sur les prix donnera lieu à une pénurie, et l'allocation obtenue est inefficace car certains échanges mutuellement bénéfiques ne peuvent pas avoir lieu. La Figure 3.6 montre l'effet du plancher sur le marché du logement locatif : par rapport à une situation sans contrainte, l'imposition du plancher diminue le nombre total de logements en location. On parle d'une situation de pénurie car le nombre de logements offerts q_O est inférieur au nombre de logements demandés q_D au prix plafond.

FIGURE 3.6 – Effets d'un plafond sur les prix, rationnement efficace

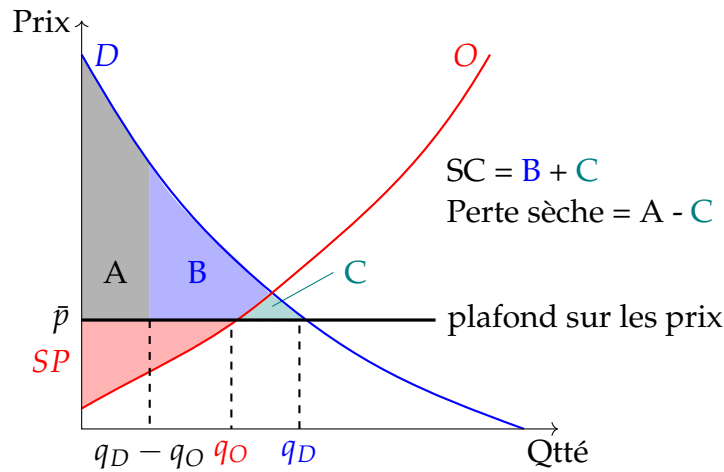


Lorsqu'un plafond est imposé, les propriétaires qui seraient prêts à mettre en location leur logement à un prix entre \bar{p} et p^* ne peuvent pas le faire. Si ces logements avaient été loués, des échanges mutuellement bénéfiques auraient engendré un surplus supplémentaire – c'est la perte sèche engendrée par la pénurie qui est représentée sur la Figure 3.6. Sur cette figure, on peut également voir l'effet du plafond sur différents groupes : les propriétaires perdent une part de leurs profits (le surplus des producteurs a baissé par rapport à la situation d'équilibre de marché), et les locataires qui parviennent à se loger gagnent (ceux là se logent à meilleur marché). Le surplus total, en revanche, a baissé.

En fait, la perte sèche donnée par la Figure 3.6 est une mesure optimiste, et il s'agit en réalité d'une limite basse sur la perte sèche réelle. En effet, nous avons considéré dans la Figure 3.6 que les q_O logements offerts en présence du plancher ont été alloués aux q_O locataires qui ont la plus haute propension à payer pour ceux-ci. Or, il y a q_D locataires potentiels qui ont une propension à payer pour un logement dépassant \bar{p} , et il n'y a aucune raison a priori de penser que c'est à ceux qui ont la plus haute propension à payer que les logements offerts seront effectivement alloués. Dans la Figure 3.7, nous considérons le scénario où le rationnement se fait de la façon la plus inefficace possible, les q_O logements en location étant alloués aux locataires potentiels avec la propension à payer la plus basse. La perte sèche correspond à la surface A, qui représente la perte de surplus de la masse $q_D - q_O$ des consommateurs qui ne parviennent pas à se loger, moins la surface C, qui correspond au nouveau surplus généré par les consommateurs ayant une faible propension à payer et qui parviennent désormais à

se loger au prix \bar{p} .

FIGURE 3.7 – Effets d'un plafond sur les prix, rationnement le plus inefficace

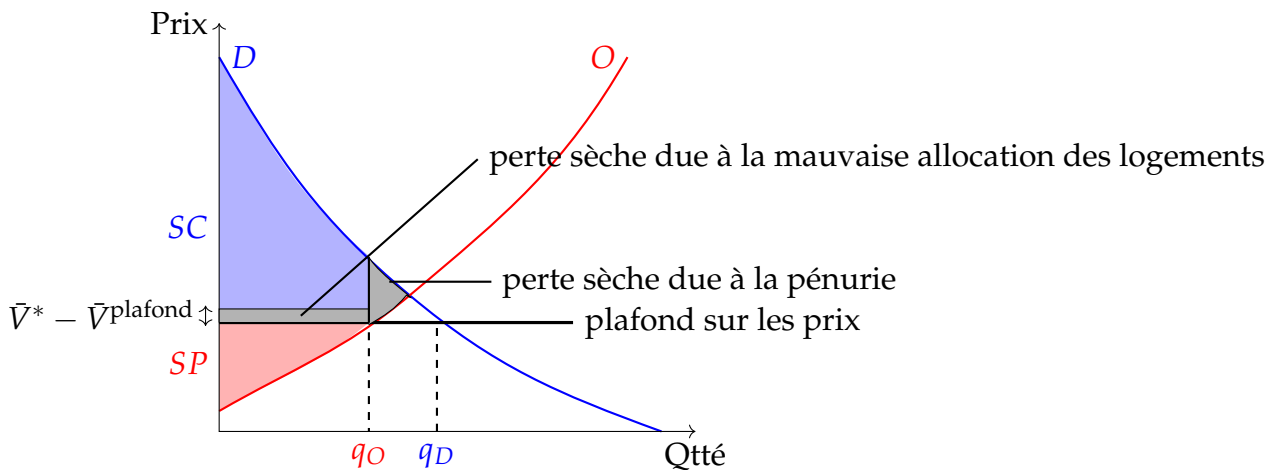


Le fond du problème est que le contrôle des prix engendre un rationnement, ce qui implique qu'une forte propension à payer ne garantit nullement l'obtention d'un logement au prix \bar{p} . La Figure 3.6 donne le cas optimiste du *rationnement efficace* (ce qui est une terminologie trompeuse car le rationnement est source d'inefficacité), tandis que la figure 3.7 donne le cas du rationnement le plus inefficace.

Une hypothèse peut-être plus raisonnable est celle dite du rationnement aléatoire, où les logements sont alloués au hasard parmi ceux qui seraient prêts à les louer au prix de \bar{p} . C'est ce qui arriverait par exemple si les logements étaient alloués selon la règle du "premier arrivé, premier servi", et que l'ordre d'arrivée des locataires dans la liste d'attente pour des logements n'était pas corrélé à leur propension à payer. Dans un tel scénario, ceux qui obtiendraient un logement auraient en moyenne une propension à payer pour celui-ci de $\bar{V}^{\text{plafond}} = \frac{1}{q_D} \int_0^{q_D} D^{-1}(q) dq$, alors qu'à l'équilibre de marché, la propension à payer moyenne était de $\bar{V}^* = \frac{1}{q^*} \int_0^{q^*} D^{-1}(q) dq$. Le surplus total des consommateurs qui parviennent à se loger est donc égal à $q_O \bar{V}^{\text{plafond}}$ si le prix d'un logement n'est que de \bar{q} , tandis que ce surplus est égal à $q_O \bar{V}^*$ si le prix s'élève à p^* . Le fait que \bar{V}^{plafond} est inférieure à \bar{V}^* reflète une mauvaise allocation des logements : lorsqu'un prix plafond est imposé, certains locataires avec une forte propension à payer n'ont pas accès un logement, alors que d'autres avec une propension à payer faible en obtiennent un. Cela engendre une perte sèche supplémentaire, égale à $q_O (\bar{V}^* - \bar{V}^{\text{plafond}})$ – voir la Figure 3.8.

Nous avons déjà évoqué plus haut les effets redistributifs du plafonnement des prix : il nuit aux propriétaires, et a tendance à bénéficier aux locataires. Plus particulièrement, il bénéficie aux locataires qui ont une propension à payer plus faible. Ainsi, à l'équilibre de marché sans contrôle des loyers, les locataires qui ont une propension à payer inférieure à p^* ne retirent pas de surplus. Or, avec un plafonnement des loyers et un rationnement aléatoire, une partie du surplus des locataires revient à des consommateurs avec une propension à payer entre \bar{p} et p^* . Typiquement, ces locataires avec une propension à payer faible seront moins fortunés que ceux qui auraient obtenu un logement en l'absence d'un plafonnement des prix. Le rationnement a ainsi un effet redistributif.

FIGURE 3.8 – Effets d'un plafond sur les prix, rationnement aléatoire



Des plafonnements des prix (associés à un rationnement organisé) sont souvent mis en place lorsque les ressources deviennent rares. C'est ainsi que pendant les deux guerres mondiales, pratiquement tous les gouvernements engagés dans le conflit ont imposé à leurs populations une consommation maximale pour la plupart des denrées alimentaires. En l'absence d'un tel rationnement, le prix de ces produits auraient été très élevés en raison d'une offre faible, et les foyers les plus pauvres n'auraient pas pu avoir accès à une alimentation suffisante. Lorsque nous évoquerons plus tard dans ce cours les enjeux liés à la taxation, nous reviendrons sur la question de la redistribution des richesses au sein de la société.

Le contrôle des loyers à San Francisco

De nombreuses villes dans le monde ont mis en place un contrôle des loyers. Dans l'ensemble, c'est une politique assez populaire, et les arguments sa faveur s'articulent généralement en deux points. D'abord, un contrôle des loyers permettrait à des foyers plus modestes d'habiter dans des villes chères. C'est une forme de redistribution, qui favoriserait en outre une diversité sociale dans chaque quartier. Ensuite, le contrôle des loyers permettrait d'éviter des phénomènes dit de gentrification, où la hausse des prix du logement pousse les foyers les plus pauvres à quitter des quartiers où ils s'étaient installés depuis longtemps et tissé des liens sociaux dont la rupture est coûteuse pour la société.

En revanche, un contrôle des loyers peut avoir plusieurs effets pervers potentiels : en plus de la perte sèche due à la mauvaise allocation des logements décrite plus haut, un plafond sur les loyers pourrait pousser les propriétaires de logements à retirer ceux-ci du marché locatif, aggravant la pénurie et rendant encore plus élevés les prix que l'on cherche à limiter. Un contrôle des loyers pourrait également limiter la rénovation des vieux logements : neuf ou délabré, un logement loué au plafond rapportera la même chose à son propriétaire, qui n'a donc aucune incitation à payer le coût de la rénovation.

Pour comprendre la pertinence de ces différents arguments théoriques, Diamond, McQuade and Qian (2019) ont étudié le cas de San Francisco. En 1994, cette ville a imposé un contrôle des loyers sur les petits immeubles construits avant 1980. Pour comprendre les effets de cette politique, Diamond, McQuade et Qian ont rassemblé des données sur les caractéristiques des foyers à San Francisco et leurs déménagements. Cela leur a permis de comparer des foyers qui ont bénéficié d'un contrôle des prix dès 1994 à des foyers similaires, mais non couverts par le plafond (car vivant dans des appartements construits après 1980).

Les foyers couverts par le contrôle des loyers ont été protégés de la forte hausse des prix du logement à San Francisco, et cela se manifeste dans les données par une plus faible propension de ces foyers à déménager.

En revanche, la politique a eu un effet pervers important : face à ces plafonds sur les loyers, les propriétaires de logements ont eu tendance à retirer leurs appartements du marché locatif ou à les démolir, pour reconstruire de nouveaux logements exempts du contrôle des prix (car construits après 1980). En réduisant l'offre de logements en location, cette réaction des propriétaires a contribué à l'augmentation des loyers à San Francisco. Par ailleurs, les nouveaux logements construits étaient en général des logements plus luxueux, loués ou vendus à des foyers aisés.

Le contrôle des loyers a donc eu un effet mitigé sur la gentrification qu'il était supposé combattre. Il a protégé les foyers modestes qui étaient déjà à San Francisco en 1994, mais a nui aux foyers modestes qui sont arrivés dans la ville après cette date (ou qui n'ont jamais pu déménager à San Francisco à cause des prix trop élevés).

3.3.4 Coûts fixes et condition d'ouverture

L'efficacité des marchés, telle que montrée par le premier théorème bien-être, repose sur l'égalisation des coûts marginaux et propensions marginales à payer. A l'équilibre, produire une unité de plus coûterait plus qu'elle ne bénéficierait aux consommateurs, et une de moins donnerait lieu à une économie moindre que la perte engendrée du point de vue des consommateurs.

Or, nous avons vu qu'à prix fixé, un producteur établit sa production au coût marginal, à condition que ces coûts excèdent les coûts moyens. Que se passe-t-il dans un cas où, pour un prix égalisant coût marginal et propension marginale à payer, la condition d'ouverture n'est pas satisfaite ? Nous allons montrer ici par un exemple que dans ce cas, il n'existe pas nécessairement d'équilibre de marché.

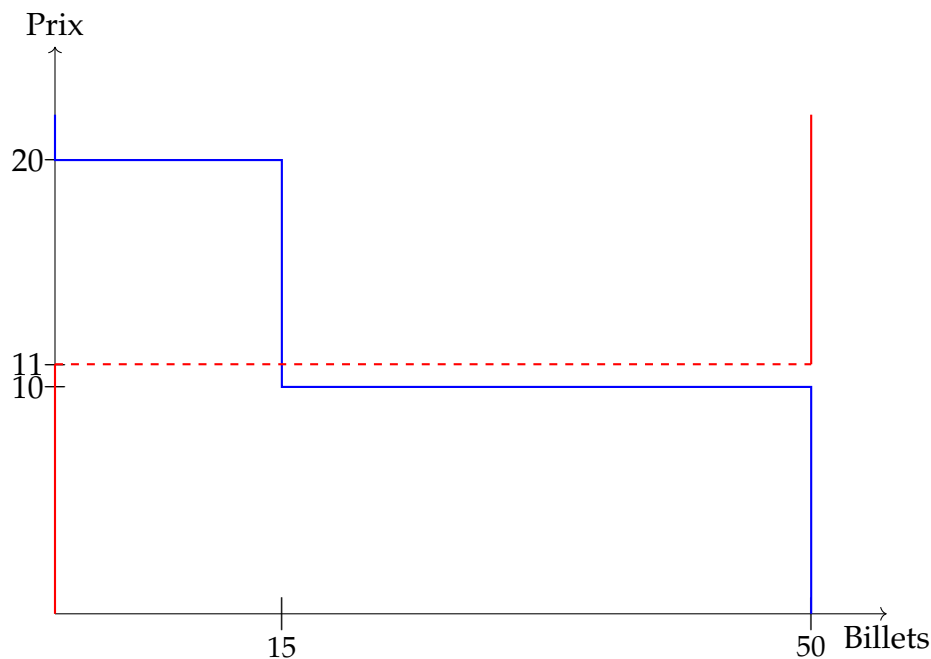
Imaginons que pour effectuer un déplacement, un groupe de personnes peut utiliser soit la voiture individuelle, soit des transports en commun. Les transports en commun (ici, un bus) ont un coût fixe de 300€ et des coûts marginaux de 5€ par passager. Le bus a la capacité de transporter 50 personnes, et le transport en voiture a un coût de 20€ par personne. Parmi les individus, 15 d'entre eux (les plus aisés) ont une propension à payer pour le trajet de 30€, tandis que les 35 autres personnes (plus modestes) ont une propension à payer pour le trajet de 10€. Analysons la demande et l'offre pour des places de bus.

Demande : pour un prix inférieur à 10€, la demande est de 50 car tous les passagers souhaitent effectuer le trajet en bus. Pour un prix compris entre 10€ et 20€, seules les personnes aisées sont prêtes à acheter un billet, et la demande est de 15. Enfin, pour un prix supérieur à 20€, la demande est nulle.

Offre : le coût marginal est de 5€ par passager jusqu'à 50 passagers, puis infini (le bus a une capacité limitée). La condition d'ouverture avec 50 passagers est que le prix satisfasse $50p \geq 300 + 5 \times 50$, soit $p \geq 11$ €. L'offre est donc nulle en dessous de 11 €, puis de 50 pour un prix supérieur.

L'offre et la demande sont représentés sur la figure 3.9. On note qu'il n'existe aucun prix permettant d'équilibrer offre et demande. En effet, au prix de 11€ et pour une demande de 15 billets, point d'intersection des courbes, la compagnie de bus effectue des pertes (à ce prix, elle est capable d'ouvrir à condition de remplir le bus, mais pas sinon).

FIGURE 3.9 – Offre (en rouge) et demande (en bleu) de billets de bus. Pour un prix de 11€, l'offre est discontinue et passe de 0 à 50 billets. La condition d'ouverture n'est pas satisfaite au point d'intersection.

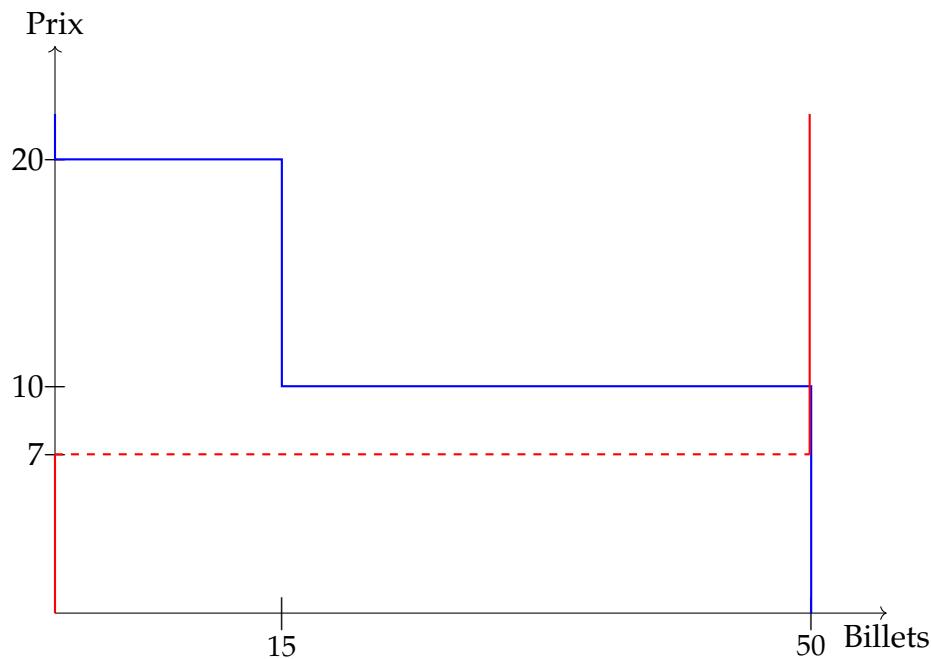


Par conséquent, il n'y a pas d'équilibre de marché dans lequel le bus opère. Le surplus total lié à l'utilisation de la voiture individuelle par les personnes aisées est de $15 \times (30 - 20) = 150$ €. Cependant, il serait plus efficace que tout le monde utilise le bus : dans ce cas, le surplus total est obtenu comme la différence entre le bénéfice pour les consommateurs et le coût, c'est à dire $15 \times 30 + 35 \times 10 - (300 + 50 \times 5) = 250$ €.

Comment permettre cette allocation efficace? Une solution est celle d'une subvention publique à la compagnie de bus financée par l'impôt. Par exemple, une subvention de 200€ par bus ouvert permet de faire baisser le coût fixe à 100€. La prix minimal d'un billet pour que la ligne de bus ne soit pas déficitaire en vendant 50 billets est alors de 7€. La figure 3.10 représente l'offre d'une compagnie subventionnée et la demande. Tous les prix situés entre

7€ et 10€ sont des prix d'équilibre de marché, conduisent à la maximisation du surplus, ainsi qu'à une allocation efficace des ressources.

FIGURE 3.10 – Offre (en rouge) d'une compagnie de bus subventionnée et demande (en bleu) de billets de bus. Tous les prix entre 7€ et 10€ sont des prix d'équilibre de marché.



Une autre solution est celle des prix différenciés (déjà proposée par Jules Dupuit en 1848) avec par exemple un tarif de 18€ pour les personnes aisées et de 8€ pour les plus modestes. À ces tarifs, chaque individu souhaite effectuer le trajet en bus. Le revenu total pour la compagnie de bus est de $15 \times 18 + 35 \times 8 = 550\text{€}$, c'est-à-dire que la compagnie reçoit juste assez pour ouvrir la ligne de bus et transporter les 50 passagers.

Pour que le marché puisse remplir son rôle d'allocation des biens dans l'économie, encore faut-il qu'un équilibre de marché existe ! Dans la lignée des travaux d'Arrow et de Debreu (notamment leur article de 1954), la théorie de l'équilibre général fournit des conditions sur l'existence d'un tel équilibre pour des économies avec un nombre arbitraire de biens échangés.

Le problème soulevé par cet exemple peut se produire concrètement dans le cas de producteurs ayant d'important coûts fixes conjointement à des faibles coûts marginaux. C'est le cas par exemple du transport ferré, ou de la production d'électricité nucléaire. Une tarification à coût marginal est efficace, mais elle conduit à des pertes, il est alors parfois nécessaire de subventionner ces productions. Une autre particularité de ces industries est que leur structure de coûts les fait tendre vers une grande concentration, nous reviendrons dessus au chapitre 6 lorsque nous étudierons les monopoles naturels.

3.3.5 Surplus et efficacité

Nous avons étudié les propriétés des marchés selon deux angles : celui de la maximisation du surplus total, et celui de l'efficacité de Pareto. Nous allons voir ici le lien entre ces

critères, et décrire une de leurs limites.

Tout d'abord, considérons une allocation de biens $(d_i)_i, (s_j)_j$ et de richesses $(t_i)_i, (r_j)_j$ avec $\sum_i d_i = \sum_j s_j$ et $\sum_i t_i = \sum_j r_j$. Nous rappelons que le surplus du consommateur i s'écrit $U_i(d_i) - t_i$ et celui du producteur j est $r_j - C_j(s_j)$. Le surplus total est donc :

$$ST = \sum_i U_i(d_i) - \sum_i C_j(s_j).$$

Par conséquent, le surplus total ne dépend que de l'allocation des biens, tandis que l'efficacité de Pareto s'applique à une allocation à la fois de biens et de richesses. En fait, les deux critères sont équivalents dans le sens suivant :

Une allocation de biens et de richesses est efficace au sens de Pareto si et seulement si l'allocation des biens maximise le surplus.

Commençons par montrer l'implication "seulement si" : pour une allocation inefficace au sens de Pareto, il en existe une qui est au moins aussi bonne pour tous les agents, et strictement meilleure pour au moins un des agents. Cette nouvelle allocation donne un surplus au moins égal à tous, et strictement supérieur à au moins l'un des agents : il constitue donc une augmentation du surplus total.

Pour l'implication "si", considérons une allocation de biens $(d_i)_i, (s_j)_j$ et de richesses $(t_i)_i, (r_j)_j$, et supposons que l'allocation de biens $(d'_i)_i, (s'_j)_j$ constitue une augmentation du surplus total par rapport à $(d_i)_i, (s_j)_j$. Soit

$$\delta = \sum_i U_i(d'_i) - \sum_i C_j(s'_j) - \sum_i U_i(d_i) + \sum_i C_j(s_j) > 0$$

le surplus libéré en passant de la première allocation de biens à la seconde, et posons $\delta' = \frac{\delta}{|I|+|J|}$. Considérons l'allocation de richesses donnée par

$$\begin{cases} t'_i = t_i + U_i(d'_i) - U_i(d_i) - \delta' & \text{pour } i \in I, \\ r'_j = r_j + C_j(s'_j) - C_j(s_j) + \delta' & \text{pour } j \in J. \end{cases}$$

On note que $\sum_i t'_i = \sum_j r'_j$ et que l'allocation de richesses $(t'_i)_i, (r'_j)_j$ est telle que tous les agents de l'économie préfèrent $(d'_i)_i, (s'_j)_j, (t'_i)_i, (r'_j)_j$ à $(d_i)_i, (s_j)_j, (t_i)_i, (r_j)_j$. La nouvelle allocation redistribue le surplus libéré également à tous les agents, de telle sorte que tous préfèrent cette nouvelle allocation à l'ancienne.

Comme nous avons vu d'une part que l'allocation issue du marché maximise le surplus, et que la maximisation du surplus est dans notre cadre équivalente au critère de Pareto, nous en déduisons le résultat suivant, connu sous le nom de second Théorème du bien-être.

Si une allocation de biens et de richesses est efficace, alors elle est allocation de marché.

Pour bien comprendre le cadre du second théorème du bien-être, il faut imaginer une

répartition initiale des biens et richesses. Soit cette allocation est efficace au sens de Pareto, et les agents n'ont pas d'intérêt mutuel à faire des échanges, soit elle ne l'est pas, et les échanges conduisent alors à une amélioration de Pareto par rapport à la répartition initiale.

Le second théorème du bien-être est important du point de vue de l'ingénierie des systèmes d'échange. Si l'on se place dans la situation d'une personne en charge de construire un tel système, alors il n'est pas possible de faire mieux que ce que à quoi aboutirait un marché dans lequel les agents peuvent échanger librement et le prix d'échange s'établit selon la loi de l'offre et de la demande. Un système alternatif soit conduirait au même résultat que le marché, soit serait inefficace au sens de Pareto.

Évidemment, le fait que les marchés permettent une maximisation du surplus, et donc une efficacité de Pareto, est une bonne chose. En revanche, les marchés ne garantissent d'aucune manière une distribution égalitaire des richesses, et de fait peuvent conduire à de grandes inégalités. Considérons une situation dans laquelle il faut partager un gâteau entre deux personnes, et que chaque personne souhaite avoir la plus grande part possible du gâteau. Tout partage du gâteau est efficace au sens de Pareto, car il n'est pas possible de donner un peu plus du gâteau à une personne sans diminuer la part de l'autre. De même, il y a une infinité d'allocations "efficaces" de ressources dans l'économie, plus ou moins inégalitaires. Dans la suite de ce cours, nous nous intéresserons aux questions de taxation : par l'impôt, l'État peut redistribuer les ressources dans la société et ainsi déterminer le niveau d'inégalité de celle-ci.

3.4 La libre entrée

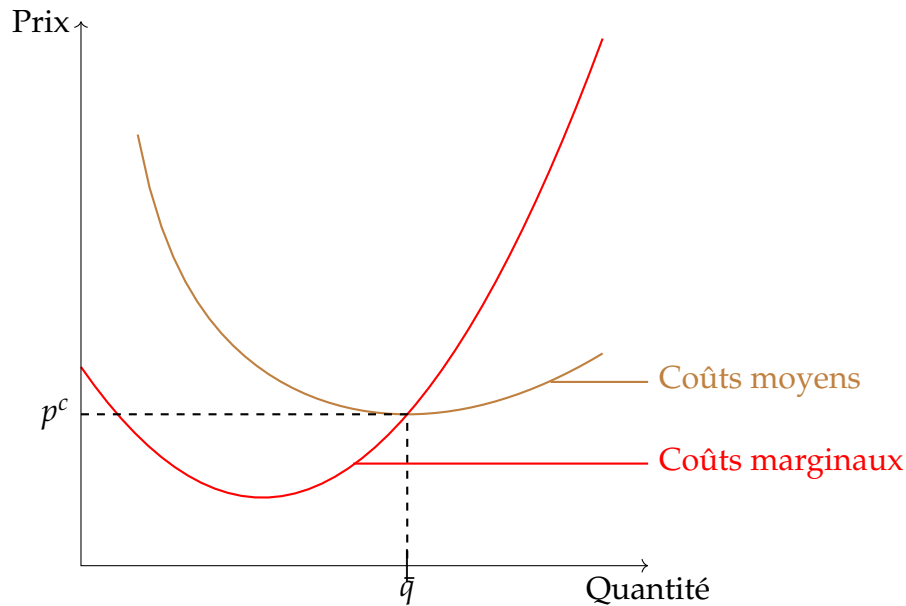
L'offre des entreprises telle que nous l'avons étudiée jusqu'ici suppose un nombre fixe de producteurs. Or, si l'ensemble des consommateurs (correspondant par exemple à la population d'un pays) ne varie pas rapidement et peut être considéré fixe, de nouvelles entreprises sont sans cesse créées et fermées. Selon l'INSEE, l'année 2021 a vu la création de plus d'un million d'entreprises en France, dont 2/3 étaient des auto-entrepreneurs.

De manière générale, une entreprise qui perd de l'argent pendant trop longtemps (on dit qu'elle est déficitaire) finira par cesser ses activités, tandis que des secteurs dans lesquelles les entreprises sont très profitables attirent de nouvelles créations. Pour comprendre l'effet de la création et destruction d'entreprises sur l'équilibre de marché, nous allons nous placer dans un cas idéalisé dit de *concurrence parfaite* dans lequel de nouvelles entreprises se créent dès que le secteur est profitable, et des entreprises ferment si le secteur est déficitaire.

Supposons que tous les producteurs d'un bien ont la même fonction de coûts totaux C , et fonction de coûts marginaux c , et font face à une demande agrégée D pour leur produit. Notons par p^c le prix d'équilibre du marché. À ce prix de marché, chaque entreprise présente produit la même quantité q qui maximise son profit. Si ces entreprises sont déficitaires, certaines d'entre elles vont fermer, ce qui conduira à une baisse de l'offre totale, donc à une hausse des prix et à une hausse des profits pour les entreprises restantes. Si en revanche au prix p^c les entreprises présentes ont un surplus strictement positif, de nouvelles entreprises vont arriver sur le marché, augmentant l'offre et faisant donc baisser les prix.

Nous voyons donc que la libre entrée et sortie des entreprises conduit à une situation dans laquelle les entreprises présentes sur le marché ont un profit nul, c'est à dire que leur coût moyen est égal au prix de marché p^c . Étant donné que la maximisation du profit par chaque entreprise à prix de vente p^c fixé impose que p^c égale le coût marginal de production, nous obtenons que le prix de marché p^c est celui auquel le coût moyen de chaque entreprise égale le coût marginal, et la quantité produite par chaque entreprise est la valeur \bar{q} donnée par l'intersection de ces courbes, comme représenté sur la figure 3.11.

FIGURE 3.11 – Prix de marché p^c et quantité produite par chaque entreprise \bar{q} sous libre entrée.

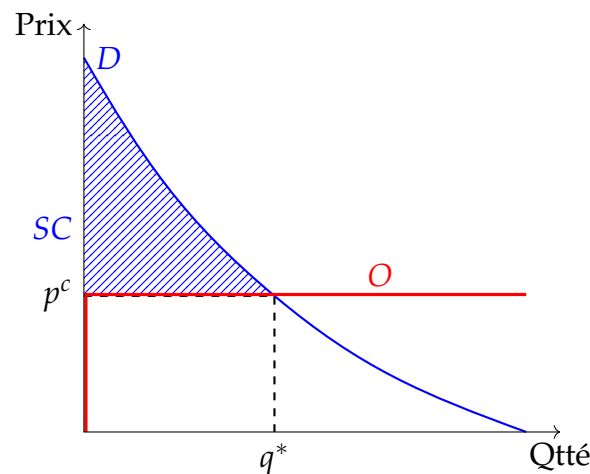


Étant donné que nous supposons que le nombre d'entreprises potentielles est illimité, toute la demande sera fournie au prix p^c . Ceci signifie que la fonction d'offre agrégée est parfaitement élastique, et son graphe est horizontal à la valeur p^c . Par ailleurs, étant donné que les entreprises font un profit nul, tout le surplus est récolté par les consommateurs, lesquels peuvent se procurer des quantités illimitées du bien au prix minimum p^c . La concurrence parfaite est donc entièrement à l'avantage des consommateurs, et nullement des producteurs. La fonction d'offre agrégée et le surplus du consommateur sont illustrés figure 3.12.

Il convient de modérer ce message par deux facteurs au moins. Tout d'abord, si la production totale du bien augmente, le nombre d'entreprises augmente en conséquence, ce qui a pour effet de faire augmenter le prix des intrants. Par exemple, une augmentation de la production totale peut créer une pénurie de main d'œuvre qualifiée, faisant ainsi augmenter les salaires, et donc les coûts de production. Dans ce cas, l'offre totale n'est pas totalement élastique et les prix sont croissants avec la quantité totale produite. La théorie de l'équilibre général étudie les relations entre l'ensemble des prix, qu'il s'agisse de celui des intrants ou des biens produits. Il permet de grandement généraliser notre analyse d'un seul prix qui suppose implicitement que les prix des intrants sont fixes.

Par ailleurs, il est important de garder à l'esprit que les mécanismes d'entrée et de sortie de firmes sur les marchés ne sont pas instantanés. Si les prix d'un bien augmentent, l'offre

FIGURE 3.12 – Concurrence parfaite : fonctions d'offre agrégée O et de demande agrégée D , surplus du consommateur SC



met du temps à s'ajuster, si bien que l'élasticité de long terme est supérieure à l'élasticité de court terme.

3.5 Défaillances des marchés

Pour établir l'efficacité des marchés, nous avons fait un certain nombre d'hypothèses, explicites ou implicites. Lorsque ces hypothèses ne sont pas satisfaites, le premier théorème du bien-être ne tient pas. On parle alors de *défaillances de marchés*. Une grande partie du travail des économistes est l'identification de ces situations, et la proposition de solutions pour y remédier (qui justifient souvent une intervention de l'État). Les défaillances reconnues comme étant majeures sont le pouvoir de marché, les externalités, les biens publics, et les asymétries d'informations. Il en existe quelques autres. Dans les chapitres suivants, nous analyserons plus précisément certaines de ces défaillances.

3.5.1 Pouvoir de marché

Nous avons pour l'instant considéré des agents économiques – consommateurs et producteurs – qui prennent leurs décisions à prix donnés. C'est une hypothèse raisonnable lorsque les agents économiques n'ont pas ou très peu d'influence sur les prix de marchés, ce qui est en pratique vérifié lorsque le marché est constitué de nombreux acteurs, chacun relativement petit par rapport à la taille totale du marché. En revanche, lorsqu'il n'existe qu'un seul producteur (on parle alors de monopole) ou un petit nombre de producteurs (ce qu'on appelle un oligopole), chaque producteur a une influence non-négligeable sur les prix, et a tendance à limiter sa production pour tirer les prix vers le haut et ainsi augmenter ses profits. Nous revenons sur les problèmes de compétition imparfaite dans le chapitre 6.

3.5.2 Externalités

Une hypothèse implicite du modèle de production et de consommation étudié est que les activités économiques des agents n'ont pas d'influence directe sur le bien-être des autres. En termes plus économiques, on dit qu'il n'y a pas d'*externalités*. On dit qu'on est en présence d'externalités positives lorsque la consommation ou la production d'un bien par un des agents augmente le bien-être des autres agents et d'externalités négatives lorsque cet impact est néfaste. Un exemple d'externalité positive est celle de la production de miel d'abeilles sur un verger attenant. On peut aussi citer les vaccins : lorsque vous utilisez un vaccin, en plus de vous protéger des maladies, vous protégez également ceux qui ne sont pas vaccinés. Un exemple d'externalité négative est celui de la pollution atmosphérique et du réchauffement climatique induit par les activités économiques. Nous étudions les externalités dans le chapitre 5 consacré à l'économie publique.

3.5.3 Biens publics

Certains biens ne peuvent pas être consommés exclusivement dans la sphère privée, comme l'éclairage urbain ou la défense nationale. On parle alors de *biens publics*. Ceux-ci ne peuvent en général pas être produits en quantité suffisante par une économie de marché laissée à elle-même. Le chapitre 5 consacré à l'économie publique traitera aussi en détail de ce cas.

3.5.4 Asymétries d'information

Pour que les consommateurs puissent faire leurs choix de manière efficace, il est important qu'ils connaissent la qualité des biens qu'ils peuvent acheter, et donc leur propension à payer. Or, de nombreuses transactions sont telles que le vendeur a plus d'information sur son produit que l'acheteur ; c'est le cas de l'hôtellerie, où la qualité du service n'est pas connue à l'avance. Des entreprises telles que TripAdvisor permettent aux consommateurs d'être mieux informés sur la qualité des biens offerts et par ricochet de rendre les marchés plus efficaces. Nous verrons dans le chapitre 7 que les d'asymétries d'information peuvent conduire à des cas extrêmes dans lesquels la défiance finit par rendre les échanges entre acheteurs et vendeurs impossibles.

3.5.5 Autres conditions du premier théorème du bien-être

Biens homogènes Lorsqu'on parle d'activités de production et de consommation d'un bien, on fait l'hypothèse implicite que tous les biens sont identiques dans leurs caractéristiques : on parle alors de bien homogènes. C'est le cas pour certaines matières premières comme le kérosène, et est une bonne approximation pour d'autres comme les baguettes de pain. Si l'on prend comme exemple les transports en taxi dans une ville, les biens ne sont pas homogènes car les courses diffèrent par leur point de départ, d'arrivée, et leurs horaires. Un système de prix homogènes (tarification à coût fixe puis à la distance parcourue

par exemple) ne permet pas d'arriver à une allocation efficace car les prix ne s'ajustent pas correctement en fonction de l'offre et de la demande. Une plateforme telle qu'Uber a pour principe de mettre en relation clients et chauffeurs et d'ajuster le prix des courses en fonction de différents paramètres. L'économie des plateformes est un sujet d'étude dynamique et relativement récent.

Prix flexibles Notre analyse de l'économie de marché a supposé jusqu'à présent que les prix sont parfaitement flexibles et s'ajustent instantanément pour égaliser l'offre et la demande. En pratique, cette hypothèse est loin d'être satisfaite, notamment à court terme. Sur le marché du travail, les salaires sont notoirement rigides et peuvent mettre plusieurs années à s'ajuster. Ces rigidités des prix génèrent des déséquilibres et des fluctuations de court terme qui sont au cœur de l'analyse macroéconomique. La politique monétaire peut jouer un rôle clef pour contrecarrer les inefficacités engendrées par ces rigidités. Nous aborderons ce sujet dans le chapitre 10 consacré à la macroéconomie.

Rationalité Dire qu'un consommateur effectue des décisions d'achat en fonction de sa propension marginale à payer pour le bien implique une hypothèse de *rationalité* : le consommateur sait ce qu'il souhaite et est capable de mettre en œuvre les moyens pour y parvenir. Les études issues de la psychologie et de l'économie expérimentale pointent un certain nombre de situations dans lesquelles la plupart des individus échouent à agir dans leur propre intérêt. Dans de telles situations, l'hypothèse de rationalité parfaite n'est pas satisfaite et l'État peut avoir un rôle à jouer afin d'améliorer l'allocation des ressources. Cette question est au cœur du domaine de *l'économie comportementale*.

Même si l'hypothèse de rationalité n'est qu'une approximation de la réalité, cela reste néanmoins une hypothèse incontournable en sciences économiques. D'abord parce que, même si les individus ne sont pas parfaitement rationnels, il est probable qu'ils cherchent à défendre leur propre intérêt et qu'ils y parviennent à peu près. Soulignons que l'intérêt individuel doit ici être défini au sens large et n'exclut pas l'altruisme.

Bibliographie

- Arrow, K. J. and Debreu, G. (1954), 'Existence of an equilibrium for a competitive economy', *Econometrica* pp. 265–290.
- Diamond, R., McQuade, T. and Qian, F. (2019), 'The effects of rent control expansion on tenants, landlords, and inequality : Evidence from san francisco', *American Economic Review* **109**(9), 3365–3394.
- Jensen, R. (2007), 'The digital provide : Information (technology), market performance, and welfare in the south indian fisheries sector', *Quarterly Journal of Economics* **122**(3), 879–924.
- Jensen, R. and Miller, N. H. (2018), 'Market integration, demand, and the growth of firms : Evidence from a natural experiment in india', *American Economic Review* **108**(12), 3583–3625.

Exercices du chapitre 3

Exercice 3.1. Discussion : La recherche de l'avantage personnel est-elle une bonne chose ?

Selon Adam Smith, *"It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own self-interest. We address ourselves not to their humanity but to their self-love, and never talk to them of our own necessities, but of their advantages"* (Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol 1)

Ce que l'on peut traduire par « *Ce n'est pas de la bienveillance du boucher, du brasseur ou du boulanger que nous attendons notre dîner, mais de leur considération pour leur propre intérêt. Nous ne nous adressons pas à leur humanité mais à leur amour de soi, et ne leur parlons jamais de nos propres nécessités, mais de leurs avantages.* »

Dans une forme plus extrême, Michael Douglas dans le film « Wall Street » nous dit que "Greed is good" (« l'avidité est une bonne chose »)

1. Dans quelle mesure pensez-vous que le premier théorème du bien-être éclaire la citation d'Adam Smith ?
2. En pratique, pensez-vous comme Michael Douglas que la cupidité conduit au bien commun ?
3. Quelles différences faites-vous entre les propositions d'Adam Smith et de Michael Douglas ?

Exercice 3.2. Étude de document

On demande de lire et commenter l'introduction de l'article de Jensen (1987) sur la page moodle (en Anglais). Il s'agit de comprendre comment l'introduction du téléphone cellulaire a pu contribuer à établir un marché là où il n'y en avait pas (ou uniquement de très locaux) et si les observations empiriques sont compatibles avec les prédictions du modèle de marché. En particulier, on pourra s'intéresser aux points suivants :

1. Loi du prix unique
2. Égalisation de l'offre et de la demande
3. Surplus du producteur, surplus du consommateur, surplus total

Exercice 3.3. Marché parfaitement concurrentiel

On s'intéresse à une économie de marché *parfaitement concurrentiel*. Ceci signifie que :

1. aucun producteur n'est assez important pour influencer les prix et
2. il y a libre entrée sur le marché, de nouveaux producteurs entrent sur le marché tant qu'ils peuvent réaliser un profit positif.

On suppose que tous les producteurs ont la même fonction de coûts C , et qu'il y a d'abord rendements croissants, puis décroissants. On pourra supposer que tous les producteurs sont infinitésimaux par rapport à la taille du marché.

1. Quelle est l'élasticité de l'offre dans ce cas ?
2. Cette prédiction du modèle vous semble-t-elle réaliste et dans quelle mesure ?
3. En pratique, pensez-vous qu'une augmentation de la demande doit faire augmenter ou diminuer les prix ?
4. Quelles hypothèses du modèle vous semblent éventuellement à revoir ?

Exercice 3.4. Subventionner l'achat de produits "vertueux" pour les pauvres

On considère une économie avec deux consommateurs, et deux biens : le premier bien (bien n) est un bien de consommation courante, et le deuxième bien (bien c) un bien de consommation culturelle. Les préférences des deux agents sont les mêmes, ils préfèrent un panier $q' = (q'_n, q'_c)$ à un autre panier $q = (q_n, q_c)$ si et seulement si $\sqrt{q'_n} + \sqrt{q'_c} > \sqrt{q_n} + \sqrt{q_c}$. On suppose par ailleurs les prix de chacun des biens normalisé à 1 :

1. Le consommateur riche a un budget de 8, et le pauvre un budget de 2. Quelle est leur choix de panier de bien ?
2. Le ministre de la culture, formé à l'école dont celui qui sort n'est pas l'inverse, considère que le consommateur pauvre n'a pas suffisamment accès à la culture. Il décide d'instaurer une subvention à la culture pour l'agent pauvre. L'agent pauvre payera désormais un prix unitaire de 0,5€ pour le bien culturel. Pour financer cette mesure, il est prévu que l'agent riche paye un prix unitaire de 1,125€ pour le bien culturel.
 - (a) en supposant que les agents n'ajustent pas leur consommation aux nouveaux prix, la mesure instaurée arrive-t-elle à s'auto-financer ? que croyez-vous qu'il arriverait ?
 - (b) sous les nouveaux prix, quel est le panier de bien consommé par le riche ? par le pauvre ? la mesure est-elle équilibrée budgétairement ?
 - (c) quel prix faudrait-il fixer pour l'accès à la culture au riche pour que la mesure soit équilibrée financièrement en prenant en compte les adaptations aux prix proposés ? s'il y a plusieurs solutions, comment l'interprétez-vous ? quelle est la meilleure solution à retenir ?
3. supposons que le ministre de la culture arrive à fixer le prix de l'accès pour les riches de manière à équilibrer ses mesures. En vous basant sur le critère de Pareto, que pouvez-vous dire de la mesure prise ?
4. quels enseignements généraux tirez-vous de cet exemple ?

Exercice 3.5. Equilibre de court terme sur le marché de l'électricité

On considère un pays avec deux types de centrales électriques : les centrales nucléaires (type N) et les centrales thermiques utilisant des énergies fossiles (type F).

Il y a 20 centrales du type N, qui ont chacune une capacité de production de 1GW=1 000MW et des coûts de production de $C_N(q) = 5\,000 + q$, où q est la quantité d'électricité produite (en MW).

Il y a d'autre part 1 000 centrales de type F ayant chacune une capacité de production de 10MW et des coûts de production de $C_F(q) = 50q$. Les deux types d'usines vendent l'électricité produite sur un réseau national.

1. Quelle est la fonction d'offre individuelle d'une entreprise de type N? d'une entreprise de type F? Quelle est la fonction d'offre de l'industrie? (Indice : représenter graphiquement ces fonctions peut vous aider à résoudre les questions suivantes).
2. Pendant les heures creuses (20h–8h), la demande est $D_1(p) = 40\,000 - 400p$. Quels sont : le prix d'équilibre, les entreprises qui produisent de l'électricité sur le marché, et la quantité produite, pendant les heures creuses?
3. Pendant les heures pleines (8h–20h), la demande est $D_2(p) = 50\,000 - 500p$. Quels sont : le prix d'équilibre, les entreprises qui produisent de l'électricité sur le marché, et la quantité produite?
4. Les coûts de production des usines de type F augmentent en raison de l'augmentation du prix des combustibles fossiles, et s'élèvent désormais à $C_F(q) = 60q$. Quel est l'impact sur les prix de marché durant les heures creuses et les heures pleines? Expliquez la différence.

Exercice 3.6. Concurrence et pollution

On considère deux pays A et B , chacun disposant d'une industrie produisant le bien X . La demande dans chaque pays pour le bien X est identique et est donnée par : $D(p) = 1\,000 - 10p$. Toutes les entreprises du secteur ont la même fonction de coût variable $C(q) = \frac{1}{2}q^2$, et un coût fixe de 50. Le marché est parfaitement concurrentiel et les firmes sont libres d'entrer/sortir du marché.

1. Chaque pays est en autarcie (pas d'échange entre pays). Déterminez les caractéristiques de l'équilibre de long terme dans chacun des pays.
2. En fait, l'activité est polluante. La désutilité sociale, mesurée en perte de surplus, imposée par une entreprise produisant une quantité q est $m(q) = \frac{3}{2}q^2 + 20q$, la convexité de $m(q)$ représente le fait qu'augmenter la production augmente la pollution par unité produite. Le gouvernement du pays A impose une taxe verte : chaque entreprise doit payer $t(q) = m(q)$ pour produire la quantité q . Quel est alors l'équilibre de long terme dans le pays A ?
3. On décide d'ouvrir le marché entre les deux pays. Il n'y a ni taxe ni coûts de transport. Seules les entreprises du pays A doivent s'acquitter de la taxe verte. Il y a 120 entreprises dans le pays A et 90 entreprises dans le pays B .
 - (a) Déterminer l'équilibre de court terme.
 - (b) Déterminer l'équilibre de long terme (le nombre d'entreprises dans chaque pays peut changer).

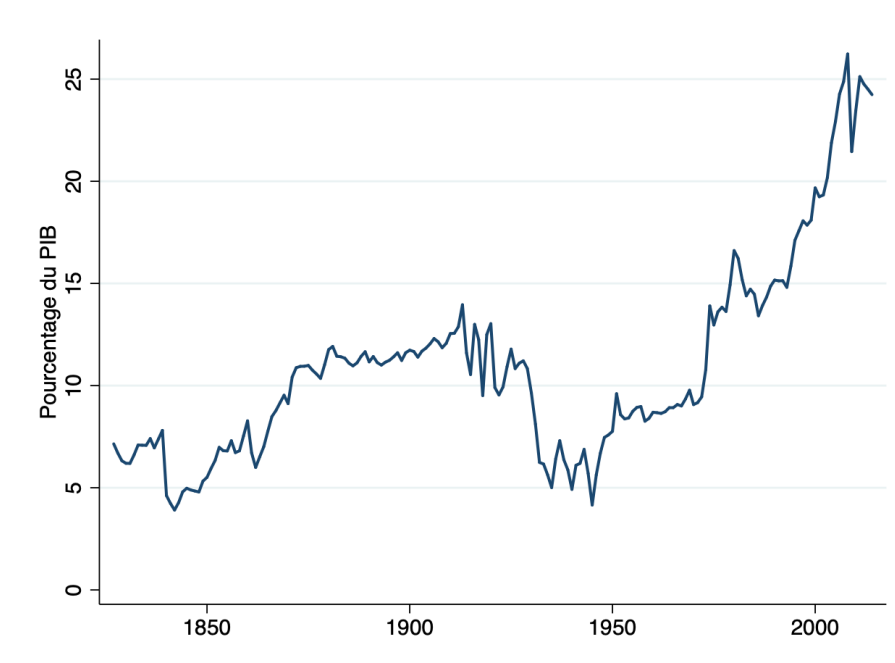
4. Comparer le surplus des consommateurs du pays B , avant et après l'ouverture (NB : n'oubliez pas l'effet de la pollution ! On déterminera la dés-utilité par entreprise, puis la dés-utilité globale en additionnant les dés-utilités des entreprises. On considérera également que les consommateurs du pays B ne sont affectés que par la pollution des entreprises du pays B). Commentez.

Chapitre 4

Le commerce international

Bien que la mondialisation des échanges soit un phénomène relativement récent, le commerce entre pays ne date pas d'hier. Depuis plusieurs millénaires, des biens circulent entre l'Asie et l'Europe en suivant la route de la soie. Le commerce a fait la prospérité de certains États, à l'image de la cité de Sriwijaya (en Indonésie moderne), de la république de Venise ou de l'Empire Britannique. Selon les situations, le commerce peut être source de conflits, comme lors des guerres de l'opium, ou au contraire de paix, comme l'explique Montesquieu selon sa conception du doux commerce.

FIGURE 4.1 – Part des exportations dans la production mondiale.

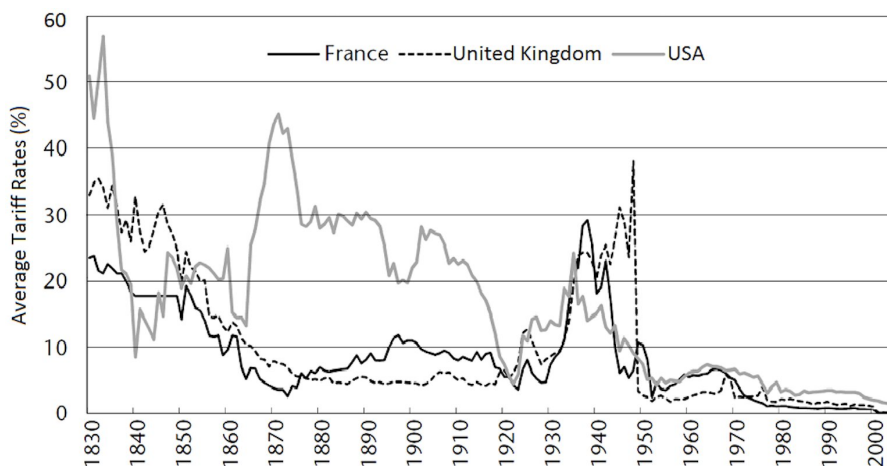


Source : Fouquin, M. and Hugot, J. (2016), Two centuries of bilateral trade and gravity data : 1827-2014

Les innovations dans le domaine du transport, la paix consécutive à la fin des guerres napoléoniennes et le développement des empires coloniaux permirent une intensification des échanges commerciaux vers la fin du XIX^e siècle, une période qualifiée de première mondialisation (voir la figure 4.1). Les deux conflits mondiaux et le retour du protectionnisme durant la Grande Dépression des années 1930 ont largement réduit la part des échanges

dans l'économie mondiale. Le commerce international s'est à nouveau développé dès la fin de la Seconde Guerre Mondiale avec l'accord général sur les tarifs douaniers et le commerce (GATT) – aujourd'hui remplacé par l'Organisation Mondiale du Commerce (OMC) – qui a entraîné la quasi-suppression de la plupart des barrières douanières, comme le montre la figure 4.2. Enfin, la deuxième mondialisation, toujours en cours, a débuté vers la fin des années 1980 avec, d'une part, l'effondrement du bloc soviétique et, d'autre part, l'intégration progressive de la Chine dans l'économie mondiale.

FIGURE 4.2 – Niveau moyen des taxes douanières sur les importations.



Source : Imlah, A.H. (2014), *Economic Elements in the Pax Britannica*, Harvard University Press.

Ces évolutions posent la question de la désirabilité du libre échange. Des sondages réalisés entre 2010 et 2012 révèlent que seuls 46% des Américains ont une opinion favorable de l'accord de libre-échange signé en 1994 entre le Canada, le Mexique et les États-Unis (le North American Free Trade Agreement, NAFTA, remplacé en 2018 par le United States-Mexico-Canada Agreement, USMCA); tandis que 95% des économistes pensent qu'un tel accord est bénéfique.¹ Dans ce chapitre, nous allons étudier en détail le modèle de Ricardo qui constitue la base de la théorie économique du commerce international et qui met en évidence les gains à l'échange.

Pour comprendre certaines oppositions au libre-échange, nous irons ensuite au delà du modèle ricardien afin de comprendre qui sont les gagnants et les perdants du commerce international. Enfin, nous passerons en revue quelques autres effets du libre-échange, qui justifient ou non l'ouverture des frontières au commerce.

4.1 Le modèle de Ricardo

Le mathématicien Stanislaw Ulam mit un jour au défi le célèbre économiste américain Paul Samuelson (lauréat du prix Nobel en 1969) en lui demandant de nommer un résultat en sciences sociales qui soit à la fois vrai et non-trivial. Samuelson revint vers lui en donnant comme réponse la notion d'*avantage comparatif*. Celle-ci était d'après lui « logiquement

1. Sapienza, P. and Zingales, L. (2013), 'Economic Experts vs. Average Americans', *American Economic Review : Papers and Proceedings*, 103(3), 636-642.

vraie [...], mais non-triviale, comme en attestent les milliers d'hommes importants et intelligents qui n'ont jamais été capables de [se l'approprier] ou d'y croire après qu'on leur ait expliquée.» Quelle est donc cette idée contre-intuitive, mais logiquement indéniable, mise en avant par Samuelson ?

La notion d'avantage comparatif découle du modèle de commerce international élaboré en 1817 par le britannique David Ricardo. Ce modèle explique pourquoi un pays a toujours intérêt à commercer avec ses voisins, quelque soit sa productivité dans chacun de ses secteurs d'activité.

4.1.1 Productivités, avantages absolus et avantages comparatifs

Une des principales raisons pour lesquelles les pays commercent entre eux est qu'ils sont asymétriques : au sein de chaque secteur d'activité, les pays ont des productivités différentes. Ainsi, la France produit plus facilement des voitures que Madagascar, et Madagascar produit plus facilement de la vanille que la France. Pour comprendre en détail les implications de ces différences de productivité, le modèle de Ricardo se focalise sur un cas simple où il n'existe que deux biens et deux pays. En outre, ces biens sont fabriqués à partir d'un unique facteur de production : le travail.

Considérons donc un scénario dans lequel il n'y a dans l'économie que la France et l'Allemagne, qui ne produisent que des bateaux et des voitures. La construction d'un bateau prend 100 heures en Allemagne et 125 heures en France, tandis que la production d'une voiture nécessite 50 heures de travail en Allemagne et 125 heures en France (voir table 4.1).

TABLE 4.1 – Nombre d'heures de travail nécessaires à la construction d'une voiture et d'un bateau dans chaque pays

	Bateau	Voiture
Allemagne	100	50
France	125	125

Ainsi, l'Allemagne est plus productive que la France dans la construction navale et l'industrie automobile : elle peut produire chacun des deux biens à moindre coût. Dans ces circonstances, on serait tenté de penser que l'ouverture au commerce serait terrible pour la France : les entreprises françaises ne pourraient résister aux prix bas proposés par l'Allemagne et seraient anéanties par cette concurrence ! Mais que se passe-t-il vraiment lorsque l'on ouvre les frontières entre les deux pays ?

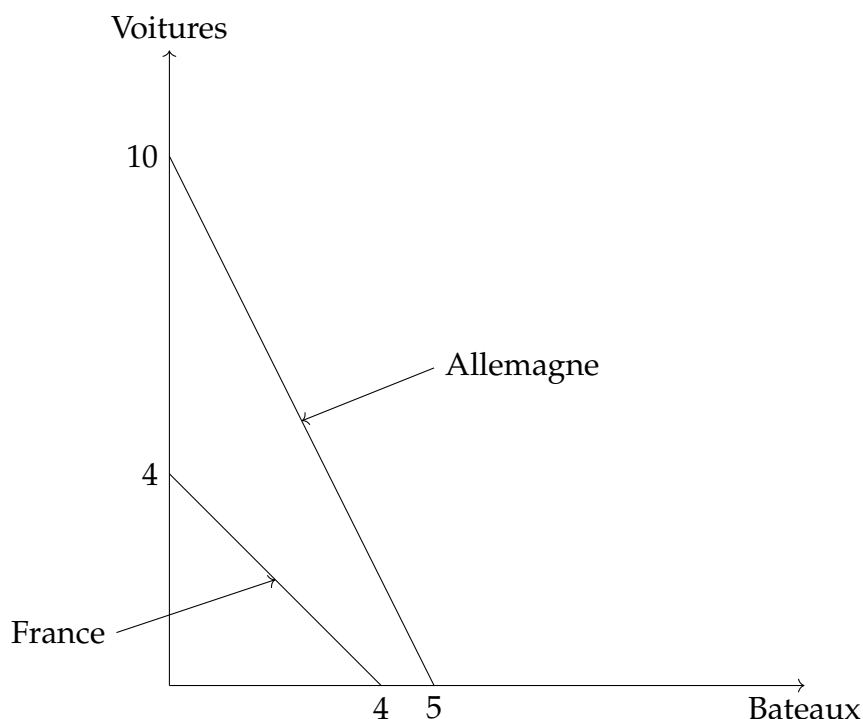
Pour analyser le problème, supposons que chaque pays est doté de 500 heures de travail. Dans ces circonstances :

- L'Allemagne peut produire 5 bateaux et 0 voitures ou 10 voitures et 0 bateaux ;
- La France peut produire 4 bateaux et 0 voitures ou 4 voitures et 0 bateaux.

Chaque pays peut également produire n'importe quelle combinaison linéaire de ces deux possibilités. Par exemple, la France peut décider de produire 2 bateaux et 2 voitures. Pour chaque pays, on peut représenter dans un plan la **frontière de production** qui correspond

à l'ensemble des possibilités de production lorsque toute la main d'œuvre disponible est utilisée (voir figure 4.3).

FIGURE 4.3 – Frontières de production de l'Allemagne et de la France.



Le fait que l'Allemagne soit plus productive que la France dans l'ensemble des secteurs d'activité fait que la frontière de production allemande est au-dessus de la frontière de production française. Soulignons qu'en autarcie, lorsque l'économie est fermée et que le commerce international est impossible, la frontière de production est également pour chaque pays la frontière de consommation, c'est-à-dire que chaque pays ne peut pas consommer un panier de biens qui dépasse sa frontière de consommation.

Les deux frontières de productivité sont décroissantes : cela signifie que pour construire un bateau supplémentaire, un pays doit renoncer à produire une certaine quantité de voitures. Ainsi, la production d'un bateau a un **coût d'opportunité**. Lorsque l'Allemagne produit un bateau supplémentaire, elle doit y consacrer 100 heures de travail, qui auraient pu être employées à la construction de deux voitures. Le coût d'opportunité d'un bateau en termes de voitures est donc de 2. Autrement dit, pour produire un bateau supplémentaire, l'Allemagne doit renoncer à l'opportunité de produire 2 voitures. La France a un coût d'opportunité plus faible : construire un bateau revient à renoncer à une voiture, soit un coût d'opportunité de 1. Notez que ce coût d'opportunité correspond à la pente de la frontière de production. De manière symétrique, on peut aussi calculer pour chaque pays le coût d'opportunité d'une voiture en termes de bateaux (voir la table 4.2).

L'idée de coût d'opportunité est une notion fondamentale en sciences économiques, qui dépasse très largement le cadre du modèle de Ricardo. De manière générale, ce coût correspond à ce à quoi il faut renoncer pour obtenir quelque chose. Ainsi, le coût d'opportunité d'une heure de loisirs est le salaire qui aurait pu être perçu en travaillant pendant une heure. D'un point de vue économique, le temps de loisir n'est donc pas gratuit ! Pour un

TABLE 4.2 – Coûts d’opportunité en Allemagne et en France

	Bateau	Voiture
Allemagne	2	1/2
France	1	1

étudiant ayant le choix entre une spécialisation en sciences économiques, en physique ou en informatique et qui hésite entre ces deux premières possibilités, le coût d’opportunité d’une spécialisation en sciences économiques est de ne pas pouvoir se spécialiser en physique. En l’occurrence, la possibilité de se spécialiser en informatique ne rentre pas dans l’analyse car elle est, pour cet étudiant, dominée. L’idée de coût d’opportunité permet de définir la notion d’avantage comparatif évoquée par Samuelson :

Un pays dispose d’un **avantage comparatif** par rapport à un autre pays dans la production d’un bien lorsque son coût d’opportunité à produire ce bien y est plus faible.

Il faut bien différencier cette notion d’avantage comparatif de celle d’avantage absolu.

Un pays dispose d’un **avantage absolu** dans la production d’un bien lorsque, à ressources égales, il peut en produire davantage.

Dans notre exemple, l’Allemagne a un avantage absolu dans la production de bateaux et dans la production de voitures, car elle est plus productive que la France dans chacun des deux secteurs d’activité. Néanmoins, c’est la France qui dispose d’un avantage *comparatif* dans la production de bateaux. Le coût d’opportunité (en termes de voitures) de la construction d’un bateau est plus faible en France, où il est de 1, qu’en Allemagne, où il s’élève à 2. L’Allemagne dispose elle d’un avantage comparatif dans la production automobile : produire une voiture correspond à renoncer à la production d’un demi-bateau, tandis qu’en France produire une voiture revient à renoncer à la production d’un bateau entier (voir la table 4.3).

TABLE 4.3 – Avantages absolus et comparatifs en Allemagne et en France

	Bateau	Voiture
Avantage absolu	Allemagne	Allemagne
Avantage comparatif	France	Allemagne

Un avantage absolu reflète une différence dans les niveaux de productivité, tandis qu’un avantage comparatif correspond à une différence *relative* de productivité. Par rapport à l’Allemagne, la France est relativement plus productive dans la construction navale que dans l’industrie automobile, tandis que l’inverse est vrai pour l’Allemagne. En conséquence, si un pays peut avoir un avantage absolu dans l’ensemble des secteurs de l’économie, comme c’est le cas de l’Allemagne dans notre exemple, il ne peut pas avoir un avantage comparatif dans tous les domaines. Si un pays est relativement plus productif qu’un autre dans

un secteur d'activité, alors il est nécessairement relativement moins productif dans un autre secteur d'activité.

4.1.2 Spécialisation et échanges

Situation en autarcie. Quelle est la conséquence de cette structure de productivités sur les échanges commerciaux entre la France et l'Allemagne? Pour le comprendre, partons d'une situation où les économies sont fermées et où les échanges internationaux sont donc impossibles. Intéressons-nous au prix des bateaux. Dans la mesure où les seuls autres biens existants sont des voitures, le prix d'un bateau n'a de sens que s'il est exprimé relativement à celui d'une voiture. Nous allons donc nous concentrer sur le ratio $p = \frac{P_{\text{Bateau}}}{P_{\text{Voiture}}}$, ce qui revient à normaliser le prix des voitures à 1 et à s'intéresser au prix p des bateaux.²

Lorsque les pays sont en autarcie, le prix de chaque bien reflète son coût de production. Par conséquent, le prix d'un bateau, relativement à celui d'une voiture, doit être égal au coût d'opportunité de la construction d'un bateau, en termes de voitures. Dans la mesure où, en Allemagne, il faut deux fois plus d'heures de travail pour produire un bateau qu'une voiture, il est naturel qu'un bateau coûte deux fois plus cher. On a donc :

$$p^{\text{Allemagne}} = \frac{P_{\text{Bateau}}^{\text{Allemagne}}}{P_{\text{Voiture}}^{\text{Allemagne}}} = 2,$$

Où $p^{\text{Allemagne}}$ correspond au prix d'un bateau en Allemagne (relativement à celui d'une voiture) en autarcie. De même, on a :

$$p^{\text{France}} = \frac{P_{\text{Bateau}}^{\text{France}}}{P_{\text{Voiture}}^{\text{France}}} = 1,$$

Où p^{France} est le prix d'un bateau en France en autarcie.

Le prix unique. Que se passe-t-il lorsqu'on ouvre les frontières en autorisant la circulation des biens entre la France et l'Allemagne? La première chose à remarquer est que le prix relatif des bateaux dans chaque pays s'équilibrera pour atteindre une valeur commune : on devra avoir $p^{\text{France}} = p^{\text{Allemagne}} = p$.

Imaginez que les prix ne s'ajustent pas, et qu'ils restent à leur valeur d'autarcie, avec $p^{\text{France}} = 1$ et $p^{\text{Allemagne}} = 2$. Un Français qui possède un bateau pourrait aller en Allemagne, l'échanger contre deux voitures, revenir en France et échanger ses deux voitures contre deux bateaux. En réitérant ce montage, il obtiendrait ensuite quatre bateaux, huit bateaux, etc. Avec des transactions peu coûteuses, il pourrait alors s'enrichir sans limite. On dit qu'il

2. Dans le chapitre précédent, nous nous sommes intéressés à un marché donné et avons exprimé les prix en euros, l'idée étant que les euros peuvent être dépensés sur les autres marchés dont nous avons fait abstraction. Il s'agissait d'une analyse dite en *équilibre partiel*. Désormais, nous étudions simultanément l'ensemble des marchés de l'économie. Nous faisons donc une analyse dite d'*équilibre général*. Or, en équilibre général, seuls les prix relatifs sont définis. On peut donc désigner un bien, les voitures, comme étant le numéraire (l'objet de référence) et se focaliser sur le prix des autres biens (en l'occurrence sur celui des bateaux). L'analyse demeure inchangée si on prend les bateaux comme numéraire.

il y a une possibilité d'arbitrage. Une telle situation ne peut pas perdurer : il y a une forte demande pour échanger des bateaux contre des voitures en Allemagne, ce qui augmente le prix relatif des voitures, et fait donc baisser $p^{\text{Allemagne}}$. En France, il y a une forte demande pour échanger des voitures contre des bateaux, ce qui fait augmenter p^{France} .

À l'équilibre, on doit nécessairement avoir l'égalité $p^{\text{France}} = p^{\text{Allemagne}} = p$, sans quoi des acheteurs pourraient faire des transactions circulaires et s'enrichir sans rien produire. Il s'agit là d'une conséquence de la loi du prix unique que nous avons vue au chapitre précédent : dans un marché, tous les échanges d'un même bien se font au même prix. En l'occurrence, une fois levées les barrières douanières entre la France et l'Allemagne, il n'existe plus qu'un seul grand marché franco-allemand pour les bateaux (ainsi que pour les voitures). Soulignons que la loi du prix unique ne tient strictement qu'en l'absence de coûts de transaction.

À propos des coûts de transaction.

L'hypothèse d'absence de coûts de transport peut paraître particulièrement forte, mais la standardisation des conteneurs et les progrès de logistique ont considérablement réduit le coût du transport de marchandises entre pays : aujourd'hui, acheminer un conteneur de 63 m³ de New-York à Marseille coûte aux alentours de 2000 dollars. Celui-ci pouvant contenir plus de 20 tonnes de marchandises, le coût de transport est souvent très faible relativement à la valeur des biens. Ceci étant, certaines marchandises sont périssables et ne peuvent être transportées que par avion tandis que certains services sont difficilement échangeables à l'international, comme les services de santé, la restauration ou les services de coiffure. L'absence d'une langue commune réduit également les flux commerciaux entre pays. Enfin, pour des produits non standardisés (comme le textile ou les jouets, par opposition au pétrole ou au blé), le commerce passe d'abord par l'instauration d'un lien de confiance entre deux entreprises, l'une exportatrice et l'autre importatrice. Ces liens sont coûteux à construire et induisent donc des frictions. L'étude de ces frictions au commerce international est aujourd'hui un enjeu de recherche majeur.

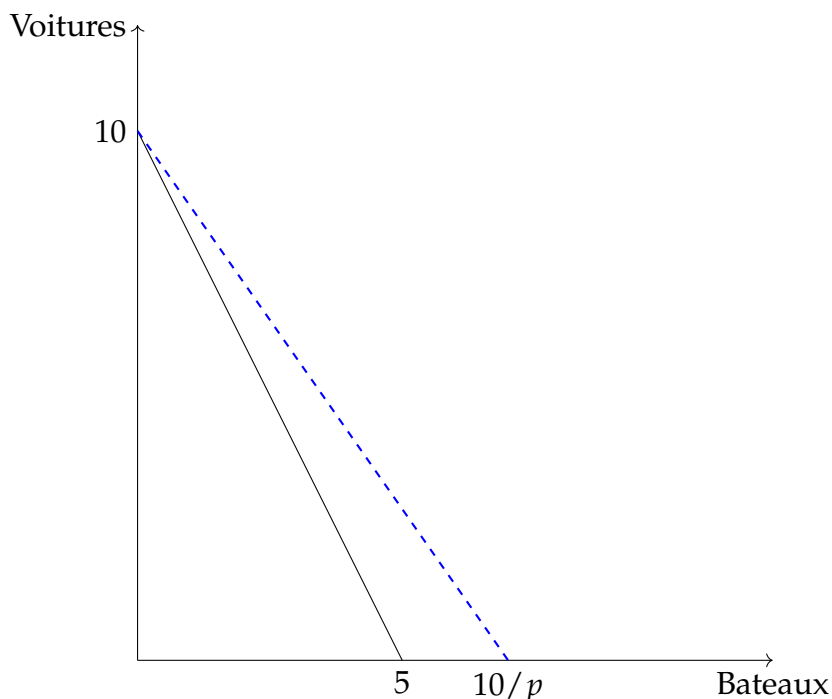
Spécialisation et gains à l'échange. Supposons que, une fois les frontières ouvertes, le prix d'un bateau se situe entre les prix qui prévalaient en autarcie en France et en Allemagne, soit $p \in (1, 2)$. Quelle est l'offre de bateaux et de voitures dans chacun des deux pays ?

Commençons par nous intéresser à l'Allemagne. Avec 100 heures de travail, le pays peut produire un bateau ou deux voitures. Or, si $p \in (1, 2)$, le prix d'un bateau est inférieur au double de celui d'une voiture. L'Allemagne a donc intérêt à se spécialiser dans l'industrie automobile. Elle produit alors 10 voitures. Chaque voiture peut être exportée en France contre l'importation de $1/p$ bateau, avec $1/p > 1/2$. Si l'Allemagne ne souhaite consommer que des bateaux, elle exporte ses 10 voitures et importe $10/p > 5$ bateaux.

Cette situation est représentée sur la figure 4.4. La ligne noire correspond à la frontière de

production de l'économie allemande (que nous avons déjà vue sur la figure 4.3). En autarcie, il s'agit également de la frontière de consommation. La ligne pointillée bleue correspond à la frontière de consommation de l'économie allemande en économie ouverte lorsque $p < 2$. Ainsi, dès lors que l'Allemagne souhaite consommer quelques bateaux, importés de France et financés par l'exportation de voitures, elle sort gagnante du libre-échange. Son niveau de consommation est supérieur à ce qui aurait été possible en économie fermée.

FIGURE 4.4 – Frontières de consommation en Allemagne en économie fermée (noir) et ouverte (bleu) lorsque $p < 2$

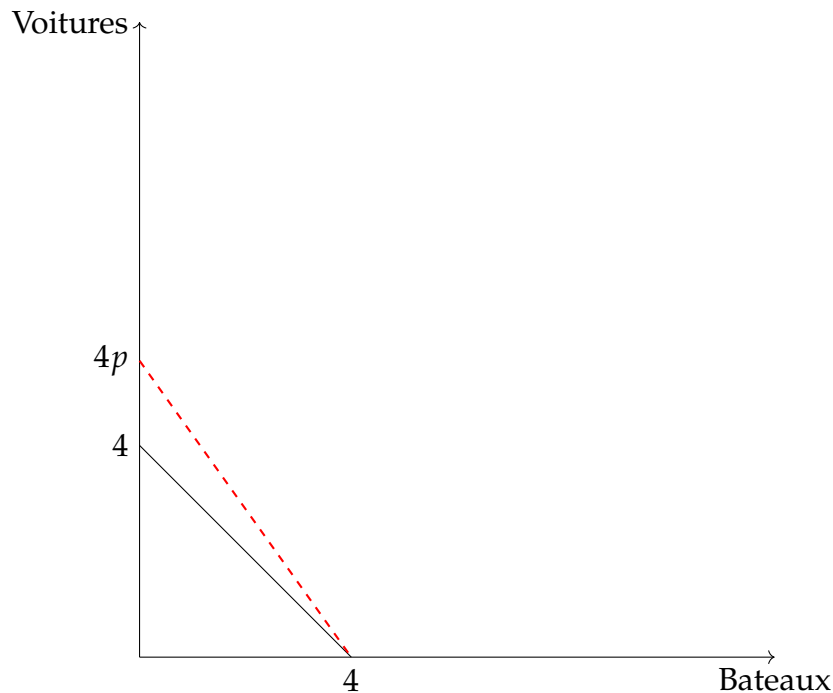


Mais, qu'en est-il de la France? Un bateau y est aussi coûteux à produire qu'une voiture. Or, si $p \in (1,2)$, un bateau a plus de valeur qu'une voiture. La France se spécialise donc dans la construction navale et produit 4 bateaux. Chaque bateau peut être échangé contre p voitures, où $p > 1$. Si la France ne souhaite consommer que des voitures, elle exporte ses 4 bateaux et importe $4p > 4$ voitures. Cette situation est représentée sur la figure 4.5. La frontière de consommation de l'économie française est supérieure en économie ouverte (ligne rouge en pointillé) qu'en économie fermée (ligne noire). Dès lors que la France consomme quelques voitures, elle bénéficie également du libre-échange!

Ainsi, chaque pays se spécialise dans le secteur où réside son avantage comparatif, ce qui permet la réalisation d'échanges mutuellement bénéfiques entre nations. La fermeture des frontières serait ici Pareto-inefficace.

Le niveau du prix p détermine la répartition des gains au commerce : lorsque p est élevé, proche de 2, les Français peuvent acheter plus de voitures et ceux-ci empochent l'essentiel des gains à l'ouverture. Lorsque p est au contraire proche de 1, ce sont les Allemands qui gagnent le plus au libre-échange. Plus l'ouverture des frontières modifie le prix des bateaux, relativement à ce qu'il aurait été en autarcie, plus le pays gagne au libre-échange!

Dans le cas limite où $p = 1$, la France ne se spécialise plus nécessairement dans la pro-

FIGURE 4.5 – Frontières de consommation en France en économie fermée (noir) et ouverte (rouge) lorsque $p > 1$ 

duction de bateaux : qu'elle produise 0 ou 4 bateaux, elle pourra consommer les mêmes quantités de biens. L'ouverture des frontières ne modifiant pas les prix relatifs, la France ne bénéficie pas du libre échange. Dans le cas où $p = 2$, c'est l'Allemagne qui cesse de se spécialiser et de bénéficier du libre-échange.

Détermination du prix unique. Nous venons de déterminer l'offre de bateaux et de voitures à prix donné. Plus précisément, lorsque $p \in]1, 2[$, l'Allemagne produit 10 voitures et la France 4 bateaux. La production totale de l'économie franco-allemande est donc indépendante de p . Lorsque $p = 1$, l'Allemagne se spécialise dans l'industrie automobile, mais la France ne se spécialise pas nécessairement. La production mondiale est donc d'au moins 10 voitures et d'au plus 4 bateaux. À l'inverse, lorsque $p = 2$, c'est la France qui se spécialise, mais pas nécessairement l'Allemagne. La production mondiale est donc inférieure ou égale à 10 voitures et supérieure ou égale à 4 bateaux. A fortiori, si $p < 1$, ni la France, ni l'Allemagne ne produisent de bateau, ce qui donne 14 voitures ; et, si $p > 2$, la production totale de voiture est nulle et celle de bateaux s'élève à 9.

Nous avons donc déterminé pour n'importe quel prix l'offre de bateaux et de voitures. Cela nous permet de construire la courbe d'offre relative de bateaux, c'est à dire le ratio $S(p) = \frac{S_{\text{Bateau}}(p)}{S_{\text{Voiture}}(p)}$ où $S_{\text{Bateau}}(p)$ est l'offre de bateaux et $S_{\text{Voiture}}(p)$ est l'offre de voitures. Ce ratio correspond à la dernière ligne de la table 4.4.

Mais comment est déterminé le prix d'équilibre p de l'économie ouverte ? De manière à ce que l'offre des entreprises soit égale à la demande des consommateurs.

Reste à construire la courbe de demande. Si $D_{\text{Bateau}}(p)$ est la demande de bateaux dans l'économie franco-allemande en fonction du prix p et $D_{\text{Voiture}}(p)$ la demande de voiture,

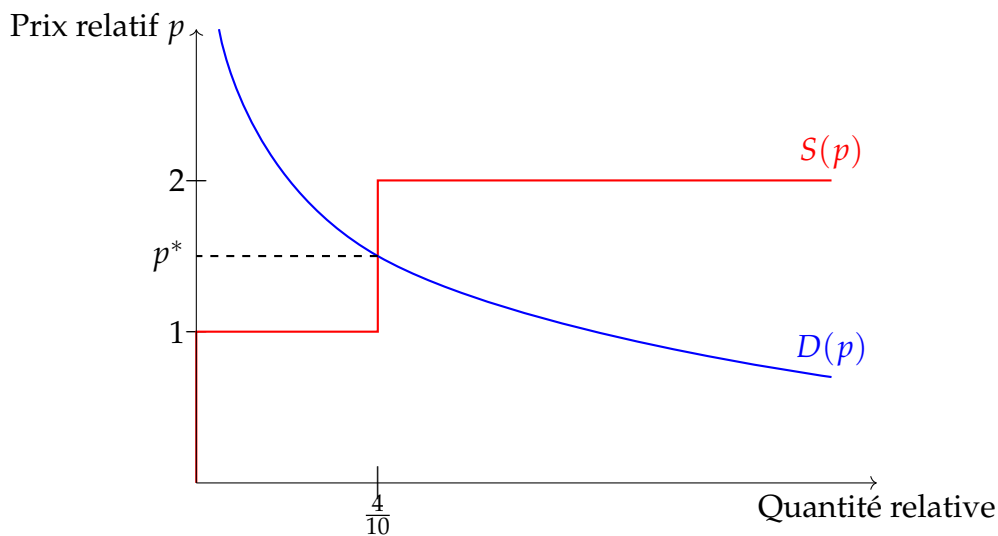
TABLE 4.4 – Offre relative de bateaux selon le prix

	$p < 1$	$p = 1$	$1 < p < 2$	$p = 2$	$p > 2$
Bateaux produits par la France	0	$\in [0, 4]$	4	4	4
Bateaux produits par l'Allemagne	0	0	0	$\in [0, 5]$	5
Voitures produites par la France	4	$\in [0, 4]$	0	0	0
Voitures produites par l'Allemagne	10	10	10	$\in [0, 10]$	0
Quantité relative de bateaux (total)	0	$\in [0, \frac{4}{10}]$	$\frac{4}{10}$	$\in [\frac{4}{10}, +\infty[$	$+\infty$

alors on peut définir la demande relative par $D(p) = \frac{D_{\text{Bateau}}(p)}{D_{\text{Voiture}}(p)}$. Naturellement, la demande de bateaux $D_{\text{Bateau}}(p)$ est une fonction décroissante du prix p des bateaux, tandis que la demande de voitures $D_{\text{Voiture}}(p)$ est une fonction croissante de p . La fonction $D(p)$ est donc décroissante en p .

La figure 4.6 représente les courbes d'offre et de demande relatives, ce qui permet de trouver le prix d'équilibre p^* et la quantité relative de bateaux produite à l'équilibre. Dès lors que le prix d'équilibre se situe entre 1 et 2, les pays sont spécialisés selon leurs avantages comparatifs.³

FIGURE 4.6 – Offre et demande relative de bateaux, et prix relatif à l'équilibre.



Soulignons qu'un accroissement de la demande de bateaux augmente leur prix relatif, ce qui est favorable à la France spécialisée dans la construction navale et défavorable à l'Allemagne spécialisée dans l'industrie automobile. Un pays bénéficie d'autant plus du libre-échange que la demande mondiale est forte pour les biens du secteur dans lequel réside son avantage comparatif.

3. Pourquoi se focaliser sur l'offre et la demande *relative*, plutôt que sur l'offre et la demande sur chacun des deux marchés? Les consommateurs n'ayant le choix qu'entre voitures et bateaux, la détermination de leur demande de bateaux pour un prix donné fixe automatiquement leur demande de voitures. Idem pour l'offre des producteurs. Par conséquent, si le marché des bateaux est à l'équilibre, alors celui des voitures doit l'être aussi! Il serait donc erroné de traiter les deux marchés séparément. Il s'agit là d'une conséquence de la loi de Walras (qui fait partie de la théorie de l'équilibre général) selon laquelle, en présence de n marchés différents, dès lors que $n - 1$ marchés sont à l'équilibre, le n -ième marché doit l'être aussi. C'est la raison fondamentale pour laquelle seuls les prix relatifs sont déterminés, ce qui nous a conduit à normaliser le prix des voitures à 1.

4.1.4 Les enseignements du modèle de Ricardo

Le modèle de Ricardo nous enseigne plusieurs choses fondamentales sur le commerce international et, plus généralement, sur le fonctionnement d'une économie de marché.

La compétitivité. Tout d'abord, les échanges sont mutuellement bénéfiques pour tous les pays qui participent au commerce international, quels que soient leurs niveaux de productivité. Cela est dû au fait que le commerce entre nations n'est pas basé sur les avantages absolus mais sur les avantages comparatifs. Comme chaque pays a au moins un secteur d'activité dans lequel il dispose d'un avantage comparatif, tous bénéficient du commerce international.

Par conséquent, un pays ne peut pas ne pas être compétitif ! La notion de compétitivité est très pertinente à l'échelle d'une entreprise : une firme qui ne l'est pas risque tout simplement de disparaître, au bénéfice de ses concurrents. Cette notion est également utile à l'échelle d'un secteur d'activité. Dans notre exemple, le commerce international anéantit l'industrie automobile française, moins compétitive que l'industrie automobile allemande. Plus généralement, les gains à l'échange sont la conséquence d'une spécialisation nationale, qui se fait forcément au détriment des industries les moins compétitives.

En revanche, la notion de compétitivité ne peut pas être transposée à l'échelle d'une économie dans son ensemble : les travailleurs de l'industrie automobile française ne disparaissent pas ; ils sont simplement réalloués vers la construction navale où ils sont attirés par de meilleurs salaires.

L'innovation technologique. Pour un pays, le commerce international s'apparente à l'introduction d'une nouvelle technologie. Dans notre exemple, du point de vue de la France, l'ouverture au commerce a le même effet que l'invention d'une machine qui permet de transformer des bateaux en voitures. Quand $p > 1$, un bateau est transformé en plus d'une voiture et, quand $p < 1$, une voiture est transformée en plus d'un bateau. Chacune de ces deux possibilités est bénéfique pour la France qui peut se spécialiser dans la production du bien que cette technologie démultiplie. Dans une certaine mesure, l'opposition au libre-échange peut donc s'apparenter à une opposition à l'innovation technologique, à l'image des "luddites" de l'Angleterre de 1811 qui, face à la révolution industrielle, brisèrent les métiers à tisser.

Ce parallèle met en lumière des similitudes entre les effets de l'innovation technologique et de l'ouverture au commerce international. Par exemple, le déclin de l'emploi dans le secteur du textile en Europe résulte à la fois des développements technologiques qui ont remplacé les travailleurs manuels par des machines et du commerce avec la Chine. Les évolutions technologiques et commerciales ont bien souvent des conséquences remarquablement similaires.

Le commerce intranational La notion d'avantage comparatif n'est pas spécifique au commerce entre pays : elle est également pertinente pour expliquer les échanges au sein d'un

pays, ou même d'une entreprise. Prenons l'exemple d'une entreprise dans laquelle deux types de tâches doivent être réalisées : des tâches managériales et des tâches administratives. Deux types d'employés y travaillent : des managers et des assistants. Leurs productivités sont décrites dans la table 4.5.

TABLE 4.5 – Nombre d'heures de travail nécessaires à différentes tâches

	Tâche managériale	Tâche administrative
Manager	2	5
Assistant	20	10

Le manager a un avantage absolu sur son assistant dans les deux types de tâches, mais il a un avantage comparatif dans les tâches managériales.⁴ Il serait donc inefficace que le manager s'occupe de tâches administratives (sauf si l'assistant, spécialisé dans les tâches administratives, est débordé). Le fonctionnement d'une économie de marché aboutit naturellement à ce que chaque travailleur se spécialise dans le domaine où réside son avantage comparatif, ce qui est Pareto-efficace.

Bien que la notion d'avantage comparatif soit généralement présentée dans le contexte du commerce international, elle décrit le fonctionnement de toute économie de marché. Cela implique que le marché est fondamentalement inclusif : chacun y a sa place, quelque soit sa productivité dans chaque secteur d'activité. Bien sûr, dans notre exemple, le manager aura une rémunération plus élevée que l'assistant. Le marché est donc efficace, mais pas forcément très égalitaire, ce qui pose la question des politiques de redistribution.

4.1.5 Remarques méthodologiques

Est-ce que le modèle de Ricardo, avec ses deux pays et ses deux biens, est réaliste ? Non, bien entendu ! Assurément, la réalité des échanges internationaux est infiniment plus complexe que la représentation simpliste que nous venons d'en faire. Pour rendre le modèle plus réaliste, il faudrait y inclure une multitude de dimensions d'hétérogénéité : augmenter le nombre de biens, le nombre de pays, autoriser chaque bien à être différent d'un pays à l'autre, distinguer les travailleurs qualifiés des non-qualifiés, prendre en compte le savoir-faire accumulé par les travailleurs au sein de leurs secteurs d'activité, inclure de multiples frictions allant des coûts de transport aux frais bancaires...

Le problème est que, même en incluant toutes ces dimensions d'hétérogénéité, notre modèle serait toujours infiniment plus simple que la réalité du monde qu'il cherche à décrire. En outre, ce modèle serait très complexe à résoudre et donc aussi obscur que la réalité qu'on cherche à comprendre. Ainsi, un bon modèle en sciences économiques n'est pas réaliste, mais lumineux, éclairant, *insightful*. En outre, même dans un modèle beaucoup plus complexe, le mécanisme des avantages comparatifs restera à l'œuvre et continuera de déterminer les flux d'échanges internationaux. Tout l'intérêt du modèle de Ricardo est d'isoler cet effet afin d'en montrer la logique et la pertinence avec une précision et une clarté mathématique.

4. Cet exemple est évidemment illustratif : nous avons tous en tête des exemples d'assistants qui seraient bien plus efficaces que leurs managers s'ils prenaient leur place.

La question intéressante n'est pas de savoir si le modèle est réaliste, mais si les résultats principaux survivraient à la prise en compte de dimensions d'hétérogénéité supplémentaires qui nous paraissent pertinentes. Autrement dit, quelles hypothèses devons-nous modifier pour invalider la logique des avantages comparatifs? C'est ainsi que se construit la réflexion scientifique dans notre discipline. Il s'agit d'un usage de la modélisation mathématique qui est moins descriptif et plus conceptuel qu'en physique.

Le cœur des sciences économiques consiste en l'aller-retour entre la réalité du monde et l'abstraction de nos modèles mathématiques. Certes, il serait dangereux de confondre un modèle avec la réalité. Mais il faut savoir tirer les enseignements d'une représentation simplifiée de la réalité. C'est là que réside toute la subtilité des sciences économiques.

Ricardo vs. les *corn laws*.

Suite à sa révolution industrielle précoce du XVIII^e siècle, l'Angleterre s'est enrichie en important des denrées agricoles qu'elle échangeait contre des biens manufacturés. Au début du XIX^e siècle, le blocus continental imposé par Napoléon entrava ces échanges, ce qui fit monter le prix des céréales. La défaite de Napoléon en 1815 aurait dû permettre la reprise des échanges, ce qui aurait été défavorable aux propriétaires terriens britanniques. Sous leur influence, le parlement instaura en 1815 les célèbres *corn laws*, mesures protectionnistes limitant fortement l'importation de céréales.

C'est alors qu'en 1817, l'économiste, financier et membre du parlement britannique David Ricardo élaborait sa théorie des avantages comparatifs afin de justifier son soutien au libre-échange et son opposition aux *corn laws*. Ricardo mourut en 1823, mais la pertinence de son analyse contribua largement à la suppression définitive des *corn laws* en 1846, ce qui marqua un tournant dans l'histoire politique britannique.

Le cas des *corn laws* montre que les propriétaires terriens auraient été des perdants du libre-échange. Mais, pour comprendre cette possibilité, il faut aller au-delà de la théorie de Ricardo.

4.2 Gagnants et perdants du libre-échange

Si le libre-échange est si bénéfique, pourquoi est-ce que certaines personnes y sont si farouchement opposées? Il est bien sûr possible que la logique des avantages comparatifs soit difficile à saisir, aboutissant à des craintes injustifiées. Une raison moins prétentieuse est que le libre-échange ne fait pas que des gagnants. Dans notre modèle de Ricardo, au sein d'un pays, tous les consommateurs sont identiques et tous les producteurs le sont également. Mais, si on introduit de l'hétérogénéité, on s'aperçoit alors que le libre-échange est défavorable à un ouvrier français spécialisé dans l'industrie automobile et dont la productivité serait nettement plus faible sur un chantier naval.

Des modèles de commerce international plus avancés permettent d'introduire de l'hétérogénéité entre consommateurs et entre producteurs. Néanmoins, ils sont plus lourds et

plus complexes à résoudre. Pour simplifier notre analyse des gagnants et des perdants du libre-échange, nous allons donc nous focaliser sur le cas d'une petite économie ouverte. Une économie est dite « petite » dès lors que l'ouverture de ses frontières ne peut pas affecter les prix mondiaux.

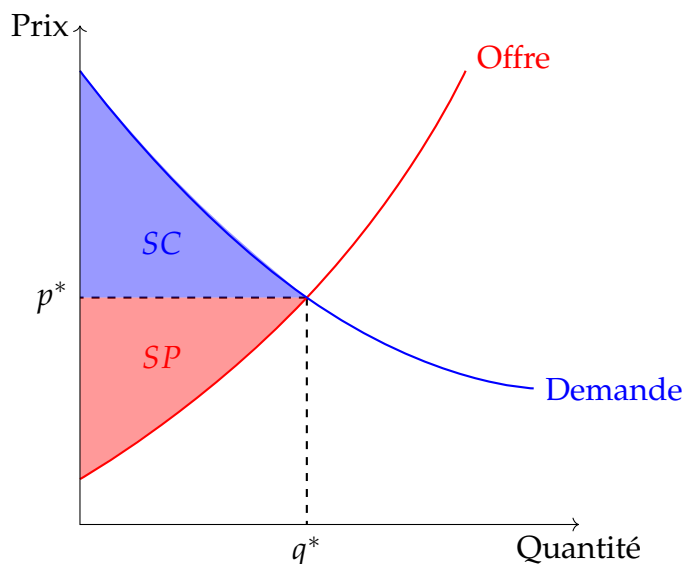
Ainsi, sur le marché du pétrole, la Belgique peut être considérée comme une petite économie ouverte : le fait que la Belgique importe plus ou moins de pétrole a un impact négligeable sur le prix mondial du pétrole. En revanche, lorsque les États-Unis changent le niveau de leurs importations et exportations de pétrole, ils peuvent avoir un impact important sur son prix : l'hypothèse de la petite économie ouverte ne serait donc pas adaptée dans ce cas.

4.2.1 Ouverture des frontières d'une petite économie

Aux chapitres précédents, nous avons introduit la notion de surplus, et avons décrit la répartition du surplus entre consommateurs et producteurs. Voyons comment adapter cette analyse au cas du commerce international.

Considérons donc une petite économie, et un marché dans celle-ci, comme par exemple le marché de l'acier. Avant l'ouverture au commerce, le prix p^* de l'acier en vigueur dans ce pays est déterminé par l'équilibre entre l'offre et la demande, tel que représenté sur la figure 4.8. La surface bleue correspond au surplus des consommateurs et la surface rouge au surplus des producteurs.

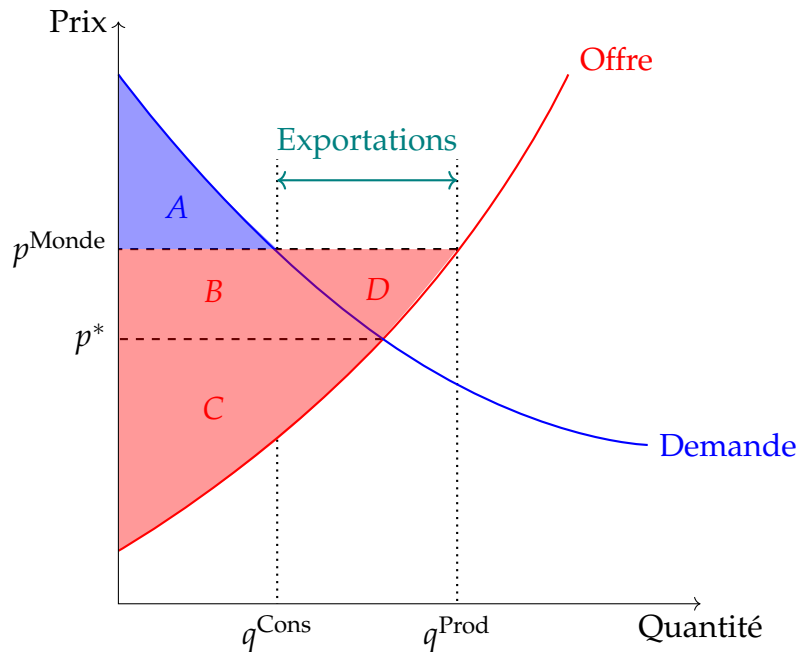
FIGURE 4.8 – Petite économie avant son ouverture au commerce international



Lorsque le pays s'ouvre au commerce, n'importe qui peut acheter ou vendre de l'acier au reste du monde au prix mondial p^{Monde} . Par la loi du prix unique, p^{Monde} doit désormais être le prix de l'acier dans l'économie domestique. En effet, aucun consommateur n'est prêt à payer plus que p^{Monde} pour de l'acier, sachant que c'est le prix auquel il peut s'en procurer à l'étranger, et aucun producteur n'est prêt à le vendre pour moins de p^{Monde} , sachant qu'il peut le vendre à ce prix à des consommateurs étrangers. Le prix p^{Monde} est donc à la fois un plafond et un plancher.

Comment réagit l'économie domestique à ce nouveau prix? Tout dépend du niveau de p^{Monde} relativement à p^* . Si $p^{\text{Monde}} > p^*$, le prix de l'acier augmente pour les consommateurs, qui réduisent donc leur demande au niveau q^{Cons} représenté sur la figure 4.9. À l'inverse, l'augmentation du prix pour les producteurs les pousse à accroître leur production, qui s'établit désormais au niveau q^{Prod} . La différence entre q^{Prod} et q^{Cons} correspond aux exportations d'acier.

FIGURE 4.9 – Petite économie ouverte avec un prix mondial supérieur au prix en autarcie



Quel est l'impact de cette hausse des prix sur le surplus? L'ouverture au commerce diminue le surplus des consommateurs d'acier, qui passe sur la figure 4.9 de $A + B$ à A . Les consommateurs sont donc les grands perdants du libre-échange. En revanche, le commerce augmente nettement le surplus des producteurs, qui passe de C à $B + C + D$. Ils sont donc les gagnants de l'ouverture des frontières. Soulignons que le surplus des producteurs augmente davantage que ne diminue le surplus des consommateurs. Ainsi, grâce au commerce, le surplus total de l'économie augmente de D , qui représente le bénéfice national du libre-échange.

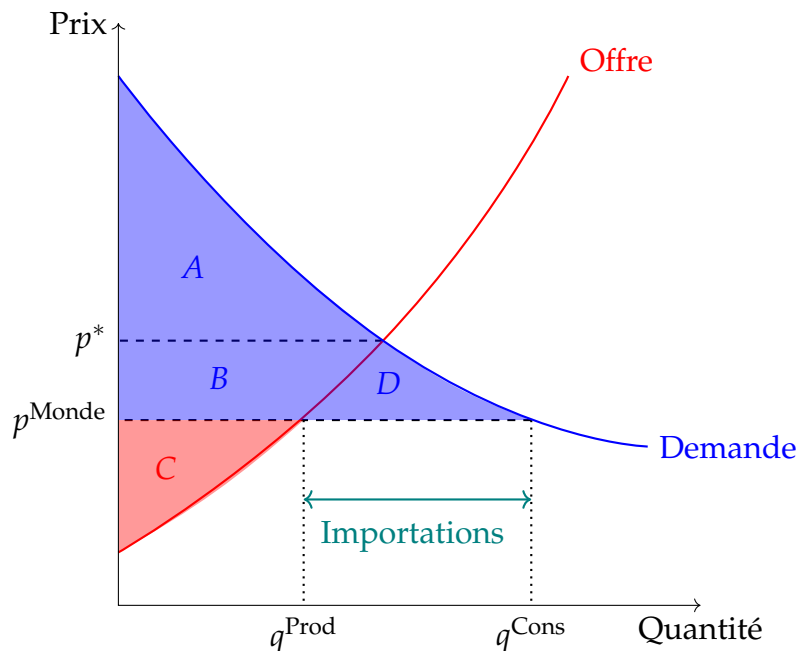
En théorie, il est donc possible d'effectuer un transfert de richesse des producteurs aux consommateurs afin de compenser ces derniers, de sorte que tout le monde gagne à l'ouverture des frontières. Il faut pour cela que le transfert soit compris entre B et $B + D$. Cette possibilité implique que le protectionnisme est Pareto-inefficace : en partant d'une situation d'autarcie, il est possible d'ouvrir les frontières et de redistribuer les gains à l'échange de telle manière que tout le monde préfère cette nouvelle situation à l'ancienne. Autrement dit, l'ouverture des frontières augmente le surplus total de l'ensemble des acteurs de l'économie, si bien qu'il serait inefficace d'y renoncer. Ce résultat explique pourquoi les économistes sont généralement si défavorables aux barrières douanières.

Ceci étant, lorsque la redistribution des gagnants vers les perdants n'est pas possible, alors le protectionnisme peut être Pareto-efficace : en partant d'une situation d'autarcie, pour

augmenter le bien-être des gagnants potentiels du libre-échange, il faut alors nécessairement diminuer celui des perdants.

Dans le cas où le prix p^{Monde} est inférieur au prix d'autarcie p^* , nous sommes dans une situation symétrique. Le prix de l'acier est désormais inférieur à ce qu'il était auparavant, ce qui pousse les entreprises à réduire leur production au niveau q^{Prod} représenté sur la figure 4.10⁵. Les consommateurs augmentent leur demande au niveau q^{Cons} . L'écart entre q^{Cons} et q^{Prod} correspond aux importations. Dans ce cas, les consommateurs sont les gagnants de l'ouverture des frontières, puisque leur surplus passe de A à $A + B + D$, tandis que les producteurs sont les perdants, leur surplus se réduisant de $B + C$ à C . Mais, à nouveau, les gagnants gagnent plus que les perdants ne perdent, si bien qu'il est possible pour les gagnants de compenser les perdants afin que tout le monde préfère le libre-échange à l'autarcie.

FIGURE 4.10 – Petite économie ouverte avec un prix mondial inférieur au prix en autarcie



Cette modélisation simple permet d'avoir une première idée des gagnants et perdants du commerce international : l'ouverture au commerce bénéficie aux consommateurs des produits dont elle diminue le prix, au détriment de leurs producteurs, et bénéficie aux fabricants des produits dont elle augmente le prix, au détriment de leurs consommateurs. Nous avons vu dans le modèle de Ricardo que l'ouverture au commerce diminue les prix des produits que le pays produit relativement moins efficacement, tandis qu'il augmente le prix des produits bénéficiant d'un avantage comparatif. Ainsi, les secteurs où résident les avantages comparatifs gagnent au libre-échange, tandis que les autres perdent. La France a un avantage comparatif dans le domaine du tourisme, du luxe, de l'industrie aéronautique... Ces secteurs se sont effectivement largement développés avec l'ouverture internationale. En

5. Cet ajustement peut s'interpréter de deux manières différentes : si on considère toutes les entreprises similaires et avec des coûts marginaux croissants, elles s'ajustent simplement en réduisant leur production. Dans le cas plus vraisemblable où les entreprises ont des productivités différentes, l'ouverture au commerce mène à la faillite des entreprises aux coûts les plus élevés.

gouvernement sont nulles, et la perte sèche est égale à l'ensemble des gains du libre échange qui ne sont donc pas réalisés.

La présence de cette perte sèche marque une nouvelle fois l'inefficacité des mesures protectionnistes. Notre analyse montre que, pour une petite économie, la politique optimale consiste à ouvrir *unilatéralement* ses frontières.

Le *China shock*

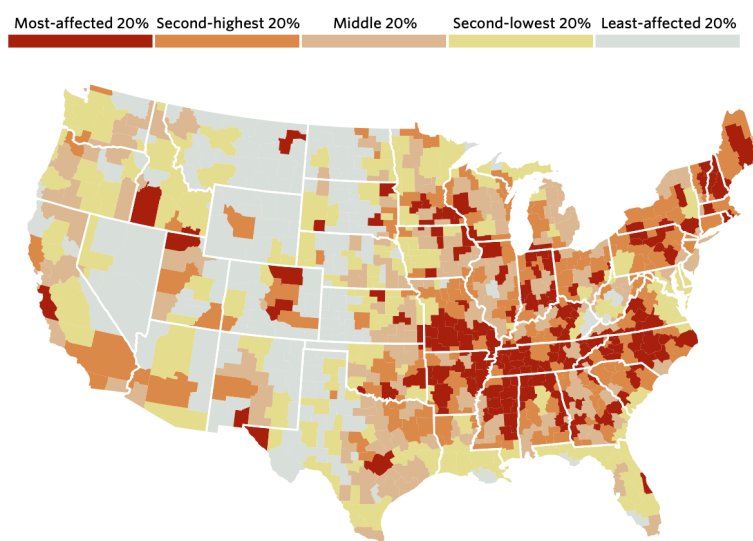
Un des phénomènes économiques les plus impressionnants des dernières décennies est la croissance fulgurante de la Chine. Entre 1990 et 2020, soit en une génération, le niveau de vie en Chine a été multiplié par près de dix. Ce développement vertigineux n'a pas seulement eu un impact sur la population chinoise : avec le commerce international, c'est le monde entier qui a été touché. Alors que la Chine ne produisait que 5% des biens manufacturés en 1990, ce chiffre atteignait près de 30% en 2020. Une grande part de cette production est exportée, en particulier depuis que la Chine a rejoint l'Organisation Mondiale du Commerce (OMC) en 2001.

Quelles ont été les conséquences de cette explosion des exportations chinoises, connue sous le nom de *China shock*? L'impact a été localisé géographiquement. En effet, les industries ont tendance à se concentrer dans certaines zones géographiques (le textile dans le Nord de la France, les nouvelles technologies dans la Silicon Valley, etc.) et les zones qui ont le plus souffert de la concurrence chinoise sont celles où étaient concentrées les industries qui avaient un désavantage comparatif par rapport à la Chine. Aux États-Unis, le *China shock* a été particulièrement important dans le Sud-Est du pays (voir la Figure 4.12). L'emploi manufacturier y a baissé significativement.

À long terme, les emplois détruits ont été recréés ailleurs : les travailleurs ont été réembauchés dans d'autres secteurs ou dans d'autres régions, lesquels ont d'ailleurs pu bénéficier de l'accroissement du commerce avec la Chine. Mais, malheureusement, cet ajustement a pris du temps. Les recherches de Autor et al. (2016) ont montré que, à court terme, cette réallocation s'est accompagnée d'une hausse du chômage, d'une baisse de la participation au marché du travail, et même d'une hausse de la mortalité (en particulier d'overdoses et des conséquences de l'alcoolisme). Les travailleurs les plus pauvres, ayant moins de mobilité géographique ou plus de difficultés à changer de secteur, ont été particulièrement touchés. Il a fallu plus d'une décennie pour que le choc soit résorbé. À court terme les coûts de cette transition ont été du même ordre de grandeur que les gains de l'ouverture au commerce. Sur le long terme, les États-Unis semblent avoir gagné de l'ouverture au commerce avec la Chine, et même en prenant en compte les coûts d'ajustement, environ 95% de la population sort gagnante du *China shock*.

L'irruption de la Chine dans le jeu du commerce mondial montre que, lorsqu'on passe d'un équilibre à un autre, les coûts d'ajustement sont substantiels. Ces coûts résultent de multiples frictions sur le marché des biens et sur le marché du travail qui freinent la réallocation des ressources vers les secteurs bénéficiant du libre échange.

FIGURE 4.12 – Exposition des différentes zones d’emploi aux États-Unis aux exportations chinoises



Source : Carte de Andrew Van Dam et Jessia Ma, utilisant les données de David Autor, David Dorn et Gordon Hanson.

4.3 Autres effets du libre-échange

Jusqu’ici, nous avons présenté deux modèles qui permettent de comprendre les principaux ressorts du commerce international ainsi que leurs conséquences pour les consommateurs et pour les producteurs. Le commerce entre nations a bien sûr d’autres effets que ceux mis en évidence par nos modèles de base. Quels sont-ils ? Que pouvons-nous en déduire pour la politique commerciale ? Petit tour d’horizon.

4.3.1 Bénéfices du libre-échange

L’accès à plus de variété. Dans les secteurs caractérisés par d’importantes économies d’échelle, une production de masse est nécessaire pour minimiser les coûts unitaires de fabrication. Par exemple, le coût de la conception d’une automobile et de la mise en place d’une chaîne de production est si élevé que le marché français peut difficilement avoir plus de deux ou trois constructeurs, avec à la clef un choix limité pour les consommateurs. L’ouverture des frontières permet à un bien plus grand nombre de constructeurs de co-exister dans le gigantesque marché mondial. Les entreprises ont alors intérêt à diversifier leurs produits, offrant beaucoup plus de choix aux consommateurs. Ainsi, un Français n’est pas réduit à choisir entre Renault et Peugeot, mais peut également acheter une BMW, une Ford ou une Toyota.

En 1979, l’américain Paul Krugman fut le premier à formaliser ce type d’échanges internationaux, ce qui lui valut le prix Nobel en 2008. Cette nouvelle théorie du commerce permet de comprendre pourquoi l’essentiel des échanges s’effectue entre des nations dont les niveaux de productivité sont similaires, ce que la théorie des avantages comparatifs de Ricardo explique difficilement. Cette théorie explique pourquoi l’Allemagne exporte des

voitures en France, alors que dans le même temps la France exporte des voitures en Allemagne. La raison est toute simple : les voitures françaises et allemandes ne sont pas les mêmes ! Certains consommateurs français préfèrent les voitures allemandes et vice versa.

Ainsi, la théorie de Krugman explique les échanges Nord-Nord, là où la théorie de Ricardo explique essentiellement les échanges Nord-Sud. L'ironie est qu'à partir des années 1980, la Chine a progressivement rejoint l'économie mondiale, générant de plus en plus d'échanges basés sur des avantages comparatifs. Soulignons que ces deux visions du commerce international ont exactement la même implication en termes de politique économique : ouvrir les frontières afin de permettre tous les échanges mutuellement bénéfiques de se réaliser.

La réduction des prix par l'intensification de la concurrence. Nous reviendrons bientôt sur une défaillance de marché majeure : la concurrence imparfaite. Nous verrons que lorsqu'un petit nombre d'entreprises contrôle une grande partie de la production, elles imposent des prix trop élevés. Le commerce international augmente la taille du marché, ce qui intensifie la concurrence et diminue les prix. Cette concurrence accrue pousse également les entreprises à innover davantage.

Face à une concurrence insuffisante, l'État cherche généralement à réguler les entreprises. Mais comme le soulignait l'économiste libéral Milton Friedman, bien souvent, la meilleure régulation est tout simplement l'ouverture des frontières.

Néanmoins, il existe des domaines dans lesquels cet argument ne tient pas, et où au contraire l'ouverture au commerce international diminue la concurrence, accentuant une défaillance de marché. C'est particulièrement le cas dans l'industrie numérique où la libéralisation du commerce a permis à des entreprises comme Google, Amazon ou Facebook de dominer le marché mondial grâce à un effet de réseau. Plus une entreprise a un réseau qui est dense, plus elle est incontournable ! La fermeture de la Chine à ces entreprises occidentales a permis l'émergence d'équivalents nationaux comme Baidu, Alibaba ou Tencent, ce qui évite aux Chinois de payer les firmes américaines pour utiliser leurs services.

La circulation des idées. L'ouverture au commerce ne permet pas seulement d'échanger des biens : avec ces biens voyagent des personnes, et avec elles des idées. La diffusion d'innovations par le commerce n'est pas nouvelle : la route de la soie a amené de nombreuses technologies chinoises en Europe, à l'image de la boussole ou de la poudre. Plus récemment, c'est bien grâce à son intégration à l'économie mondiale que la Chine a connu au cours des dernières décennies une spectaculaire hausse de sa productivité. À l'inverse, les pays victimes d'embargos internationaux, comme Cuba ou la Corée du Nord, sont condamnés à être éloignés de la frontière technologique.

4.3.2 Dangers du libre-échange

Malgré tous ces bénéfices engendrés par le libre-échange, il existe également un certain nombre de dangers ou d'effets indésirables, justifiant des restrictions au commerce interna-

tional.

L'industrie naissante. Imaginons que l'Égypte dispose d'ingénieurs talentueux, ayant le potentiel de développer un avantage comparatif dans le secteur automobile. Cependant, les premières voitures sorties des usines égyptiennes seront probablement de qualité inférieure à celles proposées par la concurrence étrangère. Plusieurs années seront nécessaires pour réduire l'écart de productivité avec les grandes multinationales du secteur. Pour permettre à l'industrie automobile égyptienne de se lancer, il serait donc souhaitable de la protéger pendant qu'elle se développe. Ainsi, les consommateurs égyptiens seraient initialement obligés d'acheter des voitures produites dans leur pays, dont le rapport qualité-prix serait médiocre. Mais, petit à petit, ces voitures s'amélioreraient, jusqu'à ce qu'elles soient en mesure d'affronter la concurrence internationale.

Cet argument de l'industrie naissante est particulièrement important pour les pays en développement et pour les industries avec un rendement d'échelle croissant. Il ne justifie néanmoins que des barrières temporaires au libre-échange.

Les difficultés de la redistribution. En théorie, le gouvernement peut mettre en place des transferts des gagnants vers les perdants du libre-échange afin que chaque individu bénéficie de l'ouverture des frontières. Le protectionnisme serait donc Pareto inefficace. En pratique, cette redistribution est difficile à effectuer. Il est souvent compliqué d'identifier quels individus ont bénéficié et quels autres ont été victimes du libre-échange. En outre, les politiques de redistribution peuvent être politiquement difficiles à mettre en place. L'ouverture des frontières fait donc nécessairement des perdants qui ne peuvent pas être compensés. Dans ce cas, le protectionnisme n'est plus nécessairement inefficace. On peut alors avoir un compromis entre la prospérité de l'économie nationale et le niveau des inégalités : la fermeture des frontières réduit cette prospérité mais, lorsque les perdants sont des travailleurs peu qualifiés, elle augmente les revenus des plus modestes.

La protection environnementale. L'Union Européenne impose des normes environnementales strictes à ses entreprises. Ces normes créent un avantage comparatif des pays étrangers dans les activités polluantes. Si la pollution est locale, alors le libre-échange ne pose aucun problème : chaque pays peut librement décider de ses propres normes environnementales en fonction du coût local de la pollution et du bénéfice de l'exportation des biens dont la production dégrade l'environnement. En revanche, si l'impact de la pollution est mondial, comme c'est le cas du réchauffement climatique, alors le libre-échange réduit considérablement l'intérêt de nos normes environnementales. Leur principal effet n'est pas de protéger la planète, mais de miner la compétitivité de nos secteurs industriels les plus polluants. Cela aboutit à ce que les gaz à effet de serre soient émis en Chine plutôt qu'en Europe – le libre-échange engendre des « fuites de carbone ». Une réponse partielle au problème consiste à mettre en place une taxe carbone aux frontières, mais cette politique est difficile à concevoir et à administrer. Ceci étant, comme nous le verrons dans le chapitre suivant, la politique

commerciale a potentiellement un rôle considérable à jouer dans la résolution planétaire du problème du réchauffement climatique.

La dépendance géostratégique. Le libre échange conduit chaque pays à se spécialiser dans certains secteurs. Les consommateurs importent donc les biens qui ne sont plus produits domestiquement. Un pays doit veiller à ne pas se retrouver en situation de dépendance face à un partenaire commercial. Ainsi l'Allemagne, qui importait près de la moitié de son gaz de Russie, s'est retrouvée en grande vulnérabilité lorsque le conflit russo-ukrainien s'est considérablement aggravé en 2022. De même que, lors de la pandémie de Covid-19, la capacité de certains pays à produire des médicaments ou des vaccins leur permettait d'en bénéficier en priorité.

Ceci étant, il est pratiquement impossible de parvenir à l'autosuffisance nationale dans tous les secteurs stratégiques. La solution consiste plutôt à diversifier ses sources d'approvisionnement afin d'éviter de se retrouver en situation de vulnérabilité et de dépendance stratégique.

La perte de souveraineté nationale. Il existe deux types de barrières douanières : les barrières tarifaires, dont nous avons déjà parlé et qui sont essentiellement des taxes sur les importations, et des barrières non-tarifaires, ou réglementaires, qui servent à limiter la concurrence étrangère. Un exemple est donné par les politiques mises en place par Colbert sous Louis XIV pour réglementer le marché du textile : pour limiter les importations, des règles arbitraires étaient imposées sur les manufactures françaises. Ainsi, les tissus de Châtillon devaient comporter 1216 fils, ceux d'Auxerre 1376, etc. Les tissus ne respectant pas ces contraintes ne pouvaient être échangés, barrant la route aux producteurs étrangers. Les traités de libre-échange sont donc généralement très longs car les différents pays harmonisent les réglementations qui s'appliquent aux produits échangés. D'ailleurs, le socle de l'Union Européenne est le "marché unique" qui définit une réglementation commune afin de permettre à chaque entreprise de vendre, sans entrave, aux 450 millions de consommateurs du marché européen.

Certaines normes ne sont pas seulement des mesures protectionnistes mais relèvent d'un choix de souveraineté nationale. Par exemple, l'Europe a opté pour une agriculture plus naturelle et moins industrielle que les États-Unis. Ainsi, l'Europe encadre plus étroitement le recours aux organismes génétiquement modifiés et interdit le bœuf aux hormones de croissance ou le lavage des poulets au chlore, pratiques répandues outre-Atlantique. Il existe donc un arbitrage à faire entre le commerce international et la souveraineté nationale en matière réglementaire.

Commerce et volatilité économique

La théorie des avantages comparatifs implique que l'ouverture des frontières pousse les pays à se spécialiser dans certaines industries. Or, nos économies sont constamment

sujets à des chocs qui affectent certains secteurs d'activité plus que d'autres, dus par exemple aux évolutions de la technologie ou à la raréfaction de certaines matières premières. On peut donc craindre qu'une économie très spécialisée soit moins résiliente et plus volatile face aux aléas de l'activité économique.

Ceci étant, le commerce est aussi une source d'assurance pour un pays. Outre les chocs sectoriels, il existe également des chocs macroéconomiques qui affectent l'ensemble de l'économie nationale, comme par exemple lors d'une crise financière ou politique. En ce cas, quand l'économie domestique ne parvient plus à produire suffisamment, le commerce permet de s'approvisionner à l'étranger. De même, quand les consommateurs domestiques ne dépensent pas suffisamment, le commerce permet de vendre à l'étranger. Ainsi, la diversification des partenaires commerciaux stabilise l'activité économique.

Caselli et al. (2020) ont utilisé une base de données regroupant 25 pays pour étudier empiriquement ces différents effets. Ils ont montré que, en moyenne, la réduction des barrières douanières entre 1972 et 2007 a *diminué* la volatilité des revenus nationaux de 36%. La diversification des échanges domine donc l'effet de spécialisation sectorielle, si bien que le commerce est un facteur de stabilité économique.

4.4 Quantifier les bénéfices du libre-échange

Quels gains observe-t-on lorsque des biens peuvent circuler librement entre plusieurs régions ? A la fin du XIX^e siècle et au début du XX^e siècle, les coûts de transport ont été drastiquement réduits, notamment grâce au développement du chemin de fer. Quelle a été la conséquence économique de cette innovation facilitant les échanges ? Nous devrions observer davantage de spécialisation et des gains liés au commerce qui en résulte. Les économistes Arnaud Costinot et Dave Donaldson se sont intéressés au cas particulier de la production agricole aux États-Unis (Costinot and Donaldson, 2016). Avec des données remontant jusqu'à 1880 sur la production agricole d'une large palette de produits dans les différentes régions américaines, ils ont vérifié que les régions américaines se sont petit à petit spécialisées dans la production d'espèces dans lesquelles elles avaient un avantage comparatif.

Si le commerce entre régions américaines n'avait pas été facilité entre 1880 et aujourd'hui par les innovations de transport, quelle serait la production agricole américaine aujourd'hui ? La différence entre cette estimation et la production agricole actuelle correspond aux gains engendrés par la baisse des coûts de transport. Costinot et Donaldson trouvent que, depuis 1880, la production agricole américaine a autant augmenté grâce à la facilitation des flux de marchandises sur le territoire que grâce aux innovations technologiques ! Les gains à la spécialisation au sein du pays ont donc été considérables. Ces gains de productivité se répercutent sur les revenus individuels : exploitant des données historiques sur le développement des réseaux de chemin de fer en Inde, Donaldson (2018) montre que les villages indiens ayant obtenu un accès au rail ont vu leur revenu moyen augmenter significativement. Ce type de gains peut expliquer pourquoi une grande partie des aides au

développement servent à développer des infrastructures.

Ces résultats empiriques nous aident à comprendre les gains au commerce *au sein* d'un pays. Qu'en est-il du commerce international à proprement parler ? Costinot and Rodríguez-Clare (2018) ont tenté d'estimer les gains au commerce extérieur des États-Unis en se posant la question suivante : combien les Américains devraient-ils déboursier pour maintenir leur niveau de vie si toutes les frontières extérieures du pays venaient à fermer ? Leur réponse est de 2 à 8% du Produit Intérieur Brut (soit 2 à 8% de la richesse créée par l'économie américaine en une année). Ces gains au commerce international sont relativement importants, mais très inférieurs aux gains engendrés par le commerce intranational. Cela s'explique par le fait que les États-Unis sont un immense pays, capable de produire un grand nombre de biens différents. Ainsi, les importations américaines ne représentaient que 15% de leur PIB en 2018. Pour un plus petit pays, comme la Belgique, dont les importations représentaient 81% du PIB en 2018, les gains au commerce sont certainement bien plus importants. En outre, ces estimations font abstraction des gains de productivité engendrés par l'échange d'idées qui accompagnent le commerce international.

4.5 Déficits commerciaux

Le mercantilisme est un courant de pensée important du XVI^e au XVIII^e siècle selon lequel une bonne politique économique vise à dégager un surplus de la balance commerciale, en ayant si nécessaire recours au protectionnisme. Jusqu'à présent, notre étude du commerce international n'apporte aucun éclairage à ce sujet car, dans le modèle de Ricardo, la seule et unique manière de financer les importations est d'exporter. Par conséquent, la balance commerciale y est toujours équilibrée.

Pour étudier les surplus et les déficits de la balance commerciale, il faut généraliser le modèle de Ricardo afin d'obtenir une représentation dynamique du commerce international. Ce type de modèle est au cœur d'une branche de notre discipline, appelée la *macroéconomie internationale*, qui s'intéresse aux liens entre les flux commerciaux et les flux financiers internationaux.

Ceci étant, même dans un contexte dynamique, les enseignements de l'approche statique ne sont pas fondamentalement remis en cause car la seule et unique manière de financer les importations est toujours d'exporter, cette année... ou plus tard. Ainsi, dans un modèle dynamique, la balance commerciale doit nécessairement être équilibrée à long terme : les surplus d'une année sont forcément compensés par des déficits futurs, et vice versa. Comme l'avait expliqué David Hume dès le XVIII^e siècle, cette simple réalité remet fondamentalement en cause la pertinence du mercantilisme.

Bibliographie

- Autor, D. H., Dorn, D. and Hanson, G. H. (2016), 'The china shock : Learning from labor-market adjustment to large changes in trade', *Annual Review of Economics* **8**, 205–240.
- Bernhofen, D. M. and Brown, J. C. (2004), 'A direct test of the theory of comparative advantage : The case of japan', *Journal of Political Economy* **112**(1), 48–67.
- Caselli, F., Koren, M., Lisicky, M. and Tenreyro, S. (2020), 'Diversification through trade', *Quarterly Journal of Economics* **135**(1), 449–502.
- Costinot, A. and Donaldson, D. (2016), How large are the gains from economic integration? theory and evidence from us agriculture, 1880-1997, Technical report, National Bureau of Economic Research.
- Costinot, A. and Rodríguez-Clare, A. (2018), 'The us gains from trade : Valuation using the demand for foreign factor services', *Journal of Economic Perspectives* **32**(2), 3–24.
- Donaldson, D. (2018), 'Railroads of the raj : Estimating the impact of transportation infrastructure', *American Economic Review* **108**(4-5), 899–934.

Exercices du chapitre 4

Exercice 1 : La conscription

Supposons qu'un pays est peuplé de 50 travailleurs qualifiés et de 50 non-qualifiés. Un travailleur qualifié peut soit produire 6 unités de biens industriels, soit 2 unités de défense nationale; un non-qualifié peut soit produire 2 unités de biens industriels, soit 1 unité de défense nationale.

A/ Montrez que les non-qualifiés possèdent un avantage comparatif dans la production de défense.

B/ Supposons que ce pays décide de produire 30 unités de défense. Ce pays aura-t-il plus de biens à consommer si la défense est produite par les qualifiés ou par les non-qualifiés? Ce résultat est-il surprenant? Expliquez.

C/ Supposons que le gouvernement instaure la conscription (le service militaire obligatoire) en désignant au hasard 10 qualifiés et 10 non-qualifiés pour la production de défense (qui, ensemble, produiront 30 unités de défense). Quelle est, dans ce cas, la production de biens industriels?

D/ Une alternative à la conscription est l'armée volontaire où l'Etat paye les non-qualifiés suffisamment pour convaincre 30 d'entre eux d'être soldat. Comparez le coût d'opportunité (en terme de biens industriels) de la production de 30 unités de défense de ces deux options : la conscription et l'armée volontaire.

E/ Le modèle de cet exercice est-il idiot?

Exercice 2 : Le modèle Ricardien

Reprenons le modèle Ricardien du cours où le nombre d'heure de travail nécessaire à la production de voitures et de bateaux est donné par :

	Voiture	Bateau
Allemagne	50	100
France	125	125

Chaque pays dispose de 500 heures de travail. On suppose qu'initialement la demande mondiale (franco-allemande) est telle que, en économie ouverte, le prix relatif des bateaux est entre 1 et 2, i.e. $P_{\text{Bateau}}^{\text{Inter}} / P_{\text{Voiture}}^{\text{Inter}} \in (1, 2)$.

A/ Calculez le salaire horaire d'un travailleur allemand et français dans chaque secteur d'activité. Dans quels secteurs vont-ils choisir de se spécialiser? Est-ce que les travailleurs allemands gagnent plus que les travailleurs français?

B/ Expliquez la construction de la figure 4.6, ci-dessus.

C/ Une forte augmentation de la demande de bateaux aboutit à ce que le prix relatif d'un bateau $P_{\text{Bateau}}^{\text{Inter}} / P_{\text{Voiture}}^{\text{Inter}}$ soit égal à 2. Quelle est l'impact de cette évolution sur chacun des deux pays? Expliquez intuitivement ce résultat.

D/ Supposons à nouveau que le prix relatif des bateaux est initialement compris entre 1 et 2, soit $P_{\text{Bateau}}^{\text{Inter}} / P_{\text{Voiture}}^{\text{Inter}} \in (1, 2)$, et que, soudainement, l'Allemagne bénéficie d'un doublement sa productivité dans la construction automobile et dans la construction navale. Quel est l'impact de cette évolution sur l'économie française? Expliquez intuitivement ce résultat.

E/ Supposons maintenant que l'Allemagne bénéficie d'un doublement de sa productivité uniquement dans la construction navale. Quel est l'impact de cette évolution sur l'économie française? Expliquez intuitivement ce résultat.

F/ Dépassons désormais le modèle Ricardien en supposant qu'il existe un tout petit nombre de travailleurs allemands (avec des productivités horaires allemandes) dans l'économie française. Ces travailleurs sont-ils favorable à l'ouverture des frontières? Expliquez.

G/ Pourquoi est-ce que le libre-échange ne fait-il jamais de perdants dans le modèle Ricardien?

Exercice 3 : Les perdants de la mondialisation

A/ L'Europe exporte des avions en Chine et importe des T-shirts. Pourquoi est-ce que, en Europe, les travailleurs peu qualifiés pâtissent du libre-échange? Quels hypothèses du modèle Ricardien devez-vous relâcher afin de répondre à cette question?

B/ Est-il possible pour les gagnants du libre-échange de compenser les perdants, de sorte que l'ouverture des frontières ne nuise à personne? Expliquez.

Exercice 4 : Compensations et efficacité

A/ L'existence de perdants du libre-échange justifie-t-elle le protectionnisme? Quelle hypothèse modifie votre conclusion? Expliquez.

B/ Le libre-échange sans compensations est-il Pareto efficace? Expliquez.

C/ Doit-on compenser les perdants du libre-échange?

Chapitre 5

Économie publique

L'économie publique est la branche des sciences économiques qui s'intéresse au rôle de l'État : quels sont les biens et services que celui-ci doit fournir aux citoyens ? Comment doit-il les fournir ? Quelles taxes et impôts doit-il lever pour financer cette action ? Comment redistribuer les richesses de manière efficace ?

Dans ce chapitre, nous allons d'abord nous intéresser à la fiscalité avec deux thèmes principaux :

1. L'**incidence fiscale** : quel est l'impact d'un impôt sur les différents acteurs de l'économie ? Autrement dit, qui paye l'impôt ?
2. Les **distorsions fiscales** : quelles sont les conséquences de la fiscalité sur l'efficacité économique ?

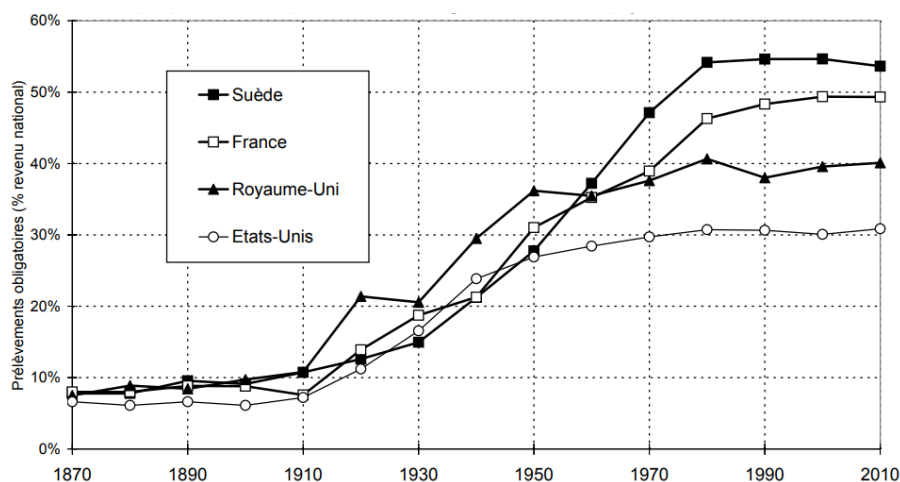
Nous étudierons ensuite deux des principales défaillances de marché : les **externalités** et les **biens publics**. Nous identifierons précisément la nature du problème, ce qui nous permettra ensuite de déterminer comment l'État peut y remédier. Nous nous intéresserons à une application importante de ces défaillances de marché : le réchauffement climatique.

5.1 Les ressources et l'action de l'État

Commençons par un bref tour d'horizon de l'importance de la puissance publique dans nos économies occidentales. Concentrés au début du XIX^e siècle sur des fonctions régaliennes, comme la défense militaire ou la justice, les États ont depuis lors joué un rôle croissant dans la vie économique et sociale des pays, assurant aujourd'hui un rôle prépondérant dans l'éducation, la santé ou la construction d'infrastructures. Les États ont aussi acquis un rôle important dans la redistribution des richesses, avec la mise en place d'impôts et de subventions conditionnés aux revenus individuels. Ce rôle accru de l'État s'est traduit par une augmentation considérable des prélèvements obligatoires au cours du XX^e siècle (voir figure 5.1).

L'impôt représente donc désormais une part importante du Produit Intérieur Brut (qui mesure la valeur de la production nationale sur une année). Aujourd'hui, les recettes fiscales s'élèvent à près de 33% du PIB dans les pays riches, avec une grande variation entre des

FIGURE 5.1 – Recettes fiscales en pourcentage du revenu national, 1870-2010.



Source : Piketty, T. (2014), *Capital in the 21st century*, Harvard University Press.

pays comme les États-Unis, où les recettes fiscales s'élevaient à 25% du PIB en 2019, et la France, où elles atteignaient 45% du PIB au même moment. Ces prélèvements obligatoires sont constitués d'une multitude d'impôts et de taxes, que l'on peut classer selon quelques grandes catégories rapportées dans la table 5.1.

TABLE 5.1 – Recette fiscales en 2019, en pourcentage du Produit Intérieur Brut.

	France	Allemagne	Royaume-Uni	États-Unis
Impôts sur le revenu	9,3%	10,5%	9,0%	10,3%
Cotisations sociales	16,6%	14,6%	6,6%	6,1%
Taxes sur la consommation	12,4%	10,2%	10,7%	4,4%
Impôts sur la propriété	3,9%	1,1%	4,1%	2,9%
Impôts sur les sociétés	2,2%	2,0%	2,3%	1,3%
Autres taxes	0,6%	0,2%	0,2%	0,0%
Total	44,9%	38,6%	32,8%	25,0%

Source : Revenue Statistics 2021, OECD.

En France, quels sont les principaux impôts et taxes qui rentrent dans chacune des ces catégories ?

1. Le premier impôt sur le revenu est la Contribution Sociale Généralisée (CSG) avec un taux unique de près de 10% payé par tous les français. Il rapporte près de 5,3% du PIB, tandis que l'Impôt sur le Revenu (IR), dont le barème est fortement progressif, ne représente que 3,1% du PIB.
2. Les cotisations sociales, également appelées charges sociales, s'apparentent à un impôt sur le revenu mais servent à financer les prestations sociales. On distingue les *cotisations contributives*, qui ouvrent des droits croissants avec le niveau de cotisations payées (comme pour les retraites ou l'assurance chômage), et les *cotisations non-contributives*, qui financent des prestations dont la générosité ne dépend pas du niveau des cotisations (comme pour la santé ou les allocations familiales). Les cotisations patronales

sont à la charge de l'employeur, et représentent près de 75% des charges, tandis que les cotisations salariales sont à la charge de l'employé. Comme le montre la table 5.1, les charges sociales sont particulièrement élevées en France.

3. La principale taxe sur la consommation est la Taxe sur la Valeur Ajoutée (TVA) qui rapporte 7,2% du PIB. Cet impôt largement répandu dans les pays riches (à l'exception notable des États-Unis) fut inventé en 1954 par Maurice Lauré, polytechnicien et haut fonctionnaire français. Il existe aussi d'autres taxes sur la consommation, notamment sur l'essence, l'alcool et le tabac.
4. Les taxes sur la propriété concernent essentiellement des taxes foncières ou des taxes d'habitation. C'est aussi dans cette catégorie que se situent l'impôt sur les donations et successions ou, encore, l'Impôt sur la Fortune Immobilière (IFI) qui fait couler beaucoup d'encre mais dont les recettes ne représentent que 0,08% du PIB.¹
5. Les entreprises doivent payer l'Impôt sur les Sociétés (IS) dès lors qu'elles réalisent des bénéfices, ce qui rapporte à l'État 1,2% du PIB. Il existe également d'autres impôts de production, payés par les sociétés.

Ces recettes fiscales servent à financer des dépenses publiques, qui sont décomposées dans la table 5.2. Les différences de dépenses publiques entre pays sont principalement dues aux dépenses sociales, qui sont nettement plus élevées en Europe que dans les autres pays riches.

TABLE 5.2 – Postes de dépense publique de 2000 à 2010, en pourcentage du PIB.

	France	Allemagne	Royaume-Uni	Etats-Unis
Dépenses sociales	34,3%	30,6%	26,2%	22,4%
Education	5,2%	4,4%	4,8%	4,7%
Santé	7,1%	7,8%	6,1%	7,7%
Retraites	12,2%	10,1%	4,8%	6,0%
Autres dépenses sociales	9,8%	8,4%	10,6%	4,0%
Autres dépenses publiques	16,7%	13,5%	15,9%	13,0%
Total	51,0%	44,1%	42,1%	35,4%

Source Piketty and Saez, 2014, 'Optimal Labor Income Taxation', *Handbook of Public Economics*, Table 1.

Les dépenses publiques sont supérieures aux prélèvements obligatoires. La différence est soit financée par l'endettement public, soit par d'autres recettes gouvernementales, comme les amendes, les recettes des établissements publics (à l'image du musée du Louvre), ou les dividendes perçus par l'État-actionnaire.

Un fort niveau de dépenses publiques engendre une fiscalité lourde ce qui, comme nous allons le voir, pose des problèmes d'efficacité économique. Soulignons néanmoins qu'un faible niveau de dépenses publiques n'est pas automatiquement un gage d'efficacité. Par exemple, en 2018, la France dépensait 11,3% de son PIB pour la santé, dont l'essentiel financé par l'État, tandis que les États-Unis dépensaient 16,9%, essentiellement financés par le secteur privé. Or, globalement, le système de santé public français semble plus efficace

1. L'ancêtre de l'IFI, l'Impôt de Solidarité sur la Fortune (ISF), rapportait environ 0,2% du PIB.

que le système privé américain, ce qui se traduit par un taux de mortalité plus faible et une espérance de vie plus élevée.

Intéressons-nous maintenant à l'impact économique de la fiscalité, en commençant par la question de l'incidence fiscale.

5.2 L'incidence fiscale

L'étude de l'incidence fiscale consiste à déterminer les conséquences d'un impôt pour les différents acteurs de l'économie. Autrement dit, cela consiste à répondre à la question : qui paye l'impôt ? Ou, plus précisément : qui supporte la charge de l'impôt ? Cette question semble triviale : il suffit de regarder qui envoie le chèque au trésor public. Par exemple, si une taxe est réglée par les vendeurs, comme c'est le cas de la TVA, c'est alors l'offre qui supporte la charge de cet impôt. Si, au contraire, c'est l'acheteur qui est imposé sur sa consommation, comme pour la taxe de séjour réglée par les vacanciers à l'issue de leurs nuits d'hôtel, c'est alors la demande qui paye. Non ?

Nous allons voir que la réalité est plus subtile et que l'incidence économique d'un impôt ne correspond pas nécessairement à son incidence légale.²

5.2.1 Imposer l'offre ou la demande ?

Supposons que l'État décide de créer une taxe sur un bien donné. Cette taxe s'élève à un montant t pour chaque unité de bien échangée.³ Il convient désormais de distinguer deux prix différents :

- p^c , le prix payé par les consommateurs lors de l'acquisition du bien ;
- p^p , le montant perçu par les producteurs lors de la vente du bien.

La différence entre ces deux prix correspond au montant de la taxe :

$$p^c = p^p + t.$$

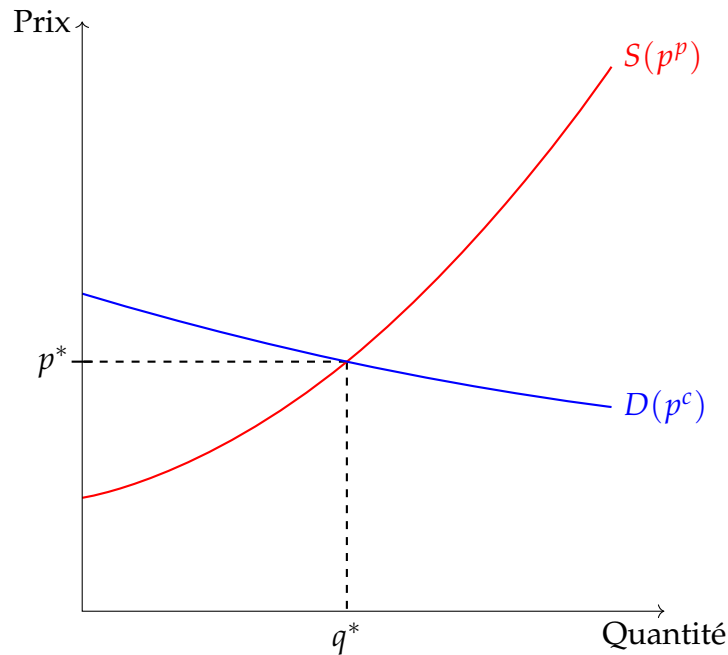
La courbe de demande $D(p^c)$ est une fonction décroissante du prix p^c payé par les consommateurs, tandis que la courbe d'offre $S(p^p)$ est une fonction croissante du prix p^p perçu par les producteurs.

Notons p le prix auquel la transaction s'effectue, c'est à dire le transfert d'argent du consommateur vers le producteur. En l'absence de taxe, on a l'équilibre habituel de la figure 5.2 où p^* et q^* sont le prix et la quantité à l'équilibre en l'absence de taxe.

2. En principe, dans la langue française, une taxe est un prélèvement affecté à une dépense publique spécifique en contrepartie d'un service rendu aux administrés (comme pour la taxe d'apprentissage), tandis qu'un impôt n'est pas affecté et ne prévoit pas de contrepartie. Par conséquent, la Taxe sur la Valeur Ajoutée (TVA) n'est pas une taxe mais un impôt. Nous utilisons ci-dessous les termes « impôt » et « taxe » de manière interchangeable, et sans se soucier d'éventuelles contreparties.

3. Nous considérons ici une taxe *accise*, c'est à dire une taxe d'un montant fixe pour chaque unité échangée. La plupart des taxes sont au contraire *ad valorem*, c'est à dire qu'elles sont proportionnelles aux prix de vente, comme pour la TVA. Tous les résultats que nous obtenons avec une accise sont également valides dans le cas *ad valorem*.

FIGURE 5.2 – Équilibre de marché sans taxe



Que se passe-t-il lorsqu'on introduit une taxe ? Considérons deux cas :

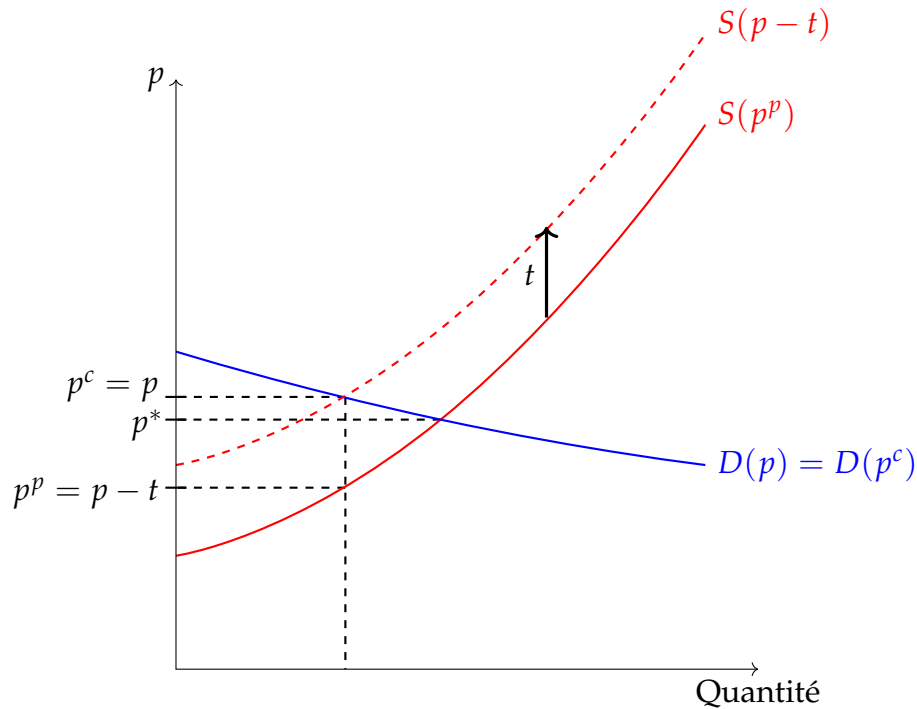
1. Dans le premier cas, ce sont les producteurs qui doivent payer la taxe. La transaction s'effectue au prix $p = p^c$ payé par les consommateurs ; tandis que les producteurs reçoivent ce prix net de la taxe dont ils doivent s'acquitter, soit $p^p = p - t$ par unité vendue.
2. Dans le second cas, ce sont les consommateurs qui sont redevables de la taxe. La transaction s'effectue au prix $p = p^p$ perçu par les producteurs ; tandis que les consommateurs payent la taxe en plus de la somme transmise aux producteurs, soit au total $p^c = p + t$ par unité achetée.

Taxe payée par les producteurs. Dans ce cas, on a $p = p^c$, donc la courbe de demande demeure inchangée. En revanche, avec $p^p = p - t$, la courbe d'offre est translatée vers le haut d'un montant t : les producteurs sont prêts à fournir une quantité $S(p^p)$ pour un prix perçu p^p , et donc pour une transaction au prix affiché $p = p^p + t$, d'où la translation vers le haut. Autrement dit, si une transaction s'effectue au prix p , alors les producteurs reçoivent $p - t$ par unité vendue et sont disposés à offrir la quantité $S(p - t)$. Cette situation est représentée sur la figure 5.3. Algébriquement, l'équilibre est donné par :

$$D(p) = S(p - t) \text{ avec } p^c = p \text{ et } p^p = p - t. \quad (\text{Prod})$$

En l'absence d'impôt, le prix d'équilibre est p^* . La taxe acquittée par les producteurs réduit le prix perçu par ceux-ci, de p^* à $p^p = p - t$. De façon plus surprenante, la taxe augmente le prix payé par les consommateurs, de p^* à $p^c = p$, bien qu'ils ne soient pas directement imposés. La charge de l'impôt est donc répartie entre les producteurs et les consommateurs.

FIGURE 5.3 – Taxe payée par les producteurs



Taxe payée par les consommateurs. Lorsque la taxe est payée par les consommateurs, on a $p = p^p$, ce qui implique que la courbe d'offre demeure inchangée. En revanche, avec $p^c = p + t$, la courbe de demande est désormais translatée vers le bas d'un montant t : les consommateurs demandent une quantité $D(p^c)$ pour un prix payé p^c , et donc pour une transaction effectuée au prix affiché $p = p^c - t$, d'où la translation vers le bas. Cette situation est représentée sur la figure 5.4. Algébriquement, l'équilibre est donné par :

$$D(p + t) = S(p) \text{ avec } p^c = p + t \text{ et } p^p = p. \quad (\text{Cons})$$

Là encore, on voit que la charge fiscale est portée à la fois par les consommateurs, dont le prix payé augmente de p^* à $p^c = p + t$, et par les producteurs, dont le prix perçu baisse de p^* à $p^p = p$.

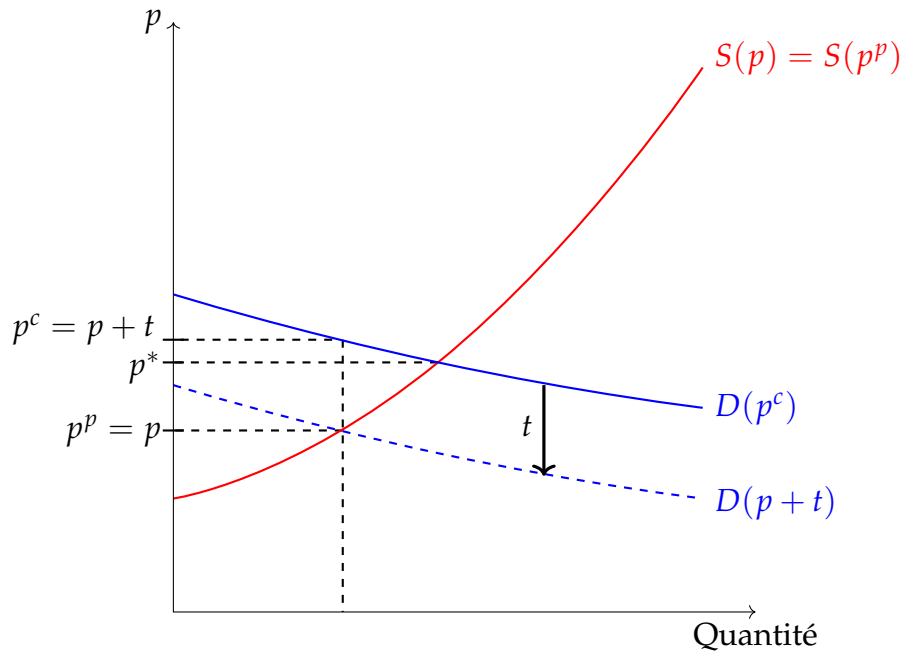
Comparaison. Que remarque-t-on en comparant ces deux politiques? Que la taxe soit payée par les producteurs ou par les consommateurs, l'équilibre économique est rigoureusement identique! Nous avons donc le résultat suivant :

Le prix p^c payé par les consommateurs, le prix p^p perçu par les producteurs et la quantité échangée dépendent uniquement du montant t de la taxe et sont indépendants de qui doit verser cet impôt au trésor public.

Ce résultat est visible graphiquement, mais également algébriquement puisque nos deux systèmes d'équations ci-dessus, (Prod) et (Cons), sont équivalents.

Au delà du formalisme mathématique, ce résultat repose sur une intuition économique

FIGURE 5.4 – Taxe payée par les consommateurs



simple et puissante. Les consommateurs ne se préoccupent que d'une seule chose : le prix p^c que leur coûte l'acquisition du bien. Les producteurs, quant à eux, ne se préoccupent que du prix p^p que leur rapporte la vente du bien. L'équilibre économique est tel que la quantité achetée au prix p^c est égale à la quantité vendue au prix p^p , où ces deux prix diffèrent du montant de la taxe, ce qui donne :

$$D(p^c) = S(p^p) \text{ avec } p^c - p^p = t.$$

Que l'impôt soit payé par l'offre ou par la demande ne change donc strictement rien à la nature du problème, ni donc à l'équilibre économique atteint.

Par conséquent, l'*incidence économique* d'un impôt est exclusivement déterminée par le jeu de l'offre et de la demande, et est indépendante de son *incidence légale*. Le législateur peut décider du niveau de l'imposition sur chaque marché, mais ne peut pas décider qui en supportera la charge.

En 2009, le gouvernement français a diminué le taux de TVA sur la restauration. Cette politique semble très différente d'une subvention aux clients, plutôt aisés, des restaurants. Et pourtant, sitôt l'équilibre économique atteint, ces deux politiques sont rigoureusement identiques!

5.2.2 Qui paye l'impôt?

Nous venons d'établir que l'incidence économique d'un impôt ne dépendait pas de son incidence légale. Mais nous n'avons pas encore complètement répondu à la question : qui paye l'impôt?

L'incidence est partagée entre les consommateurs, dont le prix payé augmente de p^* à p^c , et par les producteurs, dont le prix perçu baisse de p^* à p^p . Il est donc naturel de définir

l'incidence économique de l'impôt comme suit :

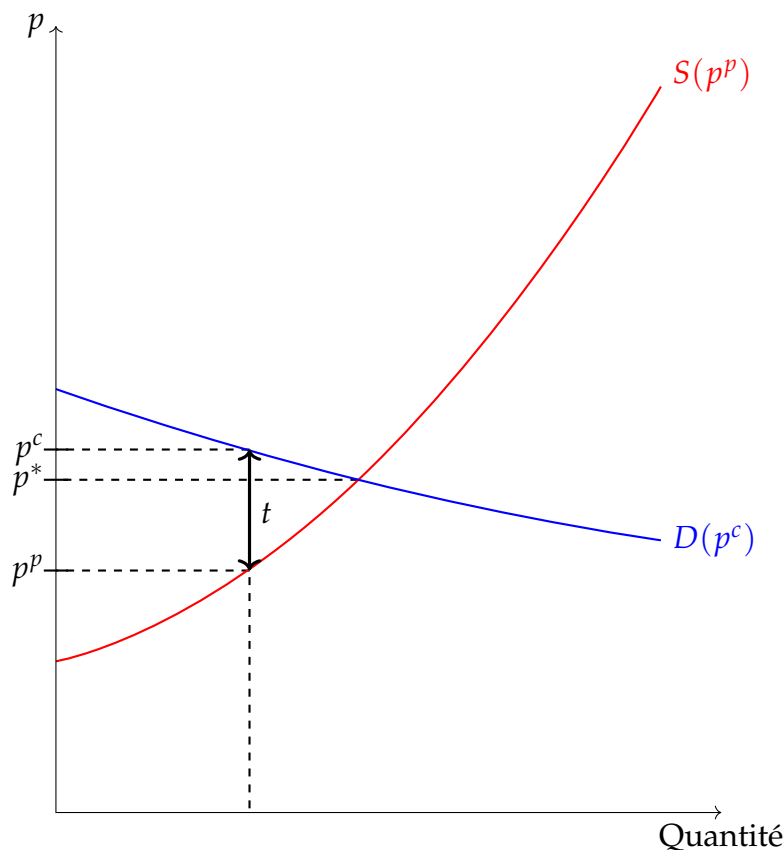
- L'incidence sur le consommateur est égale à $p^c - p^*$;
- L'incidence sur le producteur est égale à $p^* - p^p$.

Le paiement de l'impôt est ainsi réparti entre l'offre et la demande.

Que la taxe soit payée par l'offre ou par la demande, la quantité échangée à l'équilibre est tel que la demande $D(p^c)$ au prix p^c est égale à l'offre $S(p^p)$ au prix p^p avec $p^c - p^p = t$, ce qui est représenté sur la figure 5.5. On s'aperçoit qu'avec une offre relativement inélastique, qui se traduit par une courbe d'offre assez verticale, une part importante de l'impôt est payée par l'offre. De même qu'avec une demande relativement élastique, qui se traduit par une courbe de demande proche de l'horizontale, une faible part de l'impôt est payé par la demande. Cette observation illustre un résultat général concernant l'incidence fiscale.

La charge d'un impôt est principalement supportée par le côté le plus inélastique du marché.

FIGURE 5.5 – Incidence fiscale sur le consommateur et sur le producteur, avec une demande très élastique et une offre beaucoup moins élastique



Si l'offre est beaucoup plus inélastique que la demande, face à un nouvel impôt, les producteurs sont prêts à baisser fortement leur prix p^p afin d'écouler leurs marchandises, tandis que les consommateurs ne sont pas vraiment disposés à augmenter le prix p^c qu'ils payent. Dans ce cas, l'impôt est essentiellement payé par les producteurs (même lorsque se sont

les consommateurs qui envoient le chèque au trésor public). À l'inverse, si l'offre est plus élastique que la demande, les consommateurs sont prêts à payer beaucoup plus cher pour maintenir leur consommation, ce qui leur fait supporter l'essentiel de la charge de l'impôt.

Ce résultat peut facilement se démontrer mathématiquement pour une petite taxe. À l'équilibre, on $D(p^c) = S(p^c - t)$. Un développement limité de cette expression autour de $t = 0$, avec $p^c = p^*$ lorsque $t = 0$, donne

$$D'(p^*) \frac{dp^c}{dt} = S'(p^*) \left[\frac{dp^c}{dt} - 1 \right].$$

L'élasticité de l'offre et de la demande (à l'équilibre sans taxe) sont respectivement donnés par $\varepsilon^s = p^* S'(p^*) / S(p^*)$ et $\varepsilon^d = -p^* D'(p^*) / D(p^*)$. En utilisant ces définitions ainsi que la condition d'équilibre sans taxe $D(p^*) = S(p^*)$, on obtient

$$\frac{dp^c}{dt} = \frac{\varepsilon^s}{\varepsilon^s + \varepsilon^d}.$$

De même qu'avec $p^p = p^c - t$, on a

$$\frac{dp^p}{dt} = \frac{-\varepsilon^d}{\varepsilon^s + \varepsilon^d}.$$

Ainsi, plus la demande est inélastique (plus ε^d est petit), plus la taxe se traduit par une augmentation du prix p^c payé par les consommateurs plutôt que par une diminution du prix p^p perçu par les producteurs.

Ce résultat se comprend intuitivement avec quelques exemples : si l'État imposait une taxe sur l'insuline (à la demande très inélastique), l'industrie pharmaceutique pourrait augmenter le prix de vente de l'insuline du montant de la taxe sans craindre de perdre des clients. Dans cette situation, les profits des producteurs d'insuline ne sont pas changés et tout le poids de la taxe repose sur les consommateurs. Au contraire, si l'État décidait de taxer les croisières (à la demande plus élastique et à l'offre peu élastique), les compagnies de tourisme réduiraient sans doute leurs marges pour ne pas perdre de clients.

5.2.3 Application au marché du travail

Notre analyse est lourde de conséquences pour l'imposition des revenus du travail. Comme nous le verrons en détail dans le chapitre consacré à l'économie du travail, le salaire est déterminé sur le marché du travail par la rencontre entre l'offre de travail des ménages, croissante avec le salaire, et la demande de travail des entreprises, décroissante avec le coût du travail. Dans les données, on observe que l'offre de travail des ménages est relativement inélastique : la plupart des individus souhaitent travailler à plein temps quelles que soient les évolutions de leurs salaires. En revanche, la demande de travail des entreprises est très élastique, car très sensible au coût du travail.

Il existe différents impôts sur le travail : d'un côté les cotisations patronales, à la charge de l'employeur, et de l'autre les cotisations salariales et l'impôt sur le revenu, à la charge de

l'employé. La répartition des cotisations entre employeurs et employés fait parfois l'objet de débats passionnés : les syndicats plaident pour l'imposition des entreprises au nom de la défense du pouvoir d'achat, tandis que les entreprises souhaitent le contraire afin de préserver leur compétitivité.

Notre analyse de l'incidence fiscale nous apprend non seulement que ces considérations sont vaines, mais que tous ces impôts sont principalement payés par les salariés. En effet, l'offre de travail des salariés étant nettement plus inélastique que la demande de travail des entreprises, les salaires nets d'impôts baissent quasiment du montant de ces prélèvements obligatoires afin de maintenir le niveau de l'emploi, tandis que le coût du travail pour les employeurs augmente à peine !

Pour bien comprendre le mécanisme économique sous-jacent, supposons que l'État diminue les charges patronales. Le coût du travail se réduisant, la demande de travail des entreprises augmente d'autant plus fortement qu'elle est très élastique. Néanmoins, l'offre de travail des ménages étant faiblement élastique, cette hausse de la demande ne se traduit pas principalement par une hausse de l'emploi, mais plutôt par une augmentation des salaires. À l'équilibre, les salaires augmentent pratiquement du niveau de la baisse des charges patronales qui, bien que versée aux entreprises, profite donc essentiellement aux employés !

Précisons néanmoins deux limites à ce raisonnement. D'abord, nous avons ici supposé des marchés à l'équilibre. Dans le cas du marché du travail, la convergence vers l'équilibre peut prendre plusieurs années. À court terme, les salaires sont rigides et ce sont bien les entreprises qui empochent la baisse de charges. Par ailleurs, nous avons négligé le cas d'employés au salaire minimum, dont la rémunération est fixée par la législation plutôt que par un équilibre de marché : nous reviendrons sur ce cas lorsque nous étudierons spécifiquement le marché du travail.

5.2.4 L'incidence de l'imposition des entreprises

Contrairement à une idée reçue, les économistes s'intéressent exclusivement au bien-être des individus, et ne se préoccupent pas directement de la rentabilité des entreprises. Par conséquent, un principe fondamental en sciences économiques est qu'une entreprise ne paye jamais d'impôts ! Elle ne fait que les transférer aux personnes physiques qui se cachent derrière.

L'incidence de l'imposition des bénéfices des sociétés se répartit donc entre les actionnaires, dont les dividendes diminuent, les salariés, dont les salaires baissent, et les consommateurs, dont les prix augmentent. Pour quantifier cette répartition de l'incidence fiscale, il est nécessaire d'effectuer une analyse en équilibre général qui prend en compte l'interaction entre le marché financier, le marché du travail, et le marché des biens.

Une telle analyse est par nature complexe. Ceci étant, si les rendements sur investissement sont déterminés par les marchés financiers à l'échelle mondiale, alors les actionnaires ne sont que faiblement touchés par la fiscalité d'un pays donné. L'incidence de l'imposition des entreprises repose alors en grande partie sur les travailleurs et les consommateurs, dont le pouvoir d'achat baisse du montant de ces impôts. Par conséquent, la capacité des en-

treprises d'échapper à l'impôt en dégageant leurs profits dans des paradis fiscaux, comme l'Irlande ou les Bermudes, ne bénéficie vraisemblablement pas qu'à leurs actionnaires mais également à leurs salariés et à leurs consommateurs.

5.2.5 L'incidence des tarifs douaniers de Donald Trump

Pour illustrer cette question de l'incidence fiscale, intéressons nous à la guerre commerciale lancée par Donald Trump. En 2018, une part significative des produits importés aux États-Unis ont été frappés par une augmentation des tarifs douaniers. Pour les produits concernés, ces tarifs sont passés en moyenne de 2,6% à 16,6%. Cette hausse des taxes sur les importations a-t-elle été supportée par les consommateurs américains ou a-t-elle réduit les marges des entreprises étrangères ?

La figure 5.6 (a), issue de Fajgelbaum et al. (2020), montre que les importations américaines ont baissé significativement après la mise en place des tarifs, conformément à ce qui était attendu. Par ailleurs, on observe, sur le panel (b), que le prix des importations hors tarifs n'a pas baissé, ce qui indique que les entreprises étrangères n'ont pas diminué leurs marges. Le prix des importations, tarifs inclus, a lui augmenté d'une proportion similaire à la hausse des tarifs, comme le montre le panel (c). Ces résultats peuvent s'interpréter comme une faible élasticité de la demande des produits importés : ayant peu de substituts disponibles, les Américains ont complètement supporté le poids de cette taxe.

5.3 Les distorsions fiscales

Intéressons-nous désormais à l'impact de la fiscalité sur l'efficacité économique. En l'absence d'impôt, nous avons vu que l'équilibre d'une économie de marché permet à toutes les transactions mutuellement bénéfiques de se réaliser, ce qui est Pareto-efficace et engendre la maximisation du surplus total bénéficiant aux différents acteurs de l'économie, comme représenté sur la figure 5.7. Qu'en est-il en présence de taxes ?

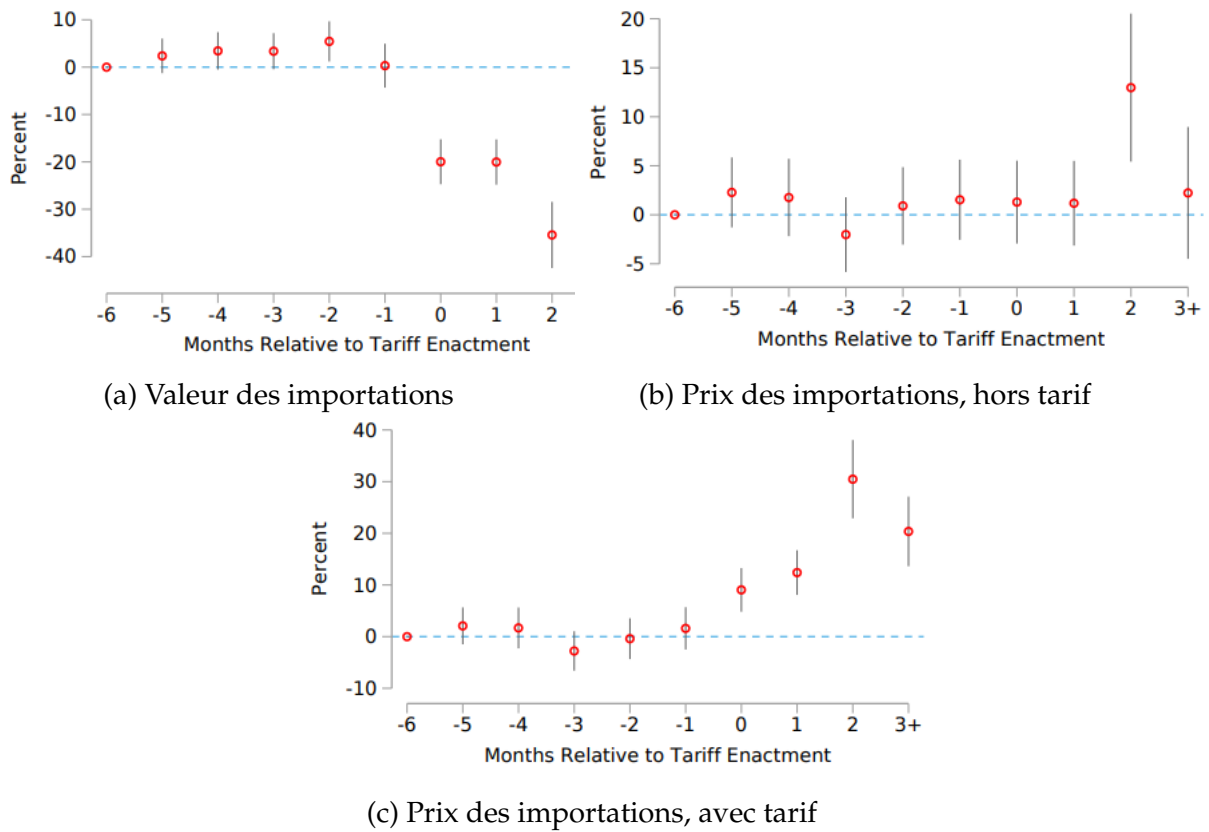
5.3.1 La perte sèche

L'introduction d'une taxe t crée un écart entre le prix p^c payé par les consommateurs, qui détermine la quantité demandée, et le prix p^p perçu par les producteurs, qui détermine la quantité offerte. Cela aboutit à la situation représentée sur la figure 5.8, où nous supposons (sans que cela n'ait d'importance) que la taxe t est payée par les producteurs.

L'augmentation du prix, de p^* à p^c , diminue le surplus des consommateurs ; de même que la diminution du prix, de p^* à p^p , diminue le surplus des producteurs. Ceci étant, tout n'est pas perdu : pour chaque unité vendue, le gouvernement touche le montant de la taxe, ce qui engendre une recette fiscale égale à $tq^* = (p^c - p^p)q^*$ représentée sur la figure 5.8 par le rectangle vert.

Reste le triangle gris. Il s'agit de la partie du surplus total qui existait avant la taxe et qui désormais ne revient ni aux consommateurs, ni aux producteurs, ni au gouvernement. C'est

FIGURE 5.6 – Évolution de la valeur totale et du prix des importations après la mise en place des tarifs douaniers. En abscisse, on a placé à 0 le mois de mise en place des tarifs.



Source : Fajgelbaum et al. (2020)

FIGURE 5.7 – Surplus du consommateur (en bleu) et du producteur (en rouge) en l'absence de taxe.

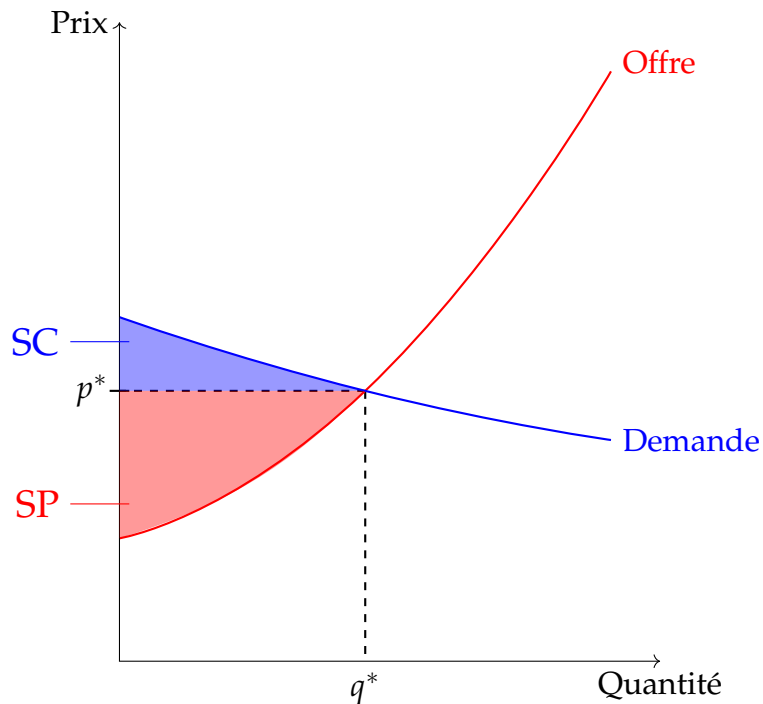
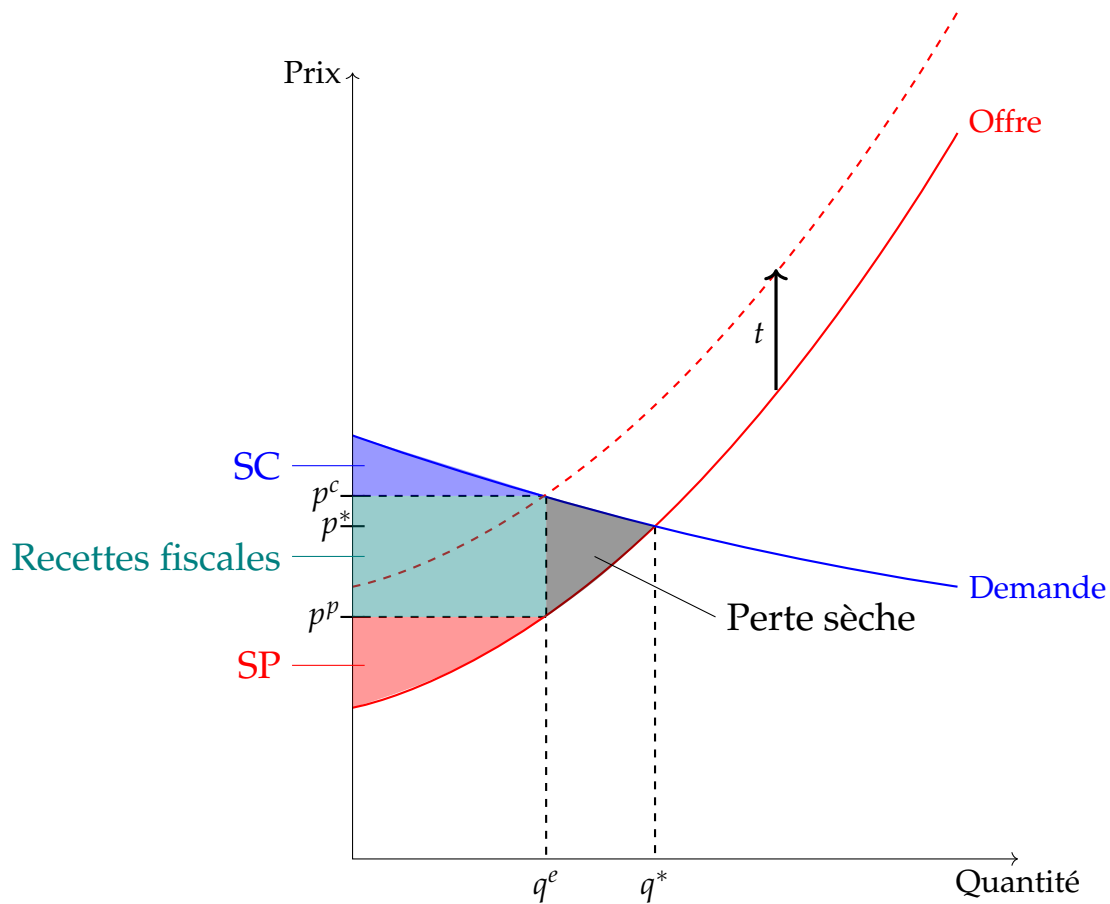


FIGURE 5.8 – Revenu fiscal (en vert), perte sèche (en noir) ainsi que le surplus du consommateur (en bleu) et le surplus du producteur (en rouge)



la **perte sèche** générée par la fiscalité. Nous avons donc le résultat suivant.

Pour récolter un euro, l'État doit nécessairement diminuer le surplus des consommateurs et des producteurs de plus d'un euro.

La perte sèche résulte des échanges mutuellement bénéfiques qui auraient du avoir lieu mais qui sont empêchés par la taxe : les consommateurs qui ont une propension marginale à payer (représentée par la courbe de demande) entre p^* et p^c souhaiteraient acheter les unités que les producteurs dont les coûts marginaux de production (représentés par la courbe d'offre) se situent entre p^p et p^* souhaiteraient leur vendre. Néanmoins, comme de telles transactions s'accompagnent d'un prélèvement de t , ces échanges mutuellement bénéfiques qui créent un surplus inférieur à t n'ont pas lieu, ce qui est Pareto-inefficace. L'aire en gris sur la figure 5.8 est donc une mesure des inefficacités économiques créées par la fiscalité. La quantité produite aurait dû être égale à q^* mais, à cause de la taxe, à l'équilibre elle se limite à q^e .

En théorie, pour remédier à cette inefficacité, on pourrait uniquement imposer la taxe sur les q^e premières unités échangées, et laisser les échanges suivants se faire sans taxe. Il est néanmoins évident qu'une telle politique est impossible à mettre en place : chacun aurait intérêt à prétendre que les transactions qu'il effectue ne font pas parties des q^e premières,

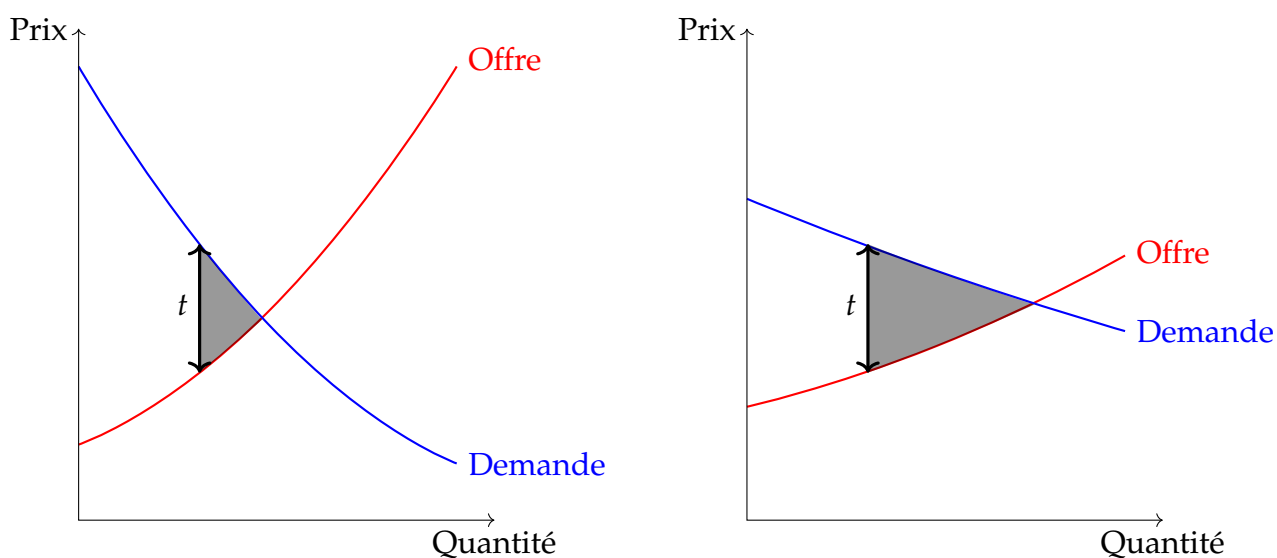
mais des dernières qui ne sont pas sujettes à la taxe. Il est donc dans la nature d'une taxe de créer une inefficacité économique mesurée par la perte sèche.

5.3.2 Les distorsions de comportement

En l'absence de taxes, le prix transmet une information qui permet d'égaliser le bénéfice marginal de la consommation au coût marginal de la production. Le prix joue donc le rôle d'un signal qui coordonne les actions des consommateurs et des producteurs de manière efficace. L'impôt *distord* le comportement de ces agents dont les choix ne reflètent plus uniquement les bénéfices et coûts marginaux, mais également le niveau de l'impôt.

Les inefficacités fiscales sont donc fondamentalement dues aux changements de comportement des consommateurs et des entreprises : plus la quantité échangée diminue suite à l'introduction de la taxe, plus la perte sèche est élevée. Par conséquent, si l'offre ou la demande est inélastique (et une des deux courbes est assez verticale), alors l'impôt ne distord que faiblement les comportements et la perte sèche est faible. C'est le cas représenté sur la figure 5.9, à gauche. En revanche, si l'offre et la demande sont tous deux très élastiques (et les deux courbes sont plutôt horizontales), alors la quantité échangée diminue considérablement, ce qui se traduit par une perte sèche élevée pour des recettes fiscales relativement basses, ce qui correspond à la figure 5.9, à droite. Il est donc préférable que le gouvernement impose les biens et services dont la consommation n'est que faiblement élastique.

FIGURE 5.9 – La perte sèche est d'autant plus élevée que l'offre et la demande sont élastiques.



Existe-il un impôt efficace, qui n'engendre aucune distorsion ? Oui, il s'agit de l'**impôt forfaitaire**. C'est un impôt qu'un individu doit payer *quel que soit son comportement*. Par exemple, un impôt annuel de 100 euros que tous les français devraient payer, indépendamment de leurs revenus, serait efficace (mais certainement pas équitable !). Cet impôt ne découragerait aucune transaction mutuellement bénéfique.⁴ En pratique, les impôts ne sont

4. Cet impôt ne modifierait les comportements qu'au travers des revenus disponibles. Or, les modifications de comportements dus aux changements de revenus ne sont pas source d'inefficacité.

jamais forfaitaires, mais reposent sur des transactions et dépendent donc du comportement des agents. Par exemple, l'impôt sur le revenu du travail décourage les ménages de travailler et les entreprises d'embaucher, ce qui est une distorsion majeure de l'activité économique.

5.3.3 L'impact du niveau d'imposition

Quel est l'impact du niveau d'imposition sur la perte sèche et sur les recettes fiscales ? Lorsque les courbes d'offre et de demande sont linéaires, la perte sèche est proportionnelle au carré du montant de la taxe. De manière plus générale :

La perte sèche est une fonction convexe du taux d'imposition.

Ce résultat provient du fait que, plus le taux d'imposition est élevé, plus les transactions dont il empêche la réalisation sont mutuellement bénéfiques. En effet, lorsque le taux d'imposition passe de t à $t + dt$, les échanges qui cessent d'avoir lieu sont caractérisés par un écart entre la propension marginale à payer des consommateurs et le coût marginal de production des fabricants égal à t ; écart d'autant plus important que t est élevé. A contrario, une petite baisse du taux d'imposition permet prioritairement aux échanges les plus bénéfiques de se réaliser.

Les recettes fiscales sont le produit du taux d'imposition par le nombre d'unités vendues. Or, plus le taux d'imposition est élevé, plus la quantité de transactions à laquelle il s'applique est faible. On a donc le résultat suivant :

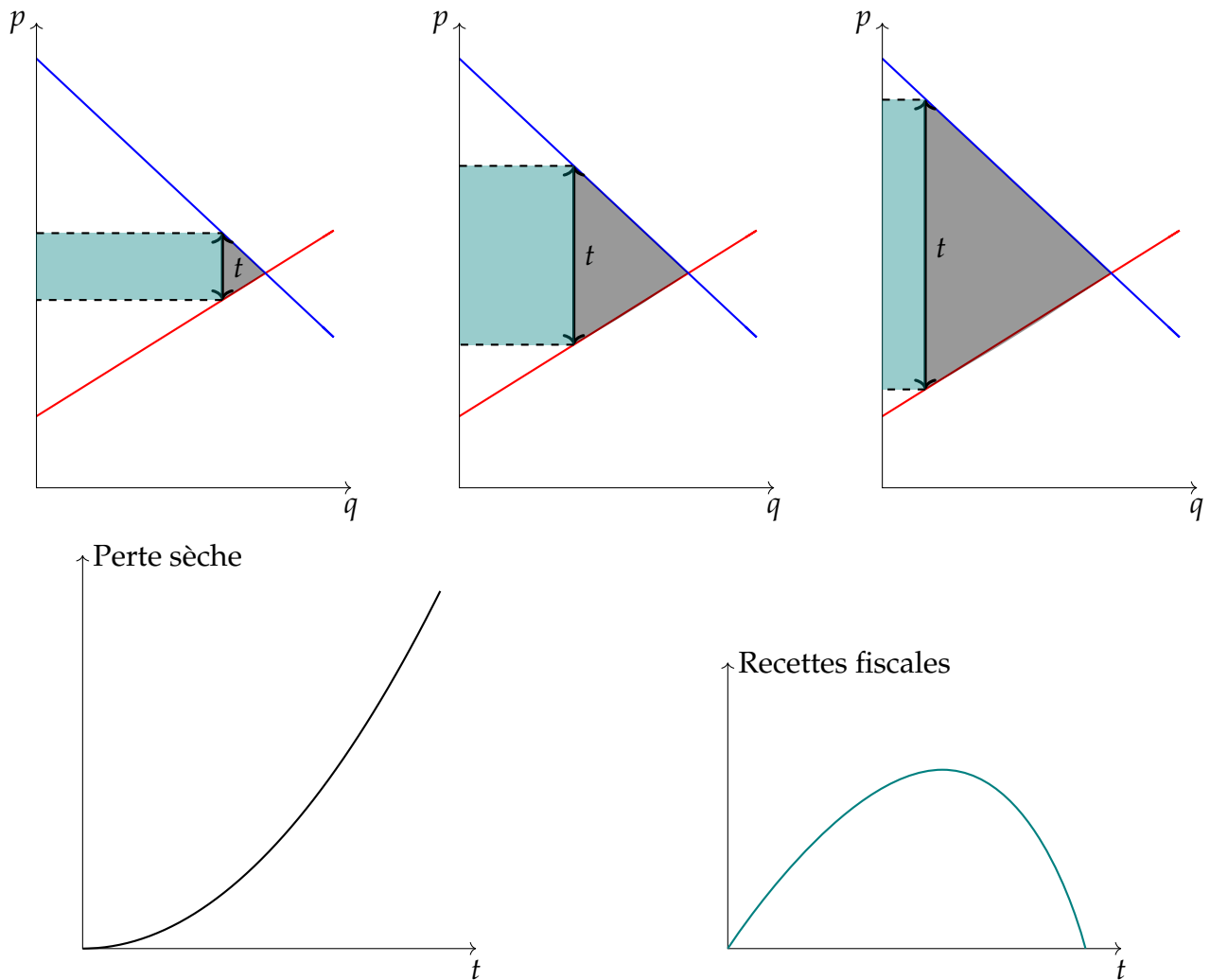
Au delà d'un certain niveau, un alourdissement de la fiscalité réduit les recettes du gouvernement.

Ces deux résultats sont représentés graphiquement sur la figure 5.10. Au fur et à mesure que le taux d'imposition augmente, la perte sèche augmente de plus en plus vite, tandis que les recettes fiscales augmentent de moins en moins vite, allant même jusqu'à diminuer au-delà d'un certain seuil.

L'idée d'une diminution des recettes est particulièrement marquante dans le cas de la fiscalité du travail : en augmentant le niveau d'imposition des revenus du travail, on décourage l'activité, réduisant ainsi l'assiette fiscale. En théorie, si le taux d'imposition atteint 100%, plus personne n'a intérêt à travailler, et les recettes fiscales sont nulles. Le revenu de l'État suit donc une courbe en cloche, comme sur la figure 5.10 en bas à droite, connue sous le nom de *courbe de Laffer*, du nom du conseiller économique du président Ronald Reagan. On devrait plutôt parler de *courbe de Dupuit*, du nom de Jules Dupuit (1804-1866, X1822) qui, en dérivant les bases de l'analyse microéconomique moderne avait naturellement obtenu cette courbe dès 1844.

Il est très important d'identifier où se situe le sommet de la courbe de Dupuit-Laffer, car il serait Pareto-inefficace de fixer un taux d'imposition supérieur. Comme nous le verrons ultérieurement, la demande de travail des entreprises étant par nature très élastique,

FIGURE 5.10 – Impact du taux d'impositions sur la perte sèche et sur les recettes fiscales.



la localisation de ce sommet dépend essentiellement de l'élasticité de l'offre de travail des ménages : plus cette élasticité est élevée, plus le sommet de la courbe de Dupuit-Laffer est bas.

De manière plus générale, lorsque l'on tente d'estimer l'impact d'une hausse d'impôt sur les recettes fiscales, il est très important de prendre en compte l'adaptation des comportements et donc la variation de l'assiette fiscale. Certes, la réaction de l'assiette dépend de l'élasticité de l'offre et de la demande, qui ne sont jamais faciles à estimer empiriquement. Mais, cela n'implique pas que la meilleure estimation soit l'absence de réaction ! Dans de nombreux pays, le ministère des finances a tendance à raisonner à assiette fixe, ce qui aboutit à une surestimation des recettes engendrées par les hausses d'impôt ou du coût budgétaire des baisses d'impôt.

5.3.4 Qu'est ce qu'une bonne fiscalité ?

L'existence de distorsions fiscales n'implique pas que les impôts sont par nature indésirables. L'État a un rôle majeur à jouer dans l'économie et doit nécessairement être financé par l'impôt. Ceci étant, ces impôts doivent être conçus de manière à minimiser les distorsions

qu'ils impliquent. Comme le disait Jean-Baptiste Colbert, contrôleur général des finances de Louis XIV : « l'art de l'imposition consiste à plumer l'oie pour obtenir le plus possible de plumes avant d'obtenir le moins possible de cris ».

Que nous enseigne notre analyse ? Premièrement, nous venons de voir que, pour chaque impôt, la perte sèche est une fonction convexe du taux d'imposition, tandis que les recettes fiscales en sont une fonction concave. Par conséquent, il vaut mieux taxer un grand nombre de produits à un taux faible, plutôt que quelques produits à un taux élevé. Autrement dit, à recette fiscale donnée, un gouvernement doit chercher à élargir les assiettes fiscales afin de réduire les taux d'imposition correspondants.

Par ailleurs, nous avons vu qu'il valait mieux taxer les biens dont l'offre ou la demande sont inélastiques. Ceci étant, il est possible que ces biens, à l'image du riz ou des pommes de terre, soient surtout consommés par les plus démunis, tandis que les biens dont la demande est élastique, à l'image des produits de luxe, soient principalement consommés par les plus fortunés. Dans ce cas, il y a un compromis entre efficacité économique et équité. À recette fiscale donnée, le gouvernement ne doit pas nécessairement chercher à minimiser la somme des pertes sèches, car une perte sèche infligée à un pauvre peut être socialement plus coûteuse qu'une perte sèche d'un même montant imposé à un riche. La valorisation de ces différentes pertes sèches relève de la philosophie politique. Il existe une branche des sciences économiques, appelée *théorie du choix social*, qui établit des liens entre des axiomes définissant des critères d'équité et les choix collectifs de la société. Plutôt que de donner des réponses définitives, cette approche met en évidence la difficulté d'appréhender les problèmes d'équité.

Ainsi, dès lors qu'on est confronté à un problème de redistribution, le gouvernement doit d'abord déterminer sa philosophie politique, plus ou moins égalitariste, afin de déterminer le poids à attacher aux pertes sèches infligées aux différents individus de l'économie. Ensuite, on peut s'appuyer sur la description du fonctionnement des marchés, et notamment sur les estimations de l'élasticité de l'offre et de la demande, afin de minimiser la somme pondérée des pertes sèches. C'est ainsi qu'on parvient au meilleur compromis entre efficacité et équité.

Plus généralement, dans l'analyse économique, on distingue l'analyse *positive*, qui consiste à étudier l'économie telle qu'elle est, et l'analyse *normative*, qui consiste à déterminer l'économie telle qu'elle devrait être. L'analyse normative consiste donc à s'appuyer sur nos connaissances positives afin d'effectuer des recommandations de politiques économiques qui minimisent les problèmes d'inefficacité et d'iniquité. L'approche normative, contrairement à l'approche positive, peut faire intervenir des jugements de valeur qu'il convient alors de poser explicitement.

5.4 Les externalités

Jusqu'ici, nous avons vu d'un bon œil la consommation des agents, puisqu'elle résulte d'échanges mutuellement bénéfiques. Mais, parfois, on voudrait plutôt que la société consomme *moins* de certains produits, comme le pétrole ou les antibiotiques. Cette possibilité

est la conséquence d'une des principales défaillances du marché : les externalités.

Une externalité se produit dès lors que l'action d'un agent affecte *directement* le bien-être des autres agents.

Par directement, nous voulons dire qu'il ne s'agit pas d'un effet qui passe par le mécanisme des prix.

5.4.1 Les externalités négatives

Commençons par étudier les externalités négatives, qui se produisent lorsque l'action d'un agent a un impact direct et négatif sur les autres agents. C'est en particulier le cas des activités polluantes. Lorsqu'une entreprise chimique pollue une rivière en amont, elle a un impact négatif sur la faune marine, et donc sur la production des pêcheurs en aval. Un fumeur exerce une externalité négative sur son entourage. Lorsqu'un automobiliste arrive dans un embouteillage, il aggrave la congestion et exerce ainsi une externalité négative sur les autres automobilistes. Lorsque nous émettons des gaz à effet de serre, nous exerçons une externalité négative sur l'ensemble de la société en contribuant au réchauffement climatique.

Lorsque les Chinois accroissent leur consommation de blé, le prix de celui-ci augmente en Europe, diminuant le bien-être des européens. Les Chinois exercent-ils une externalité négative sur les Européens? Non, car cette action s'effectue par le fonctionnement normal du mécanisme des prix au sein d'une économie de marché! Il n'y a donc pas là d'externalité, ni donc d'inefficacité.

Tentons donc de comprendre pourquoi les externalités sont une défaillance de marché. Dans le cas d'une externalité de production, il faut s'intéresser à la courbe d'offre : celle-ci reflète les coûts marginaux de production des entreprises. Lorsqu'une production engendre des externalités, chaque unité produite fait porter à la société (et non à l'entreprise) un coût supplémentaire, appelé **coût marginal externe**. La somme du coût marginal privé (porté par les entreprises) et du coût marginal externe (l'externalité, portée par le reste de la société) est égal au coût marginal social.

$$\text{Coût marginal social} = \text{coût marginal privé} + \text{coût marginal externe}$$

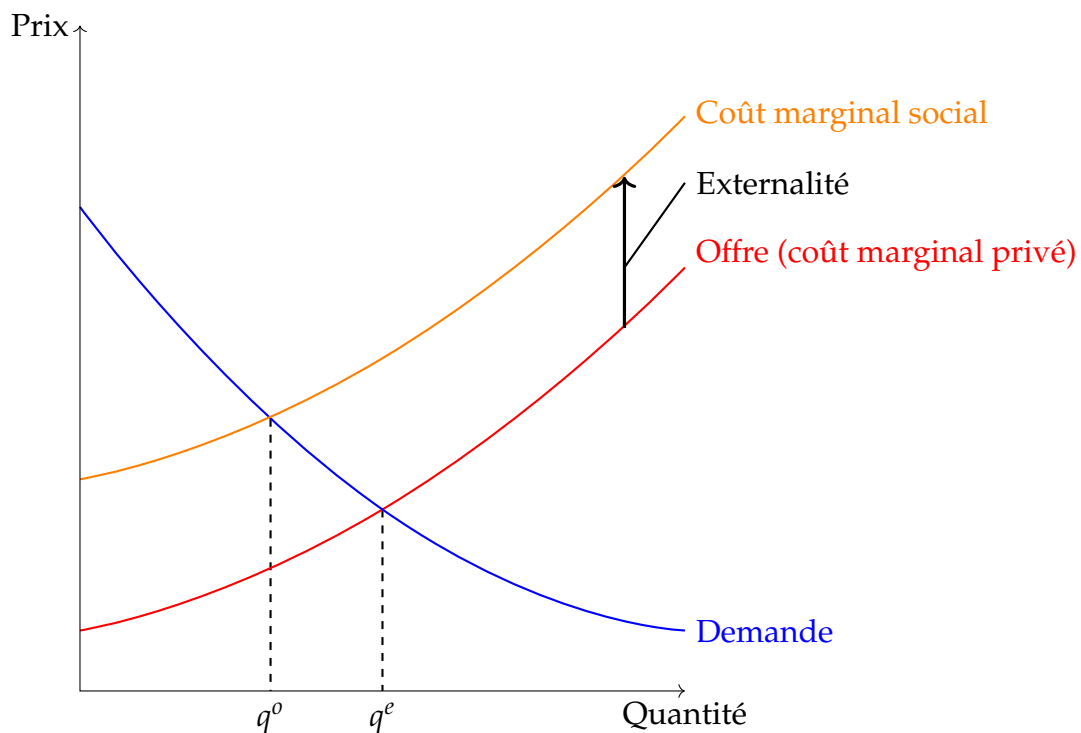
Une externalité négative a lieu dès lors que le coût marginal externe est positif, ce qui implique que le coût marginal social est plus élevé que le coût marginal privé. Par exemple, dans le cas d'une compagnie aérienne, les coûts privés comprennent le kérosène, la maintenance des appareils, les salaires des équipages, ou encore les frais liés à la logistique. Les coûts externes sont les nuisances sonores aux abords des aéroports et la contribution des émissions de gaz carbonique au réchauffement climatique.

Pour obtenir une allocation Pareto-efficace, il faut que la propension marginale à payer des consommateurs, reflétée par la courbe de demande, soit égale au coût marginal *social* de la production, ce qui correspond à la quantité q^0 de la figure 5.11. En effet, si la quantité est supérieure à q^0 , la propension à payer des consommateurs pour la dernière unité vendue

est inférieure à son coût social de production. Il serait donc souhaitable de ne pas produire cette unité. De même que, si la quantité est inférieure à q^o , il serait souhaitable d'accroître la production.

Le problème est que les producteurs déterminent leur offre en fonction de leurs coûts marginaux *privés*, et négligent les coûts marginaux externes. Par conséquent, l'équilibre de l'économie de marché correspond à la quantité q^e de la figure 5.11, qui est supérieur à la quantité optimale q^o . Il y a donc surproduction.

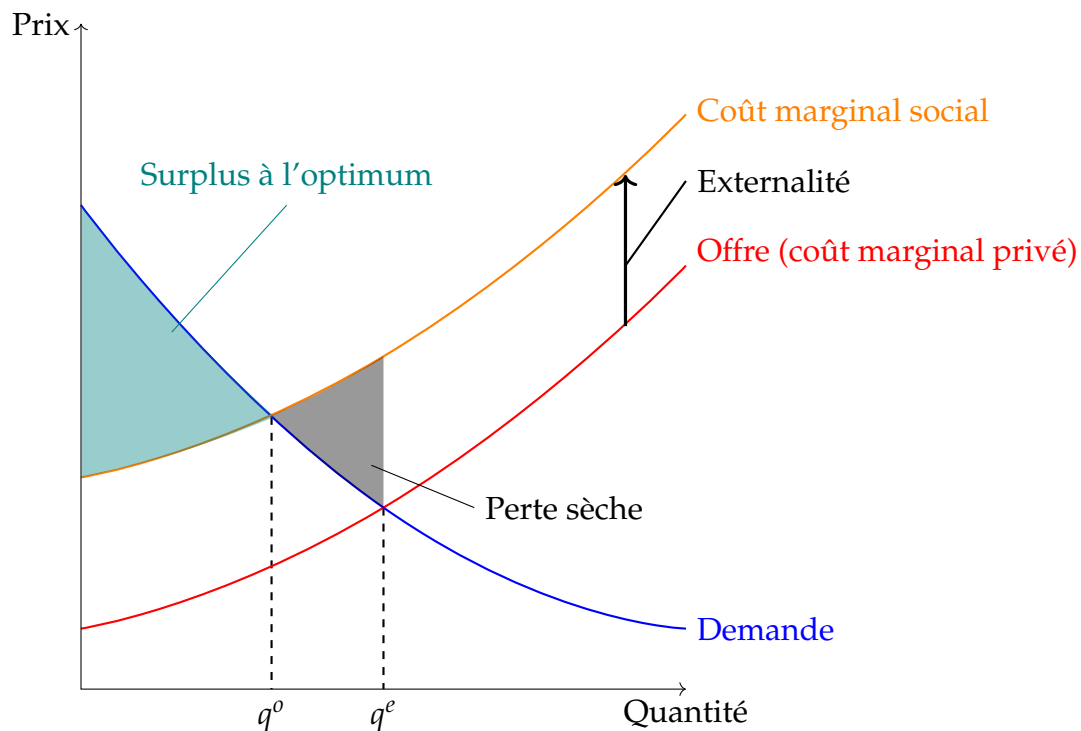
FIGURE 5.11 – Une externalité négative de production



Cette surproduction entraîne une perte sèche. Pour chaque unité produite, le surplus est égal à la différence entre la propension marginale à payer des consommateurs, reflétée par la courbe de demande, et le coût marginal *social* de la production. Jusqu'à la quantité q^o , ce surplus est positif. Il est représenté sur la figure 5.12 par la surface verte. Au delà de la quantité q^o , ce surplus est négatif car le bénéfice marginal de la consommation est inférieur au coût marginal social de la production. Ainsi, la quantité q^e produite par l'économie de marché engendre une perte sèche représentée sur la figure 5.12 par la surface grise. D'un point de vue normatif, les unités situées entre q^o et q^e n'auraient pas dû être produites. Pour reprendre notre exemple des compagnies aériennes, leur négligence totale de l'impact négatif de leur production sur les nuisances sonores et sur le réchauffement climatique engendre une surconsommation du transport aérien.

Il est également possible d'avoir une externalité négative de consommation (et non de production). Par exemple, la consommation d'antibiotiques augmente le risque d'apparition de bactéries résistantes aux antibiotiques, ce qui est nuisible au reste de la société. Dans ce

FIGURE 5.12 – Le surplus à l’optimum (en vert) et la perte sèche à l’équilibre de marché (en gris)



cas, il y a un coût marginal externe à la consommation d’antibiotiques. On a donc :

$$\text{Bénéfice marginal social} = \text{bénéfice marginal privé} - \text{coût marginal externe},$$

où le bénéfice marginal privé correspond à la courbe de demande. Par conséquent, en présence d’une externalité négative de consommation, la courbe de bénéfice marginal social se situe en dessous de la courbe de demande. À nouveau, en cas de *laissez-faire*, il y a surconsommation.

5.4.2 Les externalités positives

Il existe également des externalités positives : lorsqu’une personne se vaccine, elle ne peut plus être un vecteur de propagation de la maladie, ce qui protège les autres personnes de la société. Un jeune qui va à l’école y gagne personnellement, par exemple en améliorant ses perspectives futures sur le marché du travail, mais a également un impact positif sur la société en facilitant la transmission de connaissances, la communication entre les citoyens, voire en améliorant le fonctionnement de la démocratie. La recherche fondamentale exerce une externalité positive sur les innovateurs futurs qui pourront bâtir sur les connaissances accumulées.

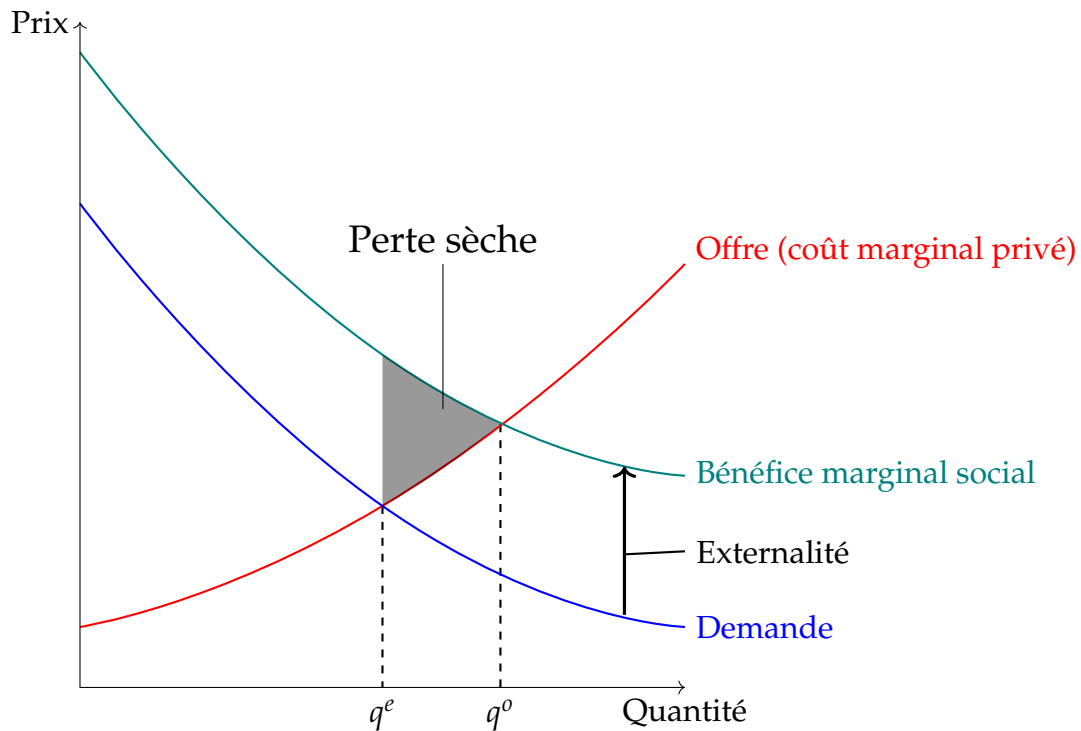
Dans le cas d’une externalité de consommation, les consommateurs décident de leur demande en fonction de leur bénéfice marginal privé, mesuré par leur propension marginale à payer et représenté par la courbe de demande. Ils ignorent totalement le bénéfice marginal

externe qui est ici positif. De manière générale :

$$\text{Bénéfice marginal social} = \text{bénéfice marginal privé} + \text{bénéfice marginal externe}$$

Pour obtenir une allocation efficace, le bénéfice marginal social, et non privé, doit être égal au coût marginal de production. En présence d'une externalité positive, le bénéfice marginal privé est inférieur au bénéfice marginal social, ce qui implique une sous-consommation du bien et donc une perte sèche, comme on peut le voir sur la figure 5.13.

FIGURE 5.13 – Une externalité positive de production



Il est également possible d'avoir des externalités positives de production. Les abeilles de l'apiculteur exercent une externalité positive sur la pollinisation des arbres de l'arboriculteur voisin, lequel contribue avec ses arbres à la qualité et à l'abondance de la production de miel. Il y a donc ici une double externalité positive de production.

5.4.3 Théorème de Coase

Pour l'économiste britannique Ronald Coase (1910–2013), pilier de l'école de Chicago, les externalités résultent d'un sous développement de la propriété privée. Prenons l'exemple de la rivière polluée en amont par une entreprise chimique, ce qui nuit en aval à des pêcheurs. Une solution serait que la rivière appartienne aux pêcheurs. Ils pourraient alors vendre des quotas de pollution à l'entreprise chimique. La création d'un droit de propriété privée permettrait des échanges mutuellement bénéfiques, ce qui garantirait l'efficacité. Ce résultat est connu sous le nom de **théorème de Coase**.

La force de ce résultat est que l'efficacité ne dépend nullement de l'identité des agents à qui sont attribués les droits de propriété privée. Ainsi, la rivière pourrait appartenir à

l'entreprise chimique. Dans ce cas, ce serait les pêcheurs qui payeraient l'entreprise pour qu'elle limite sa pollution, et cette solution serait tout aussi efficace. D'ailleurs, que la rivière appartienne aux pêcheurs ou à l'entreprise, une fois les transactions réalisées, le niveau de pollution serait le même dans ces deux cas.⁵ Bien sûr, chacun préfère que la propriété de la rivière lui soit attribuée, même si à l'équilibre cela ne change pas le niveau de la pollution.

Malheureusement, en pratique, les droits de propriété privée ne sont pas faciles à attribuer et, une fois attribués, le coût des transactions est souvent prohibitif. C'est justement à cause de ces difficultés que les externalités existent.

5.4.4 Les taxes pigouviennes

En présence d'une défaillance de marché, une question se pose inmanquablement : l'État peut-il intervenir pour restaurer l'efficacité ? Si oui, comment ?

En théorie, la solution est simple : dans le cas d'une externalité négative, il suffit d'imposer une taxe égale au coût marginal externe de production. Cela force les entreprises à *internaliser* les externalités. Cette taxe payée par le producteur a pour effet de translater vers le haut la courbe d'offre du montant de la taxe. Ainsi, sur la figure 5.11, la courbe d'offre coïncide désormais avec la courbe de coût marginal social, ce qui réduit la perte sèche à néant et restaure l'efficacité.⁶ On parle de *taxe pigouviennne*, du nom de l'économiste britannique Arthur Pigou qui fut le premier à la proposer en 1920. Non seulement cette taxe ne crée aucune distorsion nouvelle, mais elle supprime en plus les distorsions dues aux externalités ! Les taxes pigouviennes représentent donc un moyen remarquablement efficace de générer des recettes fiscales. Elles relèvent de l'application du principe du « pollueur-payeur », dont la meilleure illustration est la taxe carbone utilisée pour lutter contre le réchauffement climatique. Une taxe pigouviennne est un instrument de marché : ce n'est pas l'État qui décide du niveau de consommation et de pollution des différents agents de l'économie. Chacun est mis face à ses responsabilités et est conduit, grâce à la taxe, à internaliser l'impact environnemental de ses choix. Ainsi, une taxe pigouviennne ne supprime pas le marché, elle le corrige !

Dans le cas d'une externalité positive, la solution consiste à instaurer une subvention égale au bénéfice marginal externe de la consommation. Cela justifie par exemple que l'éducation primaire et secondaire, la recherche fondamentale ou les campagnes de vaccination soient largement financées par l'État.

5.4.5 Application au réchauffement climatique

Le réchauffement climatique, causé par un excès d'émissions de gaz à effet de serre, est une gigantesque défaillance de marché due à une externalité environnementale. Pour ré-

5. Ce résultat n'est pas vrai si les sommes échangées sont si élevées qu'elles modifient significativement la richesse totale des pêcheurs ou des entreprises. Il s'agit d'*effets de richesse* que nous supposons nuls dans ce cours d'introduction. Les effets de richesse ne sont jamais source d'inefficacité.

6. Conformément à ce que nous avons vu précédemment, cette taxe a exactement le même effet si elle est payée par le consommateur. En ce cas, graphiquement, c'est la courbe de demande qui est translaturée vers le bas aboutissant à la quantité q^0 .

soudre le problème, nous pourrions décider de cesser toute activité économique. Mais cette solution serait ruineuse. D'où la question : comment concilier la préservation indispensable de notre planète et la prospérité économique à laquelle l'écrasante majorité des individus aspirent légitimement ?

D'après les conclusions du GIEC (Groupe d'experts Intergouvernemental sur l'Évolution du Climat), pour limiter le réchauffement climatique à 1,5°C, nos émissions *nettes* de gaz à effet de serre devront être réduites d'environ 45% en 2030 par rapport à 2010, et atteindre zéro d'ici 2050. Il est donc nécessaire de réduire drastiquement nos émissions, mais il n'est pas nécessaire de les réduire à néant. Même à l'horizon 2050, pour parvenir à la neutralité carbone, les émissions resteront positives mais seront compensées par des efforts de capture du carbone, par exemple en reboisant ou en extrayant directement le carbone de l'atmosphère.

Quels sont donc les biens que nous pouvons continuer à consommer, et quels sont ceux dont nous devons nous passer ? Une réponse est donnée par la taxe carbone. Il s'agit d'une taxe pigouvienne qui augmente le prix de chaque bien proportionnellement aux émissions qu'il génère, ce qui force les différents acteurs de l'économie à internaliser l'impact climatique de leurs comportements. En 2019, le Fonds Monétaire International a estimé qu'une taxe carbone mondiale de \$75 par tonne de CO₂ permettrait de limiter le réchauffement climatique à 2°C.

On estime qu'un aller-retour en avion de Paris à New-York en classe économique produit presque deux tonnes de carbone (le double pour une place en business et le triple pour une première). Notre taxe augmenterait donc le prix du billet de \$150, soit €130. Face à cette augmentation, certains consommateurs renonceront à leur voyage, ce qui est efficace car leur propension à payer est inférieure au coût de production total de ce déplacement, en y incluant le coût de la pollution. D'autres personnes préféreront payer la taxe et maintenir leur voyage. Il est donc efficace que ces consommateurs aillent à New-York et polluent la planète car le bénéfice qu'ils tirent de ce voyage est supérieur à son coût de production, en y incluant le coût environnemental.

La taxe carbone supprime les activités polluantes les moins valorisées par les consommateurs et préserve celles qui génèrent le plus de surplus. Il s'agit donc d'un mécanisme efficace pour trouver le meilleur compromis entre prospérité économique et préservation de la planète.

S'il est si difficile de lutter contre le réchauffement climatique, c'est précisément parce que cela coûte cher. Il est donc indispensable que chaque euro investi pour la préservation de notre planète ait l'impact environnemental maximal. C'est justement ce que permet la taxe carbone, et cela explique pourquoi les économistes soutiennent presque unanimement cette solution.

Est-il juste que les plus riches polluent davantage que les plus pauvres ? Comme le suggère notre exemple du transport aérien, les plus riches ont plus de chances de maintenir leur voyage après l'instauration d'une taxe, et donc de polluer la planète, que les plus pauvres. Mais les plus riches consomment plus de ressources en général. Il serait donc inefficace de

chercher à imposer l'équité pour les seules consommations polluantes. D'ailleurs, si l'État imposait des quotas de pollution individuels, les plus pauvres seraient prêts à vendre les leurs aux plus riches, ce qui serait une amélioration de Pareto. Un excès d'inégalités doit donc se régler en redistribuant les revenus; libre ensuite à chacun de dépenser ses propres revenus (nets d'impôt et de subventions) comme il le souhaite. Cette redistribution peut d'ailleurs être en partie financée par les recettes de la taxe carbone.

Permis d'émissions échangeables. Une alternative à la mise en place d'une taxe carbone est l'utilisation d'un système de permis d'émissions. Dans un tel système, dit de *cap-and-trade*, le gouvernement décide de la quantité de gaz à effet de serre que chaque entreprise est autorisée à émettre. Si une entreprise dispose d'un quota de pollution trop élevé, elle peut revendre ses permis aux entreprises qui n'en ont pas assez. En Europe, une telle politique est mise en place depuis 2005 sous le nom de EU ETS (European Union Emission Trading Scheme). En l'absence d'incertitude, la taxe carbone est rigoureusement identique à des permis d'émissions vendus par l'État. La régulation par les prix est donc équivalente à la régulation par les quantités. En pratique, les États les plus soucieux de l'environnement mettent en place une combinaison de ces deux politiques.

Les entreprises préfèrent les permis d'émissions échangeables, à condition que les permis leurs soient initialement donnés (généralement sur la base de leurs pollutions passées). Conformément au théorème de Coase, ce transfert de richesse (par rapport à une taxe carbone ou à des permis vendus par l'État), n'entrave pas l'incitation à réduire la pollution à un niveau efficace.

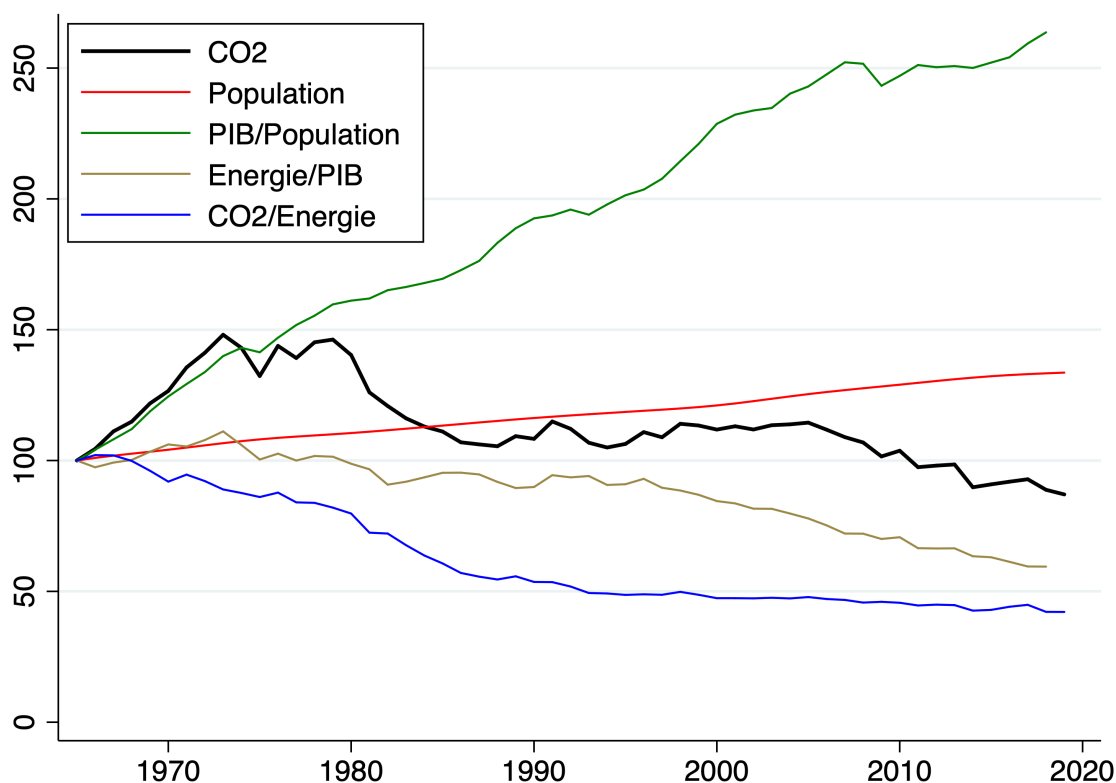
Croissance et émissions de carbone. Pour comprendre les différents chemins que nous pouvons emprunter pour réduire nos émissions de gaz à effet de serre, il est utile de voir l'identité de Kaya, due à l'économiste japonais Yoichi Kaya en 1993. :

$$\text{CO}_2 = \text{Population} \times \frac{\text{PIB}}{\text{Population}} \times \frac{\text{Énergie}}{\text{PIB}} \times \frac{\text{CO}_2}{\text{Énergie}}.$$

Pour réduire les émissions de CO₂, il nous faut réduire un ou plusieurs termes multiplicatifs de l'identité. Une première manière de réduire nos émissions est d'œuvrer pour la décroissance de la population mondiale ou du PIB par habitant. Une alternative, a priori plus efficace, consiste à réduire l'intensité énergétique de notre production ou à réduire l'intensité en carbone de notre énergie. C'est ce qu'on appelle la croissance verte! Cela est non seulement possible, mais cela se produit déjà dans de nombreux pays Européens, à l'image de la France dont le cas est représenté sur la figure 5.14. Depuis 30 ans la croissance du PIB ne s'accompagne plus d'une augmentation de notre consommation énergétique, et les énergies décarbonées (renouvelables et nucléaires) représentent une part croissante de notre mix énergétique.

Certains des biens les plus polluants que nous consommons sont importés de l'étranger. La figure 5.15 corrige cet effet : la courbe marron inclut l'énergie dépensée pour la production des biens importés, tout en soustrayant l'énergie dépensée pour produire les biens exportés.

FIGURE 5.14 – Décomposition de Kaya pour la France



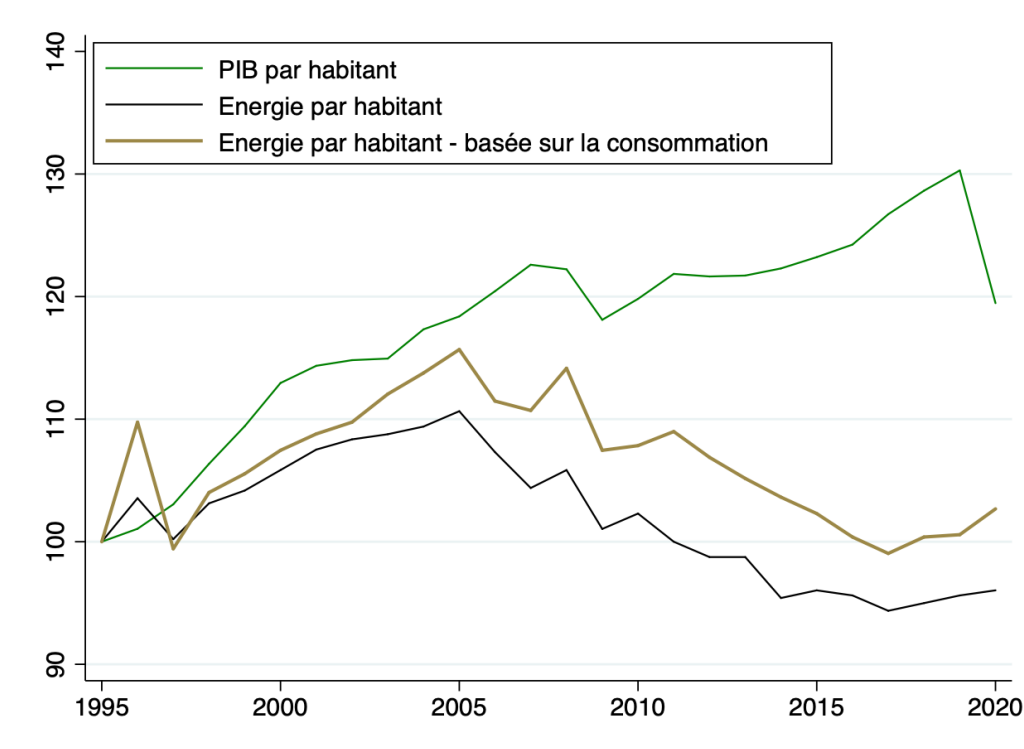
Source : Our World in Data based on Global Carbon Project.

Il s'agit donc de l'énergie dont les français ont besoin pour maintenir leur niveau de vie. On observe que, même en prenant en compte la délocalisation d'une partie de notre pollution, notre consommation énergétique a diminué au cours des dernières décennies, tandis que notre PIB a continué de croître.

La croissance de demain ne devrait pas être faite de grosses voitures et d'immenses maisons isolées et inchauffables, mais de nouveaux logiciels informatiques, de nouvelles molécules à usage pharmaceutique, d'ordinateurs plus puissants et de moyens de communication plus performants. La taxe carbone encourage naturellement cette croissance verte, et n'entraîne la décroissance que des seuls secteurs d'activité les plus polluants. Soulignons que si les innovations vertes (comme dans les énergies renouvelables) engendrent des externalités positives, contribuant par exemple à réduire les coûts de production futurs, alors des subventions à l'innovation sont justifiées en complément de la taxe carbone.

Privilégier la décroissance plutôt que la croissance verte, c'est considérablement accroître le coût de la lutte contre le réchauffement climatique. C'est même une impasse qui forcerait les pays les plus pauvres à renoncer à la préservation de la planète au nom de l'impérieuse nécessité de leur développement économique.

FIGURE 5.15 – Énergie par habitant en France, corrigée des effets du commerce international.



Source : Our World in Data.

5.4.6 Taxes vs. réglementations

Face aux problèmes de pollution, les gouvernements ont tendance à imposer un ensemble de réglementations et d'interdictions. En général, une taxe pigouvienne est plus efficace qu'une interdiction, qui correspond à une taxe infiniment élevée. La taxe préserve une certaine liberté de choix, ce qui permet aux personnes qui valorisent le plus certaines consommations polluantes de ne pas y renoncer. Par exemple, le gouvernement français a imposé en 2022 la suppression des lignes aériennes pour lesquelles existaient une alternative en train de moins de 2h30. Une taxe carbone aurait sans doute été un moyen plus efficace de réduire le trafic aérien le moins utile au regard de la pollution qu'il génère.

Un inconvénient des réglementations est qu'elles aboutissent à un effort excessif pour certaines sources de pollution et insuffisant pour d'autres. Par exemple, les subventions à la voiture électrique coûtent autour de \$250 par tonne de carbone évitée, tandis que le remplacement d'une chaudière au fioul par une pompe à chaleur ne coûte que \$50. Il est donc cinq fois plus efficace de subventionner les changements de chaudière que les changements de véhicule!

Un attrait fondamental de la taxe carbone est son étendue. Dans le cas du réchauffement climatique, la réglementation devient vite d'une complexité byzantine. Faut-il interdire les kiwi importés de Nouvelle-Zélande? Et les mangues du Pérou? Qu'en est-il des camemberts de Normandie, produits avec des vaches qui émettent du méthane en quantité? Il est quasiment impossible pour un gouvernement de mesurer l'intensité en carbone des différents biens de consommation. En revanche, à chaque étape de la production, la taxe carbone augmente

le coût de fabrication proportionnellement à ses émissions de CO₂. Conduits par cette main invisible de la taxe carbone, les consommateurs sont poussés à faire des choix efficaces.

En pratique, l'opposition politique à la taxe carbone est forte, comme l'a montré le mouvement des gilets jaunes provoqué en 2018 par une augmentation de la taxe sur les carburants. Les citoyens redoutent légitimement que la lutte contre le réchauffement climatique ne soit qu'un prétexte pour les ponctionner davantage. En pratique, il semble donc important que la taxe carbone se fasse à niveau de prélèvements obligatoires inchangé : une hausse de la taxe carbone doit mécaniquement provoquer une baisse d'autres impôts, comme par exemple la TVA ou l'impôt sur le revenu.

5.5 Les biens publics

Jusqu'à présent, nous avons vu que les marchés organisaient de manière efficace la production et l'échange des biens et des services. Mais nous avons supposé que ces biens et services étaient privés, à l'image d'une crème glacée ou d'une chemise. Il existe également des biens dits publics, pour lesquels les marchés sont défaillants.

5.5.1 Les différents types de biens

Pour comprendre la nature du problème, il convient de classer les biens de consommation selon deux dimensions : leur excluabilité et leur rivalité.

Un bien est **excluable** lorsque l'on peut empêcher quelqu'un de le consommer.

Par exemple, l'acquéreur d'une crème glacée est en droit de décider qui la mangera. Lorsqu'une entreprise crée un logiciel informatique, elle peut en interdire l'utilisation à ceux qui ne paient pas de licence. La télévision câblée est également excluable. En revanche, les poissons en mer sont non-excluables, puisque tant qu'ils sont en liberté n'importe quel pêcheur est en droit d'essayer de les attraper. De même qu'on ne peut pas empêcher qui que ce soit de bénéficier de l'éclairage urbain ou de la défense nationale. Enfin, une route est excluable si elle dispose d'un péage, mais pas sinon.

Un bien est **rival** lorsque sa consommation par une personne empêche sa consommation par une autre personne.

La crème glacée est un bien rival puisque celui qui la mange empêche les autres d'en profiter. Les poissons en mer constituent également un bien rival car lorsqu'ils sont pêchés par un marin, les réserves diminuent ce qui rend la pêche plus difficile pour les autres marins. En revanche, un logiciel informatique n'est pas un bien rival, puisqu'un utilisateur ne souffre nullement du fait que son logiciel soit également utilisé par d'autres. Idem pour la télévision câblée. Il n'y a pas non plus de rivalité dans l'éclairage urbain ou dans la défense nationale, car le bénéfice qu'un citoyen tire de ces services ne dépend pas du nombre d'autres citoyens

en bénéficiant. Enfin, une route encombrée est rivale, puisque chaque automobiliste supplémentaire aggrave la congestion, tandis qu'une route fluide est non-rivale.

Un **bien privé** est défini comme étant excluable et rival, à l'image de notre crème glacée. C'est l'opposé d'un bien public.

Un **bien public** n'est ni excluable, ni rival.

L'éclairage des rues ou la défense nationale sont donc des biens publics. Il existe enfin deux catégories intermédiaires : les **ressources communes** sont rivales mais pas excluables, à l'image des poissons en mer, et les **biens de club** sont excluables mais pas rivaux, à l'image des logiciels informatiques ou des chaînes de télévision câblée. Ces différentes possibilités sont représentées dans la table 5.3.

TABLE 5.3 – Classification des biens selon leur excluabilité et leur rivalité

	<i>Excluable</i>	<i>Non-excluable</i>
<i>Rival</i>	Bien privé Crème glacée Route à péage encombrée	Ressource commune Poissons en mer Route sans péage encombrée
<i>Non-rival</i>	Bien de club Logiciel informatique Télévision câblée Route à péage fluide	Bien public Eclairage des rues Défense nationale Route sans péage fluide

5.5.2 Le problème du passager clandestin

Le marché est incapable de produire une quantité suffisante de biens publics, pour une raison qui est simple : personne ne souhaite les financer. Dès lors que quelqu'un investit dans un bien public, tout le monde peut profiter de cet investissement. Ainsi, chacun espère que ce sera son voisin qui fera l'effort nécessaire pour financer le bien public. C'est le *problème du passager clandestin* (le *free-rider problem* en anglais).

Si l'éclairage urbain n'était pas financé par l'État, il est probable qu'il n'existerait pratiquement pas. Le coût d'installation d'un lampadaire est très élevé, alors qu'un individu en tire un bénéfice personnel assez limité. Néanmoins, comme ce lampadaire profitera à un grand nombre de personnes, il crée un surplus total important, ce qui rend désirable son installation. L'État peut donc prendre en charge le financement des biens publics, corrigeant ainsi la défaillance de marché.

5.5.3 Application au réchauffement climatique

Au-delà du problème d'externalité environnementale déjà abordé, le climat est un bien public mondial. Cela implique un grave problème de passager clandestin : chaque pays juge son propre impact relativement négligeable et espère que les autres feront les efforts nécessaires pour prévenir le changement climatique. Là est le nœud du problème ! C'est

pourquoi les accords internationaux n'ont jamais été à la hauteur du défi : soit les accords manquent d'ambition (COP 21 à Paris en 2015 ou COP 26 à Glasgow en 2021), soit ils ne sont pas signés par suffisamment de pays (Protocole de Kyoto en 1997).

À l'échelle mondiale, une grosse partie des émissions résulte de la consommation de charbon en Chine. Les efforts substantiels réalisés par les pays européens sont vains s'ils ne s'inscrivent pas dans une démarche collective à l'échelle mondiale. Face à cette dure réalité, que pouvons-nous faire ?

L'Union Européenne dispose d'un outil puissant : elle représente le premier marché au monde. Les pays européens pourraient donc s'engager, à compter d'une certaine date, à ne plus commercer avec les pays qui ne font pas leur part de l'effort nécessaire à la préservation de notre planète. L'objectif ne serait en aucun cas de déclencher une guerre commerciale à l'échelle mondiale, mais de créer une coalition mondiale pour le climat. Ainsi, il deviendrait individuellement rationnel pour chaque pays du monde de limiter drastiquement ses émissions de gaz carbonique.

Une telle stratégie est d'ailleurs préconisée par William Nordhaus, prix Nobel d'économie en 2018 pour ses travaux sur le réchauffement climatique (Nordhaus, 2015). Selon lui, la meilleure manière de surmonter le problème du passager clandestin est de former un club climatique. L'appartenance au club permettrait de bénéficier de barrières douanières très basses, mais pour y adhérer il serait nécessaire soit d'avoir une taxe carbone suffisamment élevée, soit de faire d'autres efforts équivalents en termes de réductions d'émissions.

Soulignons que le problème du passager clandestin contribue à la dégradation d'autres ressources mondiales sans frontières : les hautes mers, victimes de pollution et de surpêche, les bactéries, de plus en plus nombreuses à résister aux antibiotiques du fait de certaines pratiques médicales et agricoles, les virus, dont le développement est favorisé par le déplacement d'espèces vivantes hors de leurs habitats historiques, et l'espace en orbite basse, où les débris s'accumulent. À nouveau, la politique commerciale semble être un instrument important pour forcer tous les pays à faire les efforts nécessaires.

5.5.4 La tragédie des biens communs

Intéressons-nous désormais aux ressources communes qui sont rivales, mais ne sont pas excluables, à l'image des poissons en mer. Chaque marin décide de l'intensité de ses efforts de pêche en fonction de ses coûts privés, mais n'intègre pas dans ses choix l'externalité négative qu'il impose sur les autres pêcheurs : en réduisant la quantité de poissons disponible, il fait peser sur eux un coût direct. À l'équilibre de marché, il y a donc surpêche. La surexploitation des ressources communes est appelée **tragédie des biens communs**. Pour restaurer l'efficacité, l'État doit trouver le moyen de restreindre l'accès à la ressource commune.

En 2009, Elinor Ostrom a obtenu le prix Nobel d'économie pour ses travaux sur les manières dont différentes communautés à travers le monde parvenaient à surmonter la tragédie des biens communs. Elle a mis en évidence la capacité des petites communautés à parvenir spontanément à une exploitation efficace de la ressource commune.⁷

7. Ces principes sont décrits dans Ostrom (1990).

5.5.5 Les biens de clubs

Les biens de club sont excluables, mais non-rivaux. Il s'agit typiquement de biens dont la production engendre des coûts fixes très élevés, mais des coûts variables quasiment nuls. Par exemple, le coût de la conception d'un logiciel informatique est exorbitant, tout comme le coût de la production d'un film ou de l'élaboration d'une nouvelle molécule. Mais, une fois la première unité réalisée, les unités suivantes ne coûtent quasiment rien à produire. Il serait donc efficace que ces biens de club soient en libre accès. Or, les entreprises souhaitent vendre ces biens à des prix élevés afin d'amortir leurs coûts fixes, ce qui crée un problème de *rareté artificiellement entretenue*.

En pratique, la solution consiste souvent à décerner un brevet à l'entreprise qui a conçu ce bien de club, afin qu'elle jouisse d'un pouvoir de monopole lui permettant de rémunérer son innovation. Au bout d'un certain temps, le brevet expire et le bien tombe dans le domaine public ce qui permet à son prix de refléter enfin son coût de production marginal très faible.

5.6 Les interventions de l'État

Nous disposons désormais d'une grille d'analyse nous permettant de savoir quand une intervention de l'État dans l'activité économique est souhaitable. Dans une économie de marché, le *laissez-faire* permet la réalisation d'échanges mutuellement bénéfiques. Sous certaines conditions, le premier théorème du bien-être nous garantit que l'allocation des biens qui en résulte est Pareto-efficace.

D'un point de vue économique, il existe deux justifications possibles pour une intervention de l'État. La première est qu'au moins une des conditions du premier théorème n'est pas satisfaite, aboutissant à une allocation Pareto-inefficace. On parle alors de défaillance de marché. Les principales défaillances du marché sont : les externalités, les biens publics, la concurrence imparfaite, les asymétries d'information, et la rigidité des prix.

L'autre justification est l'existence d'un problème d'équité. Même lorsque le marché est parfaitement efficace, on peut considérer que la distribution des richesses qui en résulte est injuste. Par exemple, si les inégalités sont perçues comme étant trop élevées, l'État peut intervenir pour redistribuer les ressources des plus riches, vers les plus pauvres. Comme nous l'avons déjà souligné, les sciences économiques ne permettent pas de dire si une distribution des ressources est juste ou pas. Fondamentalement, il ne s'agit pas d'une question d'ordre scientifique, mais philosophique.

Face à une défaillance du marché ou à un problème d'équité, l'État doit-il nécessairement intervenir? Non, car il ne faut pas négliger la possibilité d'une **défaillance de l'État**, qui peut se produire pour deux raisons. D'abord, il est possible que l'État ne dispose pas de l'information ou des instruments nécessaires pour remédier à une situation indésirable. Par exemple, la redistribution des ressources est délicate et risque de décourager le travail, l'épargne et l'entreprise, ce qui limite sérieusement les possibilités en la matière.

Une défaillance plus profonde de l'État tient à sa nature. Le planificateur bienveillant est une abstraction théorique. Dans la réalité, l'action de l'État est conduite par

des hommes et des femmes qui, comme chacun d'entre nous, ont légitimement des intérêts privés à défendre : les responsables politiques cherchent à être élus et réélus ; les hauts fonctionnaires souhaitent accumuler des responsabilités et obtenir des promotions. Or, la défense de ces intérêts privés ne va pas forcément dans le sens de l'intérêt général.

En 1986, l'Américain James Buchanan a obtenu le prix Nobel d'économie pour avoir intégré ces considérations dans notre discipline. Ses travaux sont le point de départ d'une immense littérature en **économie politique** (*political economy* en anglais) qui consiste à utiliser les outils théoriques et empiriques des sciences économiques afin d'étudier le fonctionnement des systèmes politiques et leur interactions avec l'allocation des ressources. L'absence de croissance dans les pays en développement est bien souvent la conséquence de blocages politiques et institutionnels, qui sont les marqueurs de défaillances majeures de l'État.

Bibliographie

Fajgelbaum, P. D., Goldberg, P. K., Kennedy, P. J. and Khandelwal, A. K. (2020), 'The return to protectionism', *The Quarterly Journal of Economics* **135**(1), 1–55.

Nordhaus, W. (2015), 'Climate clubs : Overcoming free-riding in international climate policy', *American Economic Review* **105**(4), 1339–1370.

Ostrom, E. (1990), *Governing the Commons : The Evolution of Institutions for Collective Action*, Cambridge university press.

Exercices du chapitre 5

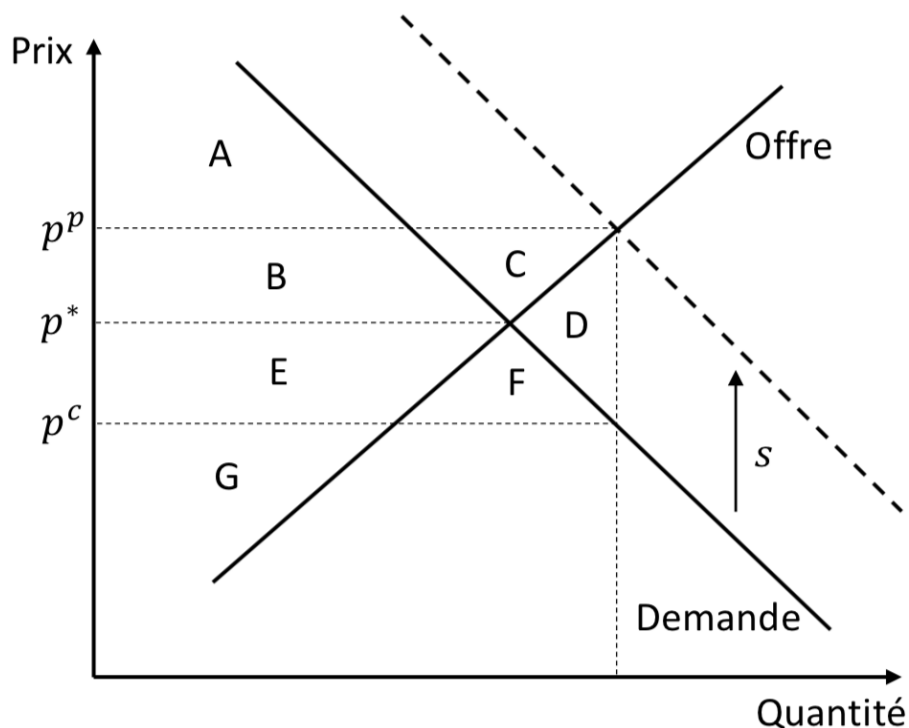
Exercice 1 : Imposer la rente foncière

La rente foncière correspond au revenu pour le propriétaire de la mise en location de la terre nue (la terre non améliorée). Quel est l'incidence et l'efficacité d'un impôt sur la rente foncière? Est-ce un bon impôt?

Exercice 2 : Subvention

Soit le marché d'un bien avec un prix d'équilibre p^* . L'Etat introduit une subvention s à la consommation de ce bien, ce qui déplace la courbe de demande vers le haut, comme indiqué sur la Figure 5.16 ci-dessous.

FIGURE 5.16 – Analyse d'une subvention.



A/ Quel est le montant du surplus du producteur et du consommateur avant l'implémentation de la subvention?

B/ Analysez l'impact de la subvention sur le bien-être des consommateurs et des producteurs ainsi que sur les recettes gouvernementales.

C/ La subvention engendre-t-elle une perte sèche? Expliquez.

D/ En France, les Aides Personnalisées au Logement (APL) sont des subventions versées aux locataires. Quels effets peut-on attendre de ces aides?

Exercice 3 : Réchauffement climatique (d'après examen 2019)

Comment résoudre efficacement le problème du réchauffement climatique? Faut-il que chaque pays fasse le même effort de réduction d'émission? Est-ce que tous les pays devraient chacun fixer leur propre taxe carbone, à des niveaux potentiellement différents? Et, si la taxe carbone était au même niveau dans l'ensemble du globe, est-ce que cela répartirait équitablement l'effort de réduction d'émission? En quoi est-ce qu'une politique de quotas d'émission diffère d'une taxe carbone? L'objectif de cet exercice est de répondre précisément à ces questions qui sont cruciales pour la conduite des négociations internationales sur le réchauffement climatique.

Le monde est composé de deux pays, N et S . Soit Q_i la quantité d'émissions en tonne de CO₂ que choisit d'émettre le pays $i \in \{N, S\}$ et \bar{Q}_i la quantité correspondante en cas de *laissez-faire*. L'effort de réduction d'émission du pays i s'élève donc à $\bar{Q}_i - Q_i$, ce qui engendre un coût pour ses citoyens de :

$$\alpha_i \frac{(\bar{Q}_i - Q_i)^2}{2}.$$

On suppose $\alpha_N > \alpha_S$.

1/ Interprétez le paramètre α_i et donnez un exemple de différence géographique entre les deux pays aboutissant à $\alpha_N > \alpha_S$.

2/ Pour préserver le climat, les scientifiques estiment que les émissions mondiales doivent diminuer d'une quantité $2E$. Un accord international oblige donc chaque pays à réduire ses propres émissions de E . Cette politique est-elle Pareto efficace? Expliquez.

3/ Supposons que les deux pays s'accordent pour instaurer une taxe carbone au taux de τ par tonne de CO₂ émise. Les citoyens du pays i doivent donc payer τQ_i à leur gouvernement afin d'émettre Q_i tonnes de CO₂.

(i) Montrez que pour réduire les émissions mondiales de $2E$, la taxe carbone doit être fixée à :

$$\tau = 2E \frac{\alpha_N \alpha_S}{\alpha_N + \alpha_S}.$$

Interprétez cette formule.

(ii) Pour chaque pays i , calculez la réduction d'émission ainsi que le coût total de cette politique. Interprétez.

(iii) Cette politique est-elle efficace? Cette politique est-elle équitable?

4/ Chaque pays i dispose désormais d'un quota d'émission C_i échangeable internationalement, tel que les émissions mondiales diminuent d'une quantité $2E$, ce qui donne $C_S + C_N = \bar{Q}_S + \bar{Q}_N - 2E$.

(i) Déterminez l'équilibre du marché d'émission, où p désigne le prix d'une tonne de CO₂. Comparez le niveau d'émission de chaque pays avec celui engendré par la taxe carbone.

(ii) Supposons que les quotas sont répartis tel que, en l'absence d'échange, chaque pays doit réduire ses propres émissions de E , ce qui implique $C_i = \bar{Q}_i - E$. Calculez le coût de cette politique pour chaque pays (avec la possibilité d'échanger les quotas). Comparez ce coût à celui de la taxe carbone. Interprétez.

5/ Quelles conclusions tirez-vous de cet exercice concernant la conduite des négociations internationales sur le réchauffement climatique?

Exercice 4 : L'éducation

Le financement de l'éducation doit-il être public? Pourquoi? Les écoles doivent-elles être publiques?

Chapitre 6

Concurrence imparfaite

Lors de l'analyse de l'équilibre de marché, nous avons supposé que les consommateurs et les producteurs sont *preneurs de prix*, c'est à dire qu'ils considèrent les prix des biens comme des données externes, et qu'ils optimisent leurs décisions d'achat et de production à prix donné. Lorsque chaque producteur et chaque consommateur a une taille négligeable par rapport au marché, chacun d'entre eux a une influence faible sur les prix ; l'approximation selon laquelle ils sont preneurs de prix est appropriée. Cependant, lorsque les marchés sont concentrés dans les mains d'un ou plusieurs grands acheteurs, ou d'un ou plusieurs producteurs, alors ces acteurs ont une influence importante sur les prix. On dit qu'ils exercent un *pouvoir de marché* et que la concurrence est imparfaite.¹

6.1 Le pouvoir de marché

6.1.1 Qu'est-ce que le pouvoir de marché ?

On distingue les cas suivants de pouvoir de marché, selon que le marché est concentré entre les mains d'un ou plusieurs acteurs dominants, et selon que ce pouvoir de marché est détenu par les acteurs en situation d'offre ou de demande.

- Dans le cas d'un *monopole*, un producteur est seul ou dominant sur un marché. Par exemple, la SNCF a bénéficié du monopole du transport ferroviaire de passagers jusqu'en fin 2020. Jusqu'aux années 2000, De Beers avait le monopole de la production et de la distribution de diamants dans le monde.
- L'*oligopole* décrit un marché concentré dans les mains d'un petit nombre de producteurs. Dans la plupart des pays, les opérateurs de téléphonie mobile forment des oligopoles. Dans le cas de deux producteurs, on parle alors de duopole. La construction aéronautique de long courriers est un duopole entre Airbus et Boeing.
- Dans le cas d'un monopole ou d'un oligopole, le pouvoir de marché est détenu par les acteurs de l'offre qui peuvent librement ajuster les prix. De façon symétrique, il existe des cas dans lesquels le pouvoir de marché est détenu par les acteurs de la demande.

1. Dès la seconde moitié du XIX^{ème} siècle, Karl Marx mettait en garde face au phénomène d'accumulation de capital pouvant donner lieu à une concentration excessive du contrôle des moyens de production.

Le *monopsonie* est la situation d'un acheteur dominant ou unique. Une telle situation arrive parfois sur le marché du travail, où les entreprises achètent le travail offert par les ménages. Lorsqu'il n'y a qu'une entreprise en activité dans une zone géographique, les ménages doivent lui fournir leur travail, donnant à cette entreprise une position de monopsonie : c'est le cas de la Compagnie des Mines de Montsou, employeur quasi-unique dans le roman *Germinal* de Zola.

- Dans un marché où un petit nombre d'acheteurs sont dominants, on parle d'un **oligopsonie**. Un exemple d'oligopsonie est celui de la transformation de lait en France, secteur dans lequel 4 acheteurs dominants représentent 60% du marché en valeur.

De manière plus générale, on peut quantifier la concentration d'un secteur économique en utilisant des indices, dont le plus connu est l'indice de Herfindahl-Hirschmann. Celui-ci est égal à la somme des carrés des parts de marché possédés par chaque entreprise ($H = \sum s_i^2$). Pour un monopole, une seule entreprise possède une part de marché $s_i = 1$, et l'indice vaut alors 1. Dans le cas où n entreprises possèdent chacune une part de marché $s_i = 1/n$, l'indice vaut $H = 1/n$, lequel tend vers 0 lorsque le nombre d'entreprises croît.

6.1.2 D'où provient le pouvoir de marché?

La concentration varie fortement d'un secteur à l'autre : la restauration est relativement peu concentrée tandis que la production automobile l'est beaucoup plus (les cinq groupes Volkswagen, Toyota, Ford, General Motors et Renault-Nissan représentent plus de la moitié de la production mondiale). Quelles sont les forces économiques qui expliquent ces différences de concentration?

Barrières légales. Dans certains cas, la concentration d'un marché provient tout simplement de la loi. Lorsqu'un écrivain, un musicien, ou un peintre produit une œuvre, celle-ci ne peut pas être copiée par d'autres pendant une longue période : dans l'Union Européenne, une production artistique est protégée jusqu'à 70 ans après la mort de son auteur, après quoi elle tombe dans le domaine public et peut être librement reproduite. Pendant toute cette période, l'auteur dispose d'un monopole sur la vente de son œuvre. De même, un inventeur peut faire breveter ses créations, et possède alors un monopole (quoique de durée limitée) sur l'exploitation de ses inventions. L'État peut parfois s'imposer comme monopole pour certains services (comme le train, l'électricité ou la poste). Lorsqu'il alloue des licences (comme pour les taxis, les fréquences radio ou téléphoniques), il agit également directement sur la structure du marché.

Structures de production. Néanmoins, la présence d'un faible nombre d'acteurs sur un marché n'est pas toujours la conséquence d'une intervention de l'État : dans certaines circonstances, une structure concentrée s'imposera naturellement. C'est notamment le cas dans les secteurs avec un coût d'entrée important, comme dans l'aviation ou dans l'automobile : amortir les énormes coûts de recherche et développement dans ces secteurs nécessite de

vendre un nombre important d'avions ou de voitures, et s'il y avait un grand nombre d'entreprises dans ce secteurs, elles ne pourraient pas toutes survivre. Pour des grandes infrastructures comme un réseau électrique ou ferroviaire, les coûts d'entrée sont tellement importants qu'une seule entreprise (ou l'État) consentira à entrer sur le marché. Une fois qu'un réseau électrique a été construit, il serait totalement inefficace qu'une entreprise en construise un second en parallèle : dans une telle situation, on parle de *monopole naturel*.

Accès aux ressources. Si de Beers avait un monopole sur la production diamantaire, c'est car l'entreprise possédait l'immense majorité des mines de diamant dans le monde : un accès limité aux ressources peut donc causer une situation de monopole.

Effets de réseau. Dans certains cas, le bénéfice que l'on tire d'un produit augmente lorsque les autres utilisent le même produit. Si tous mes amis utilisent Facebook, j'ai intérêt à utiliser Facebook plutôt que Twitter : on parle d'effet de réseau. De même, si la plupart des chauffeurs de VTC utilisent Uber, un consommateur a intérêt à utiliser le service d'Uber plutôt que celui d'un concurrent, et un chauffeur débutant aura intérêt à utiliser Uber comme plateforme, où il trouvera plus de clients. Dans cette situation, Uber développe une position dominante dans une activité qui a pourtant des coûts d'entrée faibles.

6.2 Analyse du monopole

Nous avons désormais détaillé précisément les circonstances dans lesquelles l'hypothèse d'agents preneurs de prix n'était pas vérifiée. En quoi est-ce que l'invalidité de cette hypothèse modifie le fonctionnement d'un marché tel que décrit au début de ce cours ? Que se passe-t-il lorsque nous considérons des situations où des agents ont un pouvoir de marché ?

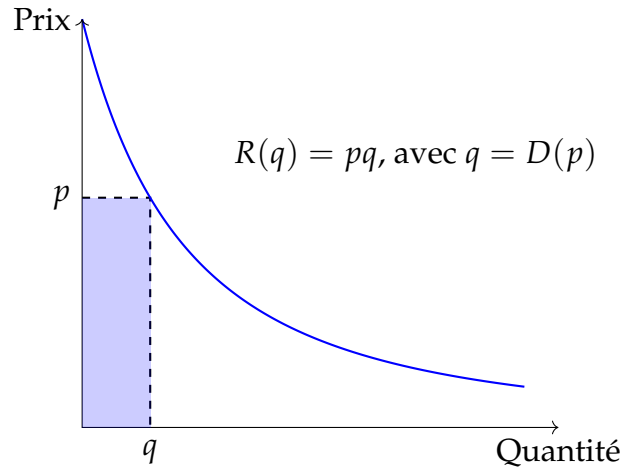
Pour comprendre les conséquences économiques d'un marché concentré, commençons par prendre un exemple extrême : celui du monopole. Considérons une entreprise privée qui est la seule à produire un bien, et qui cherche à maximiser son profit. Que se passe-t-il dans cette situation ?

Rappelons que le profit est la différence entre le revenu et les coûts de production. Le revenu du monopole est le produit de la quantité vendue q par le prix unitaire p . Les termes p et q sont liés par la fonction de demande D pour le bien : on a $q = D(p)$. Une caractéristique essentielle du monopole est de pouvoir librement choisir un prix de vente. S'il choisit p , il vend la quantité $D(p)$. Il peut aussi choisir une quantité vendue q et le prix de vente est dans ce cas $p = D^{-1}(q)$. Autrement dit, le monopole choisit un point sur la courbe de demande.

Écrivons le revenu du monopole en fonction de la quantité produite : $R(q) = qp$ avec $D(p) = q$. Graphiquement, $R(q)$ est donc l'aire du rectangle situé sous la courbe de demande, comme illustré figure 6.1.

Comme nous avons pris l'habitude de voir, le raisonnement économique est souvent un raisonnement *à la marge*. Un producteur se posera la question du revenu généré par une unité supplémentaire vendue. Ceci est illustré figure 6.2 : lorsque la production passe de q à

FIGURE 6.1 – Courbe de demande et revenu du monopole



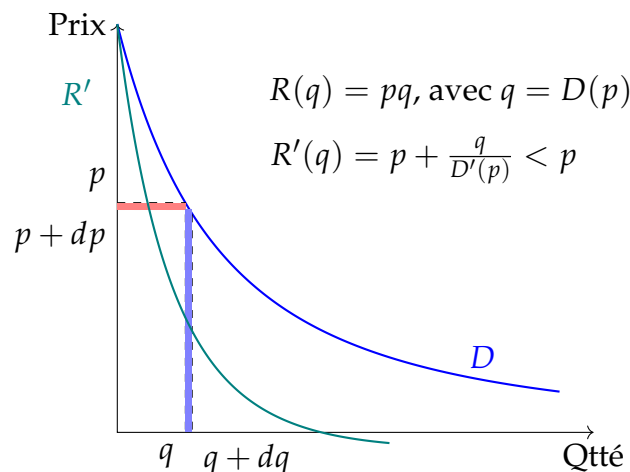
$q + dq$, le prix passe de p à $p + dp$ (avec $dp < 0$, car il y a une baisse de prix consécutive à l'augmentation des quantités vendues).

La différence de revenu lorsque l'on augmente la production de dq est la somme de deux termes : le terme positif $dq \cdot p$, aire en bleu sur le graphe, correspond à une augmentation des ventes d'une quantité dq à un prix p . Le terme négatif $dp \cdot q$, aire en rouge sur le graphe, représente la baisse de prix dp que le producteur doit concéder sur l'ensemble de la quantité produite q pour réaliser cette augmentation de la demande. Les quantités infinitésimales dq et dp sont liées par $\frac{dq}{dp} = D'(p)$ avec $D(p) = q$, soit $\frac{dp}{dq} = \frac{1}{D'(p)}$. En notant $R(q) = pq$ le revenu correspondant à une quantité q produite, le revenu marginal s'exprime comme :

$$R'(q) = p + \frac{q}{D'(p)}, \text{ avec } D(p) = q.$$

Étant donné que le second terme est négatif pour $q > 0$, on a $R'(q) < p$ si $q > 0$. La courbe des revenus marginaux se situe donc sous la courbe de demande D .

FIGURE 6.2 – Revenu marginal du monopole



Nous sommes maintenant en mesure de déterminer le choix optimal du monopole. Étant donné que le monopole maximise son bénéfice exprimé comme $R(q) - C(q)$, où C est la fonction de coûts de production, une solution q^m du problème de maximisation telle que $q^m > 0$ doit satisfaire $R'(q) = c(q)$, où c est la fonction de coûts marginaux. Ceci est intuitif : si $R' > c$, la firme a intérêt à augmenter sa production car une unité supplémentaire produite lui rapporte plus qu'elle ne coûte, et dans le cas contraire $R' < c$ elle a intérêt à baisser sa production. Graphiquement, on obtient donc la quantité optimale q^m du monopole comme l'intersection de la courbe de revenus marginaux R' et de celle de coûts marginaux c . Le prix de monopole p^m est alors obtenu par la courbe de demande : $D(p^m) = q^m$.

Est-il possible que la quantité optimale soit nulle ? C'est le cas si (et seulement si) aucune quantité produite par le monopole ne peut assurer de profit positif. Cela se passe si les coûts de production sont trop élevés, ou si la demande est trop faible. Il faut donc, en tout état de cause, vérifier que pour la quantité q^* qui égalise R' et c , on a bien $p^m q^m - C(q^m) \geq 0$, c'est à dire que le prix p^m se situe bien au dessus du coût moyen $C(q^m)/q^m$. C'est la condition d'ouverture de la firme.

6.2.1 Comparaison avec l'équilibre de marché

Comparons maintenant la solution de monopole (q^m, p^m) avec la solution de marché (q^*, p^*) dans laquelle la firme serait preneuse de prix. On se souvient que la quantité d'équilibre de marché q^* est obtenue à une intersection de la courbe de coûts marginaux c avec celle de la demande D . On sait de plus qu'au point considéré (q^*, p^*) , la fonction de coûts marginaux est nécessairement croissante et que la fonction de demande est décroissante. Par conséquent, et comme représenté figure 6.3, pour les quantités $q > q^*$, la fonction de coûts marginaux se situe au dessus de la fonction de demande D . Cela implique donc que :

- la quantité de monopole est inférieure à celle de l'équilibre de marché ;
- le prix de monopole est supérieur à celui de l'équilibre de marché.

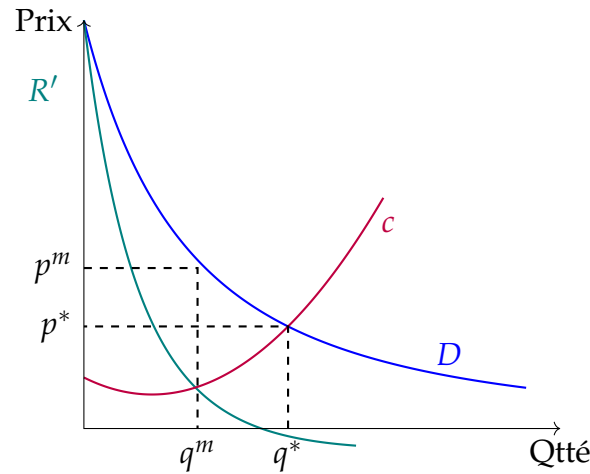
Le producteur en situation de monopole réduit sa production pour augmenter les prix, et par conséquent son profit. Pour le consommateur, le monopole est désavantageux car il conduit à des prix trop élevés.

6.2.2 Majoration des prix par le monopole

Lorsqu'une firme n'influe pas sur les prix et prend les prix de marchés comme donnés, nous savons que la quantité produite est telle le coût marginal et le prix s'égalisent : $c(q) = p$. Comparons maintenant ce prix au prix d'un monopole, cas dans lequel le coût marginal égalise le revenu marginal : $R'(q^m) = c(q^m)$. Étant donné que $R'(q^m) = p^m + \frac{D(p^m)}{D'(p^m)}$, on a

$$c(q^m) = p^m \left(1 + \frac{D(p^m)}{p^m D'(p^m)} \right) = p^m \left(1 - \frac{1}{e(p^m)} \right),$$

FIGURE 6.3 – Comparaison entre offre et prix de monopole et offre et prix de marché. Le prix de monopole est donné par l'intersection des courbes R' et c , le prix de marché par l'intersection des courbes D et c .



où $e(p^m) = -p^m D'(p^m)/D(p^m)$ représente l'élasticité-prix de la demande en p^m . Ceci nous permet d'exprimer le prix de monopole comme :

$$p^m = \frac{e(p^m)}{e(p^m) - 1} c(q^m).$$

Cette dernière expression est instructive. Tout d'abord, elle nous enseigne que, dès lors que $e(p) \in (1, \infty)$, le monopole fixe un prix supérieur à son coût marginal de production. Plus l'élasticité de la demande est faible, plus le pouvoir de marché est fort et plus le prix est élevé relativement au coût marginal. Ce résultat est intuitif, car pour une élasticité faible, le consommateur est prêt à accepter des prix plus élevés sans fortement baisser sa demande, situation dont le monopole peut aisément profiter. Ainsi, une compagnie pharmaceutique qui a l'exclusivité sur un traitement d'une maladie mortelle pourra-t-elle en tirer un revenu largement supérieur à son coût de production.

Ensuite, elle nous montre qu'un monopole ne choisit jamais un point sur la courbe de demande tel que son élasticité (en valeur absolue) est inférieure à 1. En effet, face à une élasticité si faible, le monopole a intérêt à réduire sa quantité afin d'augmenter son prix. Si l'élasticité de la demande est uniformément inférieure à 1, alors en théorie le monopole choisit de vendre une quantité infiniment faible à un prix infiniment élevé. En pratique, le budget des consommateurs n'étant pas infini, il existe toujours un prix à partir duquel l'élasticité devient supérieure à 1.

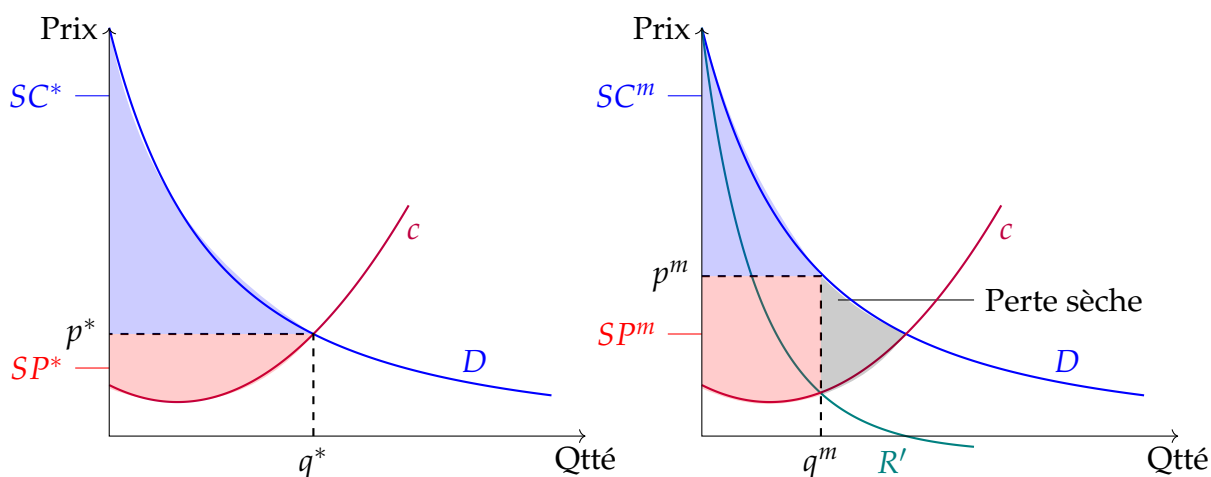
6.2.3 Effet sur le surplus

Voyons maintenant les effets du monopole en termes de surplus, en comparaison avec l'équilibre de marché. Le monopole choisit sa production afin de maximiser son propre surplus. Par conséquent, le surplus du producteur sous prix de monopole SP^m est supérieur à celui de l'équilibre de marché SP^* . Les consommateurs font, eux face à des prix plus éle-

vés en situation monopolistique qu'en équilibre de marché : le surplus des consommateurs sous monopole SC^m est donc inférieur à celui de l'équilibre de marché SC^* . Qu'en est-il du surplus total? Nous savons que le prix qui maximise le surplus total est celui de l'équilibre de marché, par conséquent le surplus total $ST^m = SP^m + SC^m$ sous monopole est inférieur au surplus total $ST^* = SP^* + SC^*$ sous équilibre de marché. La différence $ST^* - ST^m$ est appelée la *perte sèche*. Les différents surplus et la perte sèche sont représentés figure 6.4.

Pour les unités entre q^m et q^* , la propension marginale à payer des consommateurs, donnée par la courbe de demande D , est supérieure au coût marginal de production c . Toutes ces unités pourraient donc donner lieu à des échanges bénéfiques entre les consommateurs et les producteurs. Cependant, ces échanges ne se réalisent pas car le monopole refuse de vendre des biens à un prix inférieur à p^m . La présence de la perte sèche caractérise donc une situation Pareto-inefficace ; le pouvoir de marché en situation monopolistique est source d'inefficacité.

FIGURE 6.4 – Comparaison des surplus entre monopole (à droite) et marché (à gauche)



6.3 Analyse du duopole

Ainsi, un marché dans une situation de monopole n'est pas Pareto-efficace : l'hypothèse d'agents preneurs de prix était effectivement nécessaire pour obtenir le résultat d'efficience des marchés décrit au début de ce cours. Néanmoins, l'analyse du monopole reste celle d'un cas extrême, diamétralement opposé à celui du marché parfaitement compétitif que nous avons étudié plus tôt, celui-ci représentant bien des situations dans lesquelles il y a un très grand nombre d'entreprises, toutes trop petites pour avoir un impact sur les prix.

Le point commun entre ces deux situations extrêmes est que ce sont des situations sans interactions entre firmes : dans les deux cas, les entreprises prennent leurs décisions de façon indépendante, sans prendre en compte les choix des autres entreprises.

Que dire d'une situation intermédiaire, dans laquelle les entreprises ont un pouvoir de marché mais ne sont pas en situation de monopole ?

6.3.1 Le modèle de duopole de Cournot

La situation intermédiaire la plus simple que l'on puisse considérer est celle d'un duopole, dans laquelle deux entreprises dominent un marché. Nous avons déjà évoqué le duopole formé par Boeing et Airbus dans le domaine de l'aviation, mais d'autres marchés prennent la forme d'un duopole : Visa et Mastercard, iOS et Android, macOS et Windows sont d'autres bons exemples.

C'est le mathématicien et philosophe français Augustin Cournot (1801-1877) qui a en premier développé un modèle du duopole en 1838. Il considère deux firmes, que nous noterons 1 et 2, produisant des biens identiques. Les coûts de production de la firme i sont donnés par la fonction C_i , et ses coûts marginaux par la fonction c_i . Les deux firmes font face à une fonction de demande D .

De même que nous avons étudié la quantité optimale produite q^* par un monopole, nous allons rechercher des quantités produites q_1^* et q_2^* pour les firmes 1 et 2 qui soient *mutuellement optimales*. Par mutuellement optimales, on entend qu'à production donnée q_j^* de l'autre firme j , la quantité produite q_i^* est optimale pour la firme i . Le problème peut sembler circulaire à première approche, car la quantité optimale de la firme i dépend de celle de la firme j , qui dépend à son tour de celle fixée par la firme i . On parle dans ce cas d'interaction stratégique.

Pour résoudre ce problème, nous allons commencer par étudier comment chaque firme réagit aux décisions de l'autre firme.

6.3.2 Fonction de meilleure réponse

Pour fixer les idées, on se demande quelle est la quantité optimale produite $Q_2(q_1)$ par la firme 2 lorsque la firme 1 produit une quantité fixée q_1 .

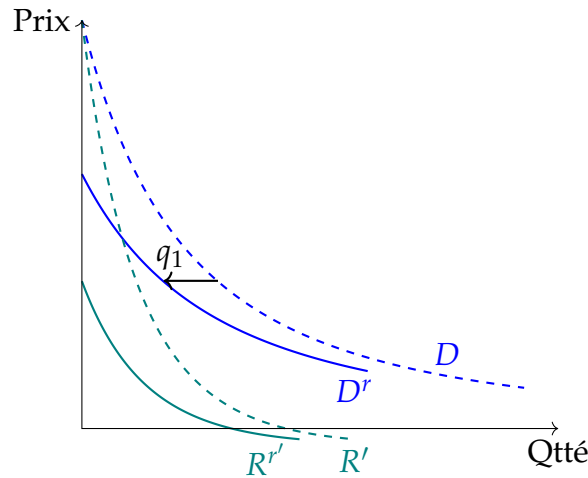
Si la firme 1 produit une quantité q_1 , la firme 2 choisit sa quantité mise en vente q_2 , et le prix p auquel ces unités sont vendues égalise offre et demande. Par conséquent p et q_2 sont liés par la relation $D(p) = q_1 + q_2$. Vu autrement, la firme 2 choisit un prix p , mais au lieu de vendre la quantité $D(p)$ qui serait celle d'un monopole, elle ne vend que la quantité $D(p) - q_1$ à ce prix. Cela signifie que la firme 2 fait face à une *demande résiduelle* D^r donnée par $D^r(p) = D(p) - q_1$. La courbe de la demande résiduelle est translatée de la courbe de demande d'une quantité q_1 vers la gauche.

Tout se passe donc, du point de vue de la firme 2, comme si elle était en situation de monopole face à la demande résiduelle D^r . Le revenu marginal résiduel $R^{r'}$ est obtenu à partir de la demande résiduelle D^r de la même manière que le revenu marginal R^r est obtenu à partir de la demande D .

De même que dans l'analyse de monopole, la firme 2 fixe une production qui égalise revenus marginaux et coûts marginaux, à ceci près que les revenus marginaux sont maintenant pris comme étant les revenus marginaux résiduels $R^{r'}$. C'est la situation de la figure 6.6.

Ainsi, pour une quantité q_1 fixée, nous avons trouvé la quantité q_2 que la firme 2 choisit de produire. Par la même occasion, nous obtenons le prix de duopole correspondant p dans la figure 6.6.

FIGURE 6.5 – Demande D , demande résiduelle D' , demande résiduelle D' et revenu marginal résiduel R'' . La demande résiduelle correspond à la demande D translatée vers la gauche de la quantité produite par la première firme, q_1 .



Pour compléter tout à fait l'analyse de la meilleure réponse $Q_2(q_1)$ de la firme 2 à la firme 1, il faut prendre en compte la condition d'ouverture de la firme 2. Supposons qu'à une production q_1 de la firme 1, il existe une production $q_2 > 0$ qui assure un bénéfice positif, et donc que la condition d'ouverture de la firme 2 soit satisfaite. Fixons une autre production $q'_1 < q_1$ de la firme 1. Les prix résultant de la production d'une quantité q_2 de la firme 2 sont supérieurs lorsque la firme 1 produit q'_1 que lorsqu'elle produit q_1 (car le prix p qui satisfait $D(p) = q_1 + q_2$ est décroissant en q_1). Par conséquent les bénéfices de la firme 2 en produisant q_2 face à q'_1 sont supérieurs à ce qu'ils sont en produisant q_2 face à q_1 , et donc la condition d'ouverture de la firme 2 est satisfaite face à q'_1 . Il y a donc deux cas possibles, soit la condition d'ouverture de la firme 2 est satisfaite pour tout $q_1 > 0$, soit il existe une valeur q_1^e en deçà de laquelle la firme 2 entre sur le marché ($Q_2(q_1) > 0$), et au delà de laquelle elle n'entre pas ($Q_2(q_1) = 0$).²

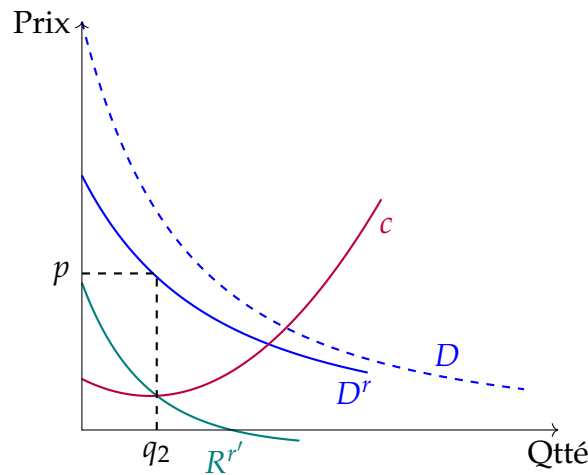
6.3.3 Équilibre de Cournot-Nash

Nous savons désormais comment la firme 2 réagit au choix de production de la firme 1. De manière symétrique, on peut trouver la fonction de meilleure réponse de la firme 1 face au choix de production de la firme 2. Ces fonctions de meilleure réponse sont l'élément fondamental permettant de décrire les interactions stratégiques entre les deux firmes, et à partir duquel nous pouvons déduire l'équilibre résultant de ces interactions.

Représentation graphique des fonctions de meilleure réponse. Pour représenter graphiquement l'interaction entre les deux firmes, on peut tracer les fonctions de meilleure réponse dans le plan (q_1, q_2) . Un exemple est tracé sur la figure 6.7 : la courbe bleue représente la meilleure réponse $Q_2(q_1)$ en fonction de la quantité q_1 produite par la firme 1, et la courbe

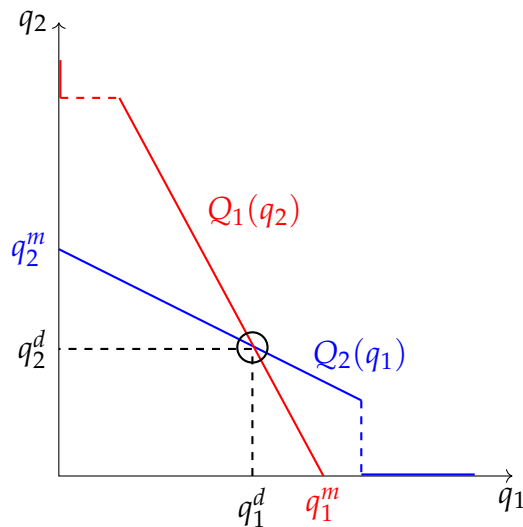
2. Le niveau précis q_1^e dépend de la structure de coût de l'entreprise. Plus précisément, plus les coûts fixes sont élevés, plus q_1^e est faible.

FIGURE 6.6 – La meilleure réponse $Q_2(q_1)$ est obtenue en égalisant coûts marginaux et revenus résiduels marginaux



rouge la meilleure réponse $Q_1(q_2)$ de la firme 1 en fonction de la quantité q_2 produite par la firme 2.

FIGURE 6.7 – Équilibre de Cournot-Nash du duopole



Comme nous pouvons le voir, dans cet exemple, la condition d'ouverture de la firme n'est pas toujours respectée : lorsque la firme 1 produit une quantité trop importante, la meilleure réponse de la firme 2 vaut $Q_2(q_1) = 0$ (et symétriquement, la firme 1 n'ouvre pas lorsque la production de la firme 2 est très importante).

Existence de l'équilibre. Les deux courbes se croisent en les quantités q_1^d, q_2^d . En ce point, la quantité q_2^d produite par la firme 2 est une meilleure réponse à celle q_1^d produite par la firme 1, et réciproquement. Le choix de chaque firme est optimal à choix donné de l'autre. Ces quantités sont appelées *équilibre de Cournot* du duopole. Elles correspondent à une situation stable dans laquelle, à choix fixé du concurrent, chaque firme maximise ses profits et

ne désire pas changer son choix de production. Cette notion d'équilibre a été plus tard généralisée par le mathématicien et économiste John Nash, raison pour laquelle on parle aussi d'équilibre de Cournot-Nash.

Convergence vers l'équilibre. Il est intéressant de noter l'existence d'un équilibre stratégique où chaque firme n'a aucune raison de modifier ses décisions, mais que se passe-t-il en dehors de cet équilibre ?

Imaginons par exemple que les firmes produisent respectivement les quantités q_1^0 et q_2^0 , distinctes des valeurs d'équilibre q_1^d et q_2^d , représentées sur la figure 6.8 en haut à gauche. Dans une telle situation, la firme 2 a intérêt à changer sa production, à la réduire plus précisément (on a $Q_2(q_1^0) < q_2^0$). Imaginons que la firme 2 ajuste effectivement son niveau de production à la baisse, au niveau $q_2' = Q_2(q_1^0)$ (en haut à droite sur la figure). Cette nouvelle situation n'est pas optimale pour la firme 1 : celle-ci ne produit pas assez face à q_2 (on a $Q_1(q_2') > q_1^0$). Si la firme 1 augmente sa production jusqu'à $q_1' = Q_1(q_2')$, on a la situation en bas à gauche. Cette nouvelle situation n'est toujours pas un équilibre car la firme 2 a encore intérêt à réduire sa production : l'issue de ce nouvel ajustement est représenté en bas à gauche de la figure 6.8.

On s'aperçoit que les quantités produites par les deux firmes se rapprochent petit à petit de l'équilibre de duopole (q_1^d, q_2^d). Lorsque les firmes prennent leurs décisions de manière séquentielle, il y a bien une convergence vers cet équilibre de Cournot-Nash qu'on peut donc qualifier de stable.

Le cas du monopole naturel. Il faut noter que les quantités de duopole q_1^d, q_2^d d'équilibre de Cournot-Nash de la figure 6.7 sont toutes deux strictement positives. C'est le cas ici car lorsqu'une firme i produit sa quantité de monopole q_i^m , la meilleure réponse de l'autre firme, j , est de produire une quantité $Q_j(q_i^m) > 0$.

La figure 6.9 illustre l'autre cas, celui dans lequel $Q_2(q_1^m) = 0$. La figure de gauche correspond à une situation symétrique où les deux firmes ont des fonctions de coût similaires, et celle de droite à une situation dans laquelle les coûts de la firme 1 sont inférieurs. Dans ces deux cas, les deux courbes de meilleures réponses ne se croisent ici pas pour deux quantités strictement positives, étant donné dans le cas symétrique que la partie en pointillés correspond à une discontinuité de la meilleure réponse. Dans le cas symétrique, nous avons deux équilibres, $(q_1^m, 0)$ et $(0, q_2^m)$ tandis que dans le cas asymétrique nous n'avons qu'un seul équilibre $(q_1^m, 0)$. Dans les deux cas nous avons soit un, soit deux équilibres de Cournot-Nash dans lesquels un des producteurs produit la quantité de monopole tandis que l'autre n'entre pas sur le marché. Il s'agit de situations dites de *monopole naturel* dans laquelle des coûts fixes importants rendent la survie simultanée de deux producteurs impossible. On a donc l'établissement d'un monopole dû à la structure de marché (coûts de production et demande pour le bien). Cette situation s'applique par exemple au réseau ferroviaire ou au réseau électrique, pour lesquels les coûts fixes colossaux ne permettent pas à deux entreprises de co-exister.

FIGURE 6.8 – Convergence vers l'équilibre de Cournot-Nash. Dans la situation initiale, en haut à gauche, les deux firmes ne sont pas à l'équilibre. Lorsqu'elles ajustent successivement leurs productions suivant leur fonction de meilleure réponse, on observe une convergence des quantités produites vers l'équilibre de duopole.

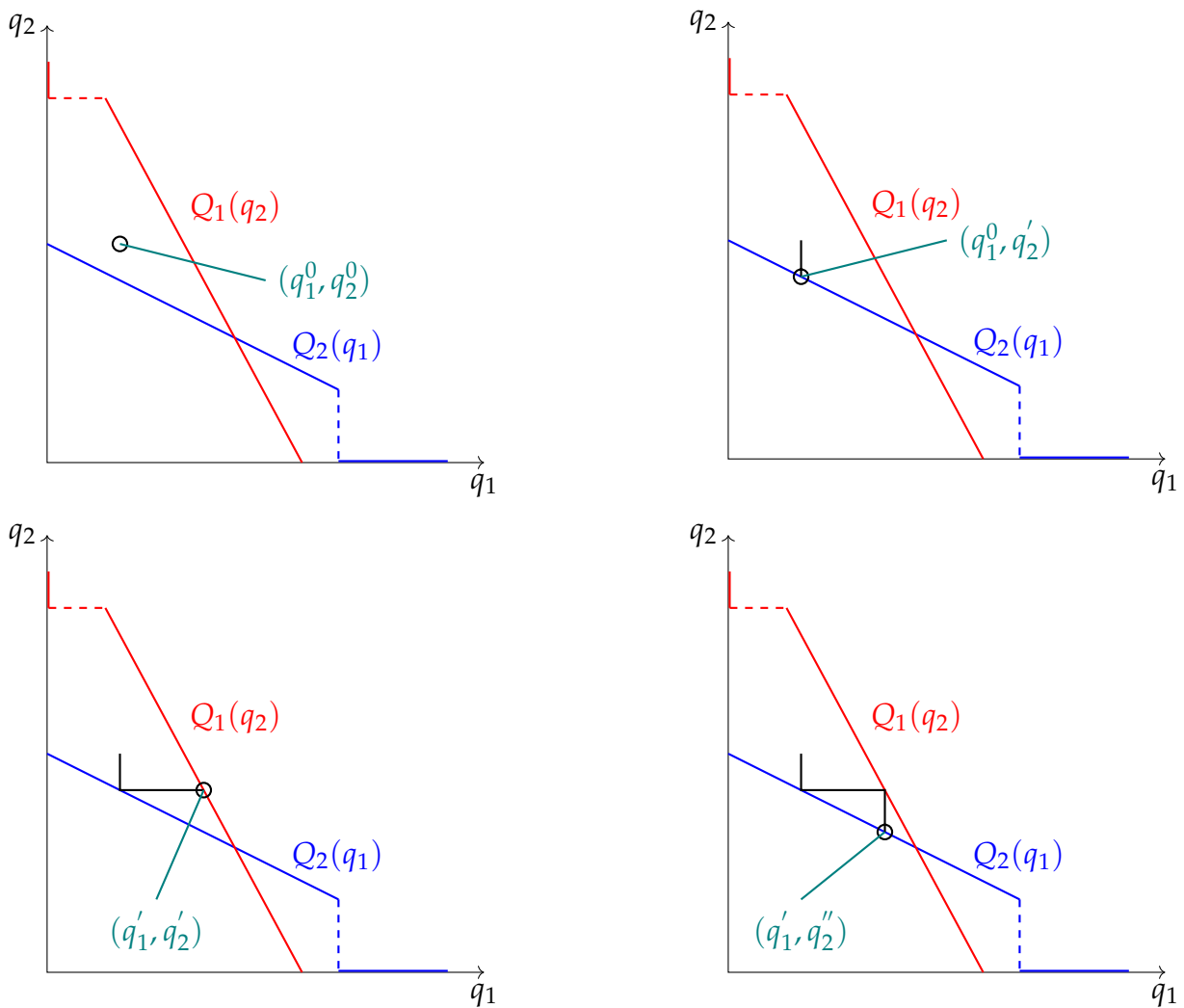
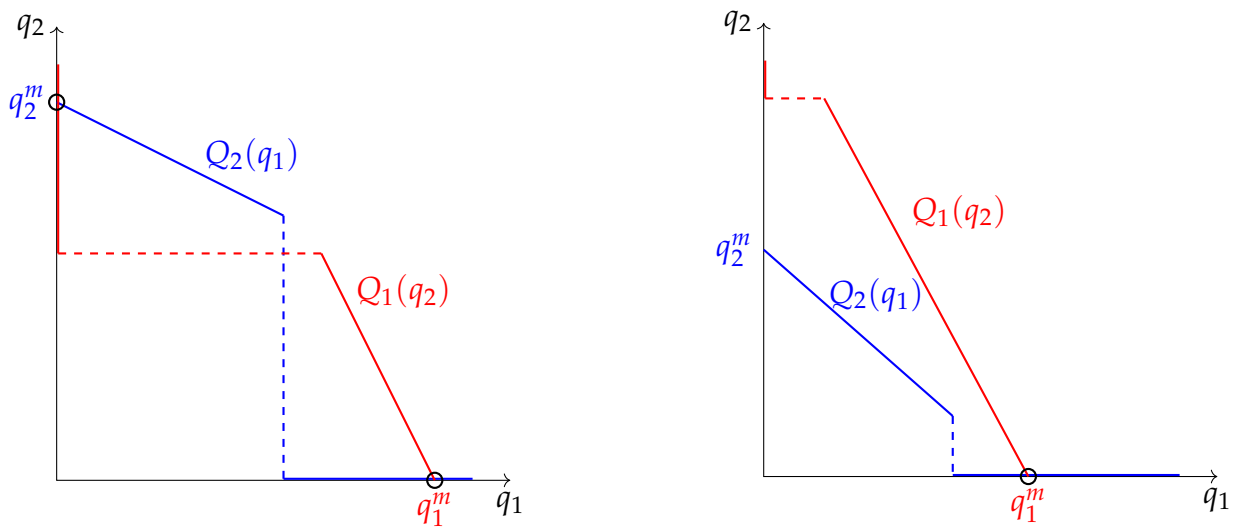


FIGURE 6.9 – Le duopole : cas du monopole naturel, avec coûts symétriques (figure de gauche) et coûts asymétriques (figure de droite). Les cercles indiquent les équilibres de Cournot-Nash.

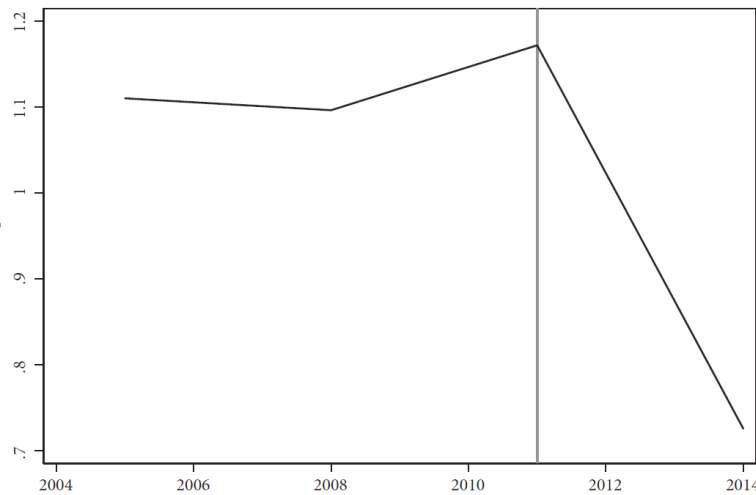


6.4 Augmentation du nombre de firmes

À l'équilibre de Cournot-Nash, la présence d'une seconde firme augmente la quantité totale produite sur le marché : l'ajout d'une nouvelle firme réduit donc la sous-production qui caractérise le monopole, et rapproche les prix de l'équilibre de marché. Que se passe-t-il lorsque le nombre de firmes présentes sur le marché augmente encore ? De manière empirique, on observe que lorsque le nombre de firmes présentes sur un marché augmente, les prix ont tendance à y diminuer.

Ainsi lorsque Free acquit une licence 4G en 2011, il rejoignit Orange, SFR et Bouygues parmi les principaux opérateurs téléphoniques en France. Cet événement causa une chute importante des prix de la téléphonie en France (voir figure 6.10)

FIGURE 6.10 – Rapport du prix de la téléphonie en France par rapport aux États-Unis. La ligne verticale correspond à l'entrée de Free sur le marché de la 4G. Source : Philippon (2019).



Lorsque le nombre de firmes augmente, le pouvoir de marché exercé par chacune, mesuré par sa capacité à influencer les prix, se trouve réduit. Il est alors intuitif que lorsque le nombre de firmes devient suffisamment important, l'équilibre d'oligopole finit par se rapprocher de l'équilibre de marché dans lequel chaque firme est preneuse de prix. Nous allons ici formaliser cette intuition.

Considérons un secteur oligopolistique avec des firmes $1, \dots, n$. Chaque firme i a une fonction de coûts C_i et une fonction de coûts marginaux c_i . La fonction de demande est D . Dans le modèle de compétition de Cournot correspondant, chaque firme i fixe une quantité produite q_i , la quantité totale est alors $q = \sum_i q_i$, le prix p est déterminé par $D(p) = q$. Le revenu de la firme i est $R_i = q_i p$, et son bénéfice est $R_i(q_i) - C_i(q_i)$.

On cherche à caractériser un équilibre de Cournot-Nash dans lequel chaque firme i produit une quantité positive q_i^0 qui maximise $R_i(q_i) - C_i(q_i)$ à production des autres firmes donnée. Les revenus résiduels marginaux de la firme i s'écrivent :

$$R'_i(q_i) = p + q_i \frac{dp}{dq_i}.$$

En utilisant l'expression de l'élasticité-prix de la demande :

$$e = -\frac{dq_i}{dp} \frac{p}{q}$$

On obtient :

$$R'_i(q_i) = p \left(1 - \frac{1}{e} \frac{q_i}{q} \right).$$

Lorsque i maximise son profit à production des autres firmes données, les revenus résiduels marginaux égalisent les coûts marginaux, et par conséquent :

$$c_i(q_i^o) = p \left(1 - \frac{1}{e} \frac{q_i^o}{q^o} \right),$$

Avec $q^o = \sum_i q_i^o$ la quantité totale produite par l'ensemble des firmes.

On s'intéresse particulièrement à la situation dans laquelle chaque producteur est petit face au marché, $q_i^o \ll q^o$. Dans ce cas les équations précédentes peuvent être approximées par :

$$c_i(q_i^o) = p,$$

lesquelles caractérisent l'équilibre de marché.

Par conséquent, lorsque toutes les firmes sont suffisamment petites par rapport à la taille globale du marché, les équilibres de Cournot-Nash se rapprochent de l'équilibre de marché, l'équilibre de marché est alors une bonne approximation de l'équilibre de Cournot-Nash.

On remarque aussi que la convergence de l'équilibre de Cournot-Nash vers l'équilibre de marché dépend de la magnitude du terme $\frac{1}{e} \frac{q_i^o}{q^o}$: plus une firme possède une part de marché importante, plus $\frac{q_i^o}{q^o}$ est élevé, et plus cette firme exerce de pouvoir de marché. Un marché comprenant un grand nombre de producteurs mais très concentré dans le sens où un faible nombre de producteurs possède une grande part de marché pourra donc être plus proche d'un équilibre d'oligopole à faible nombre de producteurs qu'à l'équilibre de marché. Aussi, plus l'élasticité-prix e est petite, moins la demande réagit à des augmentations de prix, et plus les firmes sont tentées d'exercer du pouvoir de marché pour augmenter les prix. Ces deux facteurs : marchés concentrés et élasticité-prix faible, écartent l'équilibre de l'oligopole de l'équilibre de marché.

6.5 Compétition en prix : duopole de Bertrand

Dans sa revue des travaux d'Augustin Cournot publiée en 1883, Joseph Bertrand (major de l'X en 1839) propose de reprendre l'hypothèse de compétition en quantités (chaque firme décide de la quantité produite) et d'y substituer celle de compétition en prix. Dans le modèle de *duopole de Bertrand*, chaque firme décide d'un prix de vente, et les consommateurs achètent ensuite au meilleur prix proposé.

Pour fixer les idées, supposons deux firmes produisant des biens identiques au même coût. Le coût marginal de chaque firme est constant égal à c . Par conséquent, pour chacune

des firmes, le coût de production pour q unités est cq . La demande du marché est donnée par la fonction de demande $D(p)$ continue prenant des valeurs positives.

Dans la compétition en prix, chaque firme $i \in \{1, 2\}$ annonce un prix de vente p_i . Ensuite, la demande se répartit entre les deux firmes selon les prix annoncés, si une firme est moins chère que l'autre elle reçoit toute la demande, et en cas d'égalité la demande se répartit également entre les deux firmes. En notant $Q_i(p_1, p_2)$ la quantité vendue par la firme i face à son concurrent j , on a donc :

$$\begin{cases} Q_i(p_i, p_j) = D(p_i) & \text{si } p_i < p_j \\ Q_i(p_i, p_j) = D(p_i)/2 & \text{si } p_i = p_j \\ Q_i(p_i, p_j) = 0 & \text{si } p_i > p_j \end{cases}$$

La firme i vend $Q_i(p_i, p_j)$ du bien, chaque unité est vendue au prix p et coûte q_i à produire, son bénéfice net est donc $Q_i(p_i, p_j)(p - c)$. En raisonnant par élimination, nous allons montrer qu'à l'unique équilibre possible, $p_1 = p_2 = c$.

Tout d'abord, nous ne pouvons avoir $\min\{p_1, p_2\} < c$, car dans ce cas la ou les firmes proposant le prix le plus bas feraient une perte, leur prix de vente étant inférieur au coût de production.

Peut-on avoir $\min\{p_1, p_2\} > c$? Supposons que nous soyons dans ce cas, et que la firme i ait le prix le plus élevé, éventuellement à égalité avec l'autre firme j . Son bénéfice est alors soit 0 (si $p_i > p_j$), soit $(p_i - c)D(p_i)/2$ en cas d'égalité. Alors la firme i aurait intérêt à proposer un prix très légèrement inférieur à p_j . Elle recevrait dans ce cas l'intégralité de la demande pour un prix de vente arbitrairement proche de p_j , et ferait donc un bénéfice proche de $(p_j - c)D(p_j)$, ce qui est strictement plus que $(p_j - c)D(p_j)/2$ et que 0.

On a donc forcément $\min\{p_1, p_2\} = c$. Supposons que la firme i annonce le prix $p_i = c$. Est-il alors possible que l'autre firme annonce un prix $p_j > c$? Non, car dans ce cas la firme i aurait la possibilité d'augmenter ses prix très légèrement tout en restant au dessous du prix p_j , et de faire un bénéfice positif au lieu de nul.

La seule possibilité d'équilibre est donc $p_1 = p_2 = c$, et le profit des deux firmes est nul. En partant de cette situation, si une firme augmente son prix, elle continue à avoir un bénéfice nul (car la demande reçue est alors nulle). Si elle baisse son prix, alors elle vend à perte car son prix de vente est plus bas que son coût de production.

En suivant l'analyse du modèle de Bertrand, on voit que l'équilibre est unique et satisfait $p_1 = p_2 = c$. Les firmes font donc un profit nul, et tout le surplus est extrait par le consommateur. Avec tout simplement deux firmes, nous obtenons le même résultat qu'avec un grand nombre de firmes dans le modèle de Cournot lorsqu'il y a de plus libre entrée.

Les conclusions du modèle de Bertrand sont frappantes et contrastent fortement avec celles du modèle de Cournot. Alors, en pratique, qui a raison de Cournot ou Bertrand? Il est naturel de penser que les firmes se font compétition en prix plutôt qu'en quantités, après tout, chaque producteur est libre de décider ses prix de vente et d'ajuster ensuite sa production selon les ventes. Mais il faut reconnaître que les deux modèles sont des simplifications qui ne prennent pas en compte certains aspects importants de la compétition tels que la

différentiation des prix et les choix de capacité de production.

En effet, il est très rare d'observer différents producteurs offrant exactement les mêmes produits. Chaque producteur a intérêt à se différencier des autres afin justement de pouvoir mieux satisfaire une part des consommateurs que les autres et d'attirer des parts de marché. Dans la téléphonie mobile par exemple, la qualité et la couverture du réseau, le service client, les offres liées avec des *Box* sont autant de points de différenciation entre opérateurs. Même pour un produit a priori très homogène comme l'eau minérale, l'emballage, le marketing, les réseaux de distribution sont autant de facteurs différenciants permettant d'expliquer des différences de prix importantes entre marques.

Les capacités de production sont des choix stratégiques des producteurs, effectués en amont des choix de prix. Par exemple, un producteur automobile devra mettre en place des chaînes d'assemblage avant la mise sur le marché d'une gamme auto. Une fois ces choix effectués, chaque producteur est contraint par une capacité maximale. Dans un article remarquable publié 100 ans après les travaux de Bertrand, Kreps and Scheinkman (1983) étudient un modèle dans lequel des choix de capacités de production précèdent une compétition en prix. Leur analyse montre que les équilibres de ce modèle sont les mêmes que ceux d'une compétition à la Cournot! La boucle est alors bouclée, et on comprend désormais mieux pourquoi des situations d'oligopoles conduisent à des profits (comme chez Cournot) même lorsque les firmes se font ensuite concurrence en prix (comme chez Bertrand).

6.6 La régulation du pouvoir de marché

L'organisation industrielle est la branche des sciences économiques qui s'intéresse à la concurrence imparfaite et au pouvoir de marché. Sa principale application est la régulation de la concurrence.

6.6.1 Régulation de la concurrence

Nous avons vu que la présence de firmes avec un fort pouvoir de marché avait pour effet de transférer une partie du surplus des consommateurs vers les producteurs, et engendrait une perte sèche qui caractérise une inefficacité économique.

Le premier effet explique pourquoi certaines firmes tentent d'étendre leur pouvoir de marché. Elles peuvent le faire en fusionnant avec d'autres firmes, en rachetant des potentiels concurrents avant qu'ils ne deviennent une réelle menace (ainsi, Facebook a racheté en 2014 l'entreprise WhatsApp pour un montant de 19 milliards de dollars, alors que l'entreprise comptait alors une cinquantaine de salariés), ou en vendant à des prix très bas, voir même temporairement à perte, de manière à pousser ses concurrents vers la faillite.

Le second effet est l'une des raisons pour lesquelles les États mettent en place des régulations de la concurrence pour contrer de telles actions. Ces régulations permettent de restaurer une partie de l'efficacité économique, de limiter le pouvoir économique des grandes firmes, et de protéger les consommateurs. Dans la mesure où les individus qui perçoivent les

surplus des producteurs (entre autres, les actionnaires) sont plus riches que la moyenne, ces régulations ont également des conséquences importantes pour la distribution des richesses.

La régulation de la concurrence peut prendre différentes formes. En 1911, la cour suprême américaine jugea que l'entreprise Standard Oil était un monopole illégal et ordonna sa dislocation en une série de plus petites entreprises. Standard Oil, co-fondée par John Rockefeller, avait acquis la grande majorité des activités pétrolières aux États-Unis. Cette position monopolistique a bien profité à Rockefeller, dont la richesse correspondait en 1913 à environ 2% du PIB américain! En général, la régulation de la concurrence prend la forme d'un contrôle des fusions et acquisitions. Lorsque deux grandes entreprises projettent de fusionner, des autorités de contrôle doivent au préalable estimer les effets probables de la fusion et peuvent décider de l'interdire³. Ainsi, en 2019, la commission européenne a décidé d'interdire la fusion entre les industriels ferroviaires Alstom et Siemens, jugeant que l'entité issue de cette fusion jouirait d'un trop important pouvoir de marché.

En Europe, la politique de la concurrence a été très largement influencée par les travaux de Jean Tirole sur l'organisation industrielle. Sa principale contribution, récompensée par le prix Nobel en 2014, est d'avoir systématiquement analysé en terme stratégique l'interaction entre les entreprises en s'appuyant sur la théorie des jeux qui sera l'objet du prochain chapitre.

6.6.2 Les incitations à l'innovation

Si on cherche à limiter le pouvoir de marché des entreprises, pourquoi alors crée-t-on parfois des monopoles de toutes pièces, lorsque l'on met en place un brevet par exemple? Lorsqu'un médicament est créé par une entreprise pharmaceutique et que celui-ci est protégé par un brevet, ce médicament peut être produit en quantité insuffisante et la résiliation du brevet augmenterait le surplus total créé car ce marché. Pourquoi alors mettre en place un brevet? Tout simplement pour que le marché existe en premier lieu! En l'absence de brevet ou de droits d'auteur, la production intellectuelle (une formule pharmaceutique, un livre ou une chanson) devient un bien public, et souffre alors du problème du passager clandestin. Un laboratoire pharmaceutique n'a pas intérêt à investir dans la recherche d'une nouvelle molécule s'il sait que les profits créés par celle-ci seront répartis entre l'ensemble des laboratoires. L'absence de propriété intellectuelle engendre une sous-production d'innovations technologiques et artistiques : les brevets permettent de restaurer des incitations à innover et de résoudre partiellement ce problème.

Cette idée n'est pas nouvelle : une des premières lois de *copyright*, le *Statute of Anne*, passé en 1710 en Angleterre, résume ce problème économique dans ses premières lignes :

Les imprimeurs, libraires et d'autres personnes ont récemment pris la liberté d'imprimer, copier ou publier [...] des livres et d'autres textes sans le consentement de [leurs] auteurs et propriétaires, à leur grand détriment et trop souvent

3. Le département de la justice américain considère qu'une fusion est potentiellement nuisible lorsqu'elle augmente l'indice de Herfindahl-Hirschmann défini plus haut de plus de 0.02 dans un marché à l'indice de Herfindahl-Hirschmann dépassant 0.25.

de leur ruine et celle de leurs familles : pour prévenir de telles pratiques dans le futur, et pour encourager les hommes d'esprit à composer et écrire des livres utiles...

Le *Statute of Anne* protégeait les auteurs pour une durée de quatorze ans, renouvelables. Aujourd'hui en France, les brevets accordent à leurs détenteurs un monopole pour une durée de vingt ans, à l'issue desquels l'invention entre dans le domaine public. Cette limitation de la durée des brevets permet d'éviter à long-terme la perte sèche créée par les monopoles qu'ils rendent possible : les brevets opèrent ainsi un arbitrage entre incitations à innover et limitation anti-monopolistique.

L'innovation étant le principal moteur de la croissance à long terme, il existe en sciences économiques une vaste littérature sur le sujet. En plus d'inciter à l'innovation les brevets facilitent la diffusion des connaissances. En revanche, cette propriété intellectuelle peut avoir des effets pervers lorsque des entreprises choisissent de déposer des brevets stratégiques, dans le seul but d'empêcher leurs concurrents d'explorer un sentier d'innovation prometteur.

6.6.3 Tendances récentes

L'étude du pouvoir de marché est aujourd'hui un champ de recherche très actif. De nombreux économistes documentent deux tendances des dernières décennies dans l'économie américaine : la première est une plus importante concentration des secteurs économiques, et l'augmentation du pouvoir de marché des firmes. La seconde est l'augmentation du nombre de zones dans lesquelles les travailleurs peuvent difficilement choisir leur employeur, donnant aux entreprises dans ces zones une position de monopsonne.

Pendant les trente dernières années, le nombre de fusions-acquisitions a augmenté aux États-Unis, menant à une hausse de la concentration (telle que mesurée par des indicateurs comme l'indice de Herfindahl-Hirschmann) dans différents secteurs de l'économie. En parallèle, les profits des firmes américaines ont augmenté, et celles-ci ont moins investi dans de nouveaux moyens de production. Enfin, le nombre de nouvelles firmes créées chaque année est en diminution. Ces indices suggèrent une baisse de la pression concurrentielle dans le pays (pour une description détaillée de ces tendances, voir Philippon (2019)).

Les chercheurs tentent de mieux comprendre le lien entre cette tendance de concentration des marchés aux États-Unis et d'autres tendances contemporaines dans ce pays : l'augmentation des inégalités de revenu et la réduction de la part des revenus totaux versés aux travailleurs sous la forme de salaires (contrairement à des dividendes et autres revenus du capital). Par ailleurs, le fort pouvoir de marché des géants du numérique soulève de nouvelles questions pour les régulateurs.

Il est important de noter que l'Europe suit pour l'instant une tendance différente : les marges de profit des entreprises n'y ont pas augmenté significativement, on n'observe pas de hausse similaire de la concentration des marchés, et la part des revenus du travail n'a pas baissé pendant la même période. Cette différence s'explique en partie par des différences institutionnelles, les régulateurs de la concurrence européens étant plus indépendants du

monde politique que les régulateurs américains. Très concrètement, cela se traduit notamment par des forfaits téléphoniques ou des billets d'avion nettement moins chers en Europe qu'aux États-Unis.

Bibliographie

Kreps, D. M. and Scheinkman, J. A. (1983), 'Quantity precommitment and bertrand competition yield cournot outcomes', *The Bell Journal of Economics* pp. 326–337.

Philippon, T. (2019), *The Great Reversal : How America Gave Up on Free Markets*, Harvard University Press.

Exercices du chapitre 6

Exercice 6.1. Oligopole On considère un secteur industriel composée de n firmes produisant des biens parfaitement substitués. Les coûts fixes sont supposés nuls et coûts par unité de production $c > 0$. La fonction de demande inverse (la réciproque de D) est donnée par $D^{-1}(q) = a - bq$ avec $a, b > 0$. On suppose $c < a$.

Notre première tâche est de calculer les quantités produites à l'équilibre d'oligopole.

1. Soit q_i la quantité produite par la firme i . En fonction de $q_{-i} = (q_j)_{j \neq i}$, quel est le profit $U_i(q_i, q_{-i})$, de la firme i si elle produit q_i unités du bien ?
2. Quelle est la quantité de meilleure réponse $Q_i(q_{-i})$ de la firme i ?
3. Quelles sont les quantités q_i^* produites à l'équilibre d'oligopole ?
4. Calculer la perte sèche en comparaison de l'équilibre de marché.
5. On suppose maintenant que les firmes 1 et 2 fusionnent pour donner la firme f (comme fusion). Comparer les profits de la firme f avec la somme des profits des firmes 1 et 2. Comparer les profits des firmes $i \neq 1, 2$ avant la fusion et après la fusion. Qui gagne à la fusion ?
6. À lire et commenter : article des Echos sur les possibles rapprochements dans les telecom français.

Exercice 6.2. Elasticité et pouvoir de marché

Selon vous, le pouvoir d'un marché d'un monopole est-il plus important lorsque la demande est fortement élastique, ou bien fortement inélastique ?

Pour répondre à la question de manière formelle, on s'intéresse à une fonction de demande D avec élasticité constante $e > 0$.

1. En fonction de la valeur de e , quelle est la fonction de demande $D_e(p)$ telle que l'élasticité-prix de la demande est constante et égale à e . On pourra normaliser les classes de fonctions D_e de telle manière que $D_e(1) = 1$ pour tout e .
2. On considère un monopole avec coût unitaire de production c . En fonction de c et de e , donner la quantité q^m de monopole et le prix p^m de monopole
3. Que se passe-t-il pour $e < 1$? En fait, ce résultat aurait pu être obtenu sans calculs, en utilisant les notions vues en cours. Voyez-vous comment ?
4. Pour $e > 1$, calculer la perte sèche due au monopole. Comment varie-t-elle avec e ?
5. On considère maintenant un oligopole, où n firmes ont le même prix unitaire de production c (on ignore les coûts fixes). Quel est cette fois le prix d'équilibre d'oligopole en

fonction de e, c, n ? On pourra se contenter de résoudre pour un équilibre symétrique dans lequel toutes les quantités produites sont égales.

6. A lire et commenter : article du Monde sur l'envolée des prix de l'insuline aux USA.

Exercice 6.3. Equilibre de Stackleberg

On considère un marché avec deux firmes. Les coûts de production sont de $c > 0$ par unité de bien, et la fonction de demande inverse est donnée par $D^{-1}(q) = a - bq$ avec $a, b > 0, a < c$.

1. On considère que la firme 1 entre en premier sur le marché. Cela signifie qu'elle décide de la quantité produite q_1 avant que la firme 2 décide de la quantité produite q_2 . Au moment où la firme 2 fait son choix, elle connaît la quantité q_1 décidée par la firme 1. En fonction de q_1 , quel est le meilleur choix $Q_2(q_1)$ de la firme 2?
2. En utilisant la réponse à la question précédente, donner l'expression du gain $U_1^s(q_1)$ lorsque la firme 1 produit q_1 et la firme 2 produit $Q_2(q_1)$.
3. Trouver la quantité q_1^s qui maximise $U_1^s(q_1)$, ainsi que la meilleure réponse $q_2^s = Q_2(q_1^s)$ de la firme 2.
4. Comparer q_1^s, q_2^s , ainsi que les profits obtenus lorsque la firme 1 entre avant la firme 2 aux quantités et profits de duopole (pour cette dernière partie on pourra s'appuyer sur les réponses à l'exercice 1). Commenter.

Exercice 6.4. L'objectif de cet exercice est de comprendre certains effets possibles d'une discrimination par les prix, c'est à dire d'une tarification différenciée selon les consommateurs.

Un monopole, TeleSport, vend possède les droits de retransmission par internet d'un événement sportif. Il existe deux catégories de consommateurs, les patients, qui aiment prévoir à l'avance et s'organiser, pour lesquels la fonction de demande inverse est $p = 50 - q$, et les impulsifs, qui décident au dernier moment et pour lesquels la fonction de demande inverse est $p = 100 - q$. Les coûts de retransmission de l'événement sportif pour TeleSport sont nuls.

1. Supposons que TeleSport puisse utiliser un tarif distinct pour chaque catégorie de consommateur, par exemple en faisant payer un tarif distinct selon que l'achat est effectué longtemps à l'avance ou au dernier moment.
 - a) Quel est le prix de monopole pour chacun des groupes?
 - b) Quel est le profit total de TeleSport?
 - c) Quel est le surplus des consommateurs?
2. Supposons qu'un régulateur puisse imposer une tarification à TeleSport en ayant pour objectif de maximiser le surplus total (TeleSport et consommateurs).
 - a) Quelle tarification maximise le surplus total?
 - b) Quel est alors le profit de TeleSport?
 - c) Quel est alors le surplus des consommateurs?
3. Supposons maintenant que le régulateur impose à TeleSport de ne pratiquer qu'un seul et même tarif pour les deux groupes de consommateurs.
 - a) Quelle est la fonction de demande totale pour les deux groupes?

- b) Quelle est la tarification de monopole de TeleSport ?
 - c) Quel est alors le profit de TeleSport ?
 - d) Quel est alors le surplus des consommateurs ?
4. On demande de comparer les 3 situations des questions 1/ 2/ 3/ et de commenter.
- a) Comparer les prix dans les 3 situations. Commenter.
 - b) Comparer les profits de TeleSport dans les 3 situations. Commenter. Aurait-on pu trouver cet ordre entre les 3 profits par un raisonnement direct ?
 - c) Comparer le surplus total dans les 3 situations. Commenter. Aurait-on pu trouver cet ordre entre les 3 surplus totaux par un raisonnement direct ?
 - d) Comparer le surplus des consommateurs dans les 3 situations. Commenter.
5. A la place du régulateur, seriez-vous favorable à imposer à TeleSport de pratiquer un tarif unique ?
6. A la lumière des questions précédentes, lorsqu'un monopole a la possibilité de connaître les préférences des consommateurs potentiels et de leur offrir des prix différenciés, quel est l'effet sur le surplus du monopole, sur le surplus des consommateurs et sur le surplus total en comparaison avec l'équilibre de marché ? Cette situation est-elle efficiente au sens de Pareto ?

Chapitre 7

Théorie des jeux

Au début de ce cours, nous nous sommes concentrés sur des situations dans lesquelles il n'y avait pas d'interactions stratégiques entre agents économiques : chaque firme et chaque consommateur prenait comme donné le prix de marché, et ce dernier n'était pas affecté par les choix des uns et des autres. Cette approximation permet de réfléchir à un certain nombre de problèmes économiques, mais elle devient insuffisante lorsque les choix de certains acteurs sont liés aux décisions des autres : nous avons vu au chapitre précédent que lorsqu'un petit nombre de firmes sont en compétition sur un marché, elles prennent leurs décisions de production les unes en fonction des autres. Pour modéliser ces interactions, nous avons introduit le concept de correspondance de meilleure réponse et nous sommes intéressés aux équilibres qui pouvaient résulter de ces interactions. Ces outils font partie de la théorie des jeux, qui est à la fois une branche des mathématiques appliquées et des sciences économiques, et dont les notions ont ensuite été étendues à de nombreuses autres disciplines comme la biologie ou les sciences politiques. Dans ce chapitre, nous allons généraliser les notions introduites lors de l'étude du duopole, notamment en formalisant l'idée d'équilibre de Nash, centrale en théorie des jeux. Nous avons déjà utilisé ces outils pour étudier les conséquences économiques du pouvoir de marché : nous allons également pouvoir utiliser la théorie des jeux pour comprendre d'autres défaillances de marché, notamment celles liées à des asymétries d'information.

7.1 Concept de jeu

Comme son nom l'indique, la théorie des jeux s'intéresse à des jeux ! En économie, un jeu est une situation dans laquelle des agents interagissent : chacun fait des choix, et les conséquences de ces choix dépendent des choix faits par les autres agents. La théorie des jeux étudie la façon dont les agents se comportent dans de telles situations d'interactions stratégiques. Ainsi, dans le duopole de Cournot, les agents étaient des entreprises qui choisissaient des quantités à produire. Les décisions de chaque entreprise avaient un impact direct sur les prix, affectant ainsi toutes les entreprises sur le marché : cette situation était ainsi le premier jeu rencontré dans ce cours.

7.1.1 Le dilemme du prisonnier

Pour illustrer les notions de base de la théorie des jeux, intéressons-nous à l'exemple célèbre du jeu du dilemme du prisonnier, né en 1950 à Princeton, lorsque le mathématicien Albert Tucker imagina une histoire similaire à celle-ci : deux complices, Joe et Averell, sont fait prisonniers par la police. Les enquêteurs ont établi des preuves contre les deux, mais le dossier n'est pas suffisamment accablant pour les incriminer totalement : pour cela, ils ont besoin d'obtenir des aveux. Chaque prisonnier a le choix entre avouer les méfaits, ou garder le silence. Si les deux complices avouent, ils prendront 4 ans de prison chacun. Si aucun n'avoue, les peines ne seront que de 1 an de prison chacun. Les enquêteurs interrogent les suspects dans des pièces séparées, et expliquent à chacun que s'il avoue et que l'autre garde le silence, il pourra sortir immédiatement sous statut de protection de témoin. En revanche, s'il garde le silence et que l'autre suspect l'accable en avouant, sa peine sera aggravée à 5 ans de prison.

Cette situation peut-être représentée par le tableau de la figure 7.1. Joe choisit la ligne, Averell choisit la colonne et chaque case contient, en bas à gauche, la peine de Joe, et en haut à droite, celle de Averell.

FIGURE 7.1 – Le jeu du Dilemme du Prisonnier

		Averell	
		Silence	Avouer
Joe	Silence	1 an 1 an	sortie 5 ans
	Avouer	5 ans sortie	4 ans 4 ans

Analysons les meilleurs choix possibles de chacun des prisonniers, d'un point de vue strictement personnel. Pour Joe, si Averell garde le silence, garder le silence conduit à un an de prison, tandis qu'avouer permet de sortir libre. Si Averell avoue, avouer conduit à 4 ans derrière les barreaux, et garder le silence à 5 ans. Indépendamment du choix d'Averell, Joe s'en tire mieux en avouant plutôt qu'en gardant le silence (voir les flèches verticales de la figure 7.2). La situation d'Averell est symétrique, et lui aussi a intérêt à avouer indépendamment du choix de Joe (voir les flèches horizontales de la figure 7.2). On dit dans ce cas qu'avouer est une *stratégie dominante* pour chacun des joueurs.

Cette logique conduit à ce que chaque prisonnier avoue, et passe 4 années en prison. Cela peut paraître paradoxal, car chacun a la possibilité de garder le silence, et si les deux gardent le silence les conséquences sont bien meilleures pour chacun d'entre eux. En d'autres termes, passer 1 an de prison pour chacun domine 4 ans de prison chacun au sens de Pareto, et l'issue sélectionnée par une logique de choix individuels conduit à une situation inefficace du point de vue des prisonniers.

Le dilemme du prisonnier illustre donc la possibilité que, en cas d'interactions stratégiques, la recherche par chacun de son intérêt propre est Pareto-inefficace, contrairement à l'équilibre d'une économie de marché. En général, le premier théorème fondamental ne tient pas en présence d'interactions stratégiques.

FIGURE 7.2 – Analyse du dilemme du prisonnier

		Averell	
		Silence	Avouer
Joe	Silence	1 an → sortie	5 ans
	Avouer	5 ans → sortie	4 ans

7.1.2 Externalités et dilemme du prisonnier

Face à une externalité négative, l'intérêt individuel est contraire à l'intérêt collectif ; d'où l'inefficacité. De telles situations peuvent être perçues comme relevant du dilemme du prisonnier. Pour le voir, considérons deux pays voisins, chacun ayant le choix entre réduire sa pollution due aux émissions de gaz à effet de serre, ou non. On fait les hypothèses suivantes :

- Réduire les émissions a un coût de 100 milliards d'euros pour chaque pays ;
- Un pays qui ne réduit pas ses émissions engendre une externalité négative, sous la forme d'un coût direct de 80 milliards d'euros pour chaque pays.

On calcule alors les gains de la manière suivante : si aucun pays ne réduit sa pollution, chacun subit l'externalité négative qu'il crée, ainsi que celle de son voisin, subissant alors une perte équivalente à 160 milliards d'euros. Si chaque pays réduit sa pollution, chacun ne doit payer que 100 milliards d'euros, dus au coût de la réduction de la pollution. Enfin, si un pays réduit sa pollution et l'autre non, celui qui la réduit absorbe un coût total de 180 milliards d'euros, tandis que celui qui pollue ne porte que 80 milliards d'euros d'externalités qu'il crée. La table de gains est représentée figure 7.3.

On constate ici la même logique qu'au dilemme du prisonnier. Chaque pays a intérêt à polluer, indépendamment du choix de l'autre pays, même si la meilleure issue pour les deux est obtenue lorsque les deux pays réduisent leur pollution. Nous avons vu les inefficacités engendrées par les externalités dans le cadre d'un marché, mais la même logique est à l'œuvre de manière plus générale.

De façon plus générale, de nombreuses situations sont analogues à un dilemme du prisonnier.

- Face à des entreprises mobiles, la compétition fiscale entre États consiste à diminuer ses taxes ou à offrir des incitations financières aux entreprises pour les attirer. Si chaque

FIGURE 7.3 – Jeu de pollution

		Pays 2	
		Réduire	Polluer
Pays 1	Réduire	-100	-80
	Polluer	-80	-160

Diagramme de jeu de pollution montrant les gains et les transitions :

- Transition de (-100, -100) à (-80, -80) : flèche bleue horizontale vers la droite.
- Transition de (-100, -180) à (-80, -80) : flèche bleue verticale vers le bas.
- Transition de (-180, -180) à (-160, -160) : flèche bleue horizontale vers la droite.
- Transition de (-180, -180) à (-160, -160) : flèche bleue verticale vers le bas.
- Les gains (-160, -160) sont écrits en rouge.

État riposte aux diminutions d'impôts de ses voisins par de nouvelles réductions, les recettes fiscales s'érodent sans que chaque État n'arrive à attirer beaucoup d'entreprises sur son sol.

- La tragédie des biens communs suit la même dynamique, et engendre la surexploitation des ressources naturelles : lorsque des pêcheurs en mer ne sont pas soumis à des quotas, ils ont intérêt à extraire un maximum de poissons avant leurs concurrents et accélèrent l'épuisement de la ressource.
- Lors d'un pic de pollution, on conseille à tous les automobilistes de ne pas utiliser leur voiture en attendant que les niveaux de pollution se stabilisent. Néanmoins, beaucoup de personnes, pensant qu'ils n'ont qu'un petit impact sur la pollution globale, décident d'utiliser quand même leur véhicule.
- Si une dizaine d'étudiants partagent une cuisine, on risque également de courir à la catastrophe : chaque étudiant est tenté de passer peu de temps à ranger en espérant qu'un autre fera le travail. Si les dix étudiants n'arrivent pas à s'organiser, la cuisine deviendra rapidement insalubre !

Dans *La Guerre du Péloponnèse*, l'historien athénien Thucydide explique les causes de la guerre qui déchira les cités grecques à la fin du Vème siècle avant J.C. D'après lui, une des raisons qui a poussé la Grèce à s'engouffrer dans cette guerre fratricide est sa division en petits États incapables de coopérer correctement. Il explique :

Comme [les républiques du Péloponnèse] ne forment pas un seul peuple, chacun pense à ses intérêts. Et pour l'ordinaire rien ne se termine. Les uns ont surtout en vue quelque vengeance ; les autres veulent que leurs propriétés n'aient rien à souffrir. Ils se rassemblent tard, jettent vite un coup d'œil sur les intérêts communs, et s'occupent bien plus constamment de leurs affaires personnelles. Aucun ne croit que sa négligence particulière fasse du tort au bien général : il pense qu'un autre y pourvoira pour lui ; et tous ayant séparément la même pensée, l'intérêt commun se détruit sans qu'on s'en aperçoive.

Le dilemme du prisonnier a également sa place dans les questions de concurrence vues au chapitre précédent. Imaginons deux entreprises en duopole : nous avons vu que lorsque ces entreprises sont en compétition, elles produisent plus qu'une entreprise en monopole, et

que cette augmentation de la production réduisait le profit total absorbé par les producteurs. Parfois, de telles entreprises décident de s'entendre pour réduire leur production, augmentant ainsi leurs profits et leurs marges. Lorsqu'une telle entente a lieu, on parle parfois d'un cartel. Un des cartels les plus connus est l'OPEP, qui regroupe 15 pays producteurs de pétrole qui possèdent à eux seuls plus de 80% des réserves prouvées dans le monde. Ces pays décident de quotas d'extraction pour maintenir les prix à un niveau élevé. L'OPEP est loin d'être le seul cartel dans le monde : presque tous les secteurs d'activité ont, ou ont eu, un cartel pour imposer leurs prix : c'est le cas du marché de la bière, des diamants, de l'acier, du sucre... Néanmoins, un cartel reste difficile à maintenir : en effet, les membres d'un cartel font face à un dilemme du prisonnier. Lorsque les membres d'un cartel s'engagent à limiter leur production, ou à maintenir des prix élevés, leurs profits sont tous élevés. Néanmoins, chacun des membres aurait intérêt à dévier un peu de cette entente : s'ils produisent un peu plus que ce qui a été entendu, un membre tricheur peut profiter des prix hauts causés par les restrictions des autres membres et engranger encore plus de profits. Si l'entente est réalisée sur les prix et que tous les membres du cartel s'engagent à afficher des prix très hauts, un membre tricheur peut afficher un prix un peu moins haut pour attirer plus de clients, ce qui augmente sa part du marché. Seulement, si tous les membres du cartel décident de tricher, tout se passe comme s'il n'y avait pas d'entente : c'est le retour à la concurrence. C'est pour cette raison que la coopération entre membres de l'OPEP présente des failles : plusieurs pays ont une large tendance à dépasser leurs quotas. L'ancien ministre saoudien du pétrole a d'ailleurs annoncé à propos de l'OPEP : « malheureusement, nous avons tendance à tricher » !

Ces ententes anticoncurrentielles sont en général illégales au sein d'un pays pour les raisons discutées au chapitre précédent. Les ententes entre entreprises sont ainsi sévèrement sanctionnées par les autorités du contrôle de la concurrence, et des amendes sont régulièrement infligées aux entreprises qui ne respectent pas ces lois. Pour détecter et démanteler plus facilement des cartels, les autorités de la concurrence utilisent la stratégie suivante : elles sont particulièrement clémentes avec les entreprises qui dénoncent les membres des cartels dont elles font partie. Une telle pratique place une nouvelle fois les membres du cartel face à un dilemme du prisonnier.

7.1.3 Définition formelle d'un jeu

Maintenant que nous avons vu l'exemple du dilemme du prisonnier, voyons comment formaliser un jeu en général.

On définit un *jeu sous forme normale* $G = (I, (A^i)_i, (g^i)_i)$ comme un triplet comportant :

1. Un ensemble I de *joueurs*
2. Un ensemble A^i de *stratégies* pour chaque joueur i
3. Une *fonction de gains* pour chaque joueur i ,

$$g^i: \prod_{j \in I} A^j \rightarrow \mathbb{R}.$$

Dans l'exemple du dilemme du prisonnier, nous avons deux joueurs qui ont chacun deux stratégies, « silence » ou « avouer », et les valeurs des fonctions de gain de Joe et Averell dans le tableau.

On supposera le plus souvent que les ensembles I et A^i sont tous finis, auquel cas on dit qu'il s'agit d'un jeu fini. On utilise les notations suivantes :

- $A = \prod_i A^i$. Un élément a de A est appelé un *profil de stratégies*. C'est un vecteur qui regroupe les stratégies utilisées par chaque joueur,
- $A^{-i} = \prod_{j \neq i} A^j$. Un élément a^{-i} dans cet ensemble A^{-i} décrit les choix de tous les joueurs auquel le joueur i fait face,
- $g(a) = (g^i(a))_i \in \mathbb{R}^I$, représente donc la fonction de gains vectoriels.

La notion de dominance entre stratégies permet de formaliser l'idée qu'une stratégie est meilleure qu'une autre pour un joueur indépendamment du choix des autres. Étant donné un joueur i et deux stratégies $a^i, b^i \in A^i$,

- a^i domine strictement b^i si pour tout $a^{-i} \in A^{-i}$

$$g^i(a^i, a^{-i}) > g^i(b^i, a^{-i})$$

- a^i domine faiblement b^i si pour tout $a^{-i} \in A^{-i}$

$$g^i(a^i, a^{-i}) \geq g^i(b^i, a^{-i}),$$

avec inégalité stricte pour au moins un $a^{-i} \in A^{-i}$.

De plus,

- Une stratégie qui est *dominée (strictement ou faiblement) par une autre* est dite *dominée (strictement ou faiblement)*
- Une stratégie qui *domine toutes les autres (strictement ou faiblement)* est dite *dominante (strictement ou faiblement)*.

Dans le dilemme du prisonnier, « Avouer » est une stratégie strictement dominante, et « Garder le silence » est strictement dominée par « Avouer ».

7.2 Équilibre de Nash

Lorsque nous avons étudié le duopole, nous nous sommes concentrés sur l'équilibre de Cournot-Nash et avons vu comment les joueurs convergent vers cet équilibre. Dans le cas du dilemme du prisonnier, la situation où les deux joueurs avouent est également un équilibre, car aucun des deux prisonniers n'a intérêt à changer de stratégie : se taire ne peut qu'augmenter la durée de leur peine. L'étude de tels équilibres est au cœur de la théorie des jeux. Nous allons étudier plusieurs autres exemples de jeu et formaliser la notion d'équilibre de Nash.

7.2.1 Les jeux de coordination

La chasse au cerf

Deux chasseurs peuvent décider chacun de chasser soit le cerf, soit le lapin. Chasser le lapin peut se faire seul, mais il faut être deux pour chasser le cerf. Un chasseur partant à la chasse au lapin revient avec un lapin. En revanche, un chasseur partant à la chasse au cerf n'aura de succès que si l'autre chasseur choisit lui aussi de chasser le cerf; dans le cas contraire il reviendra bredouille. Pour fixer les idées supposons qu'une chasse infructueuse équivaut à un paiement nul, revenir avec un lapin à 1, et partager un cerf à 3 chacun. Le tableau de gains du jeu est représenté figure 7.4.

FIGURE 7.4 – Le jeu de la chasse au cerf

	Cerf	Lapin
Cerf	3 ← 3	0 ← 1
Lapin	1 → 3	0 → 1

Le jeu de la chasse au cerf n'admet ni stratégie dominante, ni dominée. Chaque stratégie (cerf ou lapin) peut être le meilleur choix possible, en fonction du choix de l'autre joueur. Cependant, nous voyons qu'il existe deux issues stratégiquement stables, la première dans laquelle les deux joueurs chassent le cerf, et la seconde dans laquelle ils chassent tous deux le lapin. En effet, si l'autre joueur chasse le cerf, il vaut mieux chasser le cerf que le lapin; tandis que si l'autre joueur chasse le lapin il vaut mieux chasser le lapin. Ces deux situations où personne n'a intérêt à dévier en changeant de stratégie sont appelées *équilibres de Nash* du jeu, et généralisent bien la notion d'équilibre de Cournot-Nash vue dans le chapitre précédent. Ici, la meilleure réponse de chaque joueur à la stratégie "Cerf" est également "Cerf", tandis que la meilleure réponse à la stratégie "Lapin" est "Lapin" pour les deux joueurs. Dans les deux situations (Cerf, Cerf) et (Lapin, Lapin), les deux joueurs utilisent la meilleure réponse face à leur adversaire, c'est ce qui rend l'équilibre stable. De manière plus générale, un équilibre de Nash correspond à une situation de meilleures réponses mutuelles.

Le jeu de la chasse au cerf est un jeu dit de *coordination* : obtenir la meilleure issue dépend de la capacité des joueurs à coordonner leurs choix. De telles questions sont courantes en économie dès qu'il s'agit de faire un choix collectif.

Dans le jeu de la chasse au cerf, il y avait une issue meilleure que l'autre : mieux vaut que les deux chasseurs aillent ensemble chasser le cerf. Il semblerait étonnant que les deux joueurs se coordonnent sur la chasse au lapin : ce second équilibre n'est-il pas un peu artificiel ?

Malheureusement, il arrive que des agents se coordonnent mal : c'est typiquement le cas lors de paniques bancaires. En effet, une banque ne fonctionne pas comme une tirelire

géante : lorsque vous contractez un prêt auprès de votre banque, l'argent qui apparaît sur votre compte n'est pas présent sous la forme de pièces sonnantes et trébuchantes dans les coffres de la banque. La somme qui figure sur votre relevé de compte constitue simplement une promesse de la banque de vous fournir autant de pièces et billets si vous venez les réclamer. Comme tout le monde ne va pas vider son compte au distributeur en même temps, la banque peut tenir toutes ses promesses sans posséder sous la forme de pièces et billets l'argent contenu dans les comptes de ses clients. Usuellement, les banques ne détiennent en réserve qu'un petit pourcentage de l'argent qu'elles doivent à leurs clients, le reste étant prêté contre intérêt à des entreprises ou à des ménages qui veulent réaliser des investissements de long terme. Ce mécanisme ne fonctionne correctement que parce que les retraits des clients sont parcimonieux et espacés dans le temps. Si tous les clients d'une banque décident de venir récupérer l'argent disponible sur leur compte dans un délai très court, la banque n'est plus en mesure de fournir à ses clients l'argent dont ceux-ci disposent sur leur compte et celle-ci fait faillite. Une telle ruée des clients vers la banque pour récupérer leur argent s'appelle une panique bancaire. Cet événement peut avoir lieu lorsque les clients d'une banque pensent que celle-ci va faire faillite. Pour ne pas perdre leur argent, ils se dépêchent d'aller le chercher avant que la banque ne fasse effectivement faillite. Soulignons qu'il s'agit là de la seule attitude rationnelle d'un déposant, contribuant ainsi à amplifier la panique.

Cette situation est assez similaire à celle de la chasse au cerf, avec un « bon » équilibre (le fonctionnement normal de la banque, où les épargnants ne retirent pas leur argent en masse) et un « mauvais » équilibre, où tous se ruent vers la banque. Une simple rumeur peut déclencher une panique bancaire.

Les paniques bancaires étaient très fréquentes lors des crises financières du XX^{ème} siècle mais se font aujourd'hui plus rares. En effet, l'État intervient de nos jours pour limiter la ruée vers les distributeurs. Ainsi, pendant la crise grecque en 2015, le gouvernement limita les retraits à 60 euros par jour, empêchant les clients des banques grecques de vider leur compte d'un seul coup et ralentissant l'affaiblissement des banques. Par ailleurs, les états de l'Union Européenne doivent assurer les dépôts bancaires de leurs citoyens à hauteur de 100 000 euros par compte et par personne. Ainsi, si vous avez moins de 100 000 euros sur vos comptes en banque, vous ne perdrez pas un sou même si toutes les banques dans lesquelles vous avez un compte font faillite. Une telle régulation limite l'incitation des clients à vider leur compte en cas de crise, sachant qu'ils ne perdront pas d'argent.

Autres jeux de coordination

On observe souvent des comportements de coordination dans l'adoption de normes, par exemple dans le choix d'un sens de conduite. La majorité des pays d'Afrique de l'Ouest sont des anciennes colonies françaises qui se sont alignées sur la norme française de circulation à droite. Au cœur de ces anciens territoires français se trouvent quatre anciennes colonies anglaises qui utilisaient elles un sens de circulation à gauche : la Gambie, le Ghana, le Sierra Leone et le Nigeria. Peu après leur indépendance, ces pays ont décidé de se coordonner avec leurs voisins en changeant leur sens de circulation. Dans cet exemple, il y a bien plusieurs

équilibres potentiels (tous les pays roulent à gauche, et tous roulent à droite), mais ceux-ci sont assez équivalents : contrairement au jeu de la chasse au cerf, il n'y a plus un équilibre clairement préférable à l'autre.

Un domaine dans lequel la coordination est particulièrement visible est internet, avec le choix d'un réseau social. A l'heure actuelle, Facebook est le réseau social dominant sur le marché. Un nouvel internaute qui veut s'inscrire sur un réseau social veut en choisir un sur lequel sont inscrits ses amis : si ceux-ci sont sur Facebook, c'est cette plate-forme qu'il va rejoindre, rendant l'usage du site encore plus répandu. A l'équilibre, il est donc probable qu'un seul réseau social se démarque comme majoritaire, ou alors un par grande zone géographique. Le mécanisme de base est le même que pour le jeu du sens de circulation : le plus important pour les personnes concernées est d'imiter les choix de leurs semblables. Par ailleurs, ce choix n'est pas irrévocable : comme les modes vestimentaires qui changent régulièrement, la société peut changer de réseau social de prédilection, passant d'un équilibre à un autre. C'est ce qui s'est passé entre 2008 et 2010, lorsque Facebook devint le premier réseau social au monde, dépassant MySpace, qui fut un temps le site le plus visité aux États-Unis, et une plate-forme florissante. En l'espace de quelques mois, le marché fut complètement bouleversé et les utilisateurs de MySpace délaissèrent le site pour rejoindre Facebook.

Ces effets de réseau expliquent également pourquoi l'adoption de technologies ou de standards suit les mêmes dynamiques de coordination. Au début des années 1980, lors de l'émergence des cassettes vidéo, deux formats différents étaient en concurrence : le VHS mis au point par l'entreprise japonaise JVC et le Betamax lancé par Sony. Pendant quelques années les deux formats étaient produits et vendus, jusqu'à ce qu'une petite majorité de distributeurs (les vidéo-clubs par exemple) décident de n'utiliser que le VHS. À partir de ce moment, la part de marché de ce format augmenta rapidement jusqu'à ce que Sony abandonne le format Betamax : on a ainsi observé une coordination vers un équilibre où tous utilisaient la même technologie. Plus récemment, une même guerre de format eut lieu sur le marché des films HD, ce qui permit à Sony de prendre sa revanche : la firme japonaise lança en 2006 le disque Blu-Ray pour concurrencer le HD-DVD de Toshiba. Après quelques années de concurrence où les deux formats étaient sur le marché, Toshiba abandonna le HD-DVD en 2008 et commença à produire des lecteurs Blu-Ray.

Définition formelle d'un équilibre de Nash

Maintenant que nous avons vu plusieurs exemples d'équilibre de Nash, nous pouvons en donner une définition formelle. Un *équilibre de Nash* d'un jeu $G = (I, (A^i)_i, (g^i)_i)$ est un profil de stratégies $a \in A$ tel que

$$g^i(a^i, a^{-i}) \geq g^i(b^i, a^{-i})$$

pour tout joueur i et toute stratégie $b^i \in A^i$.

Cette notion, développée en 1951 par le mathématicien et économiste John Nash, généralise bien celle d'équilibre de Cournot décrite plus d'un siècle plus tôt. Dans un équilibre de Nash, aucun joueur n'a intérêt à changer de stratégie, à stratégies des autres joueurs don-

nées. On dit alors que chacun des joueurs joue une *meilleure réponse* au profil de stratégie des autres joueurs. Pour trouver un équilibre de Nash, on peut ainsi s'intéresser aux meilleures réponses de chaque joueur aux actions des autres, comme nous l'avons fait au chapitre précédent.

7.2.2 Stratégies mixtes, extension mixte et théorème de Nash

Un jeu sans équilibre de Nash

Le jeu de la chasse au cerf admet deux équilibres de Nash : l'équilibre de Nash d'un jeu n'est donc pas nécessairement unique. Par ailleurs, certains jeux n'admettent pas forcément d'équilibre de Nash. Considérons par exemple le jeu de « Pierre, Papier, Ciseaux ». Il s'agit d'un jeu à deux joueurs dans lequel chacun choisit entre Pierre, Papier, et Ciseaux. Ciseaux gagne contre Papier, Papier gagne contre Pierre, et Pierre gagne contre Ciseaux. C'est un jeu dit *à somme nulle* car la somme des gains des deux joueurs est nulle, il y a soit un gagnant et un perdant, soit match nul. En notant les joueurs 1 et 2, on a donc $g^2 = -g^1$. Pour les jeux à somme nulle à deux joueurs, la convention est de n'indiquer que le paiement du joueur 1 (celui qui choisit la ligne); ce joueur 1 cherche donc à maximiser le paiement indiqué. Le paiement du joueur 2 (qui choisit la colonne) est l'opposé du paiement indiqué, il cherche à maximiser son paiement, donc à minimiser le paiement du joueur 1.

Le jeu de « Pierre, Papier, Ciseaux » est représenté figure 7.5. On peut voir qu'il n'admet pas d'équilibre de Nash. En effet, comme illustré figure 7.5, dès lors que le joueur 2 joue une meilleure réponse au joueur 1, le joueur 2 remporte la partie. Et dès lors que le joueur 1 joue une meilleure réponse au joueur 2, c'est lui qui remporte la partie. Comme les deux conditions ne peuvent être vérifiées simultanément, il est impossible que les deux joueurs jouent une meilleure réponse l'un à l'autre simultanément.

FIGURE 7.5 – Le jeu de « Pierre, Papier, Ciseaux »

	Pierre	Papier	Ciseaux
Pierre	0	-1	1
Papier	1	0	-1
Ciseaux	-1	1	0

Tous ceux qui ont joué à « Pierre, Papier, Ciseaux » dans leur vie savent qu'il est important de ne pas pouvoir être prédit par l'autre joueur, il faut pour cela faire entrer de l'aléatoire dans sa manière de jouer. C'est le sens de la notion de stratégies mixtes que nous allons maintenant aborder.

Un premier exemple de stratégie mixte

Revenons à une situation dans laquelle les joueurs ne doivent faire leur choix que parmi deux stratégies. Imaginons par exemple deux armées qui s'affrontent : l'armée qui attaque

choisit de passer par la terre ou par la mer, et l'armée adverse défend soit le passage terrestre, soit le passage maritime. L'attaquant échoue face aux troupes du défenseur et ne gagne le combat que lorsqu'il passe par le côté qui n'est pas défendu. Si une victoire se traduit par un paiement de 1 et une défaite par 0, on peut représenter le jeu par le tableau ci-dessous :

FIGURE 7.6 – Un jeu d'attaque-défense

		Défense	
		Terre	Mer
Attaque	Terre	0 1	1 0
	Mer	1 0	0 1

Comme dans le jeu de pierre, papier, ciseaux, il n'y a pas ici d'équilibre de Nash : si l'un des deux joueurs est prévisible, il court à sa perte. Pour être imprévisible, les joueurs peuvent ajouter une part d'aléatoire dans leur comportement : au lieu de choisir la stratégie « terre » (T) ou « mer » (M) avec une probabilité 1, ils peuvent choisir chaque stratégie avec une certaine probabilité.

On peut par exemple construire une stratégie dans laquelle l'armée à l'attaque choisit de passer par la terre avec une probabilité p , et par la mer avec une probabilité $1 - p$, pour une valeur de p comprise entre 0 et 1, et on note cette stratégie $pT + (1 - p)M$. Une telle stratégie est appelée une *stratégie mixte*, car elle est composée de différentes *stratégies pures* du jeu, jouées avec différentes probabilités. Dans notre cas, les stratégies T et M sont les stratégies pures, et les stratégies de la forme $pT + (1 - p)M$ sont les stratégies mixtes. En reprenant les notations précédentes, on a $A^1 = A^2 = \{T, M\}$ et l'ensemble des stratégies mixtes est $S^1 = S^2 = \{pT + (1 - p)M \mid p \in [0, 1]\}$. On peut identifier chaque stratégie pure à une stratégie mixte jouant cette stratégie avec probabilité 1, et donc $A^i \subset S^i$: en introduisant les stratégies mixtes, on étend les possibilités de chaque joueur.

Le jeu que nous étudions n'admet pas d'équilibre de Nash en stratégies pures, c'est à dire qu'il n'existe pas de profil de stratégie $a \in A$ telle qu'aucun des deux joueurs n'ait intérêt à changer unilatéralement de stratégie. Est-ce toujours le cas lorsqu'on considère l'ensemble des profils de stratégies mixtes $s \in S = S^1 \times S^2$?

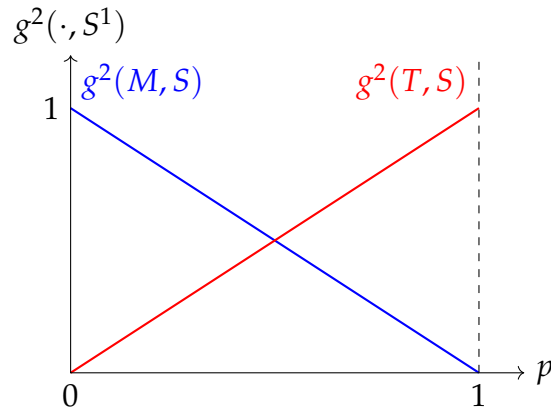
Pour cela, étudions les fonctions de meilleure réponse de chaque joueur, en commençant par le joueur 2, en défense. Si l'armée en attaque choisit la stratégie $s^1 = pT + (1 - p)M$, alors la défense empoche les gains suivants :

- En choisissant la terre : $g^2(T, S^1) = p$
- En choisissant la mer : $g^2(M, S^1) = 1 - p$

Ces gains sont représentés dans la figure 7.7. On y voit que lorsque p est strictement supérieur à 0,5 (l'attaque est plus probable par la terre), alors le défenseur a intérêt à jouer la

stratégie pure T , qui donne un gain plus élevé que M (et donc plus élevé que toute stratégie mixte $s^2 = qT + (1 - q)M$ pour $q < 1$). Lorsqu'au contraire $p < 0,5$, l'attaque est plus probable par la mer et la meilleure réponse du joueur 2 est de défendre la mer en jouant la stratégie pure M . Le cas $p = 0,5$ est un cas critique, où les deux stratégies pures M et T donnent le même gain à la défense, et celui-ci est indifférent entre toutes les stratégies parmi lesquelles il peut choisir (pures et mixtes) : elles lui donnent toutes un gain espéré de $\frac{1}{2}$.

FIGURE 7.7 – Gains de la défense pour chaque stratégie pure, en fonction de la stratégie de l'attaque $s^1 = pT + (1 - p)M$



On peut ainsi construire la correspondance de meilleure réponse du joueur 2 à la stratégie du joueur 1, qui associe à chaque stratégie du joueur 1 l'ensemble des stratégies de 2 qui maximisent son gain¹ :

$$BR^2(pT + (1 - p)M) = \begin{cases} \{M\} & \text{si } p < 0,5 \\ \{qT + (1 - q)M \mid q \in [0, 1]\} & \text{si } p = 0,5 \\ \{T\} & \text{si } p > 0,5 \end{cases}$$

Notez bien que la correspondance de meilleures réponses a pour ensemble de départ l'ensemble des stratégies du joueur 1 (S^1), mais pour ensemble d'arrivée l'ensemble des sous-ensembles de S^2 . Ici, le joueur 2 est indifférent entre toutes ses stratégies lorsque le joueur 1 joue $0,5T + 0,5M$, ce pourquoi $BR^2(0,5T + 0,5M) = S^2$.

Après avoir construit la correspondance de meilleures réponses du défenseur BR^2 , nous pouvons nous intéresser aux meilleures réponses du joueur 1. On déduit de la même manière l'ensemble des réactions optimales du joueur 1 lorsque le joueur 2 utilise la stratégie $s^2 = qT + (1 - q)M$:

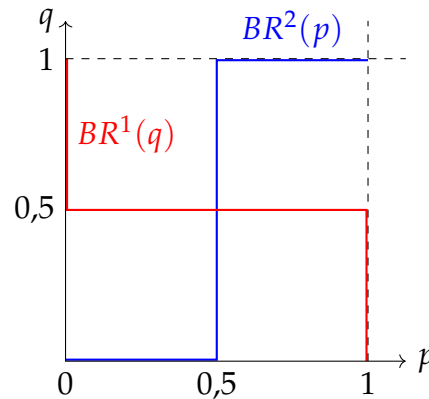
$$BR^1(qT + (1 - q)M) = \begin{cases} \{T\} & \text{si } q < 0,5 \\ \{pT + (1 - p)M \mid p \in [0, 1]\} & \text{si } q = 0,5 \\ \{M\} & \text{si } q > 0,5 \end{cases}$$

Maintenant que nous avons trouvé les correspondances de meilleures réponses de chaque joueur, nous pouvons chercher les équilibres de Nash, qui correspondent à des situations

1. La meilleure réponse est ici notée BR comme "best response" en anglais.

dans lesquelles chaque joueur joue une meilleure réponse à la stratégie de son opposant. Dans le cas du duopole de Cournot, nous avons trouvé l'équilibre de Nash en traçant les fonctions de meilleure réponse dans le plan et en trouvant leur point d'intersection. Nous pouvons faire de même pour trouver un équilibre de Nash du jeu d'attaque-défense. Si on a $s^1 = pT + (1 - p)M$ et $s^2 = qT + (1 - q)M$, alors les paramètres p et q capturent totalement le profil de stratégies (s^1, s^2) . Les fonctions BR^1 et BR^2 peuvent ainsi être tracées dans le plan (p, q) , comme dans la figure 7.8.

FIGURE 7.8 – Fonctions de meilleure réponse dans le jeu d'attaque-défense.



Les fonctions de meilleure réponse se croisent lorsque $p = q = 0,5$. Ce point d'intersection correspond à un équilibre de Nash en stratégies mixtes, qui est de plus l'équilibre unique du jeu. À l'équilibre de Nash, chaque armée choisit une stratégie aléatoirement avec une probabilité de $1/2$, et a une chance sur deux de remporter la bataille. Chaque joueur obtient un gain moyen de $0,5$. Si un joueur venait à dévier de cette stratégie, alors il deviendrait (au moins partiellement) prévisible, et donnerait à l'autre joueur l'opportunité d'augmenter ses chances de gain. Notez qu'à l'équilibre de Nash, chaque joueur est alors indifférent entre jouer sa stratégie mixte et n'importe quelle autre stratégie.

Dans le jeu « Pierre, Papier, Ciseaux », nous pouvons vérifier que le profil de stratégie mixtes dans lequel chaque joueur joue chacune des stratégies avec la même probabilité $1/3$ forme un équilibre de Nash en stratégies mixtes du jeu. En effet, si le joueur 2 joue chacune de ses 3 stratégies pures avec la même probabilité, chacune des stratégies pures du joueur 1 gagne $1/3$ du temps, perd $1/3$ du temps, et fait match nul $1/3$ du temps. Le joueur 1 est donc indifférent entre toutes les stratégies pures, et par conséquent indifférent entre toutes les stratégies mixtes. En particulier, la stratégie du joueur 1 qui choisit chaque stratégie pure avec probabilité $1/3$ est une meilleure réponse du joueur 1. Pour la même raison, cette même stratégie est aussi une meilleure réponse du joueur 2. Si un joueur dévie de ce profil, il joue une des trois stratégies avec une probabilité strictement supérieure à $1/3$, et en jouant la meilleure réponse à cette stratégie, son adversaire s'assure un gain strictement positif (alors qu'à l'équilibre de Nash il aurait eu un gain nul).

Pour aller plus loin : l'équilibre de Nash comme point fixe. Pour trouver l'équilibre de Nash dans le jeu du duopole de Cournot ou dans le jeu d'attaque-défense que nous venons

d'étudier, nous avons tracé les correspondances de meilleures réponses dans un plan, et trouvé l'équilibre à l'intersection de celles-ci. Plus généralement, il peut être utile de voir un équilibre de Nash comme un point fixe d'une correspondance. Considérons en effet la correspondance de meilleure réponse globale :

$$BR : \begin{pmatrix} s^1 \\ s^2 \end{pmatrix} \mapsto \left\{ \begin{pmatrix} t^1 \\ t^2 \end{pmatrix} \mid t^1 \in BR^1(s^2), t^2 \in BR^2(s^1) \right\}$$

Un profil de stratégie est un équilibre de Nash lorsque $\begin{pmatrix} s^1 \\ s^2 \end{pmatrix} \in BR \left(\begin{pmatrix} s^1 \\ s^2 \end{pmatrix} \right)$.

Voir ainsi un équilibre comme un point fixe permet de tirer parti des théorèmes d'analyse développés pour les étudier, comme le théorème de Kakutani ou celui de Brouwer. En particulier, le théorème d'existence d'équilibre de Nash se déduit simplement de celui d'existence d'un point fixe de Kakutani.

Formalisation de l'extension mixte d'un jeu

Notations. Considérons un jeu fini $G = (I, (A^i)_i, (g^i)_i)$. Une stratégie mixte consiste à choisir chaque stratégie pure avec une certaine probabilité : formellement, c'est donc une mesure de probabilité sur l'ensemble des stratégies A^i . On note cet ensemble de mesures de probabilité $S^i = \Delta(A^i)$. Ainsi :

- Si $A^i = \{X, Y\}$, $S^i = \{pX + (1-p)Y \mid p \in [0, 1]\}$,
- Si $A^i = \{X, Y, Z\}$, $S^i = \{pX + qY + (1-p-q)Z \mid p, q > 0, p+q < 1\}$.

Lorsque les joueurs utilisent les stratégies mixtes $s^i \in S^i$, quel est le gain espéré du joueur i ? On suppose que les choix aléatoires des différents joueurs sont indépendants, et ainsi si $s^j(a^j)$ est la probabilité que le joueur j choisisse la stratégie a^j , alors la probabilité que le profil de stratégie joué soit $a = (a^j)_j$ est $P_s(a) = \prod_{j \in I} s^j(a^j)$. On dit que $s = (s^j)_j$ induit la loi de probabilité P_s sur A .

Le gain espéré du joueur i est alors :

$$\tilde{g}^i(s) = \mathbb{E}_{P_s} g^i(a) = \sum_{a \in A} P_s(a) g^i(a).$$

Exemple. Si on reprend l'exemple de la chasse au cerf, avec :

$$\begin{cases} A^1 = \{C, L\} \\ A^2 = \{C', L'\} \\ s^1 = pC + (1-p)L \\ s^2 = qC' + (1-q)L' \end{cases}$$

Et la matrice de gains de la figure 7.4, alors on a $A = \{(C, C'), (C, L'), (L, C'), (L, L')\}$ et $s = (s^1, s^2)$ induit une probabilité sur A , avec :

$$\begin{cases} P_s((C, C')) = pq \\ P_s((C, L')) = p(1 - q) \\ P_s((L, C')) = (1 - p)q \\ P_s((L, L')) = (1 - p)(1 - q) \end{cases}$$

Et enfin :

$$\begin{cases} \tilde{g}^1(s) = 3pq + (1 - p)q + (1 - p)(1 - q) = 1 - p + 3pq \\ \tilde{g}^2(s) = 3pq + p(1 - q) + (1 - p)(1 - q) = 1 - q + 3pq \end{cases}$$

Pour caractériser tous les équilibres de Nash, en stratégies pures comme en stratégies mixtes, on calcule les correspondances de meilleures réponses définies comme dans le cas du jeu d'attaque défense. Notons p [resp. q] la probabilité que la stratégie mixte s^1 [resp. s^2] associe au choix « Cerf ». Dès lors que $q > 1/3$, le joueur 1 préfère strictement « Cerf » à « Lapin » et la seule meilleure réponse face à q est donc $p = 1$. Pour $q < 1/3$, le joueur 1 préfère strictement « Lapin » à « Cerf » et la seule meilleure réponse est $p = 0$, tandis que pour $q = 1/3$ le joueur 1 est indifférent et toute stratégie est meilleure réponse, les meilleures réponses sont donc données par l'intervalle $[0, 1]$. On a donc :

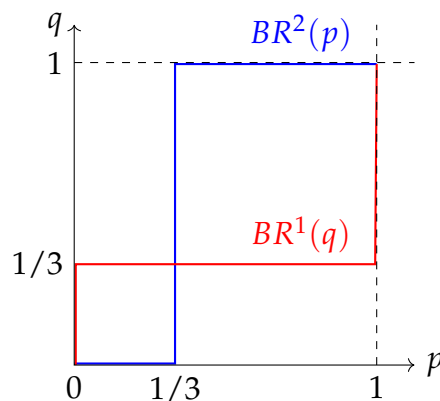
$$\begin{cases} BR^1(p) = \{1\} & \text{si } q > 1/3 \\ BR^1(p) = [0, 1] & \text{si } q = 1/3 \\ BR^1(p) = \{0\} & \text{si } q < 1/3 \end{cases}$$

et symétriquement pour le joueur 2 :

$$\begin{cases} BR^2(p) = \{1\} & \text{si } p > 1/3 \\ BR^2(p) = [0, 1] & \text{si } p = 1/3 \\ BR^2(p) = \{0\} & \text{si } p < 1/3 \end{cases}$$

Ces correspondances de meilleures réponses sont représentées dans la figure 7.9.

FIGURE 7.9 – Correspondances de meilleure réponse dans le jeu de la chasse au cerf.



Nous voyons que les correspondances de meilleures réponses s'intersectent en trois points : $(0,0)$, $(1,1)$ et $(1/3,1/3)$. Les deux premiers correspondent aux équilibres en stratégies pures déjà vus. Le troisième est un équilibre en stratégies mixtes dans lequel chaque joueur est indifférent entre les choix de « Cerf » et « Lapin ».

Extension mixte d'un jeu. Ces différents éléments nous permettent de construire l'extension mixte du jeu G , qui est simplement le nouveau jeu $\tilde{G} = (I, (S^i)_i, (\tilde{g}^i)_i)$ (cette fois-ci, ce n'est plus un jeu fini dès lors que l'un des A^i n'est pas un singleton). Un équilibre de Nash de l'extension mixte de G est appelé un *équilibre de Nash en stratégies mixtes* de G .

On remarque que toute stratégie pure $a^i \in A^i$ peut être vue comme l'élément de S^i qui associe la probabilité 1 à l'action a^i . Cela permet de voir A^i comme un sous-ensemble de S^i . On remarque que si a^i est meilleure réponse à un profil de stratégies a^{-i} parmi les stratégies pures, alors a^i est aussi une meilleure réponse parmi les stratégies mixtes. Cela implique en particulier que tout équilibre de Nash en stratégies pures est aussi un équilibre de Nash en stratégies mixtes.

Alors que l'existence d'un équilibre de Nash en stratégies pures n'était pas garanti, *tout jeu à nombre fini de stratégies pures admet un équilibre de Nash en stratégie mixtes* ! Ce résultat est connu sous le nom de théorème d'existence de Nash, démontré par John Nash en 1951². John Nash est un mathématicien américain qui a fait des contributions majeures en mathématiques (en analyse et en géométrie algébrique) ainsi qu'en théorie des jeux. Pour ses travaux, il a reçu le prix Nobel en économie en 1994 et le prix Abel en 2015.

7.2.3 Interprétation de l'équilibre de Nash

En sciences sociales, on peut interpréter un équilibre de Nash de plusieurs façons. Dans le cas des jeux de coordination vus plus haut, les équilibres de Nash forment des conventions, ou normes sociales, stables. Le choix d'un standard ou d'une mode de vêtements constitue un équilibre de Nash.

Dans le cadre du duopole de Cournot, nous avons vu comment les entreprises se rapprochent de la situation d'équilibre en ajustant progressivement leur production face aux ajustements de leur concurrent : nous pouvons interpréter cette convergence vers l'équilibre de Nash comme un processus d'apprentissage. Une expérience connue en théorie des jeux est le « jeu du 2/3 ». Il consiste à rassembler un groupe et à demander à chaque personne dans le groupe d'écrire un nombre décimal entre 0 et 100. Celui qui a écrit le nombre le plus proche des deux tiers de la moyenne des réponses gagne le jeu. Dans ce jeu, il n'y a qu'un

2. Pour démontrer le théorème de Nash, on peut utiliser le fait que c'est un point fixe de la correspondance de meilleure réponse BR . Pour chaque stratégie $-i$ des joueurs autre que i , on note $BR^i(-i)$ l'ensemble des stratégies mixtes du joueur i qui sont meilleures réponses face à $-i$, c'est à dire l'ensemble des stratégies $i \in S^i$ qui maximisent le paiement espéré du joueur i face à $-i$. On définit ensuite la correspondance de meilleure réponse BR de S vers S par $BR(s) = \prod_i BR^i(-i)$ (c'est ce que nous avons fait plus haut avec deux joueurs). En remarquant que la correspondance BR est à valeurs non vides, convexes et fermées ; que le graphe de BR est fermé ; et qu'un point fixe de BR , c'est à dire un profil de stratégies tel que $s \in BR(s)$, est un équilibre de Nash en stratégies mixtes de G , l'application du théorème de point fixe de Kakutani (1941) permet de montrer que la correspondance BR admet un point fixe, et donc G admet un équilibre de Nash en stratégies mixtes.

seul équilibre de Nash, où tous les joueurs écrivent le chiffre zéro. Néanmoins, si vous organisez effectivement ce jeu, les réponses que vous aurez ne seront probablement pas zéro : certains joueurs ne vont pas comprendre tout de suite que seule la situation où tous jouent zéro est un équilibre, et la plupart des joueurs ne s'attendent pas à ce que tous les autres joueurs jouent effectivement à l'équilibre de Nash. Néanmoins, si vous demandez au même groupe de jouer plusieurs fois au jeu, vous verrez la réponse moyenne diminuer à chaque tour et converger vers zéro : on atteint l'équilibre de Nash suite à un apprentissage collectif.

Nous avons introduit la notion de stratégie mixte dans un exemple militaire où chaque joueur cherche à surprendre son adversaire. Les hommes de guerre ne choisissent pourtant pas le sort de leurs soldats à pile ou face : l'équilibre de Nash que nous trouvons a-t-il une interprétation ? Le jeu d'attaque-défense que nous avons étudié ressemble à beaucoup de situations sportives, où les joueurs à l'attaque et à la défense choisissent un côté. Par exemple, lors d'un tir de penalty au foot, le tireur choisit généralement de viser à droite ou à gauche, et le gardien plonge d'un côté ou de l'autre de son but. Le tireur et le gardien doivent tous deux être imprévisibles, et nous sommes dans une situation très proche de celle de la figure 7.6 (avec « droite » et « gauche » à la place de « terre » et « mer »). Il faudrait y remplacer les gains (0 ou 1) par la probabilité de marquer selon que le tireur tire à droite ou à gauche, et que le gardien plonge à droite ou à gauche. À l'équilibre de Nash, on trouve la probabilité que chacun des deux joueurs choisisse « droite » ou « gauche » et la probabilité d'un but. Un chercheur a décidé de voir si les joueurs de foot tiraient aléatoirement à droite ou à gauche du but selon les prévisions de la théorie des jeux. Il analysa plus de 1400 tirs de penalty observés dans des matchs de football professionnel, trouva la matrice de gains et l'équilibre de Nash théorique associé. Le comportement prédit par la théorie des jeux était remarquablement proche des comportements observés chez les joueurs (Palacios-Huerta, 2003)! De même, au poker, les joueurs peuvent décider de bluffer : lorsqu'ils ont un jeu faible, ils peuvent miser beaucoup pour faire croire qu'ils ont un jeu fort et pousser les autres joueurs à se coucher. Au contraire, lorsqu'ils ont un jeu fort, ils peuvent miser peu pour faire croire aux autres qu'ils n'ont pas un bon jeu et les pousser à miser beaucoup. Les stratégies mixtes permettent également à certaines institutions d'être beaucoup plus efficaces, notamment lorsqu'il s'agit de lutter contre la fraude. Lorsqu'on veut éviter qu'une population agisse dans l'illégalité, il est en général impossible de contrôler les actions de tout le monde : cela prendrait beaucoup trop de temps ou d'argent. On choisit alors de vérifier les agissements d'une partie seulement de la population. Dans les aéroports, les douanes luttent contre la contrebande, et procèdent donc à des fouilles. Là aussi, des fouilles prévisibles mèneraient à leur échec. Si les contrebandiers savent que les douaniers fouillent les bagages de tel ou tel vol, ou les passagers qui viennent d'un pays en particulier, ils n'auront qu'à passer par d'autres routes. Pour éviter une telle situation, les douaniers procèdent à des contrôles aléatoires : sur tous les vols, ils contrôlent une certaine proportion des bagages et des passagers, ce qui correspond à inspecter chaque bagage ou passager avec une probabilité p . Toujours dans les aéroports, les portails de sécurité sont programmés pour déclencher des contrôles aléatoires selon des fréquences préétablies.

Enfin, un domaine dans lequel la théorie des jeux a été beaucoup appliquée est la bio-

logie, surtout pour comprendre les phénomènes évolutifs. En interprétant les traits génétiques comme des stratégies et le nombre de descendants comme des gains, une stratégie dominante caractérise des traits qui vont s'imposer au fur et à mesure des générations, et un équilibre de Nash en stratégies mixtes correspond à une situation dans lesquels plusieurs traits coexistent en différentes proportions.

Un exemple particulièrement célèbre d'application de la théorie des jeux à la biologie concerne les lézards *Uta*, communs dans l'ouest des États-Unis et le nord du Mexique (Sinervo and Lively, 1996). Au sein de cette espèce de lézards, on trouve trois types de mâles, qui ont des gorges de couleurs différentes :

- Ceux qui ont des gorges oranges sont plus gros que les autres, et vivent sur de larges territoires avec de nombreuses femelles qu'ils défendent agressivement. Néanmoins, ils ne peuvent pas surveiller toutes leurs femelles en même temps.
- Ceux qui ont des gorges jaunes sont plus sournois et profitent de cette inattention : ils ne défendent pas de territoire particulier, mais s'infiltrent discrètement sur des territoires gardés par les lézards oranges pour s'accoupler avec des femelles protégées par ceux-ci.
- Ceux qui ont des gorges bleues utilisent encore une autre stratégie : ce sont des lézards monogames qui concentrent toutes leurs ressources sur une seule femelle.

Ces trois types de lézards ont donc trois types de stratégies pour se reproduire. Laquelle est la plus efficace ? Lorsque les lézards bleus sont en majorité, ils se font battre par les oranges qui ont un territoire plus vaste et s'accaparent plus de femelles. Par contre, si les oranges étaient en majorité, les jaunes pourraient facilement se reproduire avec des femelles non surveillées par les oranges. Ces lézards sournois se développeraient donc rapidement. Enfin, si les lézards jaunes étaient en majorité, ils favoriseraient l'apparition des bleus, qui ne s'occupent que d'une femelle et qui peuvent ainsi repousser les jaunes qui essaieraient de s'en approcher. Cette situation s'apparente donc à un jeu de pierre-feuille-ciseaux, où aucune des trois stratégies n'est dominante. Au jeu de pierre-feuille-ciseaux, nous avons vu qu'à l'équilibre, il fallait choisir au hasard sa stratégie parmi les trois possibilités. Pour les lézards, il y a aussi un équilibre en stratégies mixtes : les trois types de lézards cohabitent entre eux, et si un type devient minoritaire, il dispose d'un avantage sur les deux autres types, ce qui lui permet d'augmenter en proportion.

Un autre exemple concerne les levures *saccharomyces cerevisiae*, types de levures communément utilisées par les boulangers, les viticulteurs et les brasseurs notamment. Ces levures se nourrissent de sucres simples (glucose et fructose), et peuvent transformer des sucres plus complexes (comme le sucrose) en glucose et fructose lorsqu'elles n'ont pas accès directement à ces sucres simples. Néanmoins, lors de cette transformation, les *saccharomyces cerevisiae* libèrent dans le milieu environnant la plupart des sucres simples ainsi produits, et qui peuvent être utilisés par d'autres levures. On observe dans la nature des levures qui ont la capacité de transformer des sucres complexes en sucres simples, et des levures qui ont perdu cette capacité. Ces deux types de levures correspondent à deux stratégies : les levures qui transforment les sucres ont une stratégie « altruistes », et aident leurs voisines

à croître, tandis que celles qui ont perdu la capacité de transformer les sucres exploitent le travail de leurs congénères en pompant dans leur environnement les sucres simples produits par les levures « altruistes ». On peut appeler ces levures « opportunistes ». Lorsqu'il y a principalement des levures altruistes dans la population, celles-ci produisent assez de sucres pour que des levures opportunistes puissent survivre en leur compagnie sans en produire elles-mêmes. Comme elles utilisent moins d'énergie pour se nourrir que les levures altruistes, elles peuvent se développer plus facilement que ces dernières et leur nombre augmente. Au contraire, lorsque les levures altruistes sont en minorité, les levures opportunistes ne peuvent pas trouver assez de sucres dans leur environnement pour se reproduire aussi vite que les levures altruistes, qui ne manquent pas de nourriture. Ce sont alors les levures altruistes qui se développent. À l'équilibre, il y a donc cohabitation entre levures altruistes et levures opportunistes (Gore et al., 2009).

7.3 Asymétries d'information

Jusqu'ici, nous avons mis de côté une défaillance de marché particulièrement importante : les asymétries d'information. Nous avons considéré implicitement que tous les agents ont accès à la même information. Lorsque vous achetez un litre d'essence, prendre en compte les asymétries d'information n'est pas très important car la qualité de l'essence est maintenue à un standard par des normes : il n'y a pas de risque de mauvaise surprise. Lorsque l'on s'intéresse au marché du logement ou de l'emploi, les asymétries d'information deviennent beaucoup plus importantes : des propriétaires qui veulent vendre leur maison connaissent ses vices cachés contrairement aux acheteurs potentiels. Il est difficile pour une employeuse de déceler lors d'un entretien d'embauche toutes les qualités et défauts d'un candidat à un poste. Avec l'oligopole de Cournot, les notions de théorie des jeux nous ont aidé à comprendre une défaillance de marché importante : le pouvoir de marché. La théorie des jeux donne également de nombreux outils pour modéliser les interactions entre des agents qui n'ont pas les mêmes informations à disposition. Dans ce cours, nous n'allons pas les détailler, mais simplement présenter un exemple désormais célèbre : le marché aux guimbardes, idée initialement développée par Akerlof (1978).

7.3.1 Le marché aux guimbardes

Les prix des voitures d'occasion sont inférieurs à ceux des voitures neuves, et le prix décroît avec l'ancienneté : rien de surprenant. Ce qui peut être plus étonnant est que, entre une voiture neuve et une voiture qui vient d'être achetée, la décote atteint typiquement entre 15 et 20% de la valeur ! Dit autrement, juste après l'achat d'une voiture neuve, vous avez perdu une sacrée somme d'argent.

D'un point de vue économique, rien ne justifie alors d'acheter une voiture neuve, il vaut mieux acheter une voiture ayant à peine quelques kilomètres au compteur ! Si c'était si simple, suffisamment d'acheteurs joueraient de ce bon plan, et le prix des voitures d'occasion quasi-neuves devrait remonter.

Qu'est-ce qui peut donc bien justifier que les acheteurs se détournent de modèles d'occasion quasi neufs ? Plaçons-nous du point de vue de l'acheteur face à un vendeur. La question naturelle qui se pose est : « Pourquoi ce vendeur, qui vient à peine d'acquérir ce véhicule, souhaite-t-il s'en débarrasser ? » Évidemment, il est possible que des changements soudains de plans de vie puissent motiver le vendeur. Mais il y a aussi une suspicion naturelle, celle qu'après quelque temps d'usage, le vendeur ait découvert un défaut dans le véhicule, qui ne pouvait pas être détecté au moment de l'achat.

Quelles peuvent-être les effets de telles asymétries d'information sur les échanges économiques ? Simplement des baisses de prix, ou bien le mal est-il plus grave ? Comme l'a montré Akerlof dans un article publié en 1971 sous le titre "*Market for lemons*" (ce qui peut être traduit par « marché aux guimbardes »), les asymétries d'information sont potentiellement des sources majeures de défaillance de marché, et peuvent rendre dans certains cas les échanges entre vendeurs et acheteurs tout simplement impossibles.

Considérons un marché aux voitures d'occasion. La valeur v d'un véhicule pour un vendeur, liée à sa qualité, est tirée selon une loi uniforme entre 0 et 20 000€. La valeur du même véhicule pour un acheteur est $1,5 \times v$, ce qui veut dire que l'acheteur valorise toujours le véhicule plus que le vendeur. Toute transaction à un prix d'échange situé entre v et $1,5 \times v$ est donc bénéficiaire à l'acheteur comme au vendeur. Le seul problème réside dans le fait que la valeur du véhicule n'est pas connue de l'acheteur.

Commençons par faire l'hypothèse que ni l'acheteur ni le vendeur ne connaissent la qualité du véhicule. C'est une hypothèse réaliste si, par exemple, la qualité correspond à d'éventuels problèmes mécaniques n'apparaissant qu'après un kilométrage important non encore atteint par le véhicule. Dans ce cas, du point de vue du vendeur la valeur moyenne du véhicule est donnée par l'espérance mathématique de v soit 10 000€, et la valeur moyenne du point de vue de l'acheteur est 15 000€. Tout prix d'échange entre ces deux valeurs est satisfaisant pour les deux parties et le véhicule peut-être vendu à ce prix.

Supposons maintenant qu'à la fois l'acheteur et le vendeur observent la qualité du véhicule. C'est le cas pour des facteurs facilement observables comme de la rouille éventuelle sur la carrosserie. Dans ce cas, v est connu de l'acheteur et du vendeur, et tout prix de transaction entre v et $1,5v$ est un prix d'échange satisfaisant pour le vendeur comme l'acheteur. Le prix d'échange dépend dans ce cas de la valeur du véhicule, et il est logique de penser que les véhicules de meilleure qualité (v élevé) sont échangés à des prix plus élevés que ceux de moins bonne qualité (v bas).

Passons maintenant au cas qui nous intéresse le plus, celui de l'asymétrie d'information entre l'acheteur et le vendeur. On suppose donc que le vendeur connaît v tandis que l'acheteur l'ignore. C'est une hypothèse logique si v est lié à des facteurs tels que la consommation de carburant ou d'huile ou la nécessité de réparations fréquentes par exemple.

Si tous les véhicules sont en vente indépendamment de leur qualité, la valeur moyenne pour les vendeurs d'un véhicule en vente est de 10 000€, et cette valeur moyenne est donc de 15 000€ pour un acheteur. Un acheteur qui n'est pas averse au risque acceptera donc d'acheter une voiture si son prix est inférieur à 15 000€. Si le prix est à 15 000€, tous les vendeurs qui ont des voitures ayant une valeur supérieure à 15 000€ retirent leur voiture

du marché. La valeur moyenne des voitures sur le marché chute alors à 7 500€ pour les vendeurs et à 11 250€ pour les acheteurs. Ceux-ci n'acceptent donc pas un prix supérieur à 11 250€ et si le prix s'abaisse à cette valeur, tous les vendeurs dont la voiture a une valeur entre 11 250€ et 15 000€ quittent le marché à leur tour. Ceci fait encore baisser les prix, etc. Ainsi, l'asymétrie d'information pousse tous les vendeurs à se retirer petit à petit du marché, qui n'existe plus à l'équilibre !

Formellement, supposons un prix $p > 0$ auquel s'échangent les véhicules d'occasion, prix équilibrant l'offre et la demande de ces véhicules. Pour un vendeur, il est bénéficiaire de vendre son véhicule dès lors que $v < p$, et il vaut mieux garder le véhicule que le vendre si $v > p$. Pour un véhicule d'occasion mis sur le marché, la valeur est donc uniformément répartie entre 0 et p , sa valeur moyenne pour un vendeur est donc $p/2$. Pour un acheteur, la valeur d'un véhicule en vente sur le marché d'occasion est uniformément répartie entre 0 et $1,5p$, sa valeur moyenne est donc $3/4p$. Du point de vue de l'acheteur, payer p pour un véhicule qui vaut en moyenne $3/4p$ uniquement est une perte, mieux vaut donc refuser l'achat. Le seul prix d'équilibre de marché possible est $p = 0$. A ce prix, seuls les véhicules de pire qualité s'échangent. Aucun véhicule de valeur strictement positive n'est mis en vente, alors que les acheteurs sont prêts à payer plus que les vendeurs.

7.3.2 La sélection adverse

Le problème rencontré dans le marché aux guimbardes est celui appelé **sélection adverse** : l'asymétrie d'information pousse les propriétaires de bons véhicules à se retirer du marché, et il ne reste que les mauvais véhicules. Le terme de sélection adverse est bien connu des assureurs qui sont régulièrement confrontés à ce problème. Lorsqu'ils assurent des clients, ils ne savent pas toujours si ceux-ci sont à fort ou à faible risque, et doivent alors proposer les mêmes tarifs à tous. Les clients à fort risque achètent systématiquement une assurance car il est assez probable qu'ils aient à l'utiliser. En revanche, les clients « sûrs », qui savent qu'un accident est pour eux peu probable, ne souscrivent pas à l'assurance car ils estiment que celle-ci coûte bien plus cher que ce qu'ils pourraient en gagner. Cela génère donc une sélection de clients risqués, ce qui augmente le niveau des cotisations d'assurance nécessaires au financement des remboursements à prévoir. Mais cette hausse de prix ne fait que renforcer la sélection adverse de clients risqués. Pour limiter ce problème, les assureurs peuvent par exemple enquêter sur le passé de leurs clients ou leur demander des examens médicaux dans le cas d'assurances santé. Ainsi, elles peuvent estimer le niveau de risque de leurs clients et ajuster leurs prix selon celui-ci.

La problème de la sélection adverse pose également problème lors de crises financières. Les banques font elles aussi face à un problème d'asymétrie d'information lorsqu'elles accordent des prêts à des particuliers ou à des entreprises : si les récipiendaires des prêts savent bien le niveau de risque de leurs investissements, les banques ne peuvent pas savoir à l'avance si leurs clients seront en mesure de les rembourser. Lors d'une crise, le nombre de clients à risque augmente, et il devient plus coûteux de contracter un emprunt (les taux d'intérêts associés aux prêts augmentent). La conséquence de cette évolution peut mener à un

problème similaire à celui du marché automobile. Les clients sûrs, face à l'augmentation des taux d'intérêts, peuvent décider de ne plus emprunter, estimant le coût de l'emprunt trop élevé sachant qu'ils devront le rembourser en intégralité. Les clients à risque, eux, savent qu'il est plausible qu'ils fassent défaut sur le prêt qui leur est accordé, et acceptent alors les taux d'intérêt élevés. Les banques ont alors parmi leur clients une proportion bien plus élevée de clients à risque, ce qui les pousse à augmenter encore leurs taux d'intérêt pour couvrir ce risque, et les faisant entrer dans un cercle vicieux. Dans les faits, en temps de crise, les banques limitent fortement leurs prêts et ne prêtent qu'à des clients dont elles savent qu'ils sont solvables pour éviter ce problème de sélection adverse.

7.3.3 L'aléa moral

En plus du problème de sélection adverse, les assureurs font également face à un autre risque. Imaginez qu'un assureur observe que, chaque année, 2% des utilisateurs cassent leur téléphone portable. S'il décide d'assurer 100 personnes en leur faisant payer chaque année l'équivalent de la réparation de deux téléphones et en leur promettant de remplacer le leur en cas de casse, il risque d'avoir une mauvaise surprise : le nombre de téléphones à réparer va être plus élevé que prévu ! Sachant qu'ils sont assurés, les utilisateurs font moins attention à leur téléphone et l'envoient plus souvent à la réparation. On parle alors d'**aléa moral**.

Il y a un risque d'aléa moral à chaque fois qu'il est impossible de vérifier qu'un contrat est correctement exécuté. Ainsi, une entrepreneuse qui lance une entreprise et demande de l'argent à une banque a le choix de travailler énormément et de maximiser la probabilité de son succès ou d'y consacrer moins de temps, rendant le succès commercial moins probable. La banque ne peut pas voir si l'entrepreneuse travaille beaucoup ou pas : ils ne voient que les profits de son entreprise. Si ceux-ci sont faibles, cela pourrait témoigner d'un manque de travail, mais aussi simplement d'un manque de chance. Disposant du même financement selon que sa firme ne réussisse ou pas, l'entrepreneuse peut avoir la tentation de moins travailler, et la banque pourrait augmenter les taux d'intérêt qu'elle lui impose pour se protéger de ce risque.

Pour étudier une telle situation, on distingue l'**agent** du **principal**. L'agent, par exemple l'assuré ou l'entrepreneuse, choisit les actions effectuées (faire attention ou pas, beaucoup travailler ou non). Le principal, par exemple l'assureur ou la banque, est affecté par les choix de l'agent mais ne peut pas directement les contrôler (ou bien ces choix ne sont pas de nature à servir de base à un contrat). Cependant, le principal observe des signaux (quantité produite par l'agent, retour des clients, etc.), qui peuvent servir de base à un contrat qui aligne mieux les intérêts de l'agent et du principal. Par exemple, un assureur peut réduire les frais d'assurance de ses clients qui n'ont pas eu d'accident dans le passé (système de bonus-malus) ou ne pas rembourser la totalité des dommages (on parle de franchise). Une investisseuse peut demander à un entrepreneur d'engager des fonds propres dans le projet qu'elle finance, afin que la réussite du projet profite directement à l'entrepreneur. Une entreprise peut rémunérer ses cadres sous la forme d'actions de l'entreprise (les *stock options*) pour les pousser à travailler plus. De tels contrats peuvent parfois se révéler contre-productifs :

un trader payé en bonus en fonction de ses profits de fin d'année sera incité à prendre des risques inconsidérés si le portefeuille dont il a la responsabilité effectue des pertes en cours d'année.

Bibliographie

- Akerlof, G. A. (1978), The market for “lemons” : Quality uncertainty and the market mechanism, in ‘Uncertainty in economics’, Elsevier, pp. 235–251.
- Gore, J., Youk, H. and Van Oudenaarden, A. (2009), ‘Snowdrift game dynamics and facultative cheating in yeast’, *Nature* **459**(7244), 253–256.
- Palacios-Huerta, I. (2003), ‘Professionals play minimax’, *Review of Economic Studies* **70**(2), 395–415.
- Sinervo, B. and Lively, C. M. (1996), ‘The rock–paper–scissors game and the evolution of alternative male strategies’, *Nature* **380**(6571), 240–243.

Exercices du chapitre 7

Exercice 7.1. Meilleures réponses et équilibres de Nash

1. On considère le jeu suivant dit "bataille des sexes". Dans un couple, la première personne (joueur 1) souhaite aller assister à un concert de Bach (B), et la seconde (joueur 2) à un concert de Stravinsky (S). Chacun préfère se coordonner avec l'autre plutôt que d'aller à un concert seul. La matrice de paiements est la suivante :

	B	S
B	1 2	0 0
S	0 0	2 1

On note par p la probabilité que le joueur 1 met sur la stratégie B, et par q la probabilité que le joueur 2 met sur B.

- En fonction de la stratégie mixte (représentée par p) du joueur 1, quelles sont les meilleures réponses $BR^2(p)$ du joueur 2, i.e. l'ensemble des valeurs de q qui maximisent son paiement lorsque le joueur 1 utilise p ?
 - Idem, calculer l'ensemble des meilleures réponses $BR^1(q)$ du joueur 1 à une stratégie de 2 représentée par q .
 - Dans un graphique, représenter en abscisse les valeurs de p , et en ordonnée les valeurs de q . Comme les valeurs sont dans $[0, 1]$ on obtient un carré. Représenter le graphe de BR^1 , ainsi que celui de BR^2 .
 - Donner tous les équilibres de Nash du jeu de la bataille des sexes.
2. Mêmes questions pour le jeu suivant :

	Gauche	Droite
Haut	0 1	0 -1
Bas	0 -1	1 1

Exercice 7.2. Jeu de penalties On considère un jeu de pénalties simplifié dans lequel le tireur (joueur 1) peut tirer soit à gauche (G) soit à droite (D). Le gardien de buts peut plonger soit à gauche (G) soit à droite (D).

Si le gardien plonge du côté où la balle est tirée, le tir est arrêté à coup sûr. S'il plonge de l'autre côté, le ballon entre s'il est tiré de manière réussie. Le tireur est droitier, et pour lui il est plus facile de tirer à gauche (tir croisé) qu'à droite (il faut ouvrir le pied). On suppose que la probabilité de réussir un tir à gauche est $.9$, et $.6$ à droite.

Le but du tireur est de maximiser la probabilité de marquer, et pour le gardien il s'agit de minimiser cette probabilité.

Sachant que le tireur réussit plus souvent son tir à gauche qu'à droite, pensez-vous qu'il va tirer plus souvent à gauche, ou à droite?

1. En fonction des choix a^1 et a^2 , donner la matrice de paiements du joueur 1, où le paiement est la probabilité de marquer but.
2. Soit p la probabilité de tirer à gauche. Supposons que p est connu du gardien. Quelles sont en fonction de p les meilleures réponses du joueur 2? Quel est le gain espéré $v(p)$ du joueur 1 lorsqu'il utilise la stratégie p et que le joueur 2 utilise une stratégie de meilleure réponse?
3. Quelle est la stratégie p^* du joueur 1 qui maximise $v(p)$? Quelle est la valeur correspondante \underline{v} de $v(p^*)$? On dit que le joueur 1 peut se *garantir* \underline{v} . En effet, le joueur 1 a une stratégie p^* qui assure un gain espéré de \underline{v} , indépendamment de la stratégie de 2.
4. Soit q la probabilité pour le gardien de plonger à gauche. Supposons que q est connu du tireur. Quelles sont en fonction de q les meilleures réponses du joueur 1? Quel est le gain espéré $w(q)$ du joueur 1 lorsque le joueur 2 utilise la stratégie q et que le joueur 1 utilise une stratégie de meilleure réponse?
5. Quelle est la stratégie q^* du joueur 2 qui minimise $w(q)$, et que est le gain espéré correspondant $\bar{v} = w(q^*)$ du joueur 1? On dit que le joueur 1 peut *défendre* \bar{v} . En effet, pour toute stratégie de 2, le joueur 1 a une stratégie p qui lui assure un gain espéré de \bar{v} au moins.
6. Comparer les valeurs de \underline{v} et \bar{v} .
7. De manière générale, que peut-on dire des valeurs qu'il est possible de garantir ou de défendre, laquelle est la plus grande?
8. Finalement, le tireur tire-t-il plus souvent à gauche ou à droite selon les prédictions du modèle? Commenter.

Exercice 7.3. Compétition en prix de Bertrand

Joseph Louis François Bertrand (X 1839) lut l'ouvrage d'Augustin Cournot dans lequel il propose le modèle de compétition en quantités qui porte son nom, et proposa le modèle alternatif suivant, de compétition en prix, dit aussi compétition de Bertrand.

Deux producteurs ont la possibilité de produire un bien à un coût de c par unité produite. Il n'y a donc pas de coûts fixes et le coût est linéaire en la quantité produite. La fonction de demande du bien est donnée par $D(p)$ où D est une fonction continue positive.

Chaque producteur annonce un prix : p_1 pour le producteur 1 et p_2 pour le producteur 2. Ensuite, les consommateurs ont la possibilité d'acheter les quantités désirées soit à partir du producteur 1 soit à partir du producteur 2. Comme les consommateurs privilégient le prix le plus bas pour des biens substitués :

- Si $p_1 > p_2$, le producteur 2 vend $D(p_2)$ unités au prix p_2 , le producteur 1 ne vend rien,
- Si $p_2 > p_1$, le producteur 1 vend $D(p_1)$ unités au prix p_1 , le producteur 1 ne vend rien,
- Si $p_2 = p_1$, chaque producteur vend $D(p_1)/2$ unités au prix p_1 la demande se partageant en parts égales entre les deux producteurs.

1. Écrire les fonctions de gain du jeu : $g^i(p_1, p_2)$ pour chacun des producteurs.
2. Caractériser les équilibres de Nash.
3. Comparer avec l'équilibre de Nash du duopole de Cournot et commenter.

Exercice 7.4. Compétition spatiale de Hotelling

On considère une plage, bande de sable représentée par l'intervalle $[0, 1]$. Des vacanciers sont répartis également sur la plage (distribution uniforme). Deux marchands de glace sont en compétition pour vendre des glaces. Le marchand i décide de se positionner au point $x_i \in [0, 1]$. Chaque vacancier achète une glace à prix fixé auprès du marchand le plus proche de lui. En cas d'égalité de distance la répartition des vacanciers se fait à parts égales entre les deux.

1. En supposant que le gain d'un plagiste est égal à la taille de l'ensemble des vacanciers qui sont ses clients, écrire la fonction de gains du jeu.
2. Caractériser les équilibres de Nash du jeu.
3. Ce jeu de Hotelling a longtemps été utilisé comme modèle de compétition électorale. Voyez-vous pourquoi? La prédiction du modèle électoral vous semble-t-elle pertinente? Si non, quels éléments manquent-ils selon vous au modèle?

Exercice 7.5. Enchères au second prix

On considère un objet en vente aux enchères. Il y a n participants à l'enchère. La valeur de l'objet pour le participant i est notée v_i .

L'enchère se déroule de la manière suivante :

- 1) Chaque participant écrit une enchère b_i de manière secrète sur une feuille de papier scellée dans une enveloppe.
- 2) Toutes les enveloppes sont ouvertes simultanément par le commissaire priseur. L'enchérisseur qui a l'enchère la plus élevée remporte l'enchère (en cas d'égalité, on tire au hasard

parmi les enchères les plus élevées).

3) Le prix payé par le gagnant i^* est le *second prix*, c'est-à-dire :

$$\max_{j \neq i^*} b_j$$

1. Montrer qu'enchérir $b_i = v_i$ est une stratégie faiblement dominante pour l'enchérisseur i . On pourra représenter le gain à enchérir différentes valeurs de b_i dans 3 cas a) $\max_{j \neq i} b_j < v_i$ b) $\max_{j \neq i} b_j = v_i$ et c) $\max_{j \neq i} b_j > v_i$ et montrer qu'enchérir $b_i = v_i$ maximise le paiement dans tous les cas.
2. Supposons maintenant que les valeurs de v_i sont uniformément distribuées dans l'intervalle $[0, 1]$. En supposant que tous les enchérisseurs jouent leur stratégie faiblement dominante, calculer, en fonction de n , et de v_i , la probabilité que i gagne l'objet.
3. Calculer, en fonction de n , le montant espéré $R(v_i)$ payé par i lorsque sa valuation est v_i , la moyenne R_i du montant payé par i avant de connaître sa valuation, ainsi que la somme espérée R reçue par le vendeur.

Exercice 7.6. Pierre, Papier, Ciseaux Cet exercice concerne le jeu de « Pierre, Papier, Ciseaux » vu en cours.

1. Montrer que dans tout équilibre de Nash, le paiement du joueur 1 est 0.
2. Montrer que le seul équilibre de Nash est $(1/3, 1/3, 1/3)$ pour chaque joueur.

Exercice 7.7. Concours de beauté Le grand économiste britannique John Maynard Keynes remarquait qu'une grande partie du succès des acteurs économiques résidait non pas seulement à prédire le monde, mais à comprendre ce que les autres acteurs en pensent, et ainsi de suite. Il décrivait le monde comme un concours de beauté dans lequel chaque membre du jury chercherait à se rapprocher de l'opinion exprimée par les autres membres. Le jeu suivant est une variante du jeu du "concours de beauté", proposé d'abord par la revue française "Jeux et Stratégie".

Chaque participant parmi n joueurs choisit un nombre réel compris entre 0 et 100. Chacun paye ensuite la différence en valeur absolue entre son annonce et les $2/3$ de la moyenne des annonces. Si chaque joueur j annonce x_j , le paiement du joueur i est donc $-|\frac{1}{n} \sum_j x_j - x_i|$.

1. Quelles sont les stratégies strictement dominées ?
2. Si on élimine les stratégies dominées du jeu, quelles sont les stratégies dominées du jeu restant ?
3. Quels sont les équilibres de Nash du jeu ?

Chapitre 8

L'économie du travail

Pour produire biens et services, les entreprises ont besoin de travail, de capital (usines, machines, logiciels informatiques...) et de ressources naturelles. Le travail est donc un des principaux facteurs de production.

Dans une relation d'emploi, un travailleur offre son temps et son savoir-faire contre un salaire tandis qu'un employeur, en général une entreprise, demande ces heures de travail qu'il est prêt à rémunérer. On parle donc de « marché du travail », ce qui va nous permettre d'appliquer les principes développés précédemment pour décrire le fonctionnement du marché des biens et services.

Le salaire est la principale, voir l'unique, source de revenu de la plupart des ménages. Un dysfonctionnement du marché du travail est donc potentiellement bien plus grave qu'un dysfonctionnement du marché d'un bien comme l'automobile.

Nous allons commencer par étudier le cas d'un marché du travail parfaitement compétitif, avant de nous intéresser aux inégalités et au chômage. Ce sera l'occasion d'enrichir progressivement notre description du fonctionnement du marché du travail, en soulignant les frictions et les imperfections qui le caractérisent. Nous étudierons aussi les effets des principales politiques publiques qui régissent le fonctionnement du marché du travail : l'assurance chômage, le salaire minimum et la protection de l'emploi.

8.1 Un marché du travail compétitif

Commençons par analyser un marché du travail sans frictions, auquel nous pouvons appliquer les principes vus précédemment concernant l'offre et la demande. En revanche, les positions sont désormais inversées : les ménages sont du côté de l'offre et les entreprises du côté de la demande. La quantité échangée correspond à l'emploi, mesurée en nombre d'heures travaillées, tandis que le prix de cette transaction est donné par le salaire. Ainsi, l'emploi et le salaire vont être déterminés par l'équilibre du marché du travail.

8.1.1 L'offre de travail

La fonction d'offre de travail donne le nombre d'heures que les ménages souhaitent travailler en fonction du salaire auquel ce travail est rémunéré. Comment cette fonction est-elle déterminée ?

Chaque ménage, et chaque individu au sein d'un ménage, doit répartir son temps entre travail et temps libre, que l'on peut considérer comme étant consacré aux loisirs. Un travailleur prend sa décision d'offre de travail et de demande de temps libre de manière à maximiser sa satisfaction, son bien-être. Pour un niveau de salaire donné, il souhaite donc consommer du temps libre jusqu'à ce que le bénéfice d'une heure de loisirs supplémentaire soit égale au coût de cette heure. Ainsi, l'offre de travail égalise le bénéfice marginal du loisir au coût marginal du loisir.

Quel est le coût marginal du loisir ? On pourrait imaginer que le temps libre est gratuit. Ce serait oublier son **coût d'opportunité**. En prenant une heure de temps libre en plus, un salarié renonce à l'opportunité de travailler et de gagner une heure de rémunération supplémentaire : le coût marginal du loisir est égal au salaire. Un individu rationnel détermine donc son offre de travail de telle manière que le bénéfice marginal du loisir soit égal au salaire. Cela rappelle le marché des biens où un individu choisit sa consommation tel que le bénéfice marginal de la consommation, mesurée par la propension marginale à payer, soit égale au prix.

Le bénéfice marginal du loisir est pour sa part fondamentalement subjectif, et très différent d'un individu à l'autre. Il sera élevé pour les personnes qui n'aiment pas travailler ou qui savent profiter pleinement de leur temps libre. Au contraire, il sera faible pour les individus dont le travail est la raison d'être. Ceci étant, on peut raisonnablement supposer que *le bénéfice marginal du loisir est une fonction décroissante de la consommation de loisir*. Une personne appréciera davantage une journée de vacances supplémentaire si elle travaille jour et nuit que si elle travaille à temps partiel.

Cette hypothèse implique que l'offre de travail d'un individu est une fonction croissante du salaire. En effet, un salaire plus élevé correspond à une hausse du coût marginal du loisir. Face à un loisir plus coûteux, un individu choisit d'ajuster sa consommation de loisirs de manière à en augmenter le bénéfice marginal. Mais, pour augmenter le bénéfice marginal du loisir, il faut en consommer moins et, donc, travailler plus. Ce raisonnement peut sembler abstrait, mais il s'agit de la fondation microéconomique d'un comportement qui nous semble naturel : lorsque le salaire est plus élevé, on souhaite travailler davantage !

Effet de richesse. Notre raisonnement fait l'impasse sur une subtilité importante. Une augmentation de salaire a en réalité deux effets opposés sur l'offre de travail :

- Un **effet de substitution** : une hausse de salaire rend le travail plus attractif, ce qui en augmente l'offre. C'est l'effet que nous venons de voir en détail.
- Un **effet de richesse** : une hausse de salaire rend le travailleur plus riche, ce qui lui permet de consommer plus de biens et de services mais également plus de loisirs. Dans ce cas, la hausse de salaire diminue l'offre de travail.

En pratique, l'effet de substitution domine l'effet de richesse. Il est donc raisonnable de considérer que l'offre de travail est une fonction croissante du salaire, même si pour certaines personnes les deux effets s'annulent pratiquement si bien que leur offre de travail ne change pas avec le salaire. Ceci correspond à une offre parfaitement inélastique et une courbe d'offre verticale.

Marge extensive vs marge intensive. Jusqu'à présent, nous avons supposé le travail parfaitement divisible, permettant aux salariés de choisir librement le nombre d'heures de travail effectuées. En réalité, pour beaucoup de salariés, la décision est plutôt de travailler à plein temps ou de ne pas travailler du tout. Ainsi, on distingue deux dimensions de choix pour l'offre de travail :

- La **marge extensive** : il s'agit du choix de participer ou non au marché du travail.
- La **marge intensive** : il s'agit du choix du nombre d'heures de travail effectué conditionnellement au fait de participer.

La marge extensive implique, pour chaque travailleur, l'existence d'un salaire de réserve en deçà duquel il choisit de ne pas participer. Par conséquent, plus le salaire est élevé, plus le nombre de participants au marché du travail est important. À l'échelle de la population, même en prenant en compte la marge extensive, l'offre de travail reste donc une fonction croissante du salaire.

8.1.2 La demande de travail

Tentons maintenant de déterminer la demande de travail des entreprises en fonction du salaire. Les entreprises prennent leurs décisions de production, et donc d'embauches, de manière à maximiser leurs profits. Tant que l'embauche d'un salarié supplémentaire est profitable, l'entreprise cherche à augmenter sa masse salariale ; et tant que le licenciement d'un salarié supplémentaire est profitable, l'entreprise cherche à réduire sa masse salariale. A l'optimum, le revenu marginal du travail doit être égal au coût marginal du travail.

Le coût marginal du travail correspond simplement au salaire. Le revenu marginal est quant à lui la valeur de la production générée par l'embauche d'un travailleur supplémentaire : on parle de **productivité marginale du travail**. Pour calculer cette productivité marginale du travail, une entreprise doit connaître la valeur de sa production en fonction du nombre de travailleurs embauchés. Prenons l'exemple d'une entreprise dont les possibilités sont caractérisées par la table 8.1. Passer de 0 à 1 travailleur fait progresser la production de l'entreprise de 0€ à 300€, d'où une productivité marginale du premier travailleur égale à 300€. Lorsque l'entreprise embauche un second travailleur, la production passe de 300€ à 580€, soit un gain de 280€ : c'est la productivité marginale du deuxième travailleur. Et ainsi de suite. Ici, la productivité marginale est une fonction décroissante du nombre de travailleurs.

Supposons que le salaire journalier s'élève à 200€. Quelle est la demande de travail à ce niveau de salaire ? Si l'entreprise a déjà trois salariés, l'embauche d'un quatrième augmente ses revenus de 210€, soit la productivité marginale du quatrième travailleur. Cette

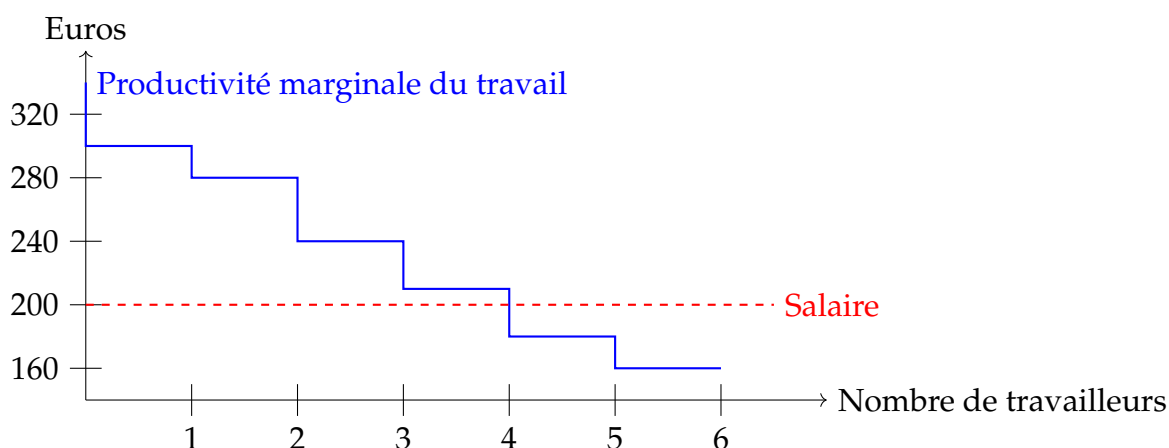
TABLE 8.1 – Production journalière d'une entreprise en fonction du nombre de travailleurs, et productivité marginale associée

Nombre de travailleurs	Production	Productivité marginale
1	300€	300€
2	580€	280€
3	820€	240€
4	1030€	210€
5	1210€	180€
6	1370€	160€

embauche augmente les coûts de l'entreprise du montant du salaire, soit 200€. Le profit journalier augmente donc de 10€, ce qui rend cette embauche profitable. A contrario, si l'entreprise a déjà cinq salariés et décide d'en licencier un, sa production baisse de 180€, soit la productivité marginale du cinquième salarié, mais elle économise 200€ de salaires. Le licenciement augmente donc le profit de 20€. Pour un salaire de 200€, la demande de travail s'élève à quatre salariés.

De manière générale, embaucher un travailleur est profitable lorsque le salaire est inférieur à la productivité marginale; tandis qu'un licenciement est profitable dès lors que le salaire est supérieur à la productivité marginale. A l'optimum, l'emploi égalise la productivité marginale du travail au salaire. La figure 8.1 est une représentation graphique de cette décision : le nombre de travailleurs demandés à l'équilibre se lit à l'intersection entre la droite horizontale du salaire et la courbe décroissante de la productivité marginale du travail.

FIGURE 8.1 – La demande de travail est déterminée par l'égalité entre le salaire (le coût marginal du travail) et la productivité marginale (le bénéfice marginal du travail)



Comme la productivité marginale est décroissante avec le nombre de travailleurs, la demande de travail d'une entreprise est une fonction décroissante du salaire : plus un salarié coûte cher, plus sa productivité marginale doit être élevée, moins le nombre de salariés est élevé. Cela correspond à ce que l'on voit sur la figure 8.1 où le nombre de travailleurs demandés augmente lorsque le salaire baisse.

La courbe de demande de travail d'une entreprise coïncide avec la courbe de producti-

tivité marginale du travail. Cela rappelle le marché des biens où nous avons vu que la courbe d'offre coïncide avec la courbe de coût marginal de production. Ces résultats parallèles résultent tous deux de la maximisation des profits par l'entreprise.

La courbe de demande de travail de l'économie est donnée par l'agrégation des courbes de demande de travail de toutes les entreprises du pays. Pour simplifier, nous supposons dorénavant que cette demande est une fonction continue du salaire.

Productivité marginale croissante. Nous avons supposé jusqu'à présent que la productivité marginale du travail est une fonction décroissante du nombre de travailleurs. Ceci étant, il est tout à fait concevable que la productivité marginale du travail soit initialement croissante. C'est par exemple le cas lorsque deux salariés ensemble produisent plus du double d'un seul salarié. Un boulanger peut difficilement faire son pain et tenir la boutique. Mais, s'il embauche un salarié, ils se répartiront le travail, rendant l'entreprise profitable. Ceci étant, au fur et à mesure que le nombre de salariés augmente, leur productivité marginale finit nécessairement par diminuer. Il s'agit d'une conséquence de la **loi des rendements décroissants** selon laquelle la productivité marginale d'un facteur de production donné (par exemple le travail ou le capital) finit par être décroissant.

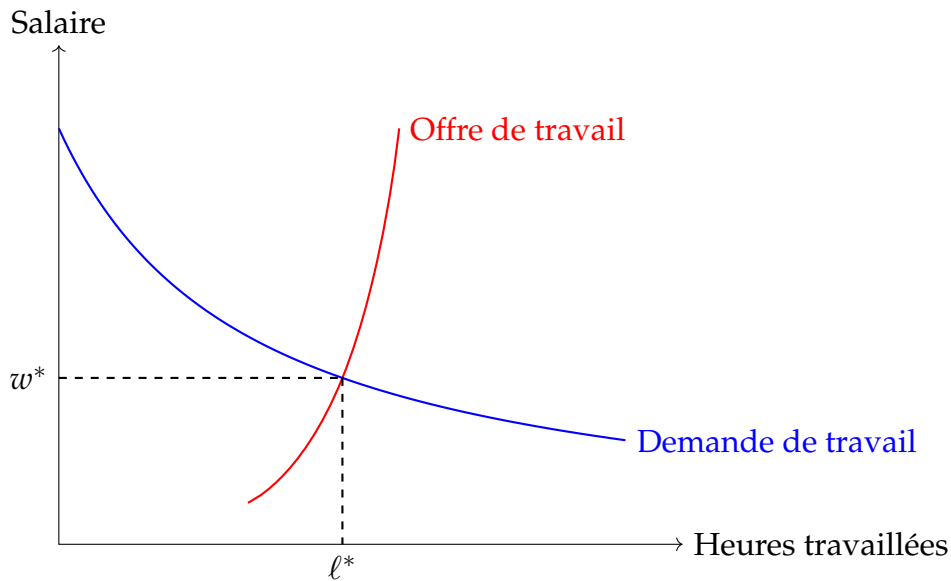
Par conséquent, il peut exister deux nombres de travailleurs tels que la productivité marginale du travail est égale au salaire. Néanmoins, une entreprise qui maximise ses profits ne choisira jamais un niveau d'embauche tel que la productivité marginale du travail y est croissante car, dans une telle situation, l'embauche d'un travailleur supplémentaire serait encore plus profitable que l'embauche du travailleur précédent. Le niveau de l'emploi est donc nécessairement déterminé sur l'intervalle où la productivité marginale du travail est décroissante. C'est exactement comme sur le marché des biens où il n'y a pas de perte de généralité à se focaliser sur la portion croissante de la fonction de coût marginal de production. Et, comme sur le marché des biens, il existe bien sûr la possibilité que les profits soient négatifs au point où la productivité marginale est égale au salaire, auquel cas l'entreprise choisit de ne pas embaucher de salariés.

8.1.3 Équilibre du marché du travail

Pour dériver les fonctions d'offre et de demande de travail, nous nous sommes posés la question suivante : à salaire fixé, quels sont les décisions prises par les travailleurs et par les entreprises. Mais comment est fixé le salaire ? Celui-ci s'ajuste de la même manière que les prix sur les marchés des biens. Lorsque la demande de travail des entreprises est supérieure à l'offre de travail des ménages, le salaire a tendance à augmenter ce qui réduit la demande de travail et augmente l'offre. À l'inverse, le salaire baisse lorsque la demande est inférieure à l'offre. À l'équilibre du marché du travail, le niveau du salaire et de l'emploi sont déterminés par l'intersection des courbes d'offre et de demande. Cet équilibre est représenté dans la figure 8.2. Le travail résulte donc d'échanges mutuellement bénéfiques entre les entreprises et les ménages.

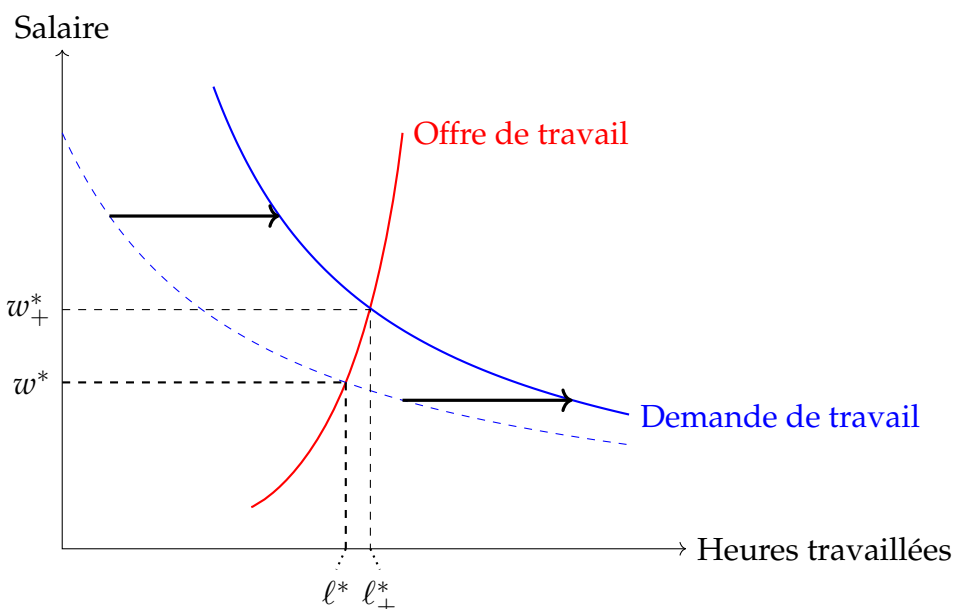
Dans un marché du travail compétitif, une hausse de la demande de travail des entre-

FIGURE 8.2 – Équilibre sur le marché du travail



prises augmente l'emploi et les salaires, comme on peut le voir sur la figure 8.3. Une réforme du marché du travail favorable aux entreprises, comme par exemple une simplification des procédures d'embauche et de licenciement, est de nature à rendre plus profitable l'embauche de salariés. Cela se traduit par une hausse de l'emploi et des salaires, ce qui bénéficie largement aux salariés. En cas de libre entrée, les profits des entreprises à l'équilibre sont toujours nuls et l'intégralité du bénéfice de la réforme revient donc aux salariés.

FIGURE 8.3 – Une hausse de la demande de travail exerce une pression à la hausse sur les salaires et sur l'emploi



Cette vision simplifiée d'un marché du travail parfaitement compétitif n'est pas totalement réaliste. Elle fait abstraction des multiples frictions qui, en pratique, enrayent cette belle mécanique. Chaque salarié étant unique et chaque entreprise étant unique, la rencontre entre l'offre et la demande prend généralement du temps. Il est donc difficile et coûteux pour

un salarié de changer d'employeur ou même pour un employeur de changer de salarié. Cette rigidité entrave la compétition, ce qui est source de conflits entre employeurs et employés, notamment sur la question des salaires.

Dans une telle situation, où est le marché du travail ? Où sont les forces de l'offre et de la demande ? À la marge ! Parmi les 100 salariés d'une entreprise, sans doute trois ou quatre d'entre eux envisagent de démissionner pour aller travailler ailleurs ; et parmi des milliers de chômeurs avoisinants peut-être que quelque uns d'entre eux souhaitent rejoindre cette entreprise. De même, l'employeur envisage peut-être de licencier deux ou trois salariés, ou au contraire d'en embaucher quelques autres. C'est là, à la marge, que se détermine le niveau de l'emploi et des salaires. Certes, il ne s'agit à court terme que des salaires des nouvelles recrues, mais ils finiront par avoir un impact sur l'ensemble des salariés de l'entreprise.

Le cas d'un marché du travail parfaitement compétitif est donc hautement simplifié, mais il permet de faire ressortir des forces incontournables. Même avec de multiples frictions venant entraver la compétition, si les entreprises veulent embaucher, les salaires et l'emploi vont avoir tendance à augmenter. Et si les ménages veulent réduire leur offre de travail, l'emploi va se contracter et les salaires vont augmenter. L'étude des imperfections du marché du travail est néanmoins essentielle pour parvenir à une analyse plus fine du marché du travail, et notamment pour comprendre le chômage.

8.1.4 Application empirique

Notre représentation du marché du travail permet de comprendre l'impact de divers chocs économiques sur l'emploi et les salaires. Imaginons par exemple que le système éducatif d'un pays s'améliore, rendant chaque travailleur plus efficace. Cela se traduit par une augmentation de la productivité marginale du travail. À salaire fixé, le nombre de travailleurs demandé par l'entreprise va donc augmenter : la courbe de demande de travail est déplacée vers la droite. À l'équilibre, l'emploi et les salaires vont donc augmenter.

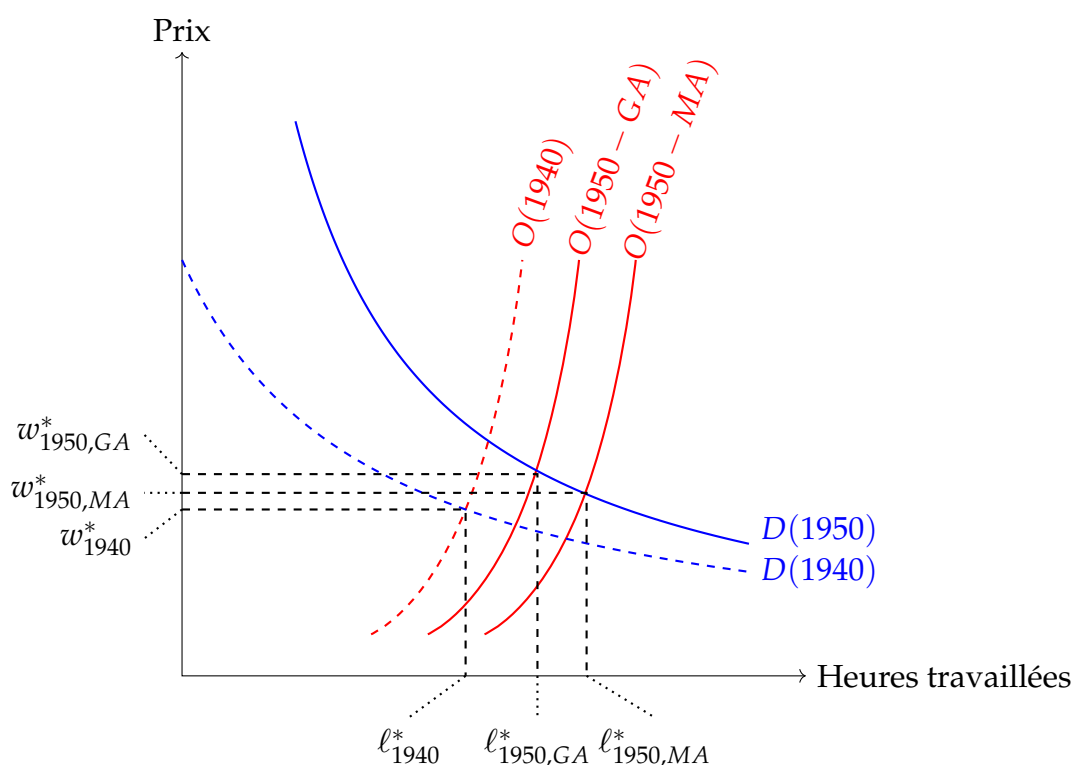
L'offre de travail change quant à elle lorsque les préférences des travailleurs changent. Le plus grand choc d'offre de travail du XX^{ème} siècle a probablement été l'entrée massive des femmes sur le marché du travail, qui peut être représenté graphiquement comme un déplacement de la courbe d'offre vers la droite.

Les économistes Daron Acemoglu, David Autor et David Lyle se sont intéressés à une manifestation de ce choc au moment de la Seconde Guerre Mondiale (Acemoglu et al., 2004). Aux États-Unis, la guerre a largement accéléré l'entrée des femmes sur le marché du travail : lorsque près de la moitié des hommes de 18 à 44 ans ont dû quitter le marché du travail pour aller combattre, les femmes les ont remplacés dans les usines et les champs. La guerre finie, les entreprises étaient habituées à employer des femmes, qui avaient elles-mêmes pris l'habitude de travailler. On estime qu'à peu près la moitié des femmes entrées sur le marché du travail à cause de la guerre y sont restées durablement. L'effort de mobilisation a été important dans tous les pays, mais on observe tout de même une variation dans les taux de mobilisation entre les différents États : la mobilisation a été la plus importante dans le Massachussets (environ 55% des hommes éligibles ont été mobilisés) et la plus faible en

Géorgie (où seuls 41% des hommes furent mobilisés).

Comment ces changements importants sur la période 1940-1950 se traduisent-ils en termes d'offre et de demande? Prenons le cas du Massachusetts et de la Géorgie. Dans ces deux États, la productivité a beaucoup augmenté pendant la guerre, déplaçant la courbe de demande de travail des entreprises vers la droite. La guerre a également eu un impact sur la courbe d'offre de travail des travailleuses, qui s'est déplacée également vers la droite : certaines femmes qui n'auraient pas travaillé si la guerre n'avait pas eu lieu ont pris des emplois et ont pris l'habitude de travailler, c'est-à-dire que leur offre de travail a durablement augmenté à cause de la guerre. Dans les États où un plus grand nombre d'hommes sont partis au front, la courbe d'offre s'est davantage déplacée vers la droite. Ces évolutions sont représentées dans la figure 8.4. On peut y voir que, à l'équilibre, le niveau de l'emploi a augmenté dans les deux États, et davantage dans le Massachusetts qu'en Géorgie. Les salaires ont eux aussi augmenté dans les deux États à cause des hausses de productivité, mais moins dans le Massachusetts qu'en Géorgie.

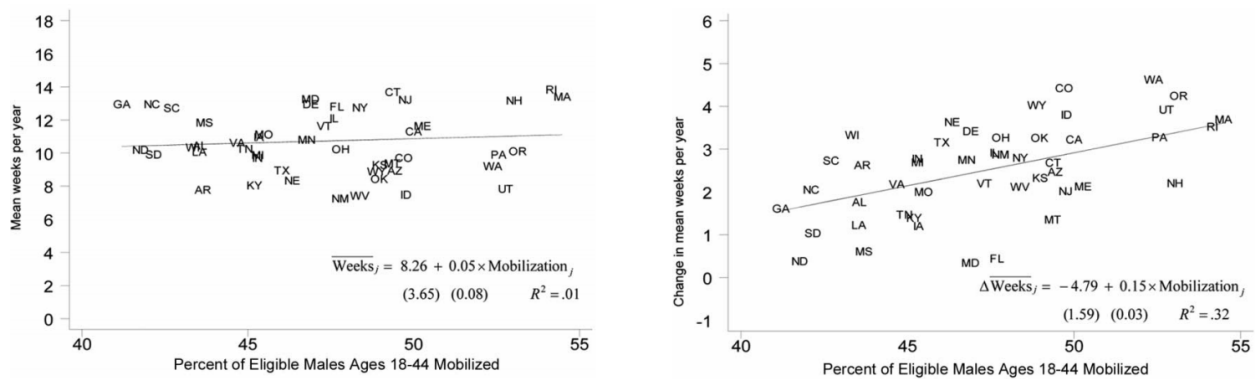
FIGURE 8.4 – Évolutions de l'offre et de la demande de travail sur la période 1940-1950. MA dénote le Massachusetts et GA la Géorgie



Ces prédictions théoriques sont-elles corroborées par les données dont nous disposons? Le graphique de gauche de la figure 8.5 montre une absence de lien entre taux de mobilisation des hommes et emploi féminin en 1940. Le graphique de droite montre que l'emploi féminin a davantage augmenté entre 1940 et 1950 dans les États qui ont plus fortement mobilisé. Cela confirme que la mobilisation des hommes en 1940 a durablement augmenté l'emploi féminin. D'après notre théorie, et conformément à la figure 8.4, les salaires ont donc dû augmenter davantage dans les États où le taux de mobilisation était faible. C'est exactement ce que l'on observe sur la figure 8.6, à gauche. Cela confirme donc notre analyse théorique :

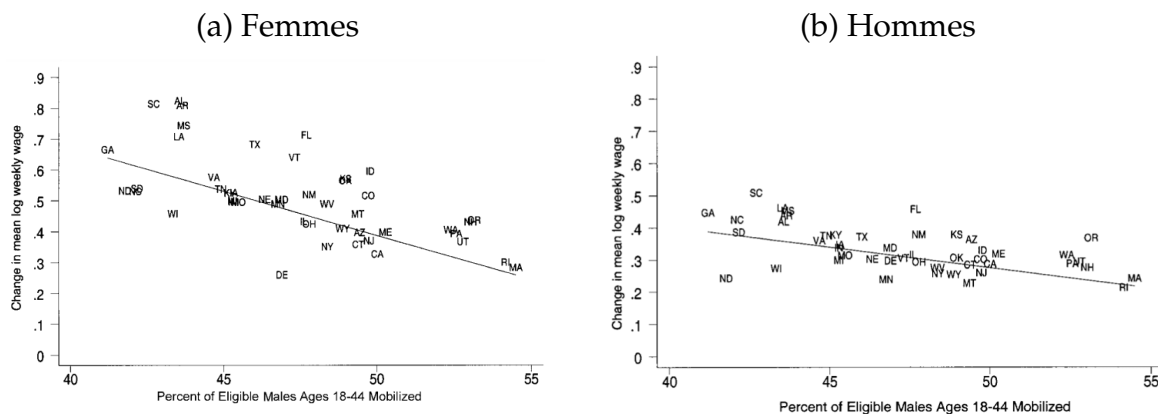
plus la mobilisation des hommes était importante, plus l'offre de travail des femmes a augmenté, moins les salaires des femmes n'ont progressé. Le graphique de droite montre le même effet du taux de mobilisation sur les salaires des hommes, mais avec une intensité moindre ce qui suggère que le marché du travail américain en 1950 était segmenté entre les hommes et les femmes : le salaire des femmes était particulièrement sensible à l'offre de travail des femmes. Cette étude confirme la prépondérance des forces de l'offre et de la demande de travail dans la détermination des salaires.

FIGURE 8.5 – A gauche, durée moyenne du travail des femmes par États en 1940 en fonction du taux de mobilisation des hommes. A droite, augmentation de la durée moyenne de travail des femmes par États entre 1940 et 1950 en fonction du taux de mobilisation des hommes.



Source : Acemoglu et al. (2004).

FIGURE 8.6 – Augmentation moyenne des salaires par État entre 1940 et 1950 en fonction du taux de mobilisation des hommes.



Source : Acemoglu et al. (2004).

8.2 Robots et emploi

La crainte de destructions massives d'emplois à cause de la robotisation ne date pas d'hier. Déjà, dans l'Angleterre du début du XIX^{ème} siècle, des ouvriers du textile appelés les

luddites détruisaient des métiers à tisser mécaniques. Plus récemment, ce sont les évolutions dans le domaine de l'intelligence artificielle qui ont relancé le débat sur l'avenir du travail. Que vont devenir les chauffeurs routiers lorsqu'un ordinateur sera capable de conduire à leur place? Même les métiers les plus qualifiés peuvent être menacés. Par exemple, l'analyse d'images médicales peut déjà être réalisée par des algorithmes qui permettent des diagnostics de grande qualité. Ceci étant, cette robotisation n'est pas forcément néfaste. Si des algorithmes vont pouvoir se substituer au travail humain, nous devrions disposer de plus de temps de loisir, avec peut-être des semaines de travail nettement plus courtes.

Tentons de réfléchir à la question de la robotisation en termes d'offre et de demande de travail. Lorsqu'une entreprise acquiert un robot, sa productivité augmente. En d'autres termes, elle est capable de produire plus pour n'importe quel nombre de travailleurs (la courbe de production totale est déplacée vers le haut). Mais quel est l'impact de cette robotisation sur la productivité *marginale* du travail?

Si robots et travailleurs sont *complémentaires*, comme par exemple lorsqu'un ordinateur permet à un salarié d'être plus performant, alors la robotisation augmente la productivité marginale du travail : la productivité d'une heure de travail supplémentaire est accrue par l'ordinateur. Cette situation est représentée sur la figure 8.7. Ici, la robotisation déplace la courbe de demande vers la droite.

A l'inverse, si robots et travailleurs sont *substituables*, comme lorsqu'une caisse automatique de supermarché remplace un caissier, alors la robotisation diminue la productivité marginale du travail, ce qui cause une contraction de la demande de travail, comme représenté sur la figure 8.8.¹

Dans les deux cas, la robotisation permet de produire plus, ce qui déplace la courbe de production totale vers le haut. La société est donc plus prospère grâce à l'automatisation. Mais dans le premier cas la productivité marginale de travail augmente, tandis que dans le second elle diminue. À l'équilibre du marché du travail, salaires et emploi augmentent donc lorsque travailleurs et robots sont complémentaires, mais diminuent lorsqu'ils sont substituables.

De manière plus générale, l'automatisation a trois effets sur l'emploi :

1. Elle crée des emplois complémentaires à la robotisation. C'est le cas des ingénieurs qui conçoivent les caisses automatiques des supermarchés, ainsi que des employés qui les font fonctionner sur place.
2. Elle détruit des emplois substituables à la robotisation. C'est le cas des caissiers que la machine remplace.
3. Enfin, elle crée des emplois dans d'autres secteurs de l'économie. La robotisation augmente la production et donc le niveau moyen de prospérité des consommateurs, ce qui génère de nouvelles demandes favorisant le développement d'autres secteurs d'activité. Il aurait sans doute été difficile d'imaginer à l'aube de la révolution industrielle

1. La productivité marginale du travail $f'(l)$ est la dérivée de la fonction de production $f(l)$. Par conséquent, comme on peut le voir sur les figures 8.7 et 8.8, la robotisation augmente le pente de la fonction de production si et seulement si la productivité marginale augmente.

FIGURE 8.7 – Impact d'un choc de robotisation sur le productivité marginale et sur la production totale lorsque robots et travailleurs sont complémentaires.

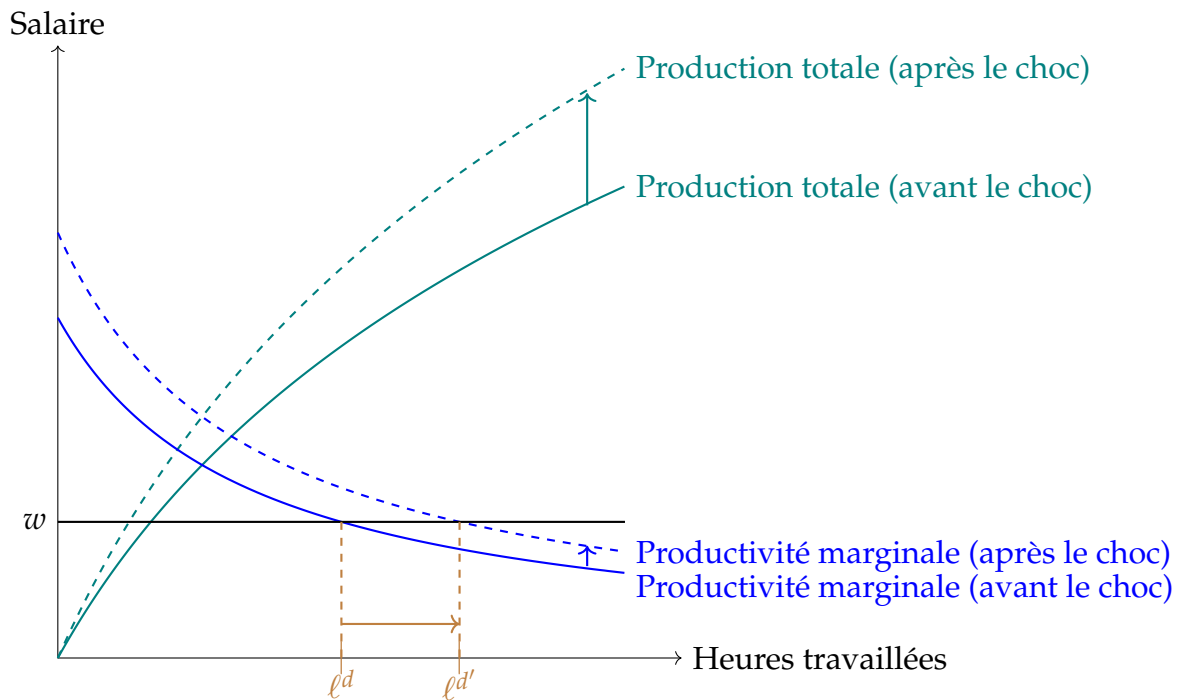
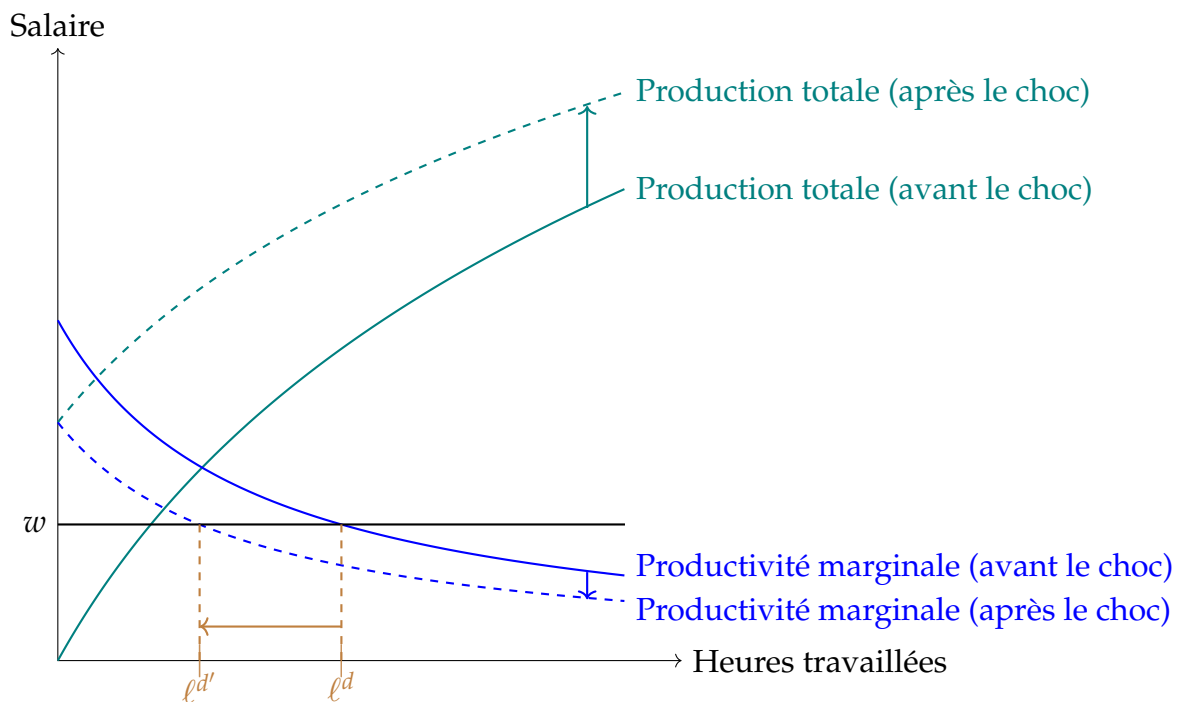


FIGURE 8.8 – Impact d'un choc de robotisation sur le productivité marginale et sur la production totale lorsque robots et travailleurs sont substituables.



le développement de métiers tels que professeur de yoga, psychologue, designer ou conseiller en communication.

Soulignons que cette dernière possibilité est un effet d'équilibre général. Ne se produisant pas là où se réalise l'automatisation, cet effet est moins visible et moins facile à quantifier que les deux premiers, mais il n'est pas moins important !

Acemoglu and Restrepo (2019) ont montré que les deux premiers effets ont été d'une intensité comparable de la fin de la Seconde Guerre Mondiale jusqu'aux années 1980 : au sein d'un secteur d'activité, la robotisation créait alors autant d'emplois qu'elle en détruisait. Néanmoins, depuis la révolution informatique amorcée dans les années 80, il semblerait que le second effet domine le premier, si bien que l'automatisation réduit la demande de travail. Il reste donc à savoir si la robotisation génère suffisamment de gains de productivité pour, grâce au troisième effet, empêcher une contraction de la demande de travail qui se traduirait par une diminution des salaires. Cette question fait toujours l'objet de recherches.

La fin du travail ?

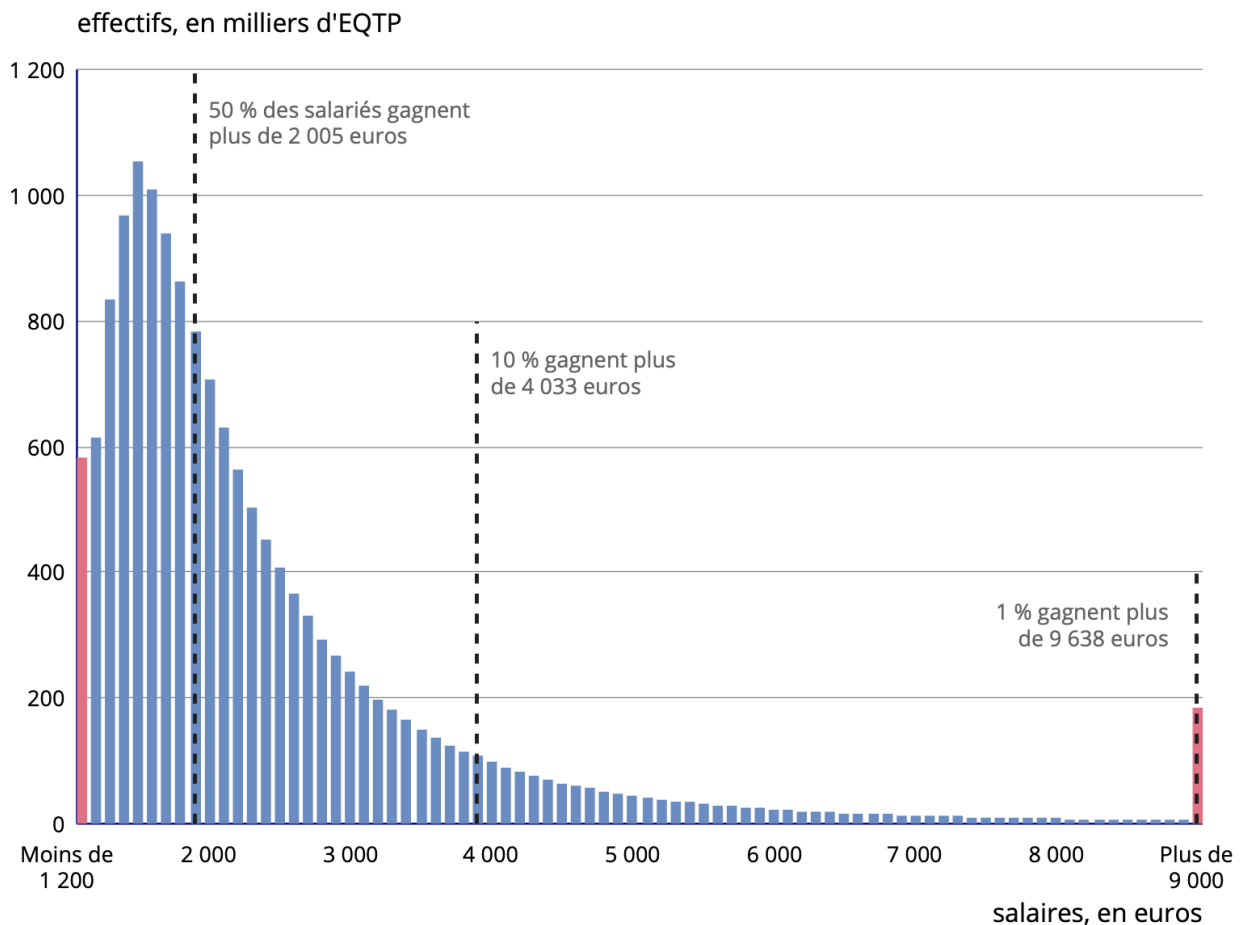
En 1932, l'économiste britannique John Maynard Keynes avait publié un essai intitulé « Perspectives économiques pour nos petits-enfants ». Dans ce texte, il fait plusieurs prédictions pour le siècle à venir. Il annonce que « le niveau de vie [dans cent ans] sera entre quatre et huit fois plus élevé qu'aujourd'hui ». Si l'on prend comme mesure du niveau de vie le PIB par habitant (ajusté à l'inflation), la prédiction de Keynes s'est réalisée : depuis qu'il a écrit cet essai il y a quatre-vingt dix ans, le niveau de vie au Royaume-Uni été multiplié par plus de cinq malgré les importantes destructions de la Seconde Guerre Mondiale. Sa seconde prédiction concerne les conséquences de cette importante croissance : selon Keynes, le progrès économique allait mener à une réduction progressive du temps passé à travailler. La production des denrées essentielles nécessitant beaucoup moins de main d'œuvre qu'auparavant, le temps autrefois utilisé pour produire de la nourriture, ou construire des infrastructures serait désormais consacré au loisir. Keynes écrit : « nous allons faire des efforts pour [répartir] autant que possible le travail qu'il restera à faire. Des journées de travail de trois heures ou des semaines de quinze heures pourraient écarter le problème pour un bon moment ». Une chose certaine est que Keynes n'a pas appliqué à sa propre vie les conseils prodigués aux générations futures : c'est en partie à cause de son acharnement au travail qu'il a quitté prématurément notre monde en 1946, à l'âge de 62 ans. Par ailleurs, les choses n'ont pas évolué après sa mort comme il le pensait. Malgré les gains importants de productivité, le nombre d'heures consacrées aux loisirs n'a presque pas changé au cours du XX^{ème} siècle. Peut-être que le travail est devenu progressivement plus plaisant, s'apparentant pour certains à un loisir.

8.3 Inégalités de revenu et redistribution

Jusqu'à présent, nous avons supposé que tous les travailleurs étaient identiques, aboutissant à un unique salaire d'équilibre sur le marché du travail. En réalité, comme on peut le constater sur la figure 8.9, il existe d'énormes dispersions de salaires. En France, la moitié des salariés sont rémunérés moins de 2 000 euros nets par mois, tandis que 10% gagnent plus de 4 000 euros, et un peu moins de 1% plus de 10 000 euros. Tentons de comprendre

cette dispersion des salaires.

FIGURE 8.9 – Distribution des salaires mensuels nets (en équivalent temps plein) en France en 2020.



Source : Insee, base Tous salariés 2020.

8.3.1 Les sources d'inégalités de salaires

Les inégalités de salaires résultent essentiellement de différences de productivité. Un travailleur dont la productivité marginale est deux fois plus élevée qu'un autre aura un salaire deux fois plus élevé. Mais quels sont les causes de ces écarts de productivité ? Il existe de multiples sources d'hétérogénéité entre individus, dont les plus importantes sont :

- La capacité intrinsèque : certains travailleurs sont plus intelligents, plus créatifs, plus habiles, ou encore plus aptes à travailler en équipe que d'autres.
- L'investissement dans le **capital humain** : les médecins ou les ingénieurs ont des capacités élevées acquises à l'issue de longues années de formation.
- Le hasard : certains travailleurs se trouvent avoir des compétences demandées par des secteurs en expansion, là où d'autres compétences sont au contraire de moins en moins valorisées du fait des innovations technologiques.

Les écarts de productivité entre individus sont considérables. Mais ils ne constituent pas l'unique source d'inégalités de salaire. Le marché du travail est caractérisé par de multiples frictions qui peuvent aboutir à des écarts entre salaires et productivité marginale du travail, notamment à court terme. Par exemple, il existe une asymétrie d'information entre les travailleurs, qui connaissent à peu près leur propre productivité, et les entreprises, qui ne l'observent qu'au bout d'un certain temps. En outre, les travailleurs et les entreprises n'étant pas parfaitement homogènes, la rencontre entre l'offre et la demande de travail ne se réalise pas instantanément. Cela génère du chômage mais également des écarts de salaire entre travailleurs identiques. Les estimations empiriques montrent néanmoins qu'à moyen et long terme les salaires reflètent essentiellement les productivités individuelles.

Productivité marginale du travail : nombre de travailleurs vs. productivité individuelle.

Nous avons vu que :

- La productivité marginale du travail est une fonction décroissante du nombre de travailleurs employés par l'entreprise ;
- Certains travailleurs ont une productivité marginale du travail plus élevée que d'autres (indépendamment du niveau de l'emploi au sein de l'entreprise).

Comment concilier ces deux dimensions de la productivité marginale du travail ?

Une représentation mathématique du problème permet de clarifier les choses. Supposons qu'il existe N types de travailleurs différents, indicés par $i \in \{1, 2, \dots, N\}$. Chaque type i est caractérisé par une efficacité productive p_i . Soit L_i la quantité de travail (en unités de temps) de type i employée par une entreprise. Au total, la quantité effective de travail employée par l'entreprise est égale à :

$$\sum_{i=1}^N p_i L_i.$$

Soit $F\left(\sum_{i=1}^N p_i L_i\right)$ la fonction de production de l'entreprise. On suppose $F'(\cdot) > 0$ et $F''(\cdot) < 0$. La seconde hypothèse correspond à la productivité marginale décroissante du travail à l'échelle de l'entreprise.

L'entreprise détermine sa demande de travail de manière à maximiser ses profits :

$$\max_{\{L_i\}_{i=1}^N} F\left(\sum_{i=1}^N p_i L_i\right) - \sum_{i=1}^N w_i L_i,$$

où w_i est le salaire (par unité de temps) des travailleurs de type i . A l'optimum, on a :

$$w_i = p_i F'\left(\sum_{j=1}^N p_j L_j\right),$$

ce qui montre que le salaire d'un travailleur i est égal à sa propre productivité marginale. Ainsi, nous avons bien simultanément les deux propriétés suivantes :

- la productivité marginale du travail est une fonction décroissante de la quantité de travail employée $\sum_{i=1}^N p_i L_i$ par l'entreprise ;

- les travailleurs ont des salaires proportionnels à leur propre efficacité productive p_i .

Soulignons que, dans le cas général, la fonction de production s'écrit $F(L_1, L_2, \dots, L_N)$. La fonction $F(\sum_{i=1}^N p_i L_i)$ correspond à un cas particulier où les différents types de travailleurs sont parfaitement substituables, à un facteur d'efficacité productive près. Ce cas est intéressant car il implique que la productivité relative d'un travailleur de type i par rapport à un travailleur de type j est déterminée par un unique paramètre p_i/p_j . Cela permet de structurer notre réflexion sur la détermination des salaires, en séparant l'impact du nombre de travailleurs employés de celui de la productivité individuelle.

Par ailleurs, cette spécification de la fonction de production nous permet de déterminer les salaires relatifs à partir de la demande de travail des entreprises. En revanche, le niveau des salaires ne peut pas être déterminé sans connaître la fonction d'offre de travail de tous les types de travailleurs.

8.3.2 Les politiques de redistribution

Face à ces inégalités, des politiques de redistribution permettent de réaliser des transferts de revenus des plus riches vers les plus pauvres. Quel est le bon niveau de redistribution? Plus fondamentalement, quel est le juste niveau des inégalités? Comme décrit précédemment, ces questions ne relèvent pas spécifiquement des sciences économiques mais de la philosophie morale et politique. D'Aristote à John Rawls, de nombreux philosophes ont développé des conceptions différentes de la juste répartition des richesses. Fondamentalement, deux visions s'opposent. La première, d'inspiration libérale, considère que chacun a un droit naturel au fruit de son travail. Par conséquent, l'État doit se contenter de corriger les défaillances de marché, sans interférer avec la répartition des richesses dans l'économie. A contrario, on peut considérer, selon une vision plus égalitariste, qu'un euro apporte plus de bonheur à un pauvre qu'à un riche. En ce cas, la redistribution améliore le bien-être social.

L'impossibilité fondamentale de mesurer et, surtout, de comparer le bonheur des uns et des autres explique pourquoi ces questions sont de nature philosophique. Ceci étant, si vous adhérez à la vision égalitariste, alors votre conception de la justice sociale permet de définir une fonction de bien-être social que le gouvernement doit chercher à maximiser. L'analyse économique consiste ensuite à caractériser la politique de redistribution qui correspond au meilleur compromis entre redistribution vers les plus pauvres et préservation des incitations à travailler et à créer des richesses. Cette question est au coeur d'une branche entière de notre discipline : la taxation optimale.

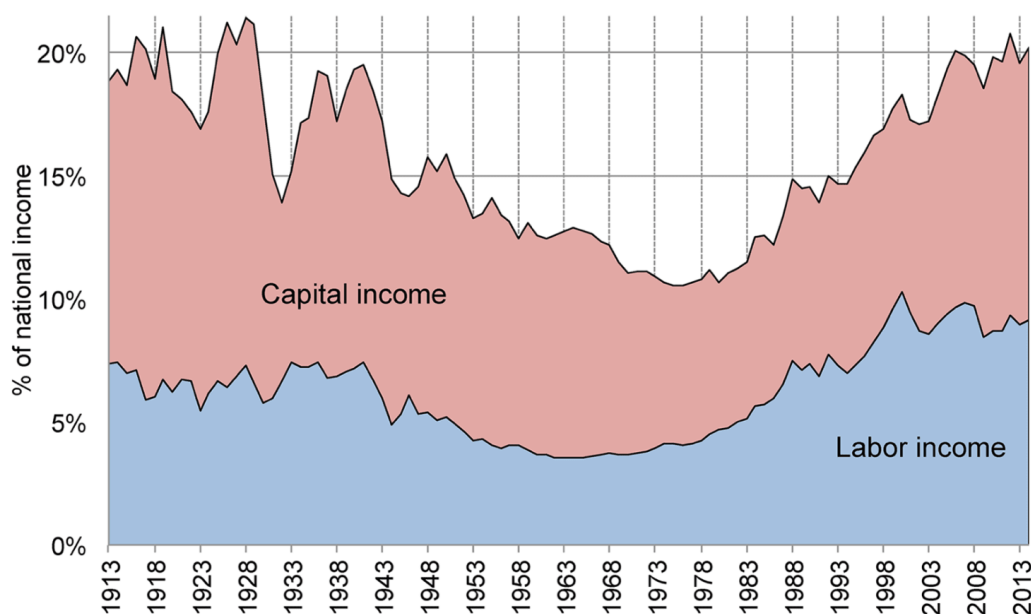
En 1971, l'économiste écossais James Mirrlees fut le premier à caractériser la forme optimale de l'impôt (non-linéaire) sur les revenus du travail, ce qui lui valut le prix Nobel en 1996. Il a trouvé que, même avec un intense désir de redistribution, la contrainte d'incitation limite fortement la progressivité de l'impôt sur le revenu. Cela est dû au fait qu'il n'est jamais souhaitable de fixer un taux d'imposition supérieur au sommet de la courbe de Dupuit-Laffer que nous avons vu au chapitre 5.3.3. En effet, si l'impôt est si progressif qu'il dépasse ce seuil, il est alors possible d'augmenter les recettes fiscales en abaissant le taux d'imposition sur les plus riches, ce qui génère une amélioration de Pareto. Le critère

de Pareto implique que l'objectif ne doit pas être de minimiser les inégalités en soi, mais d'augmenter les revenus des plus pauvres.

8.3.3 La dynamique des inégalités

Les travaux de Thomas Piketty, d'Emmanuel Saez et de Gabriel Zucman ont montré la très forte augmentation des inégalités aux Etats-Unis depuis le début des années 1980 (Piketty et al., 2018). Les part des revenus revenant au "top 1%", à savoir le 1% des ménages dont les revenus sont les plus élevés, a quasiment doublé en 40 ans, comme on peut le constater sur la figure 8.10. Aujourd'hui, près de 20% des revenus reviennent aux 1% les plus aisés. Cette augmentation des inégalités est largement due à une augmentation des inégalités des revenus du travail. Cette évolution résulte notamment des fortes rémunérations dans la finance ou au sein des directions des très grandes entreprises. Soulignons que l'accroissement de la part du « top 1% » masque une certaine mobilité sociale qui implique que ce ne sont pas toujours les mêmes ménages qui appartiennent au « top 1% » année après année.

FIGURE 8.10 – Part des revenus nationaux revenant au « top 1% » aux Etats-Unis (avant impôts).

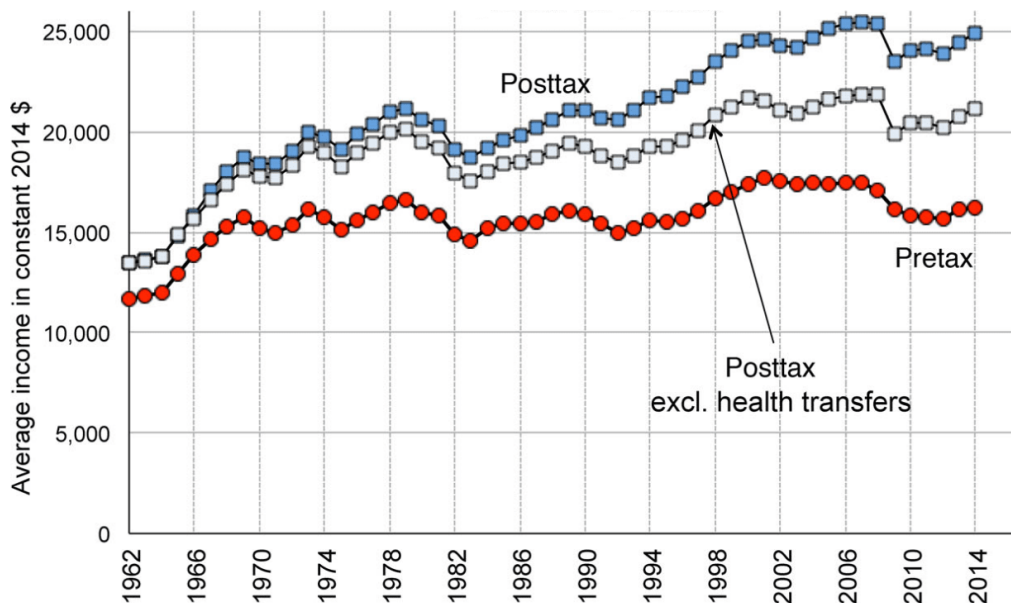


Source : Piketty et al. (2018).

La contrepartie de la très forte augmentation des hauts revenus est la quasi-stagnation des bas revenus. La figure 8.11 montre que la part des revenus capté par le « bottom 50% », soit la moitié des ménages les moins bien rémunérés, a quasiment stagné aux Etats-Unis au cours des 50 dernières années, malgré une croissance vigoureuse de l'économie américaine. Cette stagnation a été partiellement compensée par des mesures de redistribution, notamment au travers du système de santé.

En France, comme on peut le voir sur la figure 8.12, les inégalités n'ont quasiment pas évolué au cours de cette même période (Garbinti et al., 2018). L'augmentation des inégalités est donc un phénomène prépondérant aux États-Unis et, dans une moindre mesure, au

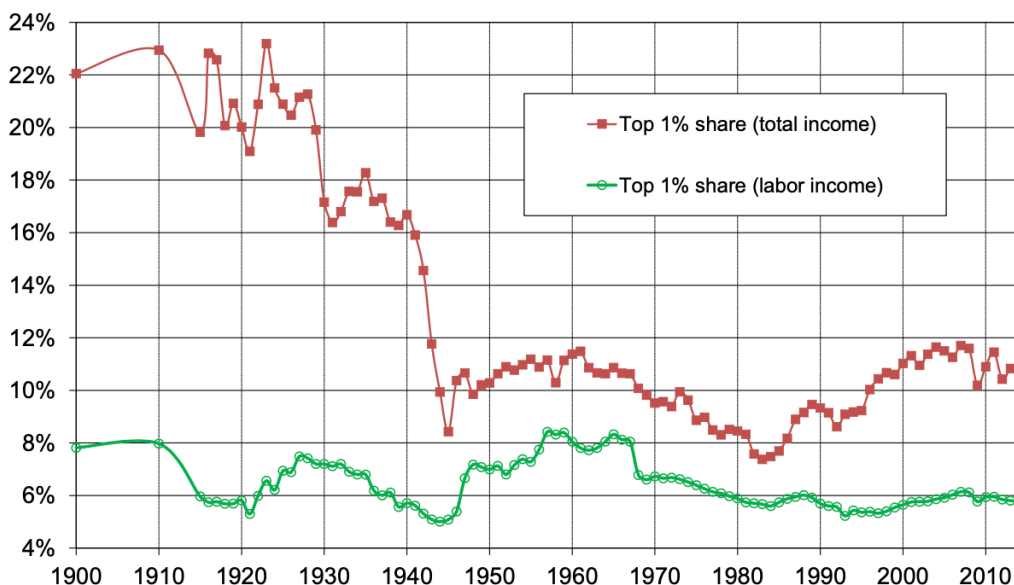
FIGURE 8.11 – Niveau moyen des revenus du « bottom 50% » aux Etats-Unis.



Source : Piketty et al. (2018).

Royaume-Uni. En revanche, la France, l'Italie et le Japon ont échappé à cette tendance. L'Allemagne est un cas intermédiaire, caractérisé par une hausse modérée des inégalités. Ces divergences entre pays sont l'objet de recherches actives en sciences économiques.

FIGURE 8.12 – Part des revenus nationaux revenant au « top 1% » en France (avant impôts).



Source : Garbinti et al. (2018).

8.3.4 Les inégalités de salaires entre les hommes et les femmes

En France, en 2018, dans le secteur privé, les salaires horaires des femmes étaient en moyenne 16,8% inférieur à celui des hommes. Un écart similaire est présent dans la plupart des pays riches et industrialisés. Comment peut-on expliquer un écart aussi massif ?

Cette question a généré une littérature empirique considérable. Le point de départ à consisté à comparer les salaires des femmes et des hommes à niveaux d'éducation, d'expérience, à secteurs industriels et à postes comparables. Comme le souligne Marianne Bertrand dans sa revue de la littérature (Bertrand, 2020), ces études ont fait ressortir deux explications principales à cet écart de salaire.

Premièrement, si les femmes font des études légèrement plus longues que les hommes, elles ont tendance à se spécialiser dans des disciplines moins bien rémunérées. Les femmes sont largement sous-représentées dans les disciplines scientifiques et dans les formations d'ingénieurs, tandis qu'elles sont sur-représentées dans les formations plus littéraires. Non seulement, les métiers d'ingénieurs sont mieux rémunérés que la moyenne, mais ils sont en plus caractérisés par des écarts de salaires quasiment nuls entre les hommes et les femmes. Une augmentation du nombre de femmes ingénieurs serait donc de nature à sérieusement réduire l'écart de salaires moyen à l'échelle de l'ensemble de l'économie.

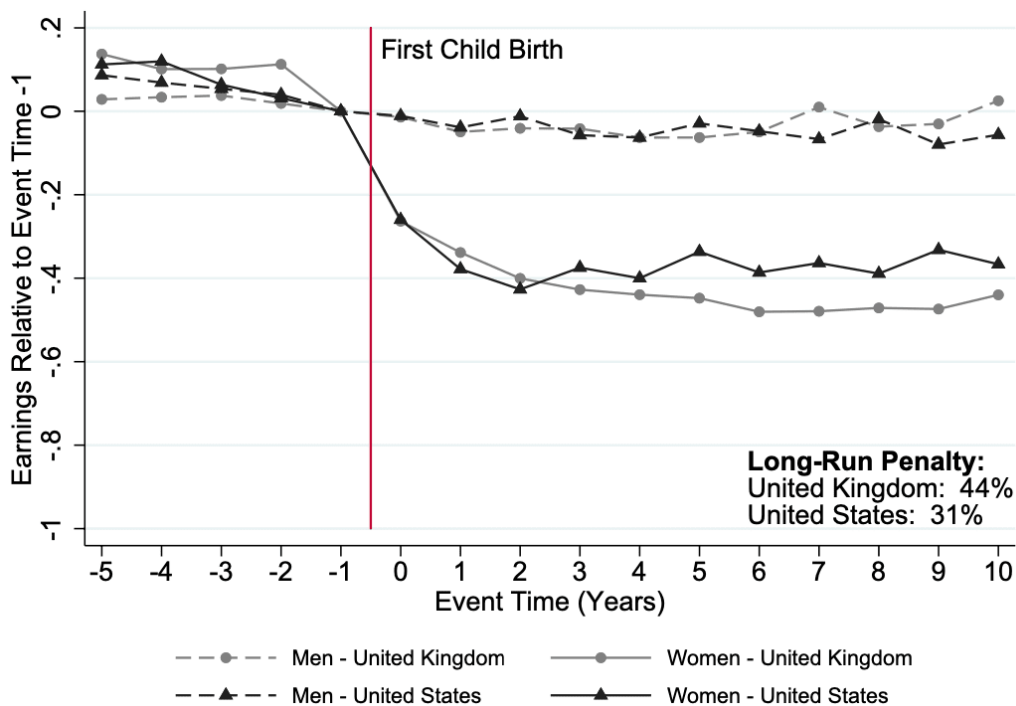
L'autre explication, encore plus importante, est le choc salarial subit par les femmes lors de l'arrivée de leur premier enfant, ce que la littérature a appelé le « *child penalty* » ou la « *motherhood penalty* ». Comme on peut le voir sur la figure 8.13 réalisée par Kleven et al. (2019), les hommes et les femmes ont des trajectoires presque identiques en terme de rémunération jusqu'à l'arrivée du premier enfant. A partir de ce moment, les mères décrochent et accumulent un retard de rémunération qu'elles ne combleront pratiquement jamais. Suite à l'arrivée du premier enfant, les femmes « choisissent » bien plus souvent que les hommes de sacrifier leur vie professionnelle pour leur vie familiale. Cela se traduit typiquement par un temps de travail réduit, voir par un emploi à temps partiel.

Les choix de spécialisation pour les études supérieures ou le choix de privilégier la vie familiale sur la vie professionnelle sont contraints par des normes sociales en vigueur et notamment par la répartition des tâches entre hommes et femmes au sein des foyers. Ces normes ne sont heureusement pas immuables, comme le suggère le fait que la « *child penalty* » est nettement plus faible dans les pays scandinaves qu'ailleurs en Europe ou aux Etats-Unis (Kleven et al., 2019).

8.4 Le chômage

Le chômage a un impact majeur sur le bien-être des individus, et peut engendrer une grave détresse sociale et psychologique. L'étude réalisée en 2008 par Andrew Clark, Ed Diener, Yannis Georgellis et Richard Lucas met particulièrement bien en évidence cet impact ravageur du chômage (Clark et al., 2008). Ils ont exploité des sondages réalisés auprès de milliers d'allemands interrogés plusieurs années consécutives. Dans ces sondages, des questions leur étaient posées à propos de leur situation professionnelle et personnelle. On leur demandait également de donner une évaluation de leur satisfaction dans la vie, sur une échelle de 0 à 10. Ces données permettent de quantifier l'impact de certains événements (mariage, divorce, naissance d'un enfant) sur cette mesure subjective du bonheur. Comme le montre la figure 8.14, les résultats sont frappants : perdre son emploi engendre une perte de bonheur presque semblable à celle de la perte d'un époux. Pire, alors que la tristesse en-

FIGURE 8.13 – Evolution du salaire des hommes et des femmes aux Etats-Unis et au Royaume-Uni avant et après la naissance du premier enfant, après avoir éliminé l'impact sur les salaires de l'expérience (mesurée par l'âge du salarié) et de l'année à laquelle on observe le salaire (pour éliminer les effets de conjoncture économique).



Source : Kleven et al. (2019).

gendrée par le veuvage s'atténue vite, le malheur est persistant dans le cas du chômage, en particulier pour les hommes.

8.4.1 Chômage et participation au marché du travail

Avant de comprendre pourquoi certaines personnes sont au chômage, commençons par définir les principaux indicateurs qui caractérisent le marché du travail. Chaque individu est classé dans une des trois catégories suivantes :

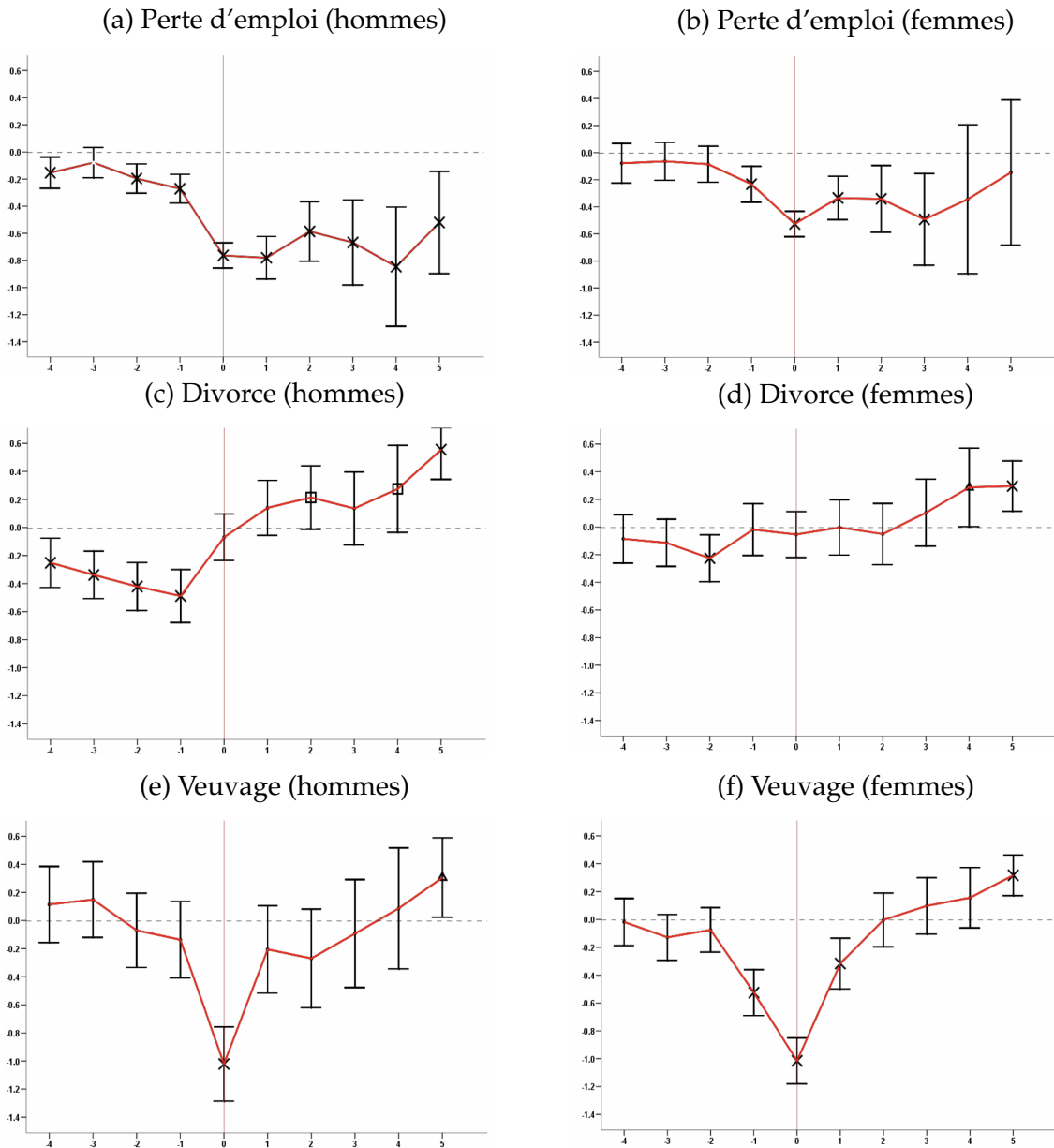
- **Employé**, lorsqu'il occupe un emploi à temps plein ou partiel ;
- **Chômeur**, lorsqu'il ne travaille pas, mais recherche activement un emploi ;
- **Inactif**, lorsqu'il n'est ni employé, ni chômeur.

La **population active** est constituée à la fois des employés et des chômeurs. Cette classification permet de définir deux statistiques essentielles pour comprendre le marché du travail. D'abord, le **taux de chômage** correspond à la part des actifs qui sont au chômage, soit

$$\text{Taux de chômage} = \frac{\text{Nombre de chômeurs}}{\text{Population active}}.$$

La figure 8.15 montre qu'aux Etats-Unis le chômage est d'abord et avant tout un phénomène cyclique, augmentant rapidement lors des périodes de récession et revenant régulièrement

FIGURE 8.14 – Impact de différents événements sur la satisfaction dans la vie. En abscisse, le temps en années, mesuré de manière relative à l'événement. En ordonnée, le niveau moyen de satisfaction (avec la même échelle pour les six graphiques). La ligne verticale représente l'année de l'événement.



Source : Clark et al. (2008).

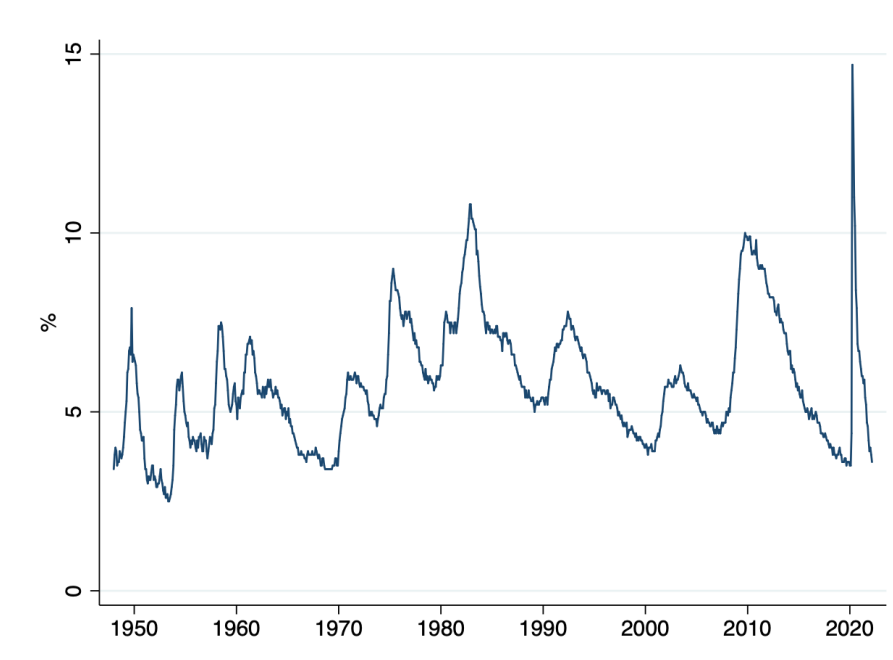
sous la barre des 5% dès lors que la conjoncture est favorable. Le chômage en France, représenté sur la figure 8.16, est nettement moins volatil et son niveau est durablement plus élevé. Le chômage y est donc un phénomène structurel dû au fonctionnement, ou plutôt au dysfonctionnement, du marché du travail.

Moins souvent évoqué mais tout aussi important, le **taux de participation** correspond à la part de la population adulte qui est active, soit

$$\text{Taux de participation} = \frac{\text{Population active}}{\text{Population adulte}}.$$

Notez que ce taux de participation n'inclut pas seulement les employés, mais également les chômeurs qui cherchent activement un emploi et qui participent ainsi au marché du travail. La figure 8.17 montre que, tout au long de la seconde moitié du XX^{ème} siècle, le taux de participation a fortement augmenté aux États-Unis, progressant au total de près de 8%. Cette évolution est le fruit de l'entrée massive des femmes sur le marché du travail. Le taux de participation a baissé depuis 2008, notamment parce que la sévérité de la crise a découragé de nombreux chômeurs, qui sont donc devenus inactifs. En outre, les seniors ayant un taux de participation plus faible que la moyenne, le vieillissement démographique a également contribué à cette baisse du taux de participation américain. En France, le taux de participation est resté à peu près constant depuis les années 70. La forte augmentation de la participation des femmes au marché du travail y a été compensée par l'avancement de l'âge de la retraite et par l'allongement des études supérieures.

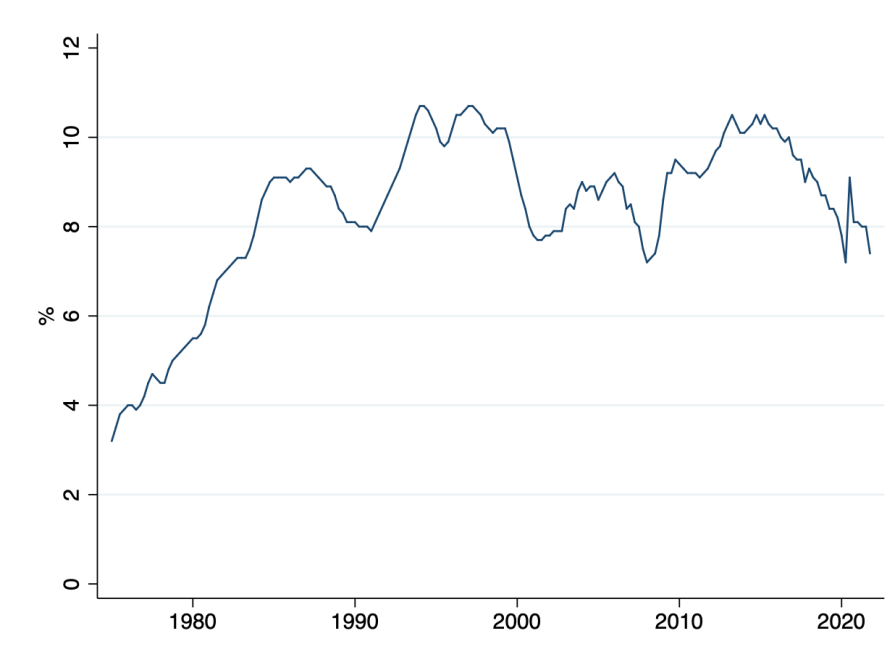
FIGURE 8.15 – Taux de chômage aux États-Unis.



Source : U.S. Bureau of Labor Statistics.

Il est également intéressant d'observer les tendances du nombre d'heures travaillées. Là aussi les divergences sont profondes entre les États-Unis et l'Europe. Comme le montre la figure 8.18, les américains n'ont pratiquement pas diminué leur temps de travail depuis

FIGURE 8.16 – Taux de chômage en France.



Source : INSEE.

1950, tandis que les français ont réduit d'un tiers le nombre d'heures passées à travailler. L'Allemagne a suivi une évolution pratiquement identique à celle de la France.

Ces évolutions expliquent l'écart de PIB par habitant entre les Etats-Unis et l'Europe. Depuis les années 1950, les américains ont un niveau de vie à peu près 30% plus élevé que les Européens. Ceci étant, l'explication de cet écart a profondément évolué. Au début de cette période, les Européens travaillaient au moins autant que les américains, mais étaient nettement moins productifs. Depuis, l'écart de productivité a disparu, mais les européens travaillent environ 30% de moins que les américains. Au fur et à mesure que la productivité européenne a rattrapé la productivité américaine, les européens en ont profité pour travailler moins, laissant l'écart de PIB par habitant à peu près inchangé.

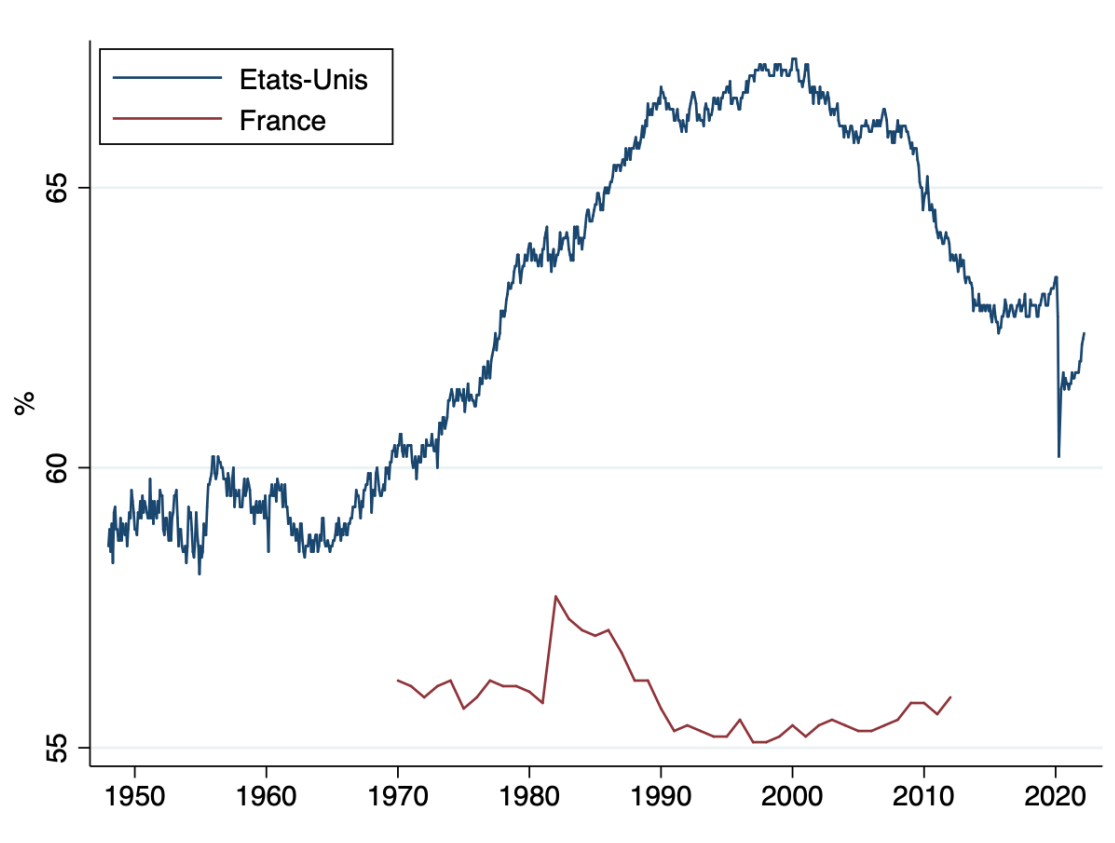
Selon Prescott (2004), cette divergence s'explique par l'augmentation de l'imposition des revenus du travail en Europe. Cette hypothèse a ouvert un vif débat académique pour savoir si l'élasticité de l'offre de travail était suffisamment élevée pour pouvoir attribuer cette divergence à la fiscalité. Au final, la méta-analyse empirique de Chetty et al. (2013) répond plutôt par l'affirmative.

8.4.2 Les causes du chômage

Pour comprendre le chômage, il faut toujours avoir à l'esprit que, dans un marché du travail parfaitement compétitif, il ne peut pas y avoir de chômage : si jamais l'offre de travail des ménages était supérieure à la demande de travail des entreprises, le niveau des salaires aurait tendance à baisser jusqu'à ce que les deux s'équilibrent à nouveau. Le chômage doit donc nécessairement s'expliquer par les frictions qui caractérisent le marché du travail.

Cette observation assez simple permet de discréditer certaines causes supposées du chô-

FIGURE 8.17 – Taux de participation (part des adultes en activité ou au chômage).

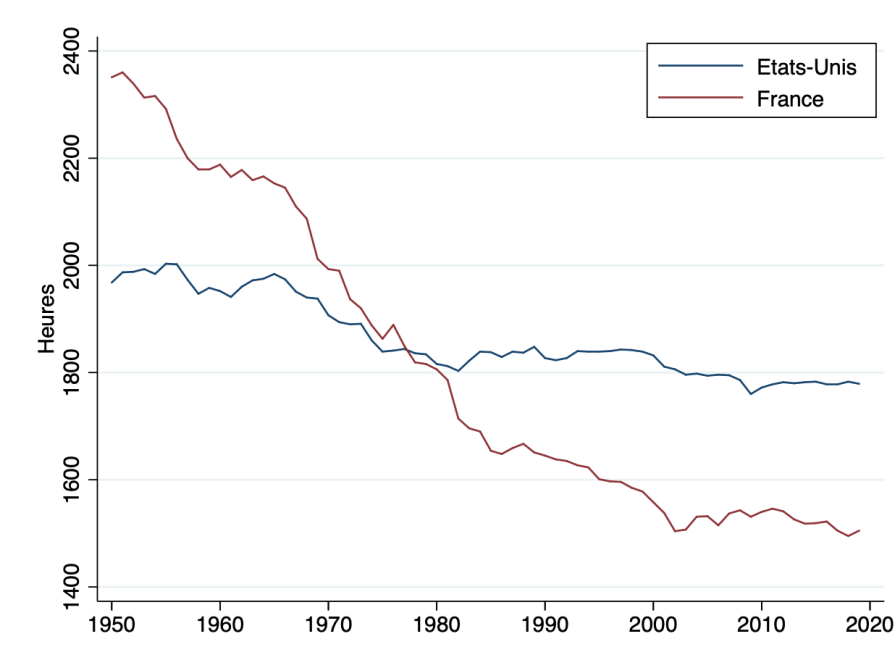


Source : U.S. Bureau of Labor Statistics.

mage. Par exemple, la désindustrialisation ou l'ouverture des frontières commerciales avec la Chine ne peuvent pas à elles seules engendrer du chômage. Si ces évolutions réduisent la demande de travail ouvrier ou non-qualifié, les personnes concernées verront leur salaire baisser jusqu'à ce que l'offre de travail soit à nouveau égale à la demande. D'ailleurs, on observe que les Etats-Unis ne se sont jamais durablement éloignés du plein emploi, alors qu'ils ont été touchés de plein fouet par la désindustrialisation et par les délocalisations.

Quels sont les principales frictions du marché du travail? Tout d'abord, chaque emploi étant unique, et chaque chômeur étant également unique, le processus d'appariement entre les deux est par nature long et coûteux. Cela implique qu'il existe toujours dans l'économie une masse de chômeurs entre deux emplois. Il s'agit là du **chômage frictionnel**. Face aux évolutions de l'économie (nouvelles technologies, changements de goûts, etc.), il y a constamment un besoin de réallocation des travailleurs d'anciens emplois peu productifs vers de nouveaux emplois plus productifs. Le rôle du marché du travail est de permettre cette réallocation, ce qui requiert d'avoir toujours un équilibre entre le nombre de postes vacants et le nombre de chômeurs. Il est donc illusoire de rechercher un taux de chômage nul; tout comme il est naïf de s'indigner devant un grand nombre de postes vacants. Ainsi, quand on parle de « plein emploi », on fait référence à un taux de chômage aux alentours de 4% ou 5%. Lorsque le taux de chômage baisse en dessous de ce niveau, les entreprises en croissance peinent à recruter, ce qui freine leur développement économique. De ce point de vue, un chômeur est donc paradoxalement utile à la société.

FIGURE 8.18 – Nombre d'heures de travail annuel par employé.



Source : OCDE.

Le chômage peut également être expliqué par des salaires trop élevés. Cela peut être une conséquence de la législation imposant un salaire minimum, cas sur lequel nous reviendrons en détail. Les syndicats ont également tendance à faire augmenter les salaires, au détriment de l'emploi dont souffrent les chômeurs qui ne sont pas syndiqués.

La présence d'asymétries d'information peut également contribuer à faire monter excessivement les salaires. Une entreprise n'observant pas parfaitement l'ardeur au travail de ses salariés peut avoir intérêt à les sur-payer tout en les menaçant de les licencier si leurs résultats ne sont pas bons. Ces **salaires d'efficience** incitent les travailleurs à donner le meilleur d'eux-mêmes afin de conserver des emplois particulièrement attractifs. Le problème est que, à ces niveaux de salaire, la demande de travail de l'ensemble des entreprises de l'économie est inférieure à l'offre des travailleurs, ce qui génère du chômage. Comme l'ont modélisé Shapiro and Stiglitz (1984), c'est justement la crainte du chômage qui encourage les travailleurs à faire des efforts, surmontant ainsi le problème d'aléa moral.

Il y a enfin la possibilité que les salaires soient rigides à la baisse. Lors d'un choc positif de demande de travail, le salaire et le niveau d'emploi à l'équilibre augmentent, et une entreprise n'aura généralement pas de mal à s'adapter, embauchant davantage et augmentant les salaires. Si au contraire une entreprise fait face à un choc adverse, diminuer les salaires est une tâche beaucoup plus ardue². Ainsi, la demande de travail des entreprises risque de baisser en dessous de l'offre de travail des employés, générant ainsi du chômage.

Enfin, un autre facteur plus controversé qui peut contribuer au chômage est la présence de « chômeurs volontaires ». Il s'agit de personnes inactives qui ne recherchent pas active-

2. S'il est souvent impossible de baisser le salaire nominal d'un employé (exprimé en euros), le salaire réel (exprimé en biens) peut baisser du fait de l'inflation. En présence de rigidité à la baisse sur les salaires nominaux, l'inflation donne donc un peu de flexibilité à la baisse aux salaires réels.

ment un emploi, mais qui jugent profitable de rester sous le statut de chômeur afin de toucher des allocations. Il est également possible qu'un chômeur refuse de travailler en dessous d'un salaire nettement plus élevé que le niveau des offres qui lui sont faites. En théorie, ces personnes devraient être catégorisées comme inactives mais, en pratique, elles contribuent à augmenter le taux de chômage.

Immigration et chômage

L'immigration augmente l'offre de travail. Mais les nouveaux arrivants consomment des biens et des services produits sur place, ce qui augmente également la demande de travail. L'effet sur les salaires et l'emploi est donc ambigu. Pour savoir quel effet domine, on peut exploiter des expériences naturelles. L'histoire regorge d'épisodes de migration massifs et non anticipés qui peuvent nous aider à comprendre leurs impacts économiques.

Un exemple particulièrement étudié est l'exode de Mariel : en 1980, suite à une succession de crises diplomatiques, Fidel Castro annonça que les cubains désirant quitter leur pays pouvaient le faire librement. Cette annonce inattendue poussa en quelques mois près de 125 000 cubains à quitter leur pays et à faire la traversée du port de Mariel vers les États-Unis, avant que Cuba ne ferme à nouveau ses frontières. Les réfugiés arrivant sur les côtes de la Floride se sont pour la plupart installés dans la zone urbaine de Miami, la ville américaine la plus proche de Cuba. Dans une célèbre étude, Card (1990) a donc choisi de prendre l'agglomération de Miami, où le nombre de travailleurs a soudainement augmenté de 7%, comme groupe de traitement; tandis que quatre autres agglomérations américaines, où le nombre de travailleur est resté inchangé, constitue le groupe de contrôle. Les résultats montrent que, malgré une arrivée massive de travailleurs étrangers, les salaires et l'emploi à Miami ont évolué de manière parallèle à ceux de villes américaines comparables qui n'avaient pas subi de chocs d'immigration.

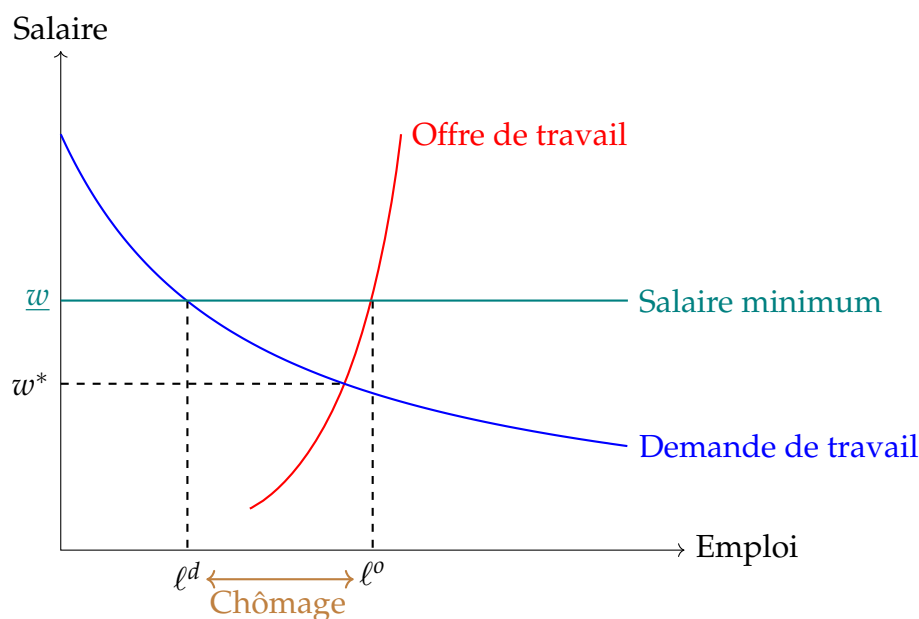
Des résultats similaires ont été obtenus par Hunt (1992) dans le cas du rapatriement en France de 900 000 personnes suite à l'indépendance de l'Algérie en 1962 et par Friedberg (2001) lorsque la population d'Israël augmenta de 12% entre 1990 et 1994 suite à l'effondrement du bloc soviétique. D'autres études ont confirmé ces résultats et n'ont trouvé d'effets économiques néfastes de l'immigration sur l'emploi ou les salaires que pour les immigrés arrivés précédemment, ce qui suggère que le marché du travail est segmenté avec un effet de congestion des immigrés dans certains secteurs d'activité.

8.4.3 Le salaire minimum

La législation fixant un salaire minimum fait partie des interventions étatiques les plus importantes sur le marché du travail. Quel est son effet sur l'emploi? Commençons par considérer le cas d'un marché du travail parfaitement compétitif. Comme le montre la fi-

gure 8.19, un salaire minimum \underline{w} supérieur au salaire d'équilibre w^* augmente l'offre de travail des ménages, jusqu'au niveau l^o , et diminue la demande de travail des entreprises, au niveau l^d . L'écart entre les deux, égal à $l^o - l^d$, correspond au chômage engendré par le salaire minimum. Ceux qui voudraient travailler à un salaire inférieur au salaire minimum ne peuvent pas le faire, et ne trouvent pas d'emploi, alors que des entreprises auraient été disposées à les embaucher. Le salaire minimum empêche donc des transactions mutuellement bénéfiques de se réaliser, ce qui est Pareto inefficace.

FIGURE 8.19 – Impact du salaire minimum sur l'emploi dans un marché du travail parfaitement compétitif.



En pratique, les salariés dont la productivité est élevée ne sont pas concernés par le salaire minimum. En revanche, les travailleurs les moins productifs risquent d'être exclus du marché du travail par un coût prohibitif de l'emploi payé au salaire minimum. Si un salarié a une productivité horaire de 10€, mais que le salaire minimum est de 12€, il aura énormément de difficulté à s'insérer sur le marché du travail. Le salaire minimum risque donc de défavoriser l'emploi de ceux qu'il vise à protéger : les travailleurs les moins qualifiés et les moins expérimentés. Selon Milton Friedman, le très libéral économiste américain de l'école de Chicago, le salaire minimum est une loi qui oblige les entreprises à discriminer les travailleurs les moins productifs.

Cet argument est valide dans un marché du travail parfaitement compétitif. Mais, y aurait-il une autre manière de concevoir le marché du travail ? Sur le marché des biens, le cas le plus extrême de compétition imparfaite est donné par le monopole : un marché dominé par un unique vendeur. De manière symétrique, on peut considérer que le marché du travail est un **monopsonne** : un marché dominé par un seul acheteur. On peut par exemple penser au cas d'une petite ville industrielle où une grosse usine est pratiquement l'unique employeur possible pour un grand nombre d'habitants.

Sur le marché des biens, nous avons vu que le monopole choisit le point de la courbe de demande qui maximise ses profits. Par rapport à un marché compétitif, le monopole dimi-

nue donc la quantité produite afin d'augmenter le prix, ce qui améliore ses marges et accroît ses profits. Dans le cas du marché du travail, c'est la situation symétrique que l'on observe. Le monopsonne choisit le point de la courbe d'offre de travail qui maximise ses profits. Il diminue donc l'emploi afin de baisser le niveau des salaires, ce qui diminue ses coûts de production et augmente donc ses profits. Dans ces conditions, imposer un salaire minimum permet d'augmenter à la fois l'emploi et les salaires ! En effet, face à un salaire minimum, le monopsonne n'a plus intérêt à comprimer l'emploi pour baisser le coût du travail. Si la notion de monopsonne n'est pas souvent évoquée dans le débat public, l'argument qu'elle justifie est régulièrement entendu : le salaire minimum protège les salariés lorsqu'ils sont en position de faiblesse dans les négociations salariales. Par ailleurs, un salaire minimum plus élevé peut pousser les employeurs à investir dans la formation de leurs employés pour que leur productivité augmente, et consolide les liens d'emploi. Soulignons néanmoins que, même dans le cas du monopsonne, si le salaire minimum est trop élevé et excède le salaire qui prévaudrait dans un marché compétitif, il créerait alors du chômage.

Quel est donc l'impact du salaire minimum sur l'emploi ? D'un côté la vision compétitive du marché du travail nous dit qu'il est néfaste, d'un autre côté la vision monopsonistique nous dit qu'il est positif. Il s'agit donc d'une question empirique. Comme les politiques de salaires minimums sont très différentes d'un pays à l'autre et changent régulièrement, de multiples évaluations ont pu être menées. Ces études ont montré que lorsque le salaire minimum est relativement faible, comme c'est le cas aux Etats-Unis, l'augmenter n'a pas un effet néfaste sur l'emploi. En revanche, en France, où il est nettement plus élevé, le salaire minimum est une cause de chômage, notamment pour les jeunes et les travailleurs peu qualifiés.

Comme son nom le suggère, le SMIC (Salaire Minimum Interprofessionnel de Croissance) est indexé sur l'inflation et la croissance. Le gouvernement peut, décider de l'augmenter au delà du minimum, et on parle alors de « coup de pouce ». En 2009, un groupe d'expert sur le SMIC a été instauré. Il est chargé de formuler une recommandation officielle, indépendamment de toute pression politique. Depuis sa création, sur la base de la littérature empirique, les experts ont toujours recommandé au gouvernement de s'abstenir de donner un coup de pouce.

Il existe néanmoins une subtilité cruciale : l'emploi n'est pas déterminé par le salaire net reçu par les salariés au SMIC, mais par le coût du travail payé au salaire minimum. Par conséquent, de nombreux économistes français plaident depuis des années pour des baisses massives de charges ciblées au niveau du SMIC, afin de réduire le coût minimum du travail tout en préservant le pouvoir d'achat des salariés les plus pauvres. Ces politiques d'allègement de charges ont été mises en oeuvre et progressivement amplifiées. On estime que cette politique est si efficace en termes de création d'emplois qu'elle ne coûte presque rien aux finances publiques : la diminution du chômage et donc des allocations versées finance quasiment la baisse de charges patronales.

8.4.4 Les allocations chômage

Pour pallier la violence d'une perte d'emploi, les gouvernements ont mis en place des politiques d'assurance chômage : les salariés cotisent afin de financer les allocations chômage versées à ceux qui ont perdu leur emploi.

Il existe néanmoins une asymétrie d'information qui limite l'efficacité de ces dispositifs : le gouvernement n'observe pas l'intensité avec laquelle un chômeur recherche un emploi. Par conséquent, les allocations chômage engendrent un problème d'aléa moral : elles découragent l'effort de recherche d'emploi et prolongent le chômage.

La politique idéale, en l'absence d'asymétrie d'information, serait de donner des allocations aussi élevées que le dernier salaire tout en obligeant les chômeurs à rechercher très activement un emploi. L'assurance serait donc parfaite et éliminerait toute chute de revenus concomitante à la perte d'un emploi. Cette politique n'est pas réalisable car elle découragerait complètement la recherche d'un nouvel emploi. Ainsi, dans la plupart des pays, le taux de remplacement des allocations chômage est fixé nettement en dessous de 100% du niveau du dernier salaire, et est souvent aux alentours de 50%. En outre, ce taux peut diminuer au cours du temps et l'éligibilité à l'assurance chômage expire au bout d'un certain temps. Cela permet d'aider les chômeurs temporairement, tout en préservant une incitation à retrouver un emploi. L'État doit ainsi trouver un équilibre entre son rôle d'assureur – il donne de l'argent aux chômeurs au moment où ils en ont le plus besoin – et le maintien des incitations à la recherche d'emploi.

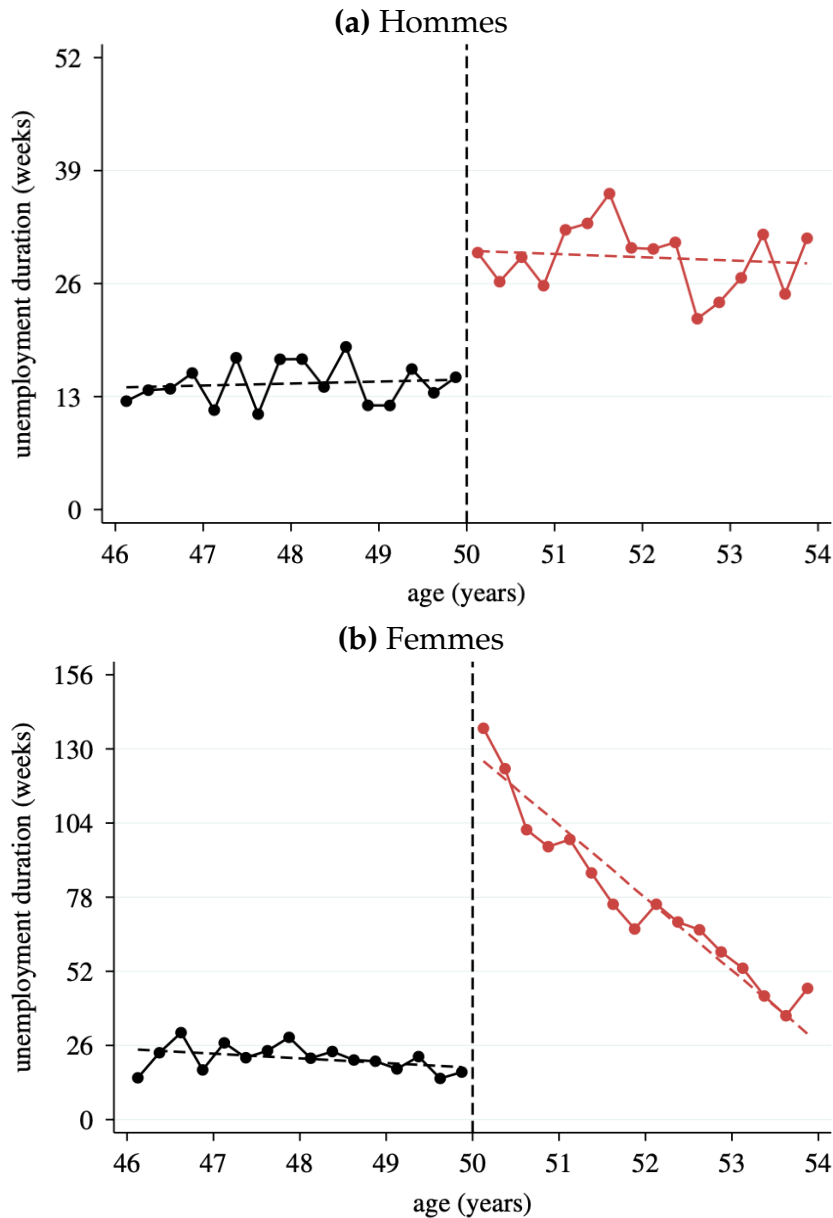
Une question empirique essentielle concerne l'ampleur de l'aléa moral : quel est l'impact de la générosité des allocations sur la durée du chômage ? Bien qu'il existe plusieurs dimensions à la générosité des allocations chômage, intéressons-nous à la durée d'indemnisation. On pourrait imaginer qu'il suffise pour cela de mesurer la corrélation entre durée d'indemnisation et durée moyenne du chômage. Le problème est qu'il existe une multitude de facteurs qui, au sein d'un pays, sont susceptibles d'affecter les deux ; sans parler de la possibilité que la persistance d'un fort taux de chômage de longue durée induise les gouvernements à prolonger la durée d'éligibilité. Cette corrélation ne nous renseignerait donc nullement sur l'impact réel, causal, de la durée d'indemnisation sur la durée du chômage. Or, c'est bien l'impact causal, et non la corrélation, qui mesure l'effet d'aléa moral généré par l'assurance chômage.

Pour surmonter cette difficulté, Lalive (2008) a exploité une expérience naturelle. En Autriche, la durée maximale d'indemnisation est de 39 semaines pour les moins de 50 ans et de 209 semaines, soit près de 4 ans, pour les plus de 50 ans. L'idée est d'utiliser cette discontinuité : les travailleurs qui ont 49 ans et 10 ou 11 mois sont en tous points similaires à ceux qui ont 50 ans et 0 ou 1 mois. La seule différence est que la durée d'indemnisation est beaucoup plus courte pour les premiers (qui peuvent être considérés comme un groupe de contrôle) que pour les seconds (qui constituent un groupe de traitement). Une différence de durée moyenne du chômage entre ces deux groupes peut donc être interprétée comme résultant de l'impact causal de l'allongement de la durée d'indemnisation.

La figure 8.20 montre (en haut) que les hommes de moins de 50 ans ont une durée

moyenne du chômage aux alentours de 15 semaines, là où les plus de 50 ans restent au chômage près de 30 semaines, soit deux fois plus longtemps. Cela nous permet de conclure que 3 années d'indemnisation supplémentaires augmentent la durée moyenne du chômage des hommes d'environ 15 semaines, soit 3 mois et demi. Pour les femmes, l'impact visible (en bas) est plus fort, puisqu'il s'élève à 100 semaines, soit près de deux années.³

FIGURE 8.20 – Durée moyenne du chômage des hommes par âge.



Source : Lalive (2008).

De nombreux autres travaux empiriques ont contribué à quantifier cet effet d'aléa moral. Ils ont permis d'établir qu'une semaine d'indemnisation en plus prolonge le chômage de 0,1 à 0,4 semaines ; tandis qu'une augmentation de 1% du taux de remplacement augmente la durée du chômage de 0,3 à 1,6% (Cahuc and Carcillo, 2014). Malgré une certaine variabi-

3. La diminution avec l'âge de la durée du chômage pour les femmes de plus de 50 ans n'est pas lié à la générosité des allocations chômage, mais plutôt au fait qu'à partir de 54 ans la plupart des femmes sans emploi quittent le chômage afin de bénéficier d'un autre dispositif d'assistance sociale.

lité des résultats selon les pays et les circonstances étudiées, ces estimations montrent que l'effet d'aléa moral est bien réel et d'une ampleur non négligeable. En outre, les résultats convergent pour montrer qu'une augmentation de la durée du chômage n'améliore pas la qualité des emplois trouvés.

8.4.5 La protection de l'emploi

En France, comme en Espagne ou en Italie, le marché du travail est caractérisé par un fort niveau de protection de l'emploi. La législation impose une indemnisation des salariés licenciés à hauteur de plusieurs mois de salaire, et de façon croissante avec l'ancienneté du salarié. Pour les salariés ayant plus de 10 ans d'ancienneté, ces indemnités peuvent représenter plus d'une année de salaire. La protection de l'emploi augmente donc considérablement le coût du licenciement pour une entreprise. Le premier effet de cette politique est de réduire les destructions d'emploi, ce qui est l'objectif recherché. Mais elle induit également un second effet : elle diminue les créations d'emplois. Les entreprises, qui sont incertaines sur leur carnet de commande à moyen terme, sont réticentes à créer de nouveaux emplois qu'elles ne pourront pas détruire en cas de difficultés financières.

Quel est donc l'impact sur le taux de chômage ? Si les destructions baissent davantage que les créations, alors la protection de l'emploi fait baisser le chômage ; et vice versa. L'effet est donc théoriquement ambigu. Pour estimer cet effet, les chercheurs ont exploité les différences de législations entre pays, ainsi que les réformes mises en oeuvre au sein de certains pays. Les résultats montrent que la protection de l'emploi n'a pas un impact important sur le taux de chômage. Les comparaisons entre pays trouvent un effet à peu près nul, tandis que l'étude des réformes suggère un effet légèrement néfaste de la protection de l'emploi sur le taux de chômage (Boeri and van Ours, 2008). Les créations d'emploi diminuent donc légèrement plus que les destructions, mais l'écart n'est pas massif.

En revanche, il existe un effet massif sur les flux de créations et de destructions d'emplois : les deux diminuent considérablement du fait de la protection de l'emploi. Cela allonge beaucoup la durée du chômage. En effet, pour un taux de chômage donné, si le flux de création est plus faible, alors un chômeur restera en moyenne plus longtemps à la recherche d'un emploi. En France, où la protection de l'emploi est particulièrement élevée, le chômage dure en moyenne 15 mois là où, aux États-Unis et au Canada, où le marché du travail est beaucoup plus flexible, le chômage ne dure que 4 ou 5 mois en moyenne.

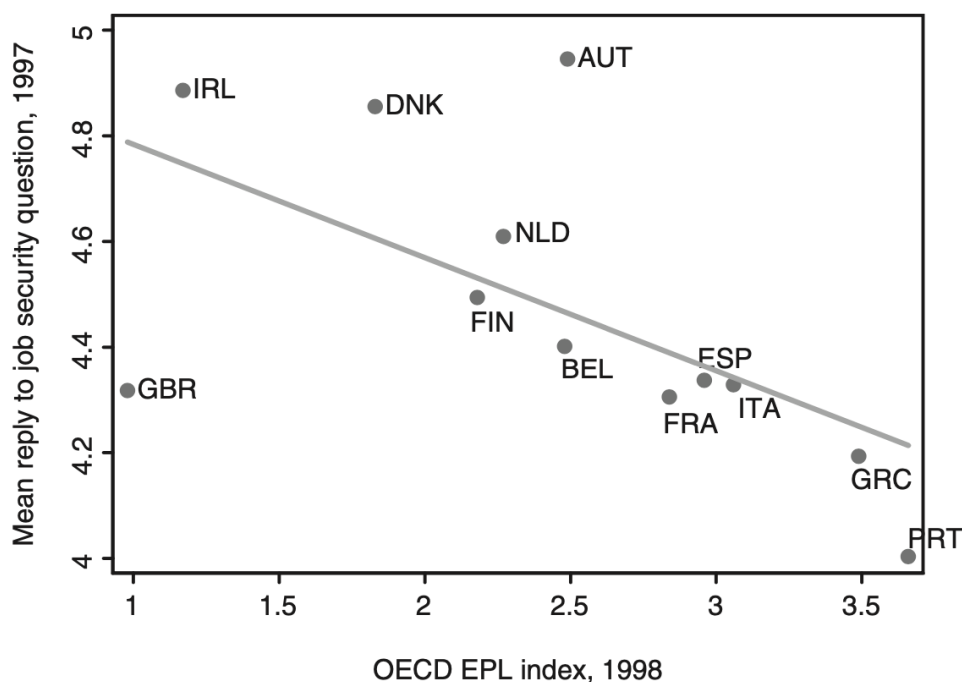
La protection de l'emploi a un autre effet indésirable : elle freine la réallocation des travailleurs de vieux emplois peu productifs vers de nouveaux emplois plus productifs. C'est la sclérose du marché du travail, où les mauvais emplois ne sont pas détruits, ce qui nuit à l'efficacité productive de l'économie. L'amplitude de cet effet sur le PIB est néanmoins difficile à quantifier empiriquement.

Pour pallier cette rigidité du marché du travail, les entreprises ont la possibilité d'embaucher en contrats courts, les Contrats à Durée Déterminée (CDD), par opposition aux contrats longs, les Contrats à Durée Indéterminée (CDI). Les premiers ne sont pratiquement pas protégés, tandis que les seconds le sont étroitement. En France, 88% des salariés (hors intérim)

sont employés en CDI et 12% en CDD; mais les CDD représentent 87% du flux des nouvelles embauches (dont les quatre-cinquième pour moins d'un mois) et les CDI seulement 13%. Les travailleurs en CDD sont donc la variable d'ajustement du marché du travail. En France, non seulement le chômage est long, mais il se termine généralement par un CDD – il est souvent nécessaire d'enchaîner plusieurs CDD avant de pouvoir enfin décrocher un CDI.

La figure 8.21 issue d'une étude de Clark and Postel-Vinay (2009) montre que, paradoxalement, plus la protection de l'emploi est élevée au sein d'un pays, plus le sentiment de précarité y est fort. Pour que cette corrélation soit si nette, il faut qu'elle soit aussi vérifiée pour les bénéficiaires d'emplois protégés! Ce résultat paradoxal s'explique par l'effet pervers de la protection de l'emploi sur la durée du chômage. Le chômage est d'autant plus redoutable qu'il dure longtemps, et qu'il se termine par un CDD. En outre, une durée élevée du chômage met les salariés en position de faiblesse face à leurs employeurs puisque, s'ils ne sont pas satisfaits de leurs emplois, il peuvent difficilement menacer de démissionner. Les salariés ne peuvent donc pas vraiment faire jouer en leur faveur la concurrence entre employeurs qui devrait exister dans un marché du travail compétitif.

FIGURE 8.21 – Niveau moyen du sentiment de sécurité de l'emploi en fonction de l'indice de sévérité de la législation sur la protection de l'emploi.



Source : Clark and Postel-Vinay (2009).

Ces effets pervers de la protection de l'emploi n'impliquent pas pour autant que la meilleure politique est le *laissez-faire*. En présence d'une politique publique d'assurance chômage, un licenciement exerce une externalité fiscale : une entreprise qui licencie impose un coût au régime d'assurance chômage. Cette externalité mérite d'être corrigée par une taxe sur les licenciements. Il s'agit là d'une taxe pigouvienne qui force l'entreprise à internaliser cette externalité. En ce cas, il est également souhaitable de subventionner les embauches. Par

ailleurs, en allongeant la durée des emplois, la protection de l'emploi incite les employeurs à investir dans la formation de leurs salariés. Dans un marché du travail plus flexible, les salariés ne seraient pas incités à investir dans les compétences spécifiques à leur employeur. Par ce biais, la protection de l'emploi peut-être source d'efficacité.

8.5 La régulation du marché du travail

Le salaire étant la principale, voire l'unique, source de revenus pour la plupart des individus, les travailleurs sont naturellement et légitimement demandeurs de protection contre le risque de rémunérations très basses et de chômage. Les gouvernements à travers le monde ont donc mis en place un certain nombre de politiques protectrices, dont les principales sont la redistribution, le salaire minimum, les allocations chômage, et la protection de l'emploi. Ces politiques peuvent être une réelle source de protection, mais en interagissant avec le fonctionnement du marché du travail elles peuvent également avoir des effets pervers, voire même parfois opposés aux objectifs recherchés en fragilisant ceux qu'elles cherchent à protéger.

Lors de la mise en place de politiques publiques, il est donc essentiel de bien comprendre comment elles affectent les comportements des travailleurs, des entreprises, et les équilibres économiques qui en résultent. On ne juge pas une politique économique en fonction de ses objectifs, mais en fonction de ses résultats !

Bibliographie

- Acemoglu, D., Autor, D. H. and Lyle, D. (2004), 'Women, war, and wages : The effect of female labor supply on the wage structure at midcentury', *Journal of Political Economy* **112**(3), 497–551.
- Acemoglu, D. and Restrepo, P. (2019), 'Automation and new tasks : How technology displaces and reinstates labor', *Journal of Economic Perspectives* **33**(2), 1–30.
- Bertrand, M. (2020), 'Gender in the twenty-first century', *AEA Papers and Proceedings* **110**, 1–24.
- Boeri, T. and van Ours, J. (2008), *The Economics of Imperfect Labor Markets*, Princeton University Press.
- Cahuc, P. and Carcillo, S. (2014), *Améliorer L'assurance Chômage*, Presses de Science Po.
- Card, D. (1990), 'The impact of the mariel boatlift on the miami labor market', *Industrial and Labor Relations Review* **43**(2), 245–257.
- Chetty, R., Guren, A., Manoli, D. and Weber, A. (2013), 'Does indivisible labor explain the difference between micro and macro elasticities? a meta-analysis of extensive margin elasticities', *NBER Macroeconomics Annual* **27**(1), 1–56.
- Clark, A. E., Diener, E., Georgellis, Y. and Lucas, R. E. (2008), 'Lags and leads in life satisfaction : A test of the baseline hypothesis', *The Economic Journal* **118**(529), F222–F243.
- Clark, A. and Postel-Vinay, F. (2009), 'Job security and job protection', *Oxford Economic Papers* **61**(2), 207–239.
- Friedberg, R. M. (2001), 'The impact of mass migration on the israeli labor market', *The Quarterly Journal of Economics* **116**(4), 1373–1408.
- Garbinti, B., Goupille-Lebret, J. and Piketty, T. (2018), 'Income inequality in france, 1900–2014 : Evidence from distributional national accounts (dina)', *Journal of Public Economics* **162**, 63–77.
- Hunt, J. (1992), 'The impact of the 1962 repatriates from algeria on the french labor market', *Industrial and Labor Relations Review* **45**(3), 556–572.

- Keynes, J. M. (1932), Economic perspectives for our grandchildren (1930), in J. M. Keynes, ed., 'Essays in Persuasion', Harcourt, Brace and Company, New York, pp. 358–373.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A. and Zweimüller, J. (2019), 'Child penalties across countries : Evidence and explanations', *AEA Papers and Proceedings* **109**, 122–26.
- Lalive, R. (2008), 'How do extended benefits affect unemployment duration? a regression discontinuity approach', *Journal of Econometrics* **142**(2), 785–806.
- Piketty, T., Saez, E. and Zucman, G. (2018), 'Distributional national accounts : Methods and estimates for the united states', *Quarterly Journal of Economics* **133**(2), 553–609.
- Prescott, E. C. (2004), 'Why do americans work so much more than europeans?', *Federal Reserve Bank of Minneapolis : Quarterly Review* **28**(1), 2–13.
- Shapiro, C. and Stiglitz, J. E. (1984), 'Equilibrium unemployment as a worker discipline device', *American Economic Review* **74**(3), 433–444.

Exercices du chapitre 8

Exercice 1 : Production viticole

Un petit pays compte 10 producteurs de vin. Chaque producteur dispose de la fonction de production suivante :

$$Q = 60L - L^2,$$

ou Q est le nombre de tonneaux de vin produit en une année and L le nombre d'employés à plein temps. Les tonneaux sont exportés pour 1000 euros l'unité. Le salaire annuel d'un employé à plein temps est W .

1/ Calculez la productivité marginale du travail (en euros). Dérivez la courbe de demande de travail de l'industrie viticole.

2/ L'économie dispose 200 travailleurs viticoles, dont l'offre de travail est parfaitement inélastique. Combien gagne annuellement un travailleur? Quel est le montant des profits engrangés par un producteur viticole?

3/ Une catastrophe naturelle anéantit deux producteurs. Comment réagissent les salaires et les profits des producteurs restants? Qu'en est-il de la production totale de vin et du niveau des profits de l'ensemble du secteur? Interprétez. [Indice : En sciences économiques, interpréter un résultat consiste à décrire avec des mots le mécanisme économique à l'œuvre. Idéalement, cette description doit être d'une précision mathématique!]

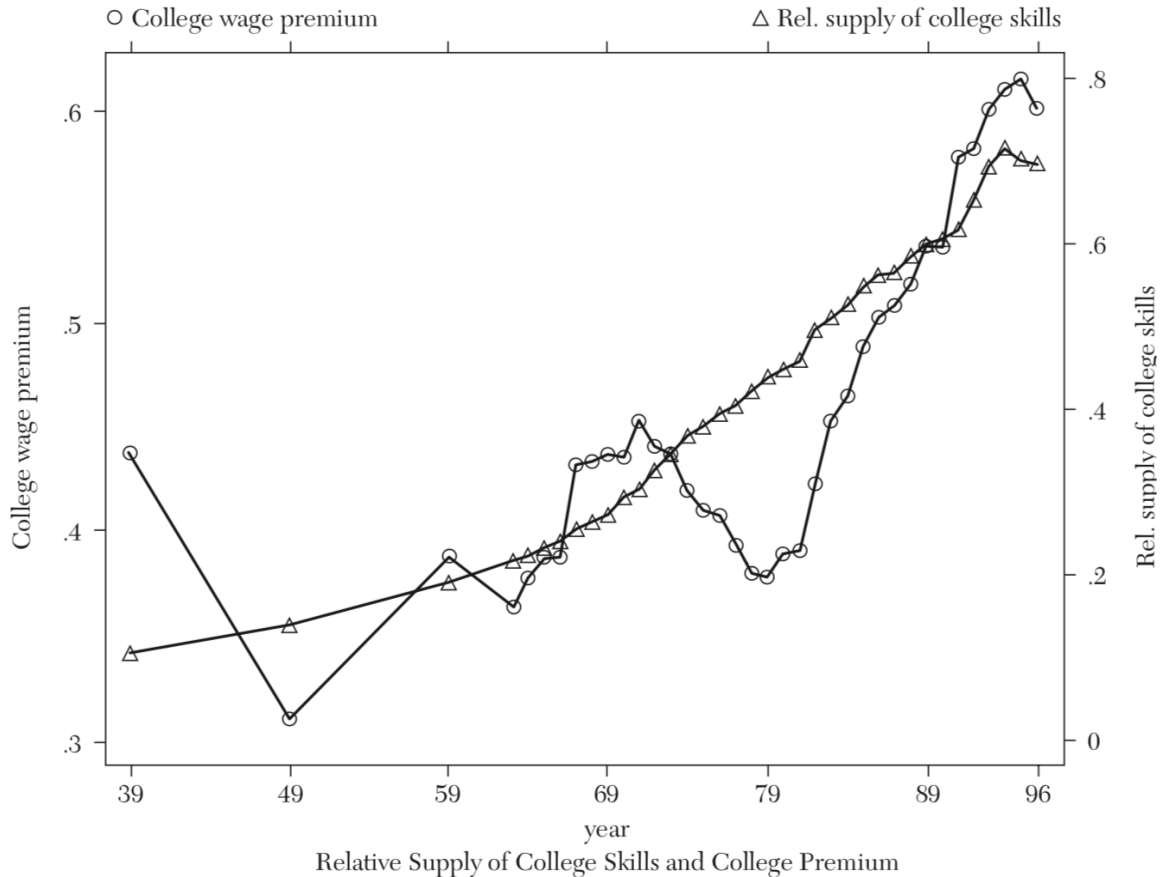
Exercice 2 : "College wage premium"

1/ La figure 8.22 ci-dessous montre une évolution clef de l'économie américaine au cours des dernières décennies. En quoi est-elle paradoxale? Comment peut-on expliquer ce paradoxe? [Indice : Le "college wage premium" est défini comme le ratio entre le salaire des travailleurs ayant été à l'université et ceux n'y ayant pas été.]

2/ Supposons que toutes les entreprises de l'économie américaine ont la fonction de production suivante :

$$Q = \left[(1 - \alpha) (A_L L)^{\frac{\epsilon-1}{\epsilon}} + \alpha (A_H H)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}},$$

FIGURE 8.22 – Nombre et salaires relatifs des travailleurs diplômés aux Etats-Unis.



où L est le nombre de travailleurs peu qualifiés (*Low education*) et H le nombre de travailleurs qualifiés (*High education*), tandis que A_L , A_H et ε sont des paramètres positifs et α un paramètre compris entre 0 et 1. Interprétez ces paramètres. [Indice : Lorsque $\varepsilon = 0$, on a $Q = \min \{A_L L, A_H H\}$, lorsque $\varepsilon = 1$, on a $Q = (A_L L)^{1-\alpha} (A_H H)^\alpha$ et, lorsque $\varepsilon \rightarrow \infty$, on a $Q = (1 - \alpha) A_L L + \alpha A_H H$.]

3/ Calculez le "college wage premium" (en supposant l'offre de travailleurs qualifiés H et non qualifiés L comme étant donnée). Quel est l'impact d'une augmentation de la proportion de travailleurs qualifiés sur le "college wage premium"? Expliquez.

4/ Au regard de ces résultats, proposez une explication plausible de l'évolution de l'économie américaine. Représentez graphiquement cette explication.

Exercice 3 : Salaire minimum

A/ Supposons le marché du travail parfaitement compétitif, avec un salaire d'équilibre w^* . Ce salaire étant jugé trop faible, le gouvernement impose un salaire minimum \underline{w} strictement supérieur à w^* . Quelles sont les conséquences de cette politique?

B/ De nombreux économistes considèrent que, loin d'être parfaitement compétitif, le marché du travail est mieux représenté par un *monopsonne*, défini comme un marché avec

un seul acheteur (tandis que le *monopole* est un marché avec un seul vendeur). Cette représentation semble particulièrement appropriée dans le cas d'une usine qui est le principal employeur au sein d'une petite ville provinciale.

Dans un marché parfaitement compétitif, le coût marginal du travail est égal au salaire. Pourquoi est-ce qu'un monopsonne fait face à un coût marginal du travail qui est plus élevé ?

C/ Représentez graphiquement l'offre de travail, le coût marginal du travail (pour un monopsonne) et la productivité marginale du travail. Sur ce graphique, représentez le niveau de l'emploi et du salaire w_M choisis par le monopsonne. Comparez cette situation au cas d'un marché parfaitement compétitif.

D/ Représentez graphiquement la perte sèche engendrée par la compétition imparfaite sur le marché du travail. Expliquez.

E/ Quelles sont les conséquences d'un salaire minimum \underline{w} strictement supérieur au salaire d'équilibre w_M du monopsonne ? [Indice : Considérez d'abord le cas où $\underline{w} < w^*$, avant de traiter le cas alternatif $\underline{w} > w^*$.]

Exercice 4 : Recettes fiscales

Supposons que, pour tous les travailleurs, la productivité marginale du travail est constante et égale à \bar{W} . La fonction d'offre de travail L^s des ménages est donnée par :

$$L^s = W^\varepsilon,$$

où W est le salaire (net d'impôts) et ε un paramètre positif.

1/ Définissez puis calculez l'élasticité de l'offre de travail.

2/ L'Etat impose les revenus du travail au taux τ . Quel taux d'imposition maximise les recettes fiscales de l'Etat ? Expliquez.

Chapitre 9

Les marchés financiers

Dans les chapitres précédents, nous nous sommes intéressés à des échanges qui avaient lieu à un moment donné. Nous allons désormais introduire la notion de temps dans nos modèles. De nombreuses décisions économiques ont une dimension temporelle, comme par exemple lorsqu'une entreprise construit une nouvelle usine qui ne commencera à produire que dans plusieurs mois ou années, lorsqu'un travailleur épargne pour s'assurer une retraite confortable, ou encore lorsqu'un État s'endette.

Comme nous allons le voir, les marchés financiers sont au cœur de ces échanges intertemporels. Pour commencer, nous allons étudier le fonctionnement du marché du crédit. Nous décrirons ensuite les principaux actifs financiers que sont les actions et les obligations. Pour terminer, nous reviendrons sur la grave crise financière de 2008 qui est lourde d'enseignements sur les sources d'instabilité financière.

9.1 Le marché du crédit

Le marché du crédit est l'exemple le plus simple et le plus emblématique d'un marché financier. Il organise l'échange entre les épargnants, qui offrent du crédit, et les emprunteurs, qui demandent du crédit.

De nombreux ménages ne dépensent pas chaque mois l'intégralité de leur revenu. L'écart entre revenus et consommation correspond à leur épargne, qui constitue leur offre de crédit. Pour produire, les entreprises ont besoin de capital : usines, machines, logiciels informatiques... Le problème est que ces biens doivent être payés au moment de leur acquisition, tandis qu'ils ne contribueront à générer des revenus qu'au fil du temps. Par conséquent, les entreprises empruntent, demandent du crédit, pour financer leurs investissements.

De nombreux jeunes ménages souhaitent également emprunter pour financer leurs études ou pour acquérir leur logement. Ils alimentent ainsi la demande de crédit. Ceci étant, pour faciliter notre exposition, nous supposerons qu'ils offrent une quantité négative de crédit ce qui nous permettra de considérer que les ménages sont du côté de l'offre et les entreprises du côté de la demande.

9.1.1 Taux d'intérêt réel et nominal

Avant de caractériser l'offre et la demande de crédit, il nous faut définir le prix sur le marché du crédit qui, en l'occurrence, est donné par le taux d'intérêt. Supposons le taux d'intérêt annuel égal à i . Si un épargnant prête 100 euros pendant une année, alors l'emprunteur doit lui rembourser $(1 + i)100$ euros au bout d'un an. Ceci étant, du fait de l'inflation, les prix des biens et services augmentent au cours de cette année. Il convient donc de définir deux notions de taux d'intérêt.

- Le **taux d'intérêt nominal** i est mesuré en unités de monnaie : par exemple, si j'emprunte 100 euros pour un an avec un taux d'intérêt nominal annuel de 3%, alors je devrai rembourser 103 euros à la fin de l'année.
- Le **taux d'intérêt réel** r est lui mesuré en terme de biens : si le taux d'intérêt réel annuel est de 3%, alors un emprunteur qui emprunte de quoi acheter 100 pommes aujourd'hui devra rembourser de quoi payer 103 pommes dans un an.

Appuyons nous sur un exemple pour établir le lien exact entre ces deux taux d'intérêt. Supposons que :

- Le taux d'intérêt nominal i est égal à 5% : un agent qui emprunte 1 euro aujourd'hui devra rembourser 1,05 euros dans un an ;
- Le taux d'inflation π est égal à 2% : une pomme qui coûte 1 euro aujourd'hui coûtera 1,02 euros dans un an.

En prêtant la valeur d'une pomme aujourd'hui, soit 1 euro, un agent récupérera dans un an de quoi acheter $1 + r = 1,05/1,02$ pommes dans un an. On a donc la relation suivante

$$1 + r = \frac{1 + i}{1 + \pi}.$$

En effectuant une approximation de premier ordre, avec des taux d'intérêt nominaux et d'inflation proches de zéro, on a $(1 + i)/(1 + \pi) \approx 1 + i - \pi$. Le taux d'intérêt réel est donc donné par :

$$r = i - \pi.$$

Cette approximation est exacte en temps continu, c'est à dire lorsque la durée d'une période est infiniment courte.

Les contrats financiers sont presque toujours écrits en termes de taux d'intérêt nominal mais, comme nous allons le voir, c'est bien le taux d'intérêt réel qui détermine le comportement des agents. C'est donc le taux réel qui détermine l'offre de crédit des ménages et la demande de crédit des entreprises. Sur le marché du crédit, c'est donc bien le taux d'intérêt réel qui joue le rôle de prix.

9.1.2 L'offre de crédit

L'offre de crédit correspond à l'épargne des ménages. Chaque ménage doit décider comment répartir ses revenus entre consommation présente et épargne, laquelle va servir à fi-

nancer la consommation future. Les individus prennent cette décision de manière à maximiser leur satisfaction, leur bien-être. Ils vont donc déterminer leur consommation et leur épargne tel que le bénéfice marginal de la consommation présente soit égale à son coût marginal.

Le coût marginal de la consommation présente, en termes de consommation future, est donné par le coût d'opportunité : pour consommer une unité de consommation en plus aujourd'hui, il faut renoncer à l'opportunité d'en consommer $(1 + r)$ dans un an. Le coût marginal est donc égal à $(1 + r)$.

Le bénéfice marginal de la consommation présente varie d'un individu à l'autre. Néanmoins, on peut supposer qu'il est décroissant du niveau de consommation présente : plus un individu consomme, moins l'unité supplémentaire de consommation lui est bénéfique.

Lorsque l'égalisation entre coût et bénéfice marginal est faite, on trouve bien que l'épargne, et donc l'offre de crédit, est une fonction croissante du taux d'intérêt réel. En effet, une augmentation du taux d'intérêt réel augmente le coût marginal de la consommation présente. En conséquence, si les ménages égalisent ce coût marginal au bénéfice marginal de la consommation présente, ils vont réduire leur consommation présente afin d'en augmenter le bénéfice marginal. Cette baisse de consommation correspond bien à une épargne plus importante. Autrement dit, si le taux d'intérêt réel passe de 2 à 10%, alors la consommation présente est plus coûteuse relativement à la consommation future, ce qui encourage à réduire cette consommation présente et, donc, à épargner davantage.

9.1.3 La demande de crédit

Qu'en est-il de la demande de crédit ? Pour produire, les entreprises ont besoin de *capital* : machines, outils, usines, logiciels informatiques... *L'investissement* d'une entreprise correspond à son acquisition de capital. Les entreprises ont besoin d'emprunter pour investir. Par exemple, si un boulanger achète un nouveau four, il doit le payer tout de suite, tandis que ce four ne générera des revenus que plus tard. Le boulanger doit donc emprunter afin de financer l'achat de ce bien d'investissement, qui lui rapportera plus tard de quoi rembourser son emprunt.

De manière plus générale, les entreprises prennent leurs décisions de manière à maximiser leurs profits. Une entreprise acquiert donc du capital jusqu'à ce que le revenu marginal du capital soit égal à son coût marginal.

Le revenu marginal du capital correspond à la production générée par l'utilisation d'une unité de capital supplémentaire : on parle de **productivité marginale du capital**. On suppose que cette productivité marginale du capital est une fonction décroissante du stock de capital : à nombre de travailleurs fixés, chaque unité supplémentaire de capital augmente moins la production que l'unité précédente. Pour produire son pain, le boulanger a absolument besoin d'un four. Un second four pourrait faciliter la production et permettrait au boulanger de diversifier son offre en proposant des croissants chauds tout au long de la journée. Un troisième four serait moins utile et ne fonctionnerait que quelques heures par jour.

Le coût marginal du capital correspond au coût de l'acquisition d'une unité supplémentaire de capital. Ce coût est une fonction croissante du taux d'intérêt réel : plus ce taux est élevé, plus la somme que l'entreprise devra rembourser à ses créiteurs sera importante.

En égalisant la productivité marginale du capital au coût marginal du capital, on trouve bien que la demande de crédit est une fonction décroissante du taux d'intérêt réel : lorsque le taux d'intérêt réel augmente, le coût marginal du capital augmente, donc la productivité marginale du capital doit augmenter, ce qui n'est possible qu'en réduisant le stock de capital, donc en réduisant l'investissement et la demande de crédit. Ce résultat semble assez intuitif : plus le taux d'intérêt réel est élevé, plus il est coûteux d'emprunter, moins les entreprises investissent.

La demande d'investissement peut également provenir des ménages, notamment lorsqu'ils souhaitent construire ou rénover leurs logements. On parle alors d'investissement et de capital résidentiel, par opposition à l'investissement et au capital productifs. Par souci de simplicité, nous faisons abstraction de l'investissement résidentiel, mais il répond aux mêmes principes que l'investissement productif.

Relation entre investissement et capital.

Le stock de capital d'une entreprise correspond à l'ensemble des biens dont elle a besoin pour produire : machines, usines, logiciels informatiques... L'investissement sur un trimestre ou une année correspond au flux d'acquisition de nouvelles unités de capital. Par conséquent, le stock de capital résulte des flux d'investissements passés. Ceci étant, le capital se déprécie au cours du temps. Par exemple, un camion ou un ordinateur deviennent de moins en moins performant au fur et à mesure qu'ils vieillissent.

Représentons par K_t le stock de capital au temps t , I_t le flux d'investissement à t , et δ le taux de dépréciation du capital. Pour simplifier l'exposition, nous supposons ici que le taux de dépréciation est constant et identique pour toutes les unités de capital. L'équation d'accumulation du capital s'écrit donc

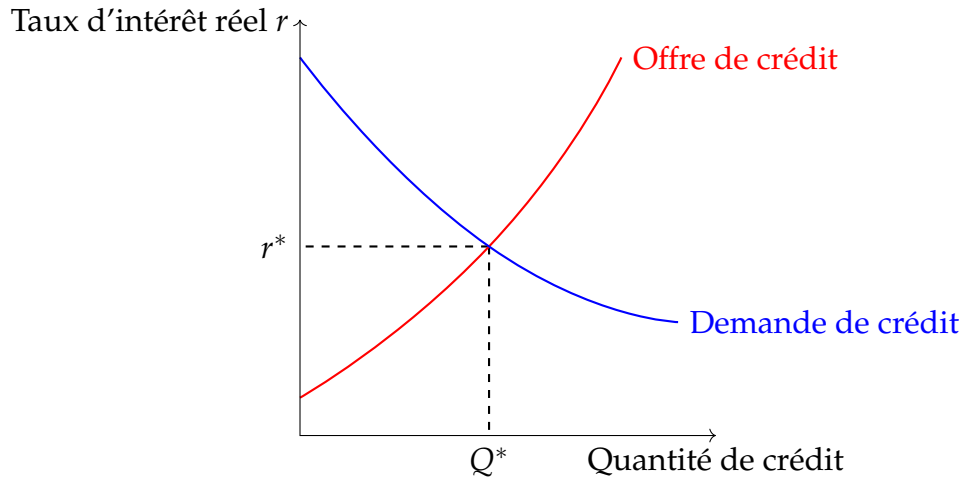
$$K_{t+1} = (1 - \delta)K_t + I_t.$$

Bien que le cœur de l'investissement et du capital soient composés de biens durables, les entreprises effectuent également des investissements *intangibles*, comme lorsqu'elles financent la Recherche et le Développement (R&D), la formation de leurs salariés, ou encore lorsqu'elles investissent dans la promotion de leur image de marque. Bien que ces unités de capital intangibles n'aient pas de réalité physique, il s'agit bien pour les entreprises de dépenses financées dans le présent afin d'augmenter leur profitabilité futur.

9.1.4 Equilibre du marché du crédit

Sur le marché du crédit, le taux d'intérêt réel est fixé par l'équilibre entre l'offre de crédit et la demande de crédit, tel que représenté figure 9.1.

FIGURE 9.1 – L'équilibre sur le marché du crédit



Le marché du crédit résulte donc de transactions mutuellement bénéfiques entre des ménages qui souhaitent différer leur consommation et des entreprises qui, au contraire, ont besoin de dépenser afin d'accroître leur capacité de production avant que cela ne s'avère profitable.

Si les entreprises souhaitent investir davantage, par exemple parce qu'elles sont optimistes sur leurs carnets de commandes futurs, alors la demande de crédit s'accroît et le taux d'intérêt augmente, ce qui incite les ménages à épargner davantage afin de financer ces investissements. Si les ménages souhaitent épargner davantage, par exemple parce qu'ils redoutent une baisse de revenus futurs, alors l'offre de crédit augmente et le taux d'intérêt baisse, ce qui encourage les entreprises à investir davantage afin d'absorber la hausse de l'épargne.

La raison d'être des marchés financiers. Au delà du marché du crédit, cette analyse nous permet de comprendre plus généralement à quoi servent les marchés financiers. Pour les entreprises, ils servent à financer l'investissement. Elon Musk n'aurait jamais pu produire de voitures Tesla sans un excellent accès aux marchés financiers. Des centaines de millions de dollars d'investissement ont été nécessaires afin de concevoir puis de mettre en place la production de cette voiture, avant même qu'elle ne génère un seul centime de revenu. Un bon développement des marchés financiers est essentiel pour l'innovation et, plus généralement, pour la prospérité d'un pays.

Pour les ménages, les marchés financiers servent à transférer la richesse dans le temps. Le premier souci d'un épargnant est de parvenir à préserver sa fortune, voire idéalement à la faire fructifier. Bien sûr, un épargnant pourrait être tenté de garder ses économies sous forme de monnaie. Mais la monnaie rapporte un taux d'intérêt nominal de 0%, et donc un taux d'intérêt réel de $r = -\pi$, soit l'érosion de la valeur réelle de la monnaie due à l'inflation.

La monnaie n'est donc pas un bon support d'épargne. Les marchés financiers offrent une bien meilleure solution : ils permettent aux épargnants de financer le capital, qui n'est pas seulement utile aux entreprises, mais qui est également un support de richesse qui traverse le temps.

En pratique, soit les épargnants prêtent aux entreprises, qui achètent du capital avant de rembourser leurs dettes avec intérêts ; soit ils achètent directement du capital, par exemple un tracteur ou un immeuble de bureaux, qu'ils louent aux entreprises. En principe, ces deux formes d'épargne des ménages doivent rapporter un même rendement, pourvu qu'ils impliquent un même niveau de risque.

L'intermédiation financière. Nous avons jusqu'ici supposé que le taux d'intérêt auquel les entreprises empruntent est identique au taux auquel les ménages épargnent. En réalité, il y a généralement un écart entre les deux. Cet écart reflète le coût de l'intermédiation financière. Les banques mettent en relation les ménages qui offrent du crédit et les entreprises qui en demandent, mais cette mise en relation ne se fait pas par magie, et chaque ménage ne passe pas en revue les entreprises qu'il connaît afin de trouver un emprunteur potentiel. Les banques qui effectuent ce travail de mise en relation sont naturellement rémunérées pour le service rendu. Les salaires des banquiers proviennent de l'écart entre le taux auquel ils prêtent de l'argent et le taux auquel ils rémunèrent les épargnants.

9.2 L'actualisation

Lorsque vous achetez une voiture, vous avez généralement le choix entre payer à la livraison ou étaler le règlement sur plusieurs années. Pour déterminer quelle option est la plus avantageuse, il est nécessaire de pouvoir comparer des sommes d'argent à différents moments dans le temps. On utilise pour cela le principe de l'actualisation. Imaginons que le taux d'intérêt nominal est de 5% par an, et que chacun peut emprunter ou épargner librement au taux de 5%. Par conséquent, 105€ dans un an ont rigoureusement la même valeur que 100€ aujourd'hui. Un agent sera donc indifférent entre recevoir 105€ dans un an ou recevoir 100€ aujourd'hui.

Plus généralement, lorsque le taux d'intérêt nominal est égal à i , alors 100€ aujourd'hui ont la même valeur que :

- $(1 + i) \cdot 100$ € dans un an ;
- $(1 + i)^2 \cdot 100$ € dans deux ans ;
- $(1 + i)^T \cdot 100$ € dans T années.

Cette règle nous permet également de déterminer la **valeur actualisée** d'un revenu futur. Par exemple, la valeur actualisée de 250 euros perçus dans T années est de

$$\frac{250}{(1 + i)^T} \text{ euros.}$$

Grâce aux marchés financiers, vous pouvez transformer $250/(1+i)^T$ euros aujourd'hui en 250 euros dans T années, et vice versa. Par conséquent, si on vous donne le choix entre ces deux sommes d'argent, vous devez être indifférent.¹

De manière plus générale, quelle est la valeur actualisée d'un flux de revenu composé de D_1 dans un an, D_2 dans deux ans, ... et D_T dans T années? C'est simplement la somme des valeurs actualisées de chacun de ces revenus :

$$\frac{D_1}{1+i} + \frac{D_2}{(1+i)^2} + \dots + \frac{D_T}{(1+i)^T}$$

Jusqu'à présent, nous avons supposé les revenus futurs en euros. Ils doivent donc être actualisés avec le taux d'intérêt nominal i . En revanche, si les revenus futurs sont spécifiés en unités de biens, ils doivent alors être actualisés avec le taux d'intérêt réel $r = i - \pi$.

9.3 Les actifs financiers

Un actif correspond à un titre de propriété ayant une certaine valeur. Certains actifs sont tangibles, comme une maison ou une œuvre d'art. D'autres sont intangibles, comme des droits de propriété intellectuelle, ou des actifs financiers. Un actif financier est un contrat qui génère des revenus pour son propriétaire. Les deux principales classes d'actifs financiers sont les obligations et les actions. Ces actifs peuvent être échangés : pour chacun de ces actifs, il existe un marché financier avec une offre, une demande et un prix.

9.3.1 Les obligations

Imaginons que Renault ait besoin d'emprunter 5 milliards d'euros pour construire une nouvelle usine d'automobiles. Une solution est de contracter un emprunt auprès d'une banque. Mais la somme est si énorme qu'une alternative consiste à emprunter directement sur les marchés financiers. Pour cela, l'entreprise Renault vend des obligations (en anglais, *bonds*).

Supposons que Renault ait besoin d'emprunter le 1^{er} janvier 2022 pour une durée de 5 ans. A cette date, elle vend alors une série de morceaux de papier sur lesquels est écrit : "l'entreprise Renault s'engage 100€ au propriétaire de ce morceau de papier le 1^{er} janvier 2027". Chaque morceau de papier est une obligation. Il s'agit d'un contrat de dette qui *oblige* juridiquement l'entreprise Renault à honorer son engagement financier. Ces obligations sont vendues sur les marchés financiers. Si leur prix initial est de 80 euros, alors leurs acquéreurs versent 80€ à l'entreprise Renault au 1^{er} janvier 2022. En échange, Renault versera 100€ le 1^{er} janvier 2027 au propriétaire de chacune de ces obligations.

1. Si, au lieu d'être constant, le taux d'intérêt est égal à i_t entre le temps $t-1$ et le temps t , alors la valeur actualisée, au temps t , de 250 euros perçus au temps $t+T$ s'élève à

$$\frac{250}{\prod_{s=1}^T (1+i_{t+s})} \text{ euros.}$$

Le propriétaire de l'obligation n'est pas forcé de la détenir jusqu'au 1^{er} janvier 2027, date de maturité de l'obligation. À n'importe quel moment, il peut la vendre sur le marché obligataire. De toutes les façons, Renault devra payer 100€ au 1^{er} janvier 2027 au propriétaire de l'obligation, qui ne sera pas nécessairement celui qui l'avait initialement acheté.

En appliquant nos formules d'actualisation, on peut déduire le taux d'intérêt associé à une obligation. Notre exemple implique que, au 1^{er} janvier 2022, 80€ est la valeur actualisée de 100€ perçus le 1^{er} janvier 2027. On a donc

$$80 = \frac{100}{(1+i)^5},$$

ce qui nous permet de déduire que le taux d'intérêt (nominal) correspondant est égal à $i = (100/80)^{1/5} - 1 = 4,56\%$. Il y a donc deux manières de faire référence au prix d'une obligation : soit en donnant le prix auquel elle est échangée, soit en donnant le taux d'intérêt correspondant.

Si l'obligation est initialement vendue 90€, alors Renault s'endette au taux d'intérêt $(100/90)^{1/5} - 1 = 2,13\%$. Si le prix de l'obligation est de 100€, alors le taux d'intérêt correspondant est nul. Cela illustre un principe important en finance :

Plus une obligation vaut cher, moins le taux d'intérêt correspondant est élevé.

Il est toujours utile d'avoir à l'esprit le cas d'une obligation toute simple qui rapporte 1 euro dans un an, et plus rien ensuite. Le lien entre son prix et son taux d'intérêt est donné par :

$$P = \frac{1}{1+i}.$$

Si le prix de cette obligation est de 95 centimes, alors son taux d'intérêt est pratiquement de 5%. Si son prix est de 99 centimes, son taux est de 1%.

En pratique, les obligations peuvent être très complexes, spécifiant différents paiements à différentes dates. Ces paiements peuvent même être conditionnés à la survenue de tel ou tel événement, comme par exemple au déclenchement d'une crise financière ou d'une catastrophe naturelle. La créativité de l'industrie financière est sans limite pour pouvoir satisfaire les besoins, parfois très spécifiques, de leurs clients.

Les entreprises émettent des obligations appelées les *corporate bonds*. Lorsque un État dépense plus que ses recettes fiscales, il doit emprunter pour financer ses déficits budgétaires. Il le fait en émettant des obligations, qu'on appelle parfois les « bons du trésor ». Il s'agit de dette souveraine (*sovereign bonds*).

Les obligations sont-elles sûres ? En principe oui, sauf si son émetteur est incapable d'honorer sa dette, auquel cas il fait **défaut**. Une entreprise en défaut de paiement est placée en liquidation. Ses créanciers deviennent alors propriétaires de son capital, qui est liquidé au meilleur prix afin de rembourser en partie la somme qui leur est due.

Plus une obligation est risquée, plus son prix est faible, et plus le rendement qui lui est

associé est élevé. Cela illustre un principe fondamental en finance : *high returns only come with high risks!* Si un banquier vous propose de placer votre épargne pour un rendement anormalement élevé, méfiez-vous : il y a probablement un risque caché qui explique le niveau élevé de ce rendement.

Les obligations souveraines. Malgré des niveaux vertigineux d'endettement public, les dettes souveraines sont généralement perçues comme étant très sûres, ce qui se traduit par des taux d'intérêt plus faibles que sur les autres obligations. Cela tient à deux raisons. D'abord, pour rembourser leurs dettes, les Etats peuvent avoir recours à la coercition fiscale, en levant taxes et impôts. Ensuite, certains Etats s'endettent dans leur propre monnaie. En principe, les Etats-Unis d'Amérique peuvent toujours imprimer des dollars pour éviter un défaut souverain sur leur dette nationale. Bien sûr que les Etats évitent d'en arriver à de tels extrêmes qui engendreraient un gros risque inflationniste ; mais cette possibilité d'imprimer de la monnaie permet, en période de crise, de tuer dans l'œuf toute panique sur la dette nationale. En pratique, les bons du trésor américains sont perçus comme étant l'actif refuge par excellence. Ainsi, lors de la grave crise financière de 2008, le prix des obligations souveraines américaines a augmenté et leur rendement diminué, alors qu'au même moment la plupart des prix des actifs financiers faisaient des plongées vertigineuses.

Soulignons que, depuis la création de l'euro, la France ou la Grèce ne s'endettent plus dans leur propre monnaie et doivent pouvoir compter sur le soutien de la Banque Centrale Européenne (BCE) en cas de crise obligataire. Les investisseurs perçoivent la dette française comme étant très sûre car elle bénéficie du soutien implicite de la BCE, qui ne peut pas laisser tomber un pays tel que la France sans mettre en danger la survie de l'euro. La France s'endette donc à des taux très bas. Idem pour l'Allemagne. La situation n'est pas aussi claire dans le cas de la Grèce, comme l'a montré la crise de la dette européenne qui s'est déroulée de 2010 à 2013 pendant laquelle les taux sur la dette grecque se sont envolés, avant que la Grèce ne soit obligée de restructurer sa dette à l'issue d'après négociations avec ses créanciers afin d'étaler ses remboursements futurs pour éviter un défaut pur et simple.

Obligations et coupons.

Auparavant, les obligations prenaient réellement la forme d'un morceau de papier, une sorte de gros billet comme on peut le voir sur la figure 9.2. Lorsque l'obligation arrivait à maturité, dix ans dans notre exemple, le détenteur de ce morceau de papier pouvait se présenter auprès de l'émetteur pour lui réclamer son dû.

En général, le détenteur d'une obligation ne reçoit pas de l'argent uniquement lorsque l'obligation arrive à maturité. Attachés à l'obligation se trouvent en général des petits coupons, comme sur la figure 9.3, qui permettent au détenteur de l'obligation de percevoir de l'argent à intervalles réguliers. Aujourd'hui, les obligations sont dématérialisées, mais on appelle toujours « coupons » ces versements réguliers d'argent!

FIGURE 9.2 – Une obligation américaine émise en 1943. Les premières lignes écrites sur cette obligation expliquent bien son fonctionnement : "Les États-Unis d'Amérique, pour [en avoir reçu la somme], promettent de payer au détenteur [de cette obligation] mille dollars, le 15 décembre 1969".

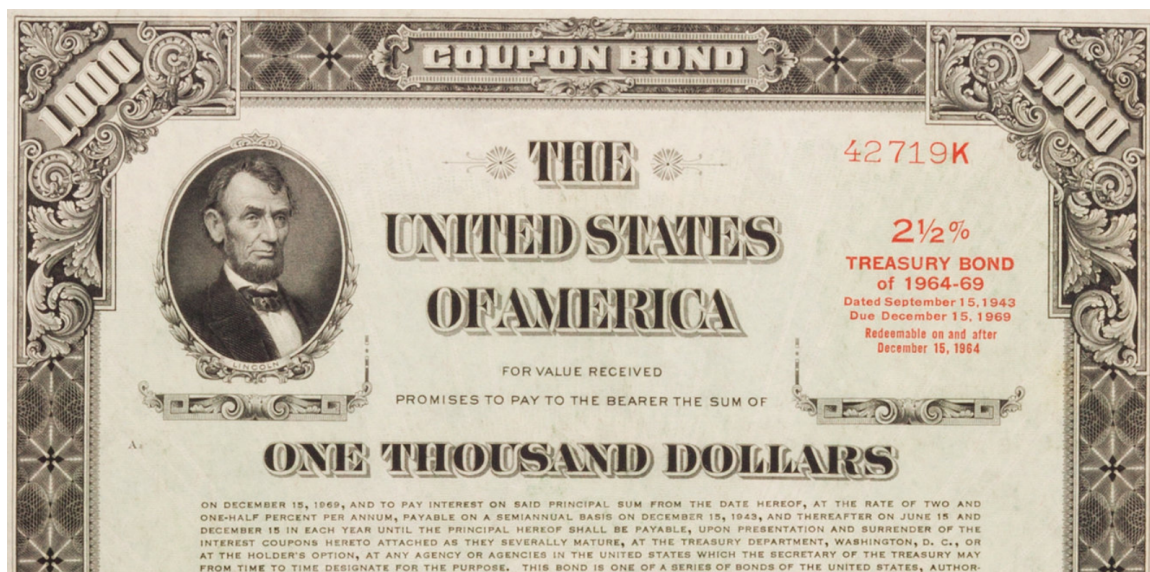


FIGURE 9.3 – Les coupons associés à l'obligation de la figure précédente. Sur chaque coupon figure une date à laquelle leur détenteur peut recevoir une petite somme d'argent. Ici, chaque coupon permet de réclamer 12,50 dollars.



9.3.2 Les actions

Une action (en anglais, *share, equity*, ou encore *stock*) est une part de la propriété d'une entreprise. Si une entreprise est constituée de 1000 actions et que vous en possédez 10, vous êtes alors propriétaire de 1% de l'entreprise. Cela vous permet de toucher 1% des dividendes versés par l'entreprise et vous donne droit à 1% des droits de vote lors de son assemblée générale. Les actions peuvent s'échanger sur le marché boursier.

Lors d'une assemblée générale, les actionnaires nomment le conseil d'administration de l'entreprise, qui nomme à son tour ses principaux dirigeants. Si un actionnaire dispose de plus de 50% des actions d'une entreprise, il disposera d'une majorité de voix au conseil d'administration et pourra donc exercer un contrôle total sur l'entreprise. Ce pouvoir de décision distingue les actionnaires des porteurs d'obligations, qui n'ont pas leur mot à dire dans la gestion de l'entreprise.

Lorsqu'une entreprise a fini de payer ses salariés, ses fournisseurs et ses créanciers, elle dégage un profit. Le conseil d'administration décide alors soit de réinvestir ces profits, soit

de les distribuer aux actionnaires sous forme de dividendes. Chaque actionnaire reçoit donc une part de dividendes égale à la fraction de l'entreprise dont il est propriétaire.

Fondamentalement, une action n'a de valeur que parce qu'elle permet de toucher des dividendes. Par conséquent, la **valeur fondamentale** d'une action est égale à la valeur actualisée du flux de dividendes perçu par l'actionnaire. Si on anticipe un flux de dividendes $\{D_1, D_2, \dots\}$ par action, alors la valeur d'une action doit être égale à

$$\sum_{t=1}^{\infty} \frac{D_t}{(1+i)^t}.$$

En pratique, toute la difficulté est de prendre en compte l'incertitude sur ce flux de dividendes futurs, ce qui fait intervenir des calculs probabilistes sophistiqués.

Une **bulle** correspond à une situation où les épargnants achètent une action, non pas pour les dividendes qu'elle génère, mais uniquement parce qu'ils pensent pouvoir la revendre encore plus cher. Autrement dit, une bulle existe dès lors que le prix d'une action est supérieur à sa valeur fondamentale. Une bulle est donc vouée à enfler jusqu'à ce que le prix de l'action devienne ridiculement élevé. A ce moment, plus personne n'est disposé à acheter cette action, la bulle éclate et le prix chute.

9.3.3 Hypothèse des marchés efficients

Une notion fondamentale en finance est l'*hypothèse des marchés financiers efficients* (en anglais, *efficient market hypothesis*). Cette hypothèse implique qu'à chaque instant, le prix d'une action reflète l'ensemble des informations disponibles sur le montant des dividendes futurs : les marchés financiers seraient alors **informationnellement efficaces**. Cette notion d'efficacité est différente de l'efficacité de Pareto sur laquelle nous nous sommes concentrés auparavant.

L'hypothèse des marchés financiers efficients implique qu'il est quasiment impossible de prédire les évolutions de la bourse : tout ce qui permettrait de faire des prévisions est déjà incorporé dans le prix de marché. En d'autres termes, il n'est pas possible de « battre le marché ».

Soulignons que cette hypothèse est vraie dans la mesure où beaucoup d'investisseurs la croient fautive. En effet, des milliers de financiers scrutent les évolutions de la bourse afin de tirer parti du moindre écart entre le prix d'une action et les informations disponibles sur les dividendes futurs de l'entreprise correspondante. La vigilance et l'âpreté au gain de ces financiers implique que le prix des actions reflète quasiment en temps réel l'ensemble des informations disponibles sur la profitabilité des entreprises.

À l'inverse, l'hypothèse des marchés efficients serait fautive si tout le monde la croyait vraie. En effet, si les financiers n'avaient rien à gagner à scruter les évolutions de la bourse, alors un écart entre prix des actions et flux de dividendes pourrait s'installer. Cette hypothèse des marchés efficients montre donc l'interaction entre théorie économique et comportements individuels qui est propre aux sciences économiques en général, et à la finance en particulier.

La notion de marchés financiers efficients a été formulée pour la première fois en 1900

par le mathématicien français Louis Bachelier (1870–1946) dans sa thèse de doctorat intitulée « théorie de spéculation ». Eugene Fama, de l'Université de Chicago, a obtenu le prix Nobel en 2013 pour avoir montré empiriquement que le prix des actifs est quasiment impossible à prédire à court terme. Ce prix Nobel a été partagé avec Robert Shiller, de Yale, qui a montré qu'à long terme, on peut s'attendre à ce que les excès des marchés financiers soient corrigés, ce qui permet de prédire que le prix d'une action a tendance à revenir vers sa valeur fondamentale.

Faut-il chercher à battre le marché ?

Pour placer son épargne en bourse, on distingue habituellement deux stratégies : une passive et une active. La stratégie passive consiste à acheter l'ensemble des actions cotées en bourse, avec une pondération égale à la valeur totale de chacune des entreprises au sein du marché. Par exemple, si un fond dispose de 1% du volume de l'épargne à placer, il achète 1% de chacune des entreprises cotées. Autrement dit, cela consiste à acheter le *portefeuille de marché*. Une stratégie active consiste à composer son portefeuille en fonction des prévisions des performances boursières de chacune des entreprises. Par rapport au portefeuille de marché, cela implique de sur-pondérer les actions dont on pense qu'elles vont battre le marché et de sous-pondérer celles qui vont sous-performer le marché. Les stratégies actives sont-elles préférables aux stratégies passives ?

Selon William Sharpe, lauréat du prix Nobel 1990 pour ses travaux sur la théorie du portefeuille, la réponse est non. Dans un court article de 1991, il en fait une démonstration simple et puissante. Par construction, les stratégies passives font aussi bien que le marché dans son ensemble. Or, le marché est constitué des investisseurs actifs et passifs. Donc, en moyenne, les stratégies actives ont également une performance égale au marché dans son ensemble ! Certes, certains investisseurs de génie, à l'image de Warren Buffet, vont sur-performer le marché ; mais cela doit nécessairement se faire au détriment d'autres investisseurs *actifs* moins talentueux.

Le problème est que les stratégies actives sont beaucoup plus coûteuses à suivre que les passives. Elles requièrent de rémunérer des financiers qualifiés et avisés qui vont analyser attentivement l'évolution de l'économie et suivre les entreprises une par une pour savoir lesquelles ont le plus de potentiel. Les fonds de placement ayant une stratégie active vont forcément avoir des frais de gestions supérieurs à ceux ayant une stratégie passive. Or, en moyenne, ces deux catégories de fonds ont la même performance. Donc, net de frais, les stratégies passives sur-performent les actives. Cela explique sans doute le succès phénoménal des fonds indiciels (ou *index tracker*) qui se contentent de répliquer passivement la performance d'un indice boursier, comme le CAC 40 à la bourse de Paris ou le S&P 500 à celle de New York.

9.3.4 Actions ou obligations

Via les banques, les particuliers peuvent facilement accéder aux marchés financiers pour placer leur épargne. Est-il préférable de posséder des actions et des obligations? Comme le montre la table 9.1, le rendement réel (net d'inflation) des actions aux Etats-Unis a été en moyenne de 7,6% entre 1947 et 1996, tandis qu'il a été inférieur à 1% pour les obligations souveraines, considérées comme sans risque. Les actions semblent donc être un bien meilleur placement que les obligations. Mais, elles sont également beaucoup plus risquées! L'écart-type du rendement réel moyen sur une année est de 15,5% pour les actions, contre 1,8% pour les obligations.²

TABLE 9.1 – Rendements réels et leurs variations aux Etats-Unis de 1947 à 1996.

	Rendement réel moyen	Ecart-type du rendement réel moyen
Obligations sans risque	0,8%	1,8%
Actions	7,6%	15,5%

Pour financer l'acquisition de capital, comme par exemple pour construire une usine, une entreprise peut soit émettre des obligations, soit émettre de nouvelles actions en effectuant une **augmentation de capital**. Légalemment, les revenus de l'entreprise doivent prioritairement être utilisés pour payer les salariés, les fournisseurs et les créanciers. Les actionnaires ne reçoivent donc des dividendes que si les profits de l'entreprise sont suffisamment élevés. Cette structure de priorité explique pourquoi les actions sont, par nature, beaucoup plus risquées que les obligations. Leurs rendements sont donc logiquement plus élevés pour compenser la prise de risque que leur détention implique.

Quelles recommandations peut-on faire en matière de placements financiers? La sagesse recommande de suivre les quatre préceptes suivants :

- Un bon portefeuille financier contient à la fois des actions et des obligations. La fraction d'actions et d'obligations doit être ajustée en fonction de la tolérance au risque : plus un épargnant est averse au risque, plus la fraction d'obligations doit être forte. De même, plus il est prévu de conserver le portefeuille longtemps, plus la part d'actions pourra être élevée, les bonnes années pouvant sur le long terme compenser les mauvaises.
- Un portefeuille d'actions doit être diversifié. Pour ne pas être surexposé au risque encouru par une entreprise, il convient de ne pas mettre tous ses œufs dans le même panier. Un portefeuille mal diversifié aura, en moyenne, le même rendement qu'un portefeuille diversifié, mais avec un niveau de risque plus élevé. Si "*high returns only come with high risks*", la réciproque n'est pas vraie! Une solution simple et efficace consiste à acheter des parts d'un fond indicel qui se contente de suivre l'évolution d'un indice boursier, comme le CAC 40 à la bourse de Paris. En principe, l'idéal consiste même à

2. En l'occurrence, les obligations « sans risques » sont légèrement risquées en termes réels du fait de la variabilité de l'inflation. Ceci étant, cette variabilité est en partie prévisible si bien que l'incertitude sur le rendement réel de ces obligations souveraines est inférieur à 1,8%.

acheter des fonds indiciels de différents pays afin d'avoir un portefeuille diversifié à l'international.

- L'hypothèse des marchés financiers efficients implique qu'il est vain de tenter de spéculer à la hausse ou à la baisse sur certaines actions. Le seul cas où ce genre de spéculation a de bonnes chances d'être profitable est lorsqu'un épargnant dispose d'informations privées sur une entreprise, qui ne sont pas encore reflétées dans les cours de bourse. Le problème est que ce type de spéculation représente un *délit d'initié*, strictement interdit par la loi.
- Malgré l'hypothèse des marchés financiers efficients, les marchés sont sujets aux comportements moutonniers qui font s'écarter les prix de leurs valeurs fondamentales. Le célèbre investisseur américain Warren Buffet, qui fut le deuxième homme le plus riche du monde, a réalisé une partie de sa fortune sur la maxime suivante : *be fearful when others are greedy, and be greedy when others are fearful*. Lorsque les marchés financiers sont euphoriques et ont une confiance inébranlable dans l'avenir, c'est le moment de se méfier et de vendre des actions pour acheter des obligations. À l'inverse, lorsque tout le monde panique, c'est le moment d'acheter. Autrement dit, en matière de placements boursiers, il faut savoir nager à contre-courant !

L'usage des termes investissement et capital.

Jusqu'à présent, nous avons défini l'investissement comme étant l'acquisition de biens nécessaires à la production, comme l'achat de machines ou la construction d'usines. Le stock de capital résulte du cumul de tous ces flux d'investissement passés, nets de leur dépréciation. Il s'agit là des définitions les plus rigoureuses de ces deux termes, que nous continuerons à utiliser dans le chapitre suivant. Mais une confusion vient du fait que ces termes sont utilisés dans bien d'autres situations. Tentons d'en comprendre la logique sous-jacente (en mettant entre guillemets tous les usages abusifs des termes investissement et capital).

Un « investisseur » est en réalité quelqu'un qui place son épargne, ou celle d'autres personnes. Sur les marchés financiers, les « investisseurs » sont donc du côté de l'offre d'épargne des ménages et non de la demande d'investissement des entreprises. En décidant où allouer leur épargne, ces « investisseurs » déterminent quelles entreprises vont pouvoir investir, bien qu'ils ne le fassent pas directement eux-mêmes.

Le « capital » d'une entreprise correspond à l'ensemble de ses actions. La valeur d'une entreprise est donnée par sa « capitalisation » boursière, qui reflète en principe la valeur actualisée du flux de dividendes à venir. Cette terminologie résulte du fait que les actionnaires sont les propriétaires de l'ensemble du capital tangible (machines, usines...) et intangible (brevets, image de marque...) de l'entreprise, qui sont à l'origine de tous les profits futurs.

Le marché des « capitaux » n'est pas un marché où s'échangent des machines, mais où s'échangent des actifs financiers, comme les actions ou les obligations. De même,

lorsqu'on parle de la mobilité internationale du « capital », ce ne sont pas les machines qui traversent les océans, mais uniquement les titres de propriété des entreprises qui les possèdent.

Quant au « capitalisme », nul ne sait précisément de quoi il s'agit tant ce terme a été utilisé pour décrire des réalités différentes !

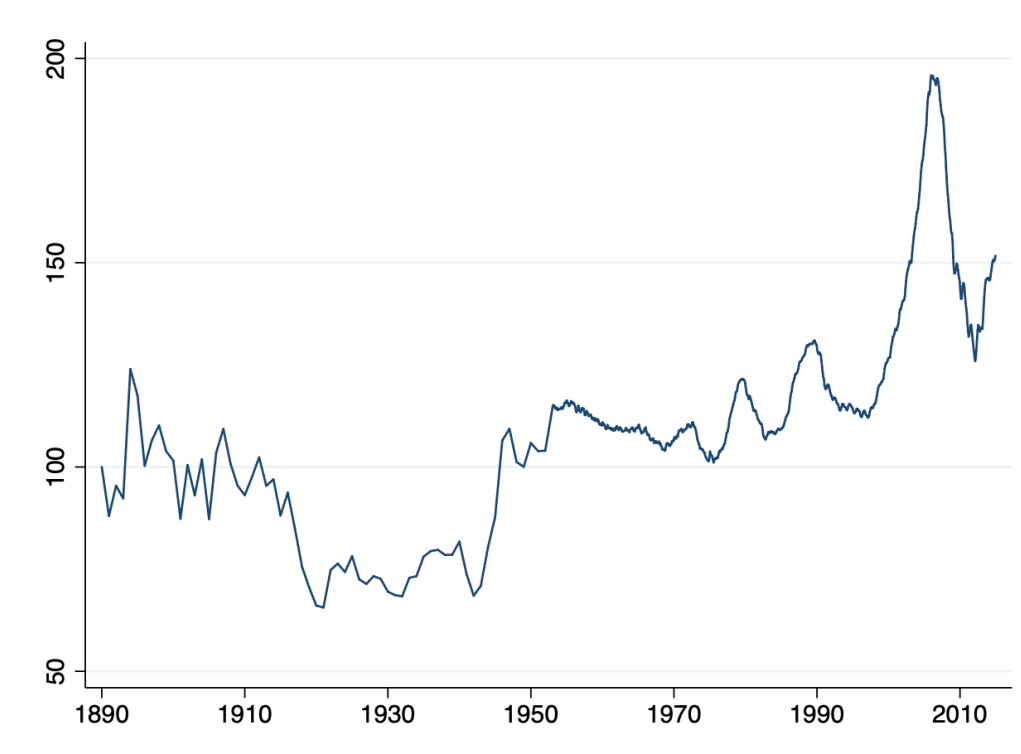
9.4 La crise financière de 2008

En 2008, le monde a été secoué par la pire crise financière mondiale depuis la Grande Dépression des années 1930. Tentons d'en comprendre les mécanismes, et notamment le rôle clef joué par le surendettement que l'on retrouve au cœur de nombreuses crises financières.

9.4.1 La bulle immobilière

Pendant les années qui ont précédé la crise s'est développée une gigantesque bulle spéculative sur le marché immobilier américain. Pendant une trentaine d'années consécutives, les prix de l'immobilier n'ont cessé d'augmenter, comme on peut le voir sur la figure 9.4. La poursuite de cette hausse paraissait inéluctable, jusqu'à ce qu'en 2007 la bulle éclate.

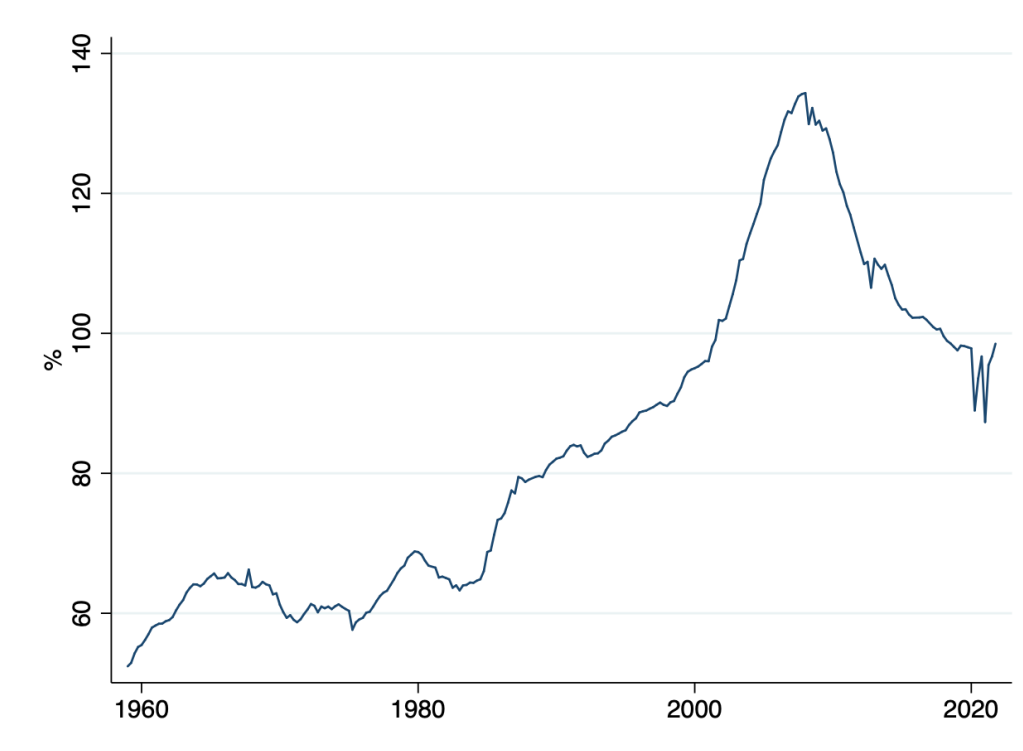
FIGURE 9.4 – Indice Case-Shiller des prix de l'immobilier aux États-Unis



Source : Shiller, R.J. (2015), *Irrational Exuberance*, 3rd Edition, Princeton University Press.

La clef du problème est que cette bulle était basée sur une spirale d'endettement massif des ménages américains, illustrée sur la figure 9.5. L'anticipation d'une hausse continue des prix a poussé les ménages à emprunter et les banques à prêter, même aux ménages les moins

FIGURE 9.5 – Dette des ménages aux Etats-Unis relativement au revenu annuel disponible.



Source : Board of Governors of the Federal Reserve System and U.S. Bureau of Economic Analysis.

solvables. Au pire, pensaient les banques, en cas d'incapacité à rembourser ses dettes, un ménage pourra vendre sa maison, ce qui lui permettra non seulement de rembourser sa banque mais aussi, avec un peu de chance, d'engranger une plus-value ! L'abondance du crédit et l'optimisme général ont créé une dynamique de renforcement mutuel entre les prix de l'immobilier et l'endettement des ménages.

Lorsque la bulle a éclaté, l'économie américaine est entrée dans un cercle vicieux. Craignant de ne pas être remboursées, les banques ont arrêté de prêter et le crédit s'est donc asséché. Ne pouvant plus renouveler leurs emprunts, les ménages surendettés se sont retrouvés en défaut de paiement. Leurs maisons ont été saisies et vendues, amplifiant la baisse des prix de l'immobilier.

Mais comment est-ce qu'une baisse de 25% des prix de l'immobilier a provoqué une crise menaçant tout le système financier américain de s'effondrer ? La réponse à cette question réside dans l'*effet de levier*.

9.4.2 L'effet de levier

Si un nouveau propriétaire achète une maison, il peut la financer de deux manières : il peut utiliser ses économies, on parle de *capitaux propres*, ou il peut s'endetter. Les capitaux propres correspondent au montant de la richesse du propriétaire qui réside dans cette maison :

$$\text{Capitaux propres} = \text{Valeur de l'actif} - \text{Dette}$$

On définit alors le levier comme le ratio de la valeur de l'actif sur les capitaux propres :

$$\text{Lever} = \frac{\text{Valeur de l'actif}}{\text{Capitaux propres}}$$

Un levier important correspond donc à un investissement largement financé par la dette. Par exemple, si une maison vaut €600 000 et est financée en empruntant €500 000, alors les capitaux propres du propriétaire sont de €100 000 et le levier s'élève à 6 : la maison vaut 6 fois plus cher que les économies que le propriétaire y a investi.

Si le prix de la maison augmente de 10%, passant à €660 000, alors les capitaux propres du propriétaire passent de 100 000 à 660 000 - 500 000 = 160 000 euros. Les capitaux propres augmentent donc de 60%. Au contraire, si le prix de la maison baisse de 10%, les capitaux propres du propriétaire sont réduits à €540 000 - 500 000 = 40 000 : ils baissent de 60%.

L'effet de levier multiplie donc à la fois les gains et les pertes engendrées par les variations du prix de l'actif. Cela est dû à ce que ces variations de prix ne modifient pas l'endettement et sont donc absorbées à 100% par les capitaux propres. Si vous trouvez une bonne affaire, l'effet de levier permet d'avoir recours à l'endettement afin d'augmenter le montant de votre investissement, ce qui démultipliera vos gains. Mais, attention, l'effet de levier peut vous ruiner si cette affaire s'avère être moins profitable que prévu !

9.4.3 La crise bancaire

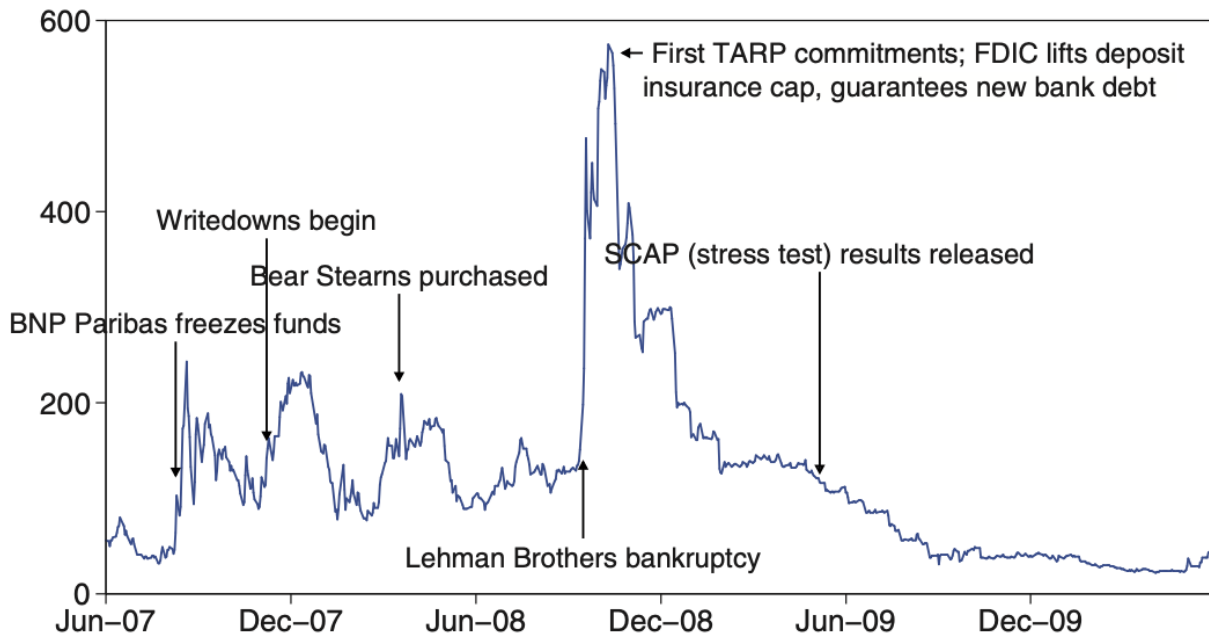
En 2008, les ménages américains étaient très endettés, et l'effet de levier était important. La chute des prix a donc provoqué la ruine de nombreux ménages, qui ont dû mettre la clef sous la porte. Mais les ménages n'étaient pas les seuls à être surendettés : les banques l'étaient bien davantage !

Pour ne pas arranger les choses, au lieu de prêter l'argent de leurs déposants, comme c'est traditionnellement le cas, les banques avaient structuré des obligations spéciales afin de pouvoir échanger sur les marchés financiers l'argent qu'elles prêtaient aux ménages. Ces nouveaux instruments financiers étaient à la fois complexes et opaques. Les acheteurs de ces obligations avaient beaucoup moins d'informations sur leur fiabilité que les vendeurs. Lorsque la crise a éclaté, cette opacité a généré un énorme problème de *sélection adverse*, qui a entièrement gelé tout une partie du système financier.

Sous la pression combinée de la baisse de l'immobilier, de l'effet de levier et du gel de certains marchés obligataires, le 15 septembre 2008, l'impensable s'est réalisé : la banque Lehman Brothers, une des plus importantes institutions financières au monde, a déposé son bilan. L'interconnexion financière était si forte que c'est alors l'ensemble du système bancaire américain qui menaçait de s'écrouler. Un tel effondrement aurait été dramatique car les entreprises industrielles ont besoin du financement des banques pour pouvoir fonctionner.

Les autorités américaines ont alors décidé d'intervenir massivement afin de stopper l'hémorragie. Afin de renflouer le secteur financier, un plan gigantesque de 700 milliards de dollars a été mis en place pour racheter des actifs immobiliers et pour prêter aux banques.

FIGURE 9.6 – Écart de taux d'intérêt (mesuré en points de base, où 100 points de base = 1%) entre les prêts interbancaires et les bons du trésor américains d'une maturité de 3 mois.



Source : Chodorow-Reich (2014).

Ce plan a été financé par l'émission de bons du trésor, l'actif sûr par excellence en cas de crise dont les épargnants du monde entier raffolent.

La figure 9.6 montre l'écart entre le taux d'intérêt auquel les banques se prêtent entre elles et celui auquel le Ministère des Finances emprunte sur les marchés financiers. Cet écart est un indicateur du niveau de stress sur les marchés financiers. En temps normal, la probabilité d'une faillite bancaire est quasiment nulle, si bien qu'une banque est disposée à prêter à une autre banque à un taux aussi faible qu'à l'État américain. Pendant la crise, la confiance s'est évaporée et les banques refusaient de se prêter entre elles, sauf à des taux très élevés.

L'intervention du gouvernement a marqué un coup d'arrêt à la crise financière, mais n'a pas permis d'endiguer la crise économique, beaucoup plus durable. Aux États-Unis, le taux de chômage a mis près de six années à revenir à la normale, tandis qu'en Europe, la reprise a été encore plus difficile. Derrière cette crise financière, se cachaient des déséquilibres macroéconomiques difficilement résorbables.

Bibliographie

Chodorow-Reich, G. (2014), 'The employment effects of credit market disruptions : Firm-level evidence from the 2008–9 financial crisis', *Quarterly Journal of Economics* **129**(1), 1–59.

Sharpe, W. F. (1991), 'The arithmetic of active management', *Financial Analysts Journal* **47**(1), 7–9.

Exercices du chapitre 9

Exercice 1 : Projet d'investissement

1/ Une entreprise envisage de réaliser un projet d'investissement qui requiert un financement de 100 euros aujourd'hui et de 50 euros dans un an. Ce projet générera un revenu de 55 euros dans chacune des trois prochaines années (soit 55 euros dans un an, 55 euros dans deux ans et 55 euros dans trois ans). Est-il souhaitable de financer ce projet lorsque le taux d'intérêt nominal est de 10% ?

2/ Quel est la conséquence d'une diminution du taux d'intérêt nominal à 5% ? Expliquez intuitivement pourquoi votre résultat est désormais différent.

Exercice 2 : Investissements immobiliers

1/ Pour acheter un bien immobilier, vous empruntez la somme de E sur n année au taux d'intérêt annuel $r > 0$. Pour rembourser votre emprunt, vous devez payer la somme de y par an, de l'année 1 à n . Déterminez y en fonction de E , n et r .³

2/ Soit un bien immobilier qui génère des loyers annuels (net de frais d'entretiens) égaux à d . Supposons que le rendement sur les investissements immobiliers s'élève à x .⁴ Pourquoi ce rendement immobilier x peut-il être différent du taux d'intérêt réel r ? Quel doit être le prix P du bien immobilier ?

3/ On suppose dorénavant le rendement immobilier x plus élevé que le taux réel r .

Supposons que vous soyez propriétaire du bien immobilier d'une valeur de P et que vous réinvestissiez dans l'immobilier les loyers d que vous percevez chaque année. Quelle est votre fortune F_n au bout de n années ?

4/ En supposant que vous puissiez vous endetter au taux r sur n années, décrivez une meilleure manière de faire fructifier votre fortune P . Quelle est alors votre fortune L_n au bout de n années ? Comparez F_n et L_n pour $r = 1\%$, $x = 4\%$, et $n = 15$ ans.

3. Tout au long de l'exercice, on considère que les sommes, comme y ou E , sont des variables réelles exprimées en unités de biens. Ainsi, r est le taux d'intérêt réel. Bien sûr, dans le cas où l'inflation est nulle, les variables réelles sont également des variables nominales.

4. Lorsqu'on parle d'"investissements immobiliers", il s'agit en fait d'épargne placée dans l'immobilier. On est donc du côté de l'offre d'épargne, et non de la demande d'investissement qui correspond soit à la construction soit à la rénovation immobilière. Bien sûr, à l'équilibre, si l'offre d'épargne immobilière est abondante, alors les prix de l'immobilier montent et l'investissement en construction ou en rénovation augmente.

5/ Sous quelle condition est-ce que L_n est infiniment grand ? Interprétez.

Exercice 3 : Le déficit budgétaire

1/ Le gouvernement décide d'augmenter le déficit budgétaire de zéro à un milliard d'euros afin de financer une baisse d'impôts. Comment réagit le taux d'intérêt ? Quel effet cela a-t-il sur la demande de crédit des entreprises ? Représentez graphiquement cet effet.

2/ Les ménages anticipent que l'accroissement du déficit budgétaire engendre une hausse de un milliard d'euros de la valeur actuelle des prélèvements fiscaux. Comment cela affecte-t-il votre réponse à la question précédente ?

Exercice 4 : Fonds propres

L'entreprise Apple est tellement prospère qu'elle dispose de près de 250 milliards de dollars de fonds propres. Apple n'a donc pas besoin d'emprunter pour financer ses investissements (en recherche et développement). Cela implique-t-il que la demande d'investissement (en recherche et développement) d'Apple est indépendante du taux d'intérêt ?

Exercice 5 : Valeur fondamentale

1/ Une obligation perpétuelle n'a pas d'échéance. Elle prévoit le versement chaque année d'un revenu fixe, et cela pour l'éternité. En supposant un taux d'intérêt nominal constant, égal à $i > 0$, calculez la valeur fondamentale p^* d'une obligation perpétuelle dont le coupon annuel s'élève à d euros.

2/ Soit p_t le prix de l'obligation perpétuelle dans t années. Sous quelle condition est-ce qu'un ménage souhaitant épargner de la date t à $t + 1$ est indifférent entre détenir cette obligation perpétuelle et effectuer un autre placement rapportant le taux d'intérêt i ? Que se passe-t-il si cette condition n'est pas satisfaite ?

3/ Déduire le prix p_t de l'obligation perpétuelle à la date t . Interprétez. Parmi les solutions trouvées, laquelle vous semble la plus plausible ?

4/ Supposons que chaque année il y a une probabilité q que le prix p_t de l'obligation revienne à sa valeur fondamentale p^* . Quel est désormais le prix de l'obligation perpétuelle à la date t ? Expliquez.

Chapitre 10

Macroéconomie

La crise financière de 2008 et la pandémie de coronavirus en 2020 ont entraîné des récessions d'une ampleur sans précédent dans la plupart des économies du globe. Comment les pouvoirs publics doivent-ils répondre à de tels chocs ? La branche des sciences économiques qui s'intéresse à cette question, et plus généralement aux évolutions de l'économie à grande échelle s'appelle la macroéconomie.

Par définition, la microéconomie consiste à analyser les comportements individuels des ménages, des entreprises et leurs interactions. C'est ce que nous avons fait jusqu'à présent. La macroéconomie consiste au contraire à étudier l'économie dans son ensemble. Comme nous allons le voir, un certain nombre de relations existent à l'échelle d'un pays entre la consommation, l'investissement, la production, les revenus, les taux d'intérêt ou encore l'inflation, ce qui va nous permettre de dégager une vision d'ensemble du fonctionnement de l'économie. Cette approche permet d'expliquer les fluctuations de l'activité économique, ainsi que le rôle des politiques monétaires et budgétaires afin de gérer au mieux ces aléas.

Dans ce chapitre d'introduction à la macroéconomie, nous commencerons par étudier les égalités de comptabilité nationale, qui sont le socle de l'analyse macroéconomique. Nous nous intéresserons ensuite aux idées révolutionnaires de John Maynard Keynes qui fut le premier à donner une analyse cohérente du fonctionnement de l'économie dans son ensemble. Puis, nous verrons quel est l'impact de la monnaie sur l'activité économique, avant de déterminer le rôle de la politique monétaire. Enfin, nous reviendrons sur la Grande Récession de 2008, riche d'enseignements pour l'analyse macroéconomique.

10.1 Les agrégats macroéconomiques

L'analyse macroéconomique consiste à étudier des agrégats : consommation, investissement, production, revenus... Une variable agrégée combine des données individuelles de ménages ou d'entreprises en un seul indicateur pour l'ensemble de l'économie.

10.1.1 Le Produit Intérieur Brut

L'agrégat le plus important est le Produit Intérieur Brut ou PIB (en anglais, *Gross Domestic Product* ou *GDP*). Il s'agit de la valeur de l'ensemble de la production de l'économie au cours d'une période donnée ; typiquement un trimestre pour le PIB trimestriel ou une année pour le PIB annuel. Plus précisément :

Le Produit Intérieur Brut est égal à la valeur de tous les biens et services *finaux* produits dans l'économie au cours d'une période donnée.

Un bien final est défini comme un bien acheté par son utilisateur ultime. Il peut s'agir d'un bien de consommation acheté par un ménage ou d'un bien d'investissement acheté par une entreprise. A l'inverse, un bien intermédiaire est transformé dans le processus de production. Par exemple, une plaque de beurre est un bien final si elle est achetée par un ménage, mais un bien intermédiaire si elle est achetée par un fabricant de biscuits.

Comment mesurer le PIB si on peut difficilement distinguer les biens finaux des biens intermédiaires ? En utilisant la notion de *valeur ajoutée*. Pour une entreprise, la valeur ajoutée est égale à la valeur de sa production diminuée de la valeur de biens intermédiaires qu'elle a achetés. Le PIB peut être mesuré en additionnant la valeur ajoutée de l'ensemble des entreprises du pays.

On distingue deux notions du PIB :

- Le **PIB nominal**, défini comme la valeur de marché des biens et services finaux *aux prix courants*. Pour calculer le PIB nominal en 2018, on multiplie le volume de biens et services finaux produits en 2018 par leurs prix en 2018. La croissance du PIB nominal d'une année à l'autre peut donc être causée par une augmentation du *volume* de biens produits ou par une augmentation des *prix* de ces biens. Si l'on souhaite mesurer l'accroissement de la prospérité économique, seul le premier de ces deux effets, l'effet volume, importe.
- Le **PIB réel** est défini comme la valeur de marché des biens et services finaux *à prix constants*. Pour calculer le PIB réel, on fixe les prix à leur niveau lors d'une année de référence, par exemple 2011. Ensuite, pour obtenir le PIB réel en 2018, on multiplie le volume de biens et services finaux produits en 2018 par leurs prix en 2011. La croissance du PIB réel permet donc de mesurer l'accroissement du volume de la production.

La croissance du PIB nominal est à peu près égale à la somme de la croissance du PIB réel et du taux d'inflation (qui correspond à la croissance du niveau moyen des prix). Les taux de croissance économique rapportés par la presse font toujours référence au PIB réel.

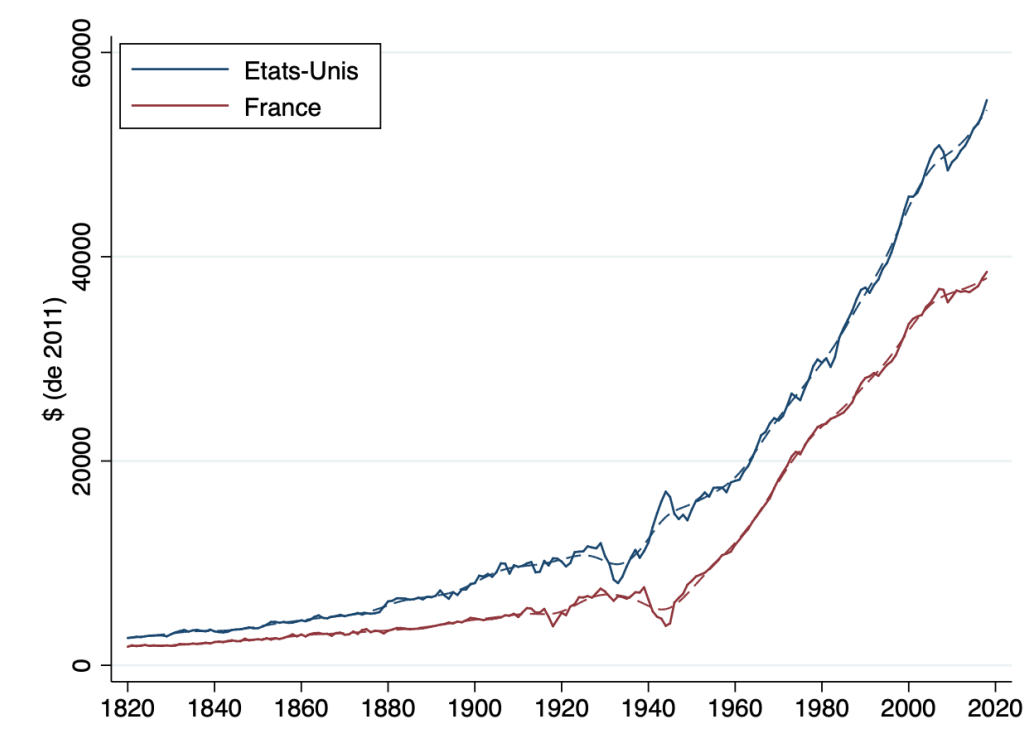
Le choix de l'année de référence pour le calcul du PIB réel.

Les prix de différents biens ne croissent pas tous au même rythme. Au cours des dernières décennies, les prix ont fortement augmenté dans le luxe, tandis qu'ils ont dimi-

nué dans l'électroménager. Le choix de l'année de référence pour calculer le PIB réel n'est donc pas neutre. Pour surmonter cette difficulté, l'évolution du PIB réel entre l'année t et $t + 1$ est généralement calculée en prenant successivement l'année t puis l'année $t + 1$ comme année de référence. La moyenne entre ces deux mesures donne la croissance du PIB réel entre t et $t + 1$. Pour déterminer le PIB réel chaque année, il ne reste plus qu'à fixer le niveau du PIB réel à un moment donné. On choisit pour cela une année de référence pour laquelle le PIB réel est normalisé au niveau du PIB nominal. Cette méthode, dite « en chaîne », permet de gommer l'impact du choix de l'année de référence sur la mesure du taux de croissance du PIB réel.

La figure 10.1 montre l'évolution du PIB réel par habitant depuis 1820. On observe deux choses. La première est l'accroissement extraordinaire du niveau de la production au cours des 150 dernières années. Depuis 1870, le taux de croissance annuel du PIB réel par habitant s'est élevé à environ 1,7% en France et aux États-Unis. À ce rythme, le niveau de vie double tous les 41 ans! Cela nous permet d'atteindre aujourd'hui des niveaux de vie qui auraient été difficilement imaginables il y a un siècle et demi.

FIGURE 10.1 – PIB réel par habitant, mesuré en dollars de 2011.



Source : The Maddison Project Database. La tendance, en pointillés, est obtenu avec un filtre statistique, dit de Hodrick-Prescott.

La seconde observation est l'irrégularité du rythme de la croissance économique, qui se fait en dents de scie. Certaines années, l'économie est en plein boom et le PIB croît fortement tandis que, d'autres années, l'économie est en récession et le PIB se contracte.

Ces deux observations sont liées à deux sujets qui structurent l'analyse macroéconomique :

- La croissance : la macroéconomie de long terme cherche à comprendre quels sont les ressorts de la croissance du PIB.
- Les cycles : la macroéconomie de court et moyen terme s'intéresse aux fluctuations du PIB.

Le principal moteur de la croissance à long terme est l'innovation. C'est grâce aux innovations technologiques que l'on peut produire plus à partir d'une même quantité de ressources. Néanmoins, l'innovation ne tombe pas du ciel, mais est généralement le fruit d'un environnement qui lui est favorable. L'innovation se fait souvent par *destruction créatrice* : les nouvelles technologies remplacent les précédentes, ce qui fait et défait les fortunes, parfois de manière douloureuse, et notamment pour les leaders des technologies sur le point d'être dépassées. Il est donc primordial que les institutions économiques, politiques et sociales d'un pays n'entravent pas ce processus. Si les Etats-Unis sont leaders technologiques dans de nombreux domaines, c'est parce que les institutions y sont favorables à l'innovation, notamment parce qu'elles garantissent la compétition et la propriété privée. Cela encourage les entrepreneurs à investir dans l'innovation, sans craindre d'être spoliés.

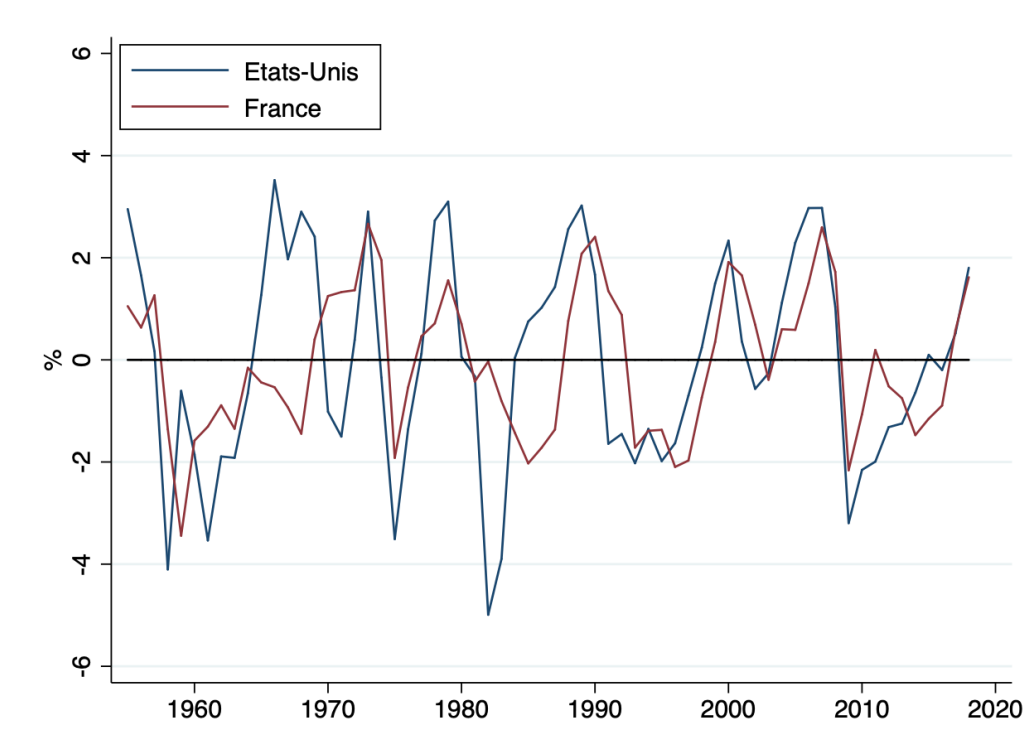
Ce chapitre est essentiellement consacré à la question des cycles, qui a historiquement constitué le coeur de l'analyse macroéconomique depuis les travaux du britannique John Maynard Keynes (1883–1946) dans les années 1930. On pourrait croire que les fluctuations du PIB sont négligeables par rapport à la tendance de long-terme. Il y a néanmoins trois raisons de s'y intéresser. D'abord, ces fluctuations sont associées à des fluctuations du taux de chômage : les contractions du PIB réel, appelées récessions, sont caractérisées par un chômage de masse qui peut être dévastateur. Ensuite, ces fluctuations sont directement influencées par la politique macroéconomique, constituée de deux branches :

- La **politique monétaire**, qui consiste à ajuster la quantité de monnaie dans l'économie ;
- La **politique budgétaire**, qui consiste à ajuster le niveau des prélèvements et des dépenses de l'État.

L'analyse macroéconomique cherche à comprendre comment utiliser ces leviers pour lisser l'activité économique afin d'éviter les récessions ou, pire, les dépressions. Enfin, les dépressions économiques peuvent être persistantes, au point de perdurer bien au-delà de la durée d'un cycle : ce fut le cas de la Grande Dépression américaine qui dura du début des années 1930 jusqu'à la Seconde Guerre mondiale.

La figure 10.2 montre l'écart entre le PIB réel par habitant et la tendance de long-terme représentée en pointillés sur la figure 10.1. Cet écart correspond à la composante cyclique de l'activité économique. Ce cycle (en anglais, *business cycle*) est stochastique, et il est possible que l'économie subisse coup sur coup deux chocs négatifs, ou deux chocs positifs. En d'autres termes, il est difficile de prédire l'état de l'économie dans un futur de moyen terme. On considère néanmoins que la période des cycles économiques est de 2 à 10 ans, pour une moyenne de 6 ou 7 ans. Le seul cycle déterministe de l'économie est la saisonnalité : la consommation de gaz est plus élevée en hiver et la consommation de séjours en bord de mer plus élevée en été.

FIGURE 10.2 – Ecart entre le PIB réel par habitant et la tendance de long-terme (représentée en pointillés sur la Figure 10.1)



Doit-on chercher à maximiser le PIB par habitant ?

Le PIB par habitant est la principale mesure de la prospérité économique d'un pays. Ceci étant, cette statistique est loin d'être parfaite et il serait tout à fait malvenu de chercher à la maximiser à tout prix. En particulier, le PIB réel fait abstraction des éléments suivants :

- Le PIB ne prend pas en compte la dépréciation du capital, du fait de son vieillissement et de son obsolescence progressive. Le Produit Intérieur Net, PIN, corrige ce problème en retranchant la dépréciation au Produit Intérieur Brut, PIB. Ceci étant, la dépréciation est toujours difficile à quantifier, ce qui implique des erreurs de mesure beaucoup plus importantes sur le PIN que sur le PIB.
- La production domestique, comme la cuisine, la garde des enfants, ou encore le jardinage, ne résultent pas de transactions marchandes et ne sont donc pas intégrées dans le PIB. Pourtant, il s'agit bien d'une production de richesses.
- L'économie souterraine n'est pas non plus comptabilisée dans le PIB. Dans les pays en développement, ces activités sont une part essentielle de l'activité économique. Certaines mesures du PIB essaient de corriger ce biais, mais forcément avec des marges d'erreurs importantes.
- La détérioration de l'environnement ou l'épuisement des ressources naturelles ne sont pas négativement prise en compte dans le calcul du PIB.

- La valeur du temps libre n'est pas non plus prise en compte. On pourrait facilement accroître notre PIB en augmentant notre temps de travail, mais cela ne serait pas forcément souhaitable.

En sciences économiques, on considère habituellement que l'objectif de la politique économique doit être de maximiser le bien-être des citoyens. Comme expliqué précédemment, des considérations philosophiques sont nécessaires pour agréger le bien-être d'individus hétérogènes. Ceci étant, quelque soit la conclusion à laquelle on parvient, il est clair que l'objectif ne doit pas être de maximiser le PIB par habitant.

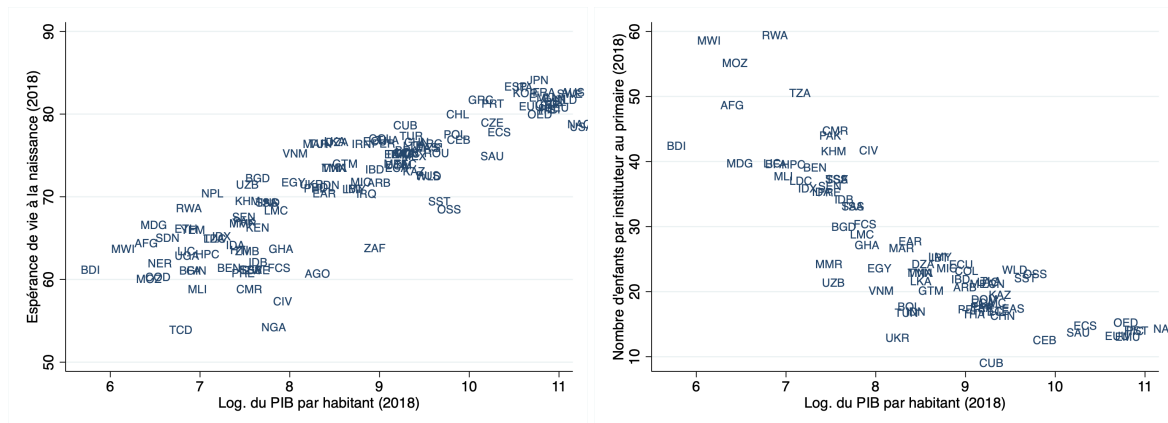
Le PIB par habitant reste néanmoins un bon moyen de comparer le niveau de vie entre pays, comme le montre la table 10.1. En outre, la figure 10.3 met en évidence un lien étroit entre le PIB par habitant et la santé ou l'éducation, tandis que la figure 10.4 montre une forte corrélation entre le PIB par habitant et le sentiment de satisfaction dans la vie. Malgré toutes ses imperfections, le PIB reste un indicateur essentiel de la prospérité économique.

TABLE 10.1 – PIB par habitant, en parité de pouvoir d'achat. Source : FMI

	PIB par habitant (2020)
USA	63 051
France	45 454
Russie	27 394
Kenya	4 993
Haiti	1 728

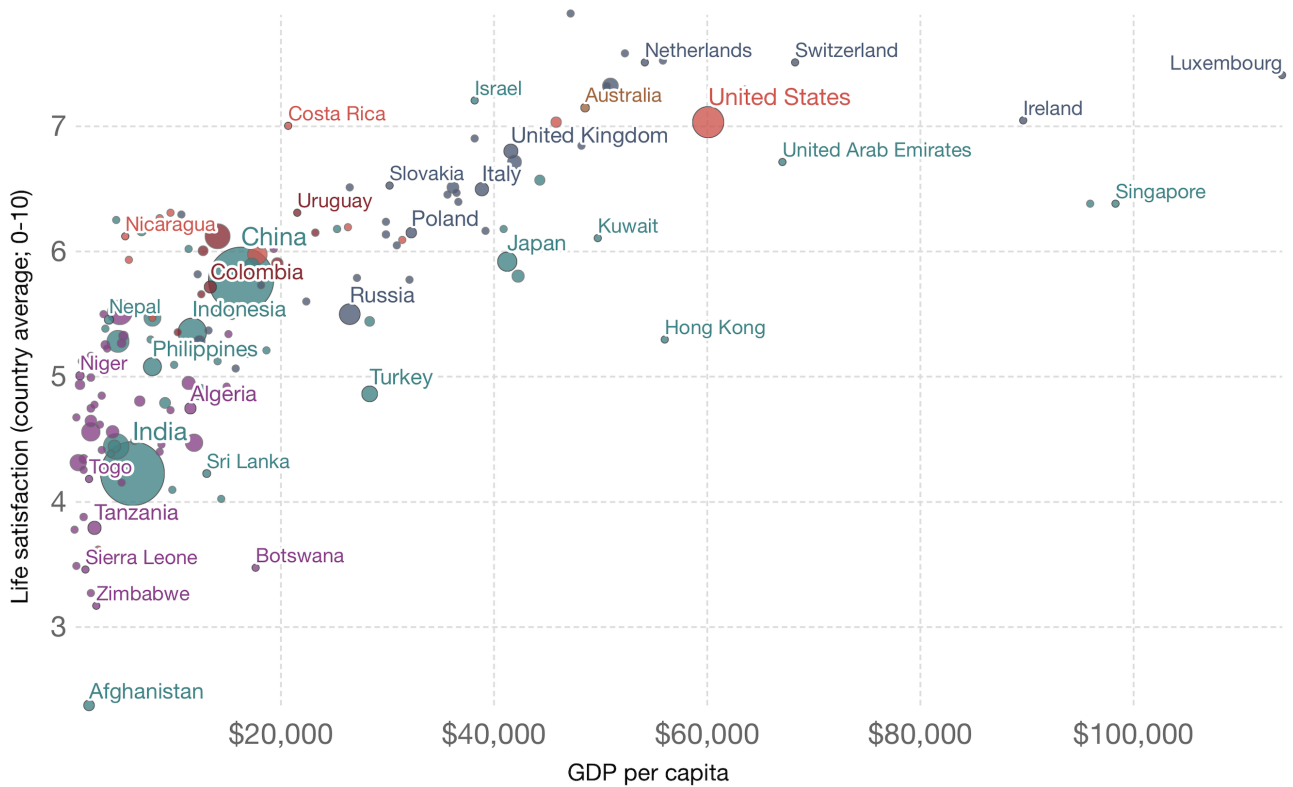
Source : FMI

FIGURE 10.3 – Espérance de vie et nombre d'enfants par instituteur en fonction du PIB par habitant dans différents pays.



Source : Banque Mondiale.

FIGURE 10.4 – Satisfaction dans la vie en 2020 en fonction du PIB par habitant en dollars.



Source : World Happiness Report 2021.

10.1.2 L'inflation

Une des variables clef de l'analyse macroéconomique est l'inflation, qui mesure la hausse du niveau général des prix. Pour mesurer l'inflation, il est nécessaire d'avoir un indice du niveau des prix. La difficulté de cette mesure tient au fait que les prix évoluent de manière très hétérogène d'un bien à l'autre.

Une première manière de mesurer l'évolution du niveau moyen des prix est d'exploiter l'écart entre le PIB nominal et le PIB réel. L'indice des prix ainsi donc donné par le **déflateur du PIB** dont l'expression est :

$$\text{Déflateur du PIB} = \frac{\text{PIB nominal}}{\text{PIB réel}} \times 100.$$

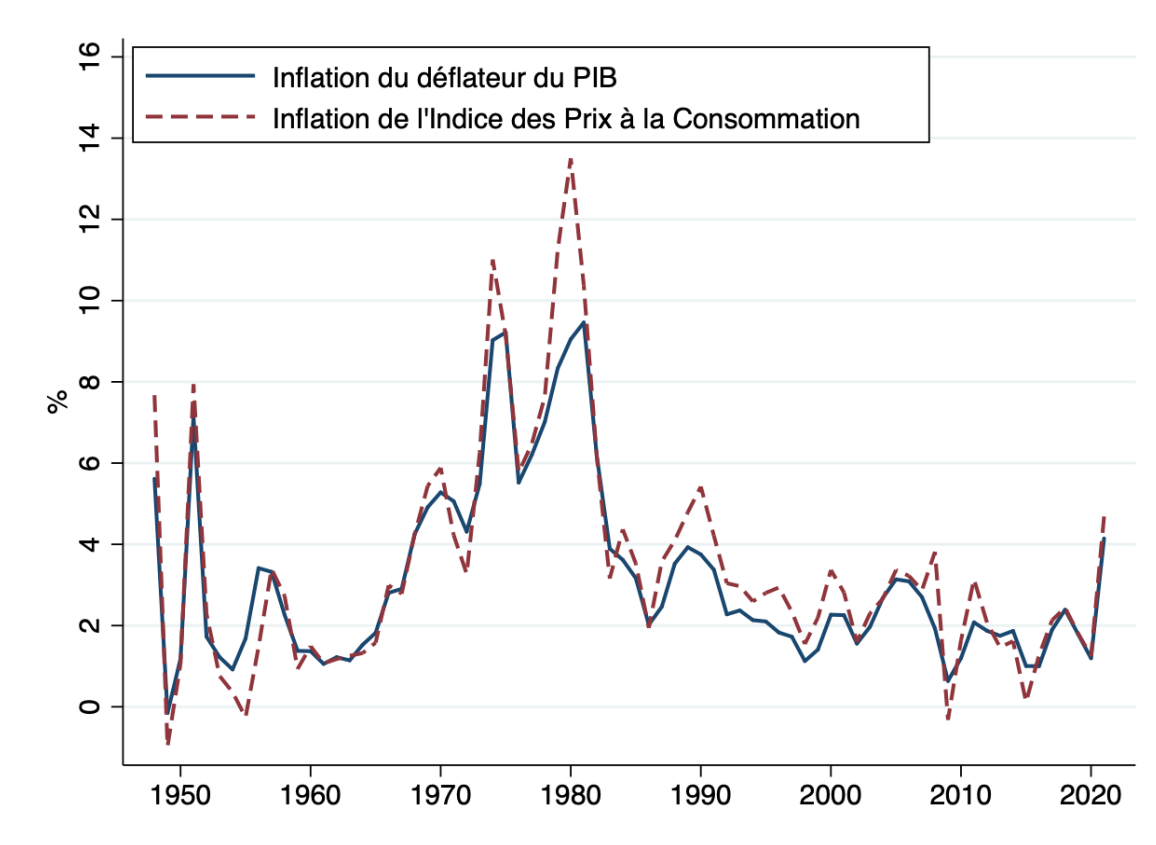
Par construction, le déflateur est donc normalisé à 100 l'année de référence utilisée pour le calcul du PIB réel. On a donc :

$$\text{Inflation} = \frac{(\text{Déflateur du PIB à l'année } t) - (\text{Déflateur du PIB à l'année } t - 1)}{\text{Déflateur du PIB à l'année } t - 1}.$$

La croissance du déflateur mesure la hausse du PIB nominal qui n'est pas dû à une augmentation du PIB réel, ce qui nous donne bien une mesure de l'inflation.

Une autre manière de mesurer le niveau des prix consiste à calculer l'**Indice des Prix à la Consommation** ou IPC (en anglais, *Consumer Price Index* ou *CPI*). Pour construire cet indice, on définit un panier représentatif constitué de biens et services de consommation, pondérés

FIGURE 10.5 – Inflation annuelle aux Etats-Unis du déflateur du PIB et de l'Indice des Prix à la Consommation.



Source : U.S. Bureau of Economic Analysis and U.S. Bureau of Labor Statistics.

par leur importance dans le budget des ménages. L'indice est normalisé à 100 pour une année de référence. On a alors :

$$\text{Inflation} = \frac{(\text{IPC à l'année } t) - (\text{IPC à l'année } t - 1)}{\text{IPC à l'année } t - 1}.$$

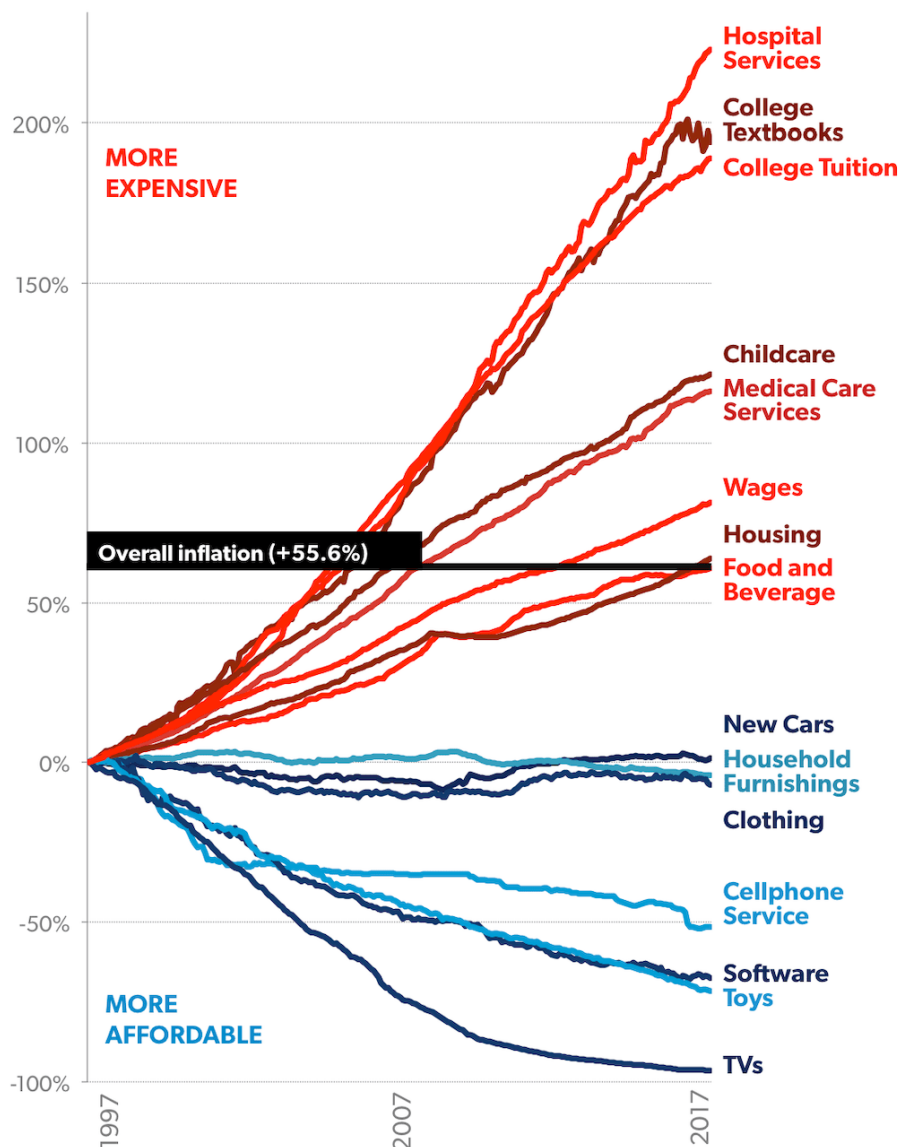
Les deux indices aboutissent généralement à des mesures très proches du taux d'inflation, comme le montre la figure 10.5. Il existe néanmoins quelques différences entre les deux : seul le déflateur du PIB prend en compte les prix des biens d'investissement, ceux achetés par l'Etat, et ceux qui sont exportés. À l'inverse, seul l'indice des prix à la consommation prend en compte les prix des biens importés. En pratique, les banques centrales (dont nous reparlerons) utilisent plutôt l'indice des prix à la consommation car il est calculé bien plus fréquemment que le déflateur du PIB.

Par ailleurs, un indicateur suivi de très près par les banques centrales est celui de l'**inflation sous-jacente** (en anglais, *core inflation*). Il s'agit d'une mesure de l'inflation basée sur un indice des prix à la consommation qui exclut l'énergie et la nourriture. Les prix des matières premières énergétiques et agricoles sont en effet soumis à de fortes fluctuations à court terme qui résultent davantage de facteurs internationaux que de l'activité économique nationale. La politique monétaire tente donc de ne pas surréagir à ces fluctuations relativement indépendantes de l'activité économique nationale, d'où son usage de l'indicateur d'inflation

sous-jacente.

La figure 10.6 illustre dans le cas des Etats-Unis l'énorme hétérogénéité de l'évolution des prix des différents biens et services. L'éducation et la santé sont devenus de plus en plus chers, alors que les prix de l'électroménager n'ont cessé de diminuer. Ceci étant, ces écarts de prix relatifs entre différents biens sont un phénomène microéconomique. Du point de vue de l'analyse macroéconomique qui est l'objet de ce chapitre, c'est bien le niveau moyen de l'inflation qui nous intéresse, car c'est lui qui est lié aux fluctuations de l'activité économique et à la politique monétaire que nous allons étudier.

FIGURE 10.6 – Evolution des prix de différents biens de consommation et des salaires aux Etats-Unis.



Source : American Enterprise Institute and Bureau of Labor Statistics.

10.1.3 Les revenus

Intéressons nous maintenant aux revenus à l'échelle de l'ensemble de l'économie. Ils se décomposent en trois catégories :

- Les revenus du **travail** s'élèvent à WL , où W est le salaire réel moyen et L l'offre de travail agrégée, soit le nombre total d'heures de travail effectuées par l'ensemble des travailleurs du pays.
- Les revenus du **capital** s'élèvent à RK , où R est le rendement brut (réel) du capital, c'est à dire ce qu'une entreprise paye pour la location d'une unité de capital à ses propriétaires, tandis que K est le stock agrégé de capital.
- Les **profits économiques** des entreprises, que les ménages possèdent. Ceux-ci s'élèvent à $\Pi = Y - WL - RK$ où Y est le PIB réel. Les profits correspondent à la différence entre la valeur de la production Y (mesurée en unités de biens) et les dépenses de salaire WL et de location du capital RK des entreprises.¹

Profits économiques vs. profits comptables.

Il est important de bien distinguer le *profit économique* du *profit comptable*. Les deux coïncident, et s'élèvent à Π , lorsque les ménages sont propriétaires du capital (machines, usines...), ce qui implique que les entreprises louent ce capital auprès des ménages. Lorsque, à l'inverse, tout le capital est possédé par les entreprises, alors celles-ci réalisent un profit comptable de $\Pi + RK$, tandis que le profit économique reste par définition égal à Π . De manière plus générale, le profit économique d'une entreprise est défini comme étant égal à la différence entre son profit comptable et le coût d'opportunité du capital qu'elle possède et qu'elle utilise dans son processus de production.

Prenons l'exemple d'une compagnie aérienne : si le marché du transport aérien est compétitif, et que la compagnie est locataire des avions qu'elle exploite, ses profits comptables sont nuls. Au contraire, si la compagnie est propriétaire des avions, alors les profits comptables incluent le rendement du capital RK , soit les revenus générés par la location à elle-même de ses propres avions. Dans les faits, les entreprises sont toujours propriétaires d'une partie du capital qu'elles exploitent.

De manière générale, si l'économie est parfaitement compétitive avec libre entrée, alors les profits économiques doivent être nuls. A l'inverse, si les profits économiques sont positifs, c'est parce que la concurrence doit être imparfaite. Les profits économiques, aussi appelés « surprofits », sont donc une mesure du pouvoir de marché des entreprises.

L'analyse macroéconomique repose sur des identités de comptabilité nationale. Il s'agit de relations entre variables qui doivent nécessairement être satisfaites à l'échelle de l'économie dans son ensemble. Une première identité de comptabilité nationale concerne le lien entre production et revenus.

1. Si l'on s'intéressait à une entreprise isolée, il faudrait aussi prendre en compte ce qu'elle achète à ses fournisseurs : néanmoins, le prix de ces intrants correspond à des salaires et des dépenses de capital en amont de la chaîne de valeur. À l'échelle de l'économie entière, on a donc bien $\Pi = Y - WL - RK$.

A l'échelle macroéconomique, la production est toujours égale aux revenus, soit

$$Y = WL + RK + \Pi.$$

La somme des revenus générés par une économie doit nécessairement être rigoureusement identique à sa production. La raison est simple : chaque bien ou service final produit doit nécessairement appartenir à quelqu'un, ce qui en fait une source de revenus. Les biens et services sont soit donnés aux travailleurs sous forme de salaire, soit aux propriétaires du capital sous forme de rente, soit aux actionnaires des entreprises ayant produit ces biens sous forme de profits.

Deux statistiques importantes sont la **part du travail** et la **part du capital** dans l'économie. La part du travail correspond à la fraction des revenus versés sous forme de salaires. La part du capital correspond à la fraction des revenus versés aux propriétaires du capital loué aux entreprises ou aux propriétaires des entreprises elle-mêmes. Un fait empirique remarquable est que, ces parts sont restées à peu près constantes depuis la révolution industrielle avec :

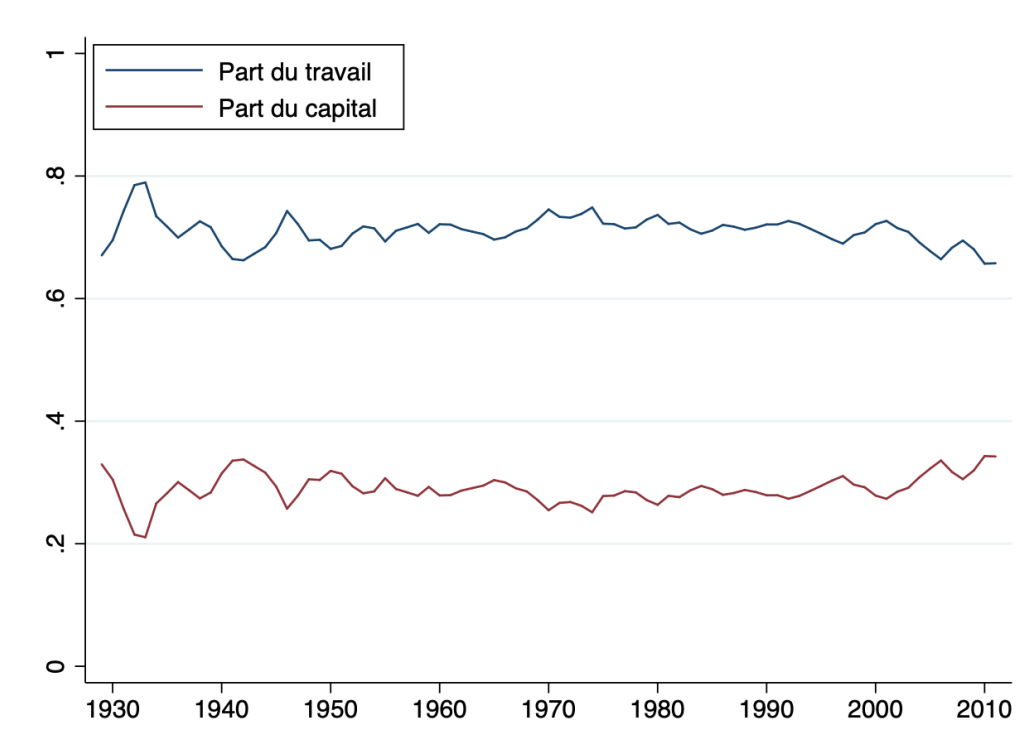
$$\begin{aligned} \text{Part du travail} &= \frac{WL}{Y} \approx \frac{2}{3} \\ \text{Part du capital} &= \frac{RK + \Pi}{Y} \approx \frac{1}{3} \end{aligned}$$

Cela signifie que les deux tiers de la valeur produite dans l'économie revient aux travailleurs. Le tiers restant constitue la rémunération des propriétaires de capital. La figure 10.7 illustre la stabilité de la part du travail et du capital dans l'économie américaine depuis près d'un siècle.

Aux États-Unis, depuis le tournant du siècle, la part du travail a sensiblement diminué. De nombreux travaux en macroéconomie cherchent à expliquer cette rupture de tendance. Une hypothèse privilégiée est l'affaiblissement de la concurrence qui aurait engendré une hausse des profits économiques Π , ce qui aurait réduit d'autant la part du travail, mais sans augmenter les revenus « normaux » du capital RK .

Il convient de contraster le cas d'une économie fermée, qui n'échange pas avec le reste du monde, de celui d'une économie ouverte, qui échange des biens et des actifs financiers avec d'autres pays. En économie fermée, la production est égale à la somme des revenus de tous les ménages de l'économie. En revanche, en économie ouverte, les ménages résidant dans le pays peuvent toucher des revenus du capital qu'ils possèdent à l'étranger, tandis que certains revenus générés par l'économie nationale sont versés à des propriétaires du capital national qui résident à l'étranger. Ainsi, la production est toujours égale à la somme des revenus générés par l'économie nationale, mais pas nécessairement aux revenus des ménages résidants au sein du pays.

FIGURE 10.7 – Part du travail et du capital aux Etats-Unis.



Source : Piketty, T. and Saez, E. (2003), 'Income Inequality in the United States, 1913-1998', *Quarterly Journal of Economics*, 118(1), 1-39; updated data.

10.1.4 Les dépenses

Intéressons nous désormais aux dépenses agrégées de l'économie. Celles-ci sont composées :

- De la consommation des ménages (C);
- De l'investissement des entreprises (I);
- Des dépenses gouvernementales (G);
- Des exportations nettes des importations (NX).

Notre seconde identité de comptabilité nationale concerne le lien entre production et dépenses.

À l'échelle macroéconomique, la production est toujours égale aux dépenses, soit

$$Y = C + I + G + NX.$$

La somme des dépenses de l'économie est identique à sa production. La raison est simple : chaque bien ou service final produit dans l'économie doit nécessairement être acheté par quelqu'un. Les biens et services sont soit achetés par des consommateurs, soit par des entreprises qui investissent afin d'accroître leur stock de capital, soit par le gouvernement, ou encore par l'étranger.

Concernant ce dernier point, les exportations nettes NX sont égales à la différence entre les exportations X et les importations IM . On peut donc écrire notre identité de comptabilité nationale $Y + IM = C + I + G + X$, ce qui met en évidence le fait que certains biens consommés au sein du pays par les consommateurs, les entreprises et le gouvernement sont importés, tandis que certains biens produits dans le pays sont achetés par des étrangers et donc exportés.

Que se passe-t-il lorsque la production est légèrement supérieure aux dépenses? Les biens périssables non-vendus disparaissent et ne font donc pas partie du PIB réel Y . Les biens non-vendus qui ne sont pas périssables sont stockés par leurs producteurs. Ils constituent des inventaires qui sont une composante de l'investissement I des entreprises.

La table 10.2 donne la décomposition du PIB des États-Unis entre les différentes catégories de dépenses – les chiffres sont très similaires dans les autres pays riches et industrialisés. La consommation représente la plus grosse part des dépenses. L'investissement est quantitativement moins important, mais beaucoup plus volatil et joue donc un rôle majeur dans les fluctuations économiques. L'investissement comporte trois composantes : l'investissement non-résidentiel, qui correspond aux investissements des entreprises (par exemple à l'achat de nouvelles machines); l'investissement résidentiel, constitué des dépenses de construction et de rénovation de nos logements; et enfin les inventaires, qui correspondent aux variations du stock de produits finis qui résultent de l'écart entre la production des entreprises et leurs ventes. Les dépenses gouvernementales sont constituées de l'achat de biens et de services par le gouvernement. Elles incluent donc les dépenses militaires, d'éducation ou de santé. En revanche, elles n'incluent pas les transferts, qui correspondent à une redistribution des ressources entre ménages : des riches vers les pauvres, des actifs vers les retraités, ou encore des employés vers les chômeurs. Enfin, à long terme, les exportations et les importations sont à peu près égales, ce qui explique la faible part des exportations nettes dans le PIB.

TABLE 10.2 – Décomposition du PIB américain (1947-2012)

	Part du PIB
Consommation	65.0%
Investissement	15.9%
Non résidentiel	10.6%
Résidentiel	4.6%
Inventaires	0.6%
Dépenses gouvernementales	20.3%
Exportations nettes	-1.1%
Exportations	7.9%
Importations	-9.0%

Quel est le lien entre le capital K , qui détermine certains revenus, et l'investissement I , qui est un des principaux composants des dépenses agrégées? Souvenez-vous que, conformément à ce que nous avons vu au chapitre précédent, les deux sont liés par l'équation d'accumulation du capital :

$$K_{t+1} = K_t - \delta K_t + I_t,$$

Où δ est le taux de dépréciation du capital. Le *stock* de capital correspond donc à l'accumulation des *flux* d'investissements passés, net de la dépréciation du stock de capital.

10.1.5 Trois mesures du PIB

Grâce à nos deux identités de comptabilité nationale, nous savons que la production, les revenus dégagés par l'économie nationale et les dépenses sont nécessairement identiques :

$$Y = WL + RK + \Pi = C + I + G + NX.$$

L'intuition économique derrière cette relation est simple : chaque unité de bien produit doit nécessairement appartenir à quelqu'un (première égalité) et doit être achetée par quelqu'un (seconde égalité).

Cela nous donne trois manières de mesurer le PIB. La première consiste à cumuler la valeur ajoutée par l'ensemble des entreprises du pays afin de calculer directement le niveau de la production ; la seconde consiste à additionner l'ensemble des revenus ; et la troisième consiste à agréger les dépenses nationales ajustées du niveau des exportations nettes. En pratique, les instituts de statistique publique obtiennent des petits écarts entre ces trois évaluations du PIB à cause d'erreurs de mesures.

10.2 L'analyse macroéconomique

Durant les années 1930, les États-Unis, suivis d'une bonne partie du reste du monde, ont été confrontés à la pire crise économique des temps modernes : la Grande Dépression. Le taux de chômage américain est passé de 3,2% en 1929 à 24,9% en 1933 ; tandis que le PIB réel s'est contracté de 30% au cours de ces quatre années. Pire, cette dépression était persistante, à tel point que le chômage est resté supérieur à 15% et le PIB inférieur à son niveau de 1929 pratiquement jusqu'au début de la Seconde Guerre mondiale.

Comment expliquer une telle crise alors que l'économie américaine n'avait pas perdu sa capacité à produire ? Les chômeurs ne demandaient qu'à travailler, les usines qu'à fonctionner, et les savoir-faire technologiques auraient dû alimenter une croissance robuste et durable de l'activité économique. La situation semblait inexplicable.

En 1936, le célèbre économiste britannique John Maynard Keynes publia la *Théorie générale de l'emploi, de l'intérêt et de la monnaie* (Keynes, 1936) dans laquelle il parvint à expliquer cette dépression économique grâce à deux idées révolutionnaires. Ce faisant, il inventa l'analyse macroéconomique telle que nous la pratiquons depuis.

Sa première idée était que pour comprendre les fluctuations de la production Y , il ne faut pas s'intéresser aux processus de production eux-mêmes, impliquant travail et capital, mais aux dépenses des ménages C , des entreprises I et du gouvernement G . Autrement dit, il faut étudier le PIB par le prisme de notre identité de comptabilité nationale selon laquelle, dans une économie fermée (où les importations et les exportations sont nulles) nous avons

l'égalité :

$$Y = C + I + G.$$

Par conséquent, la dépression de Y doit nécessairement provenir d'une dépression de C , de I ou de G . L'hypothèse d'une économie fermée est pertinente dans le cas d'une grande économie comme les Etats-Unis où une crise aussi sévère que la Grande Dépression ne peut pas principalement s'expliquer par des causes externes.

La seconde idée de Keynes était qu'à court terme, la quantité de biens produits – l'offre – est déterminée par le niveau des dépenses – la demande. L'idée est que, les prix étant rigides, c'est l'offre qui s'ajuste pour coïncider avec la demande lorsque cette dernière fluctue. Par exemple, si un constructeur automobile lance un nouveau modèle de voiture, il fixe un prix de vente, fait la publicité de son modèle, et attend ensuite les commandes. Si le succès commercial est au rendez-vous, les usines tournent jour et nuit. Sinon, la production reste faible. De même, un restaurant décide du niveau de ses prix qu'il imprime sur son menu. Ensuite, le nombre de repas qu'il produit dépend du nombre de clients qui se présentent à ce restaurant. À plus long terme, si les clients se font trop rares, le constructeur automobile et le restaurateur baisseront leurs prix. Néanmoins, cet ajustement des prix ne se fait pas à court terme, et ce sont donc bien les variations de la demande qui causent les variations de la production.

Ces deux idées simples amènent Keynes à la conclusion suivante.

En présence de prix fixes ou rigides, le niveau de la production est déterminée par la *demande* de consommation des ménages, par la *demande* d'investissement des entreprises et par la *demande* de dépenses gouvernementales.

Les dépressions économiques sont donc dues à une insuffisance de la demande agrégée! Si l'économie produit moins de richesses qu'elle n'est capable de le faire, c'est parce que ni les ménages, ni les entreprises, ni le gouvernement ne sont disposés à acheter la totalité des richesses qui pourraient être produites.

Que dire de cette demande agrégée? Tout d'abord, elle subit des chocs, et est donc volatile. Parfois, l'optimisme domine : les banques prêtent facilement, les ménages consomment et les entreprises investissent. A d'autres moments, le pessimisme l'emporte et la demande agrégée est déprimée. Par ailleurs, elle dépend du taux d'intérêt réel.

La demande agrégée est une fonction décroissante du taux d'intérêt réel.

Lorsque le taux d'intérêt réel augmente, nous avons vu au chapitre précédent que les ménages épargnaient davantage, et dépensaient donc moins. De même, une hausse du taux d'intérêt réel dissuade les entreprises d'emprunter pour investir. Ainsi, comme la demande de consommation des ménages et la demande d'investissement des entreprises sont toutes deux décroissantes du taux d'intérêt réel, la demande agrégée est bien une fonction décroissante du taux d'intérêt réel.

Une récession, soit une contraction du PIB réel, s'explique par une demande agrégée trop faible, et donc par un taux d'intérêt réel trop élevé. Soulignons que la hausse du chômage qui accompagne généralement les récessions résulte d'un phénomène d'équilibre général provoqué par l'interaction entre le marché financier, le marché des biens et le marché du travail. Plus précisément, le taux d'intérêt réel est trop élevé sur le marché financier, ce qui déprime la demande de biens des consommateurs et des entreprises, ce qui dissuade les entreprises d'embaucher. En effet, lorsque les carnets de commandes se vident, les entreprises ont tendance à licencier leurs salariés. Le chômage peut donc être le fruit d'un phénomène d'équilibre général, qui ne s'explique pas exclusivement par un dysfonctionnement du marché du travail. L'analyse keynésienne est donc fondamentalement différente des approches microéconomiques que nous avons vues jusqu'à présent.

Mais comment le taux d'intérêt réel peut-il être trop élevé? Souvenez-vous que le taux d'intérêt réel r est égal à la différence entre le taux d'intérêt nominal i et l'inflation π :

$$r = i - \pi$$

À court terme, nous avons vu que les prix étaient rigides. L'inflation π l'est donc également. Par conséquent, le taux d'intérêt réel est déterminé par le taux d'intérêt nominal i , lequel est lui-même déterminé par la banque centrale au travers de sa politique monétaire. Avant de nous pencher sur la politique monétaire, arrêtons nous sur quelques implications paradoxales de l'analyse keynésienne.

10.3 Paradoxes de la macroéconomie keynésienne

L'analyse macroéconomique est divisée en deux grandes écoles de pensées :

- La **macroéconomie keynésienne** qui suppose que les prix, les salaires et le taux d'intérêt réel sont rigides.
- La **macroéconomie néoclassique** qui suppose que les prix, les salaires et le taux d'intérêt sont parfaitement flexibles et que les marchés sont donc toujours à l'équilibre.

On considère habituellement que les prix sont rigides à court terme, qu'ils s'ajustent progressivement dans le moyen terme et qu'ils sont complètement ajustés dans le long terme. L'approche keynésienne, que nous avons exposée dans la section précédente, est donc largement dominante pour étudier le court et le moyen terme, notamment au sein des banques centrales. Ceci étant, pour bien comprendre sa portée, il est utile de comparer les mécanismes néoclassiques et les mécanismes keynésiens. Dans l'analyse qui suit, nous supposons que, dans l'économie keynésienne, le taux d'intérêt nominal i est fixe ce qui implique que le taux d'intérêt réel $r = i - \pi$ l'est également; tandis que dans l'économie néoclassique le taux d'intérêt réel s'ajuste toujours instantanément.

10.3.1 Le paradoxe de l'épargne

Supposons que les ménages veulent soudainement épargner davantage. Dans une économie néoclassique, l'augmentation de l'offre d'épargne des ménages diminue le taux d'intérêt réel, ce qui encourage les entreprises à investir davantage, conformément à ce que nous avons vu dans le chapitre précédent. L'augmentation de l'épargne accroît donc l'accumulation de capital. En outre, la baisse du taux d'intérêt réel encourage la consommation et modère l'accroissement de l'épargne.

Dans une économie keynésienne, une hausse de la propension à épargner n'entraîne pas de diminution du taux d'intérêt réel. Son principal effet est de contracter la demande de consommation des ménages. Mais, si les ménages consomment moins, la demande d'investissement des entreprises diminue également. En effet, les entreprises n'ont pas intérêt à augmenter leurs capacités de production au moment où leurs carnets de commande se vident. Seule une baisse du taux d'intérêt réel aurait permis de contrecarrer cet effet en encourageant les ménages à consommer et les entreprises à investir.

Par conséquent, dans une économie keynésienne, la hausse de la propension à épargner implique une *diminution* de l'investissement et donc une *réduction* du stock de capital. Or, la richesse accumulée par les ménages correspond précisément au stock de capital. Paradoxalement, la hausse de la propension à épargner diminue le volume de l'épargne accumulée par les ménages ! En réalité, les ménages souhaitent épargner davantage, mais dans une économie plus petite du fait de la contraction du PIB que ce désir d'épargne engendre. Le volume de l'épargne diminue donc. Ce paradoxe de l'épargne montre que les raisonnements keynésiens sont fondamentalement différents des raisonnements néoclassiques.

10.3.2 Le paradoxe de la flexibilité

Il semblerait que la principale source d'inefficacité de l'économie keynésienne soit la rigidité des prix. Pour s'en assurer, supposons une économie keynésienne déprimée à cause d'une insuffisance de la demande. Quel est alors l'impact d'une hausse de la flexibilité des prix ?

Si les prix sont plus flexibles, alors, face à une insuffisance de la demande, ils baissent davantage. Cela diminue le taux d'inflation π , ce qui augmente le taux d'intérêt réel $r = i - \pi$. La demande de consommation des ménages et d'investissement des entreprises se contracte davantage et l'économie est donc encore plus déprimée ! C'est la paradoxe de la flexibilité. Le résultat est le même si les salaires sont rigides, car les prix reflètent alors le niveau des salaires.

Si l'insuffisance de la demande est due à un surendettement des ménages, alors la baisse des prix et des salaires augmente la valeur réelle de la dette, ce qui réduit encore plus la demande. Le surendettement renforce donc le paradoxe de la flexibilité.

Comme nous l'avons vu dans la section précédente, le chômage keynésien est un phénomène d'équilibre général : le taux d'intérêt réel est trop élevé sur les marchés financiers, ce qui déprime la demande sur le marché des biens, ce qui dissuade les entreprises d'embaucher sur le marché du travail. Le fond du problème n'est donc pas le niveau excessivement

élevé des prix et des salaires, mais celui du taux d'intérêt réel.

Le paradoxe de la flexibilité montre que le problème soulevé par l'économie keynésienne est si profond qu'il ne se corrige pas naturellement par l'ajustement progressif du niveau des prix ou des salaires. La politique monétaire a donc un rôle crucial à jouer afin de fixer le taux d'intérêt nominal de manière à maintenir l'économie au plein emploi.

10.4 Le rôle de la monnaie

Avant d'étudier comment la banque centrale peut utiliser la politique monétaire pour stabiliser l'activité économique, tâchons dans cette section de caractériser l'impact de la monnaie sur différentes variables de l'économie.

10.4.1 La monnaie

La monnaie remplit trois fonctions. C'est :

- Un **moyen d'échange** : sur un marché, les biens peuvent être échangés contre de la monnaie. Cela surmonte la difficulté du troc qui requiert une double coïncidence des besoins.
- Une **unité de compte** : en déterminant l'unité dans laquelle sont exprimés les prix (en l'occurrence, les euros), la monnaie permet de mesurer et de comparer la valeur des biens et des services.
- Une **réserve de valeur** : la monnaie peut être un support d'épargne puisqu'elle conserve l'essentiel de sa valeur au cours du temps.

Dans la plupart des pays riches, la monnaie prend la forme de pièces et de billets qui sont imprimés, et donc émis, par la **banque centrale**. Aujourd'hui, outre les pièces et les billets, les banques centrales émettent de la monnaie sous forme électronique en créditant des comptes bancaires. Ces sommes dématérialisées peuvent être échangées contre des pièces ou des billets auprès de la banque centrale.

L'État peut, au travers de sa banque centrale, modifier la masse monétaire, ce qui constitue un des principaux leviers de la politique macroéconomique.

Les supports de monnaie.

Auparavant, des métaux précieux, à commencer par l'or et l'argent, étaient utilisés comme monnaie. Ces métaux avaient l'avantage d'être non-périssables, ductiles, fongibles, faciles à transporter, relativement rares, et surtout brillants! Ils remplissaient donc toutes les fonctions attendues d'une monnaie.

Au travers de l'histoire, d'autres objets que les pièces métalliques ont pu faire figure de monnaie. Depuis des temps anciens, des coquillages ont rempli cette fonction. Au XVII^{ème}, dans la Nouvelle-France (le Canada français), des fourrures de castor étaient

utilisés comme monnaie. Aujourd'hui, dans les prisons, où les détenus n'ont souvent le droit de détenir ni pièces, ni billets, d'autres objets non-périssables servent de monnaie : des cigarettes, des timbres poste, ou même des sachets de ramen lyophilisés.

Des billets en papier ont progressivement été utilisés en parallèle des métaux précieux. En Chine, des billets ont été émis dès la fin du premier millénaire, et Marco Polo les décrit dans ses récits. En Europe, des billets ont été utilisés de façon intermittente depuis le XV^{ème} siècle. Dans la grande majorité des cas, ces premiers billets ne constituaient qu'un moyen simplifié pour transporter de l'or ou de l'argent : un billet correspondait alors à une somme de métal précieux que l'on pouvait récupérer sur présentation du billet, à l'image des 20 dollars de la figure 10.8.

En présence d'une monnaie métallique, il est pratiquement impossible pour l'État de mener une politique monétaire, puisque les variations de l'offre de monnaie sont déterminées par les quantités d'or ou d'argent extraites des sous-sols à travers le monde. Durant le XIX^{ème} et une bonne partie du XX^{ème} siècle, la convertibilité de la monnaie en or ou en argent impliquait que les chocs économiques ne pouvaient pas être contre-carrés par la politique monétaire, ce qui a engendré de fortes fluctuations de l'activité économique.

En 1971, sous la présidence de Nixon, les États-Unis décidèrent d'abandonner la convertibilité du dollar en or : le gouvernement américain pouvait alors émettre autant de dollars qu'il le désirait, indépendamment de ses réserves d'or. Le dollar est donc aujourd'hui une monnaie dite « fiduciaire », dans le sens où un billet de vingt dollars n'a aucune valeur intrinsèque. Les monnaies fiduciaires n'ont que la valeur qui leur est communément admise (*fiducia* signifie « confiance » en latin). L'abandon de l'étalon or a considérablement augmenté la puissance des banques centrales et leur capacité à gérer les fluctuations de l'activité. Malgré la persistance d'une certaine nostalgie pour la monnaie or, rares sont les économistes qui plaident pour un retour à ce système qui implique l'abandon de toute politique monétaire.

FIGURE 10.8 – Un billet américain échangeable contre de l'or (*gold certificate*), émis en 1905. Il est inscrit sur le billet : « Ceci certifie qu'il a été déposé au département du Trésor des États-Unis vingt dollars en pièces d'or, payables au porteur à la demande ».



10.4.2 Le rôle de la monnaie à long terme : la théorie quantitative de la monnaie

Dans l'analyse macroéconomique, il est important de bien distinguer les variables réelles des variables nominales. Les variables réelles se rapportent à des unités de biens : le PIB réel, la consommation des ménages, l'investissement des entreprises, le stock de capital accumulé, les salaires réels (en unités de biens) ou encore le taux d'intérêt réel. Les variables nominales se rapportent quant à elles à des unités de monnaie : les prix, les salaires nominaux (en euros), l'inflation, ou encore le taux d'intérêt nominal.

On considère habituellement qu'à long terme, la monnaie est **neutre**, c'est à dire que la quantité totale de monnaie n'a aucun impact sur les variables réelles. La prospérité d'une économie est déterminée par son stock de capital, par l'ardeur au travail de ses habitants, et par leurs connaissances techniques et technologiques, mais pas par la quantité de monnaie en circulation. Si la banque centrale double la masse monétaire, à terme, cela double le niveau des prix sans rien changer à la prospérité de l'économie.

Cette idée est formalisée par la **théorie quantitative de la monnaie**, dont l'intuition fut évoquée dès le XVI^{ème} siècle par Nicolas Copernic, avant d'être développée par David Hume au XVIII^{ème} siècle, et surtout par Milton Friedman dans les années 1950 et 1960 qui en fit le cheval de bataille des monétaristes de l'École de Chicago. Le point de départ de cette théorie est le concept de *vélocité de la monnaie*, dénoté par V , qui correspond à la vitesse à laquelle la monnaie circule dans l'économie. Soit M la masse monétaire, la quantité de monnaie émise par la banque centrale. Si on suppose la quantité de transactions (en euros) égale au PIB nominal PY , on a

$$V = \frac{PY}{M}.$$

Ainsi, si le PIB nominal est de 1000 milliards d'euros et la masse monétaire égale à 200 milliards d'euros, alors chaque euro réalise en moyenne 5 transactions chaque année. En réalité, la quantité de transactions n'est pas exactement égale au PIB nominal : certaines transactions, comme la vente d'une voiture d'occasion ou d'un appartement, requièrent de la monnaie sans contribuer au PIB, puisqu'elles n'impliquent pas de production nouvelle. Ceci étant, cela n'affecte notre expression de la *vélocité de la monnaie* qu'à un facteur de proportionnalité près, sans modifier fondamentalement l'analyse qui suit.

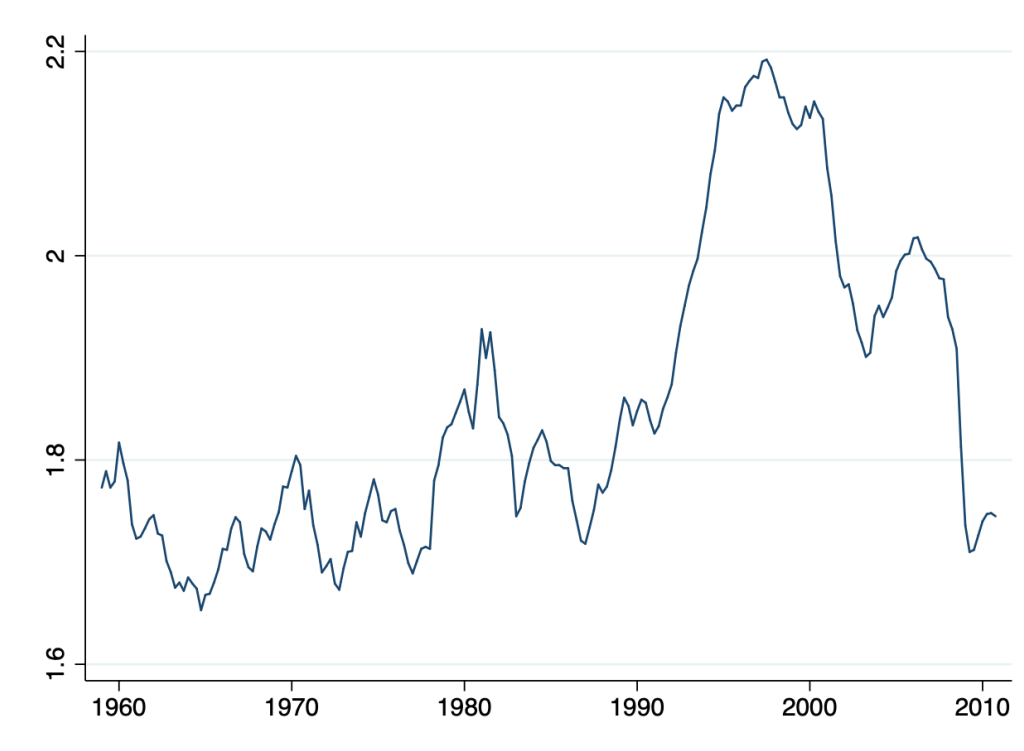
La théorie quantitative de la monnaie part du constat que la *vélocité de la monnaie* est stable et exogène, c'est à dire indépendante de la politique économique et, notamment, de la politique monétaire, ce qui semble cohérent avec la figure 10.9. Seules des innovations technologiques affectant les moyens de paiement, comme l'invention de la carte de crédit, du distributeur de billets ou encore du paiement sans contact, sont de nature à modifier la *vélocité de la monnaie*, qui est donc relativement indépendante de la politique monétaire.

Or, notre définition de la *vélocité de la monnaie* implique que

$$P = \frac{MV}{Y}.$$

Si la *vélocité* V est exogène et pratiquement constante, on parvient à la conclusion suivante

FIGURE 10.9 – Vélocité de la monnaie aux Etats-Unis.



Source : Federal Reserve Bank of St. Louis.

qui constitue le coeur de la théorie quantitative.

Le taux d'inflation est égal à la différence entre le taux de croissance de la masse monétaire et le taux de croissance du PIB réel.

En outre, si l'on considère qu'à long terme la monnaie est neutre, alors le PIB réel Y est également exogène. Cela implique que le niveau des prix P est proportionnel à la masse monétaire M . La seule raison pour laquelle un pays peut être confronté à une inflation trop élevée est un rythme trop rapide de la croissance de la masse monétaire. Milton Friedman résumait ce diagnostic par sa célèbre formule :

L'inflation est partout et toujours un phénomène monétaire!

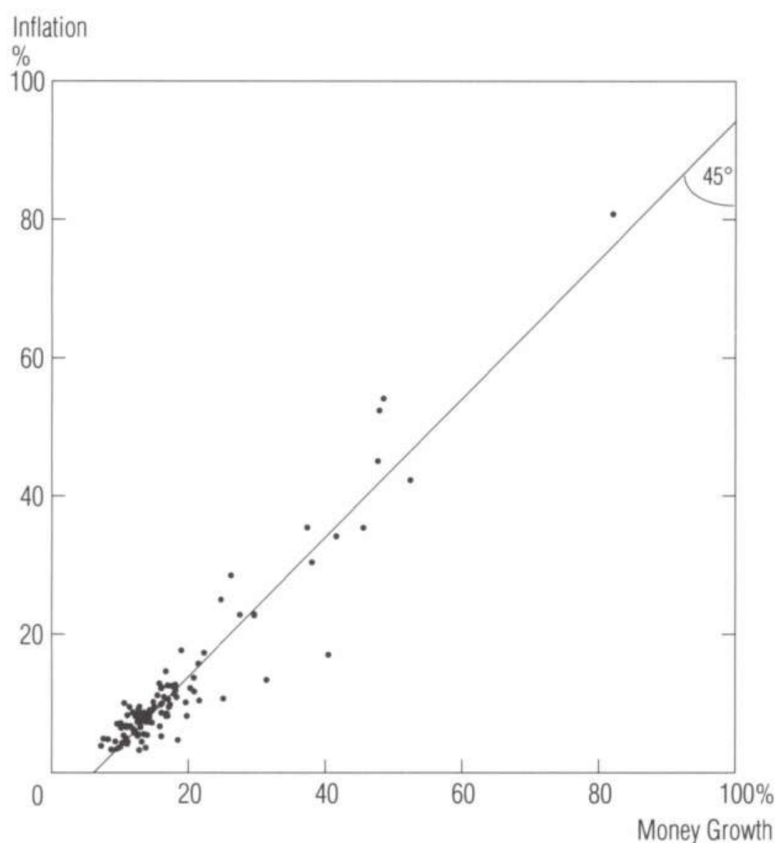
La figure 10.10 montre que le taux de croissance de la masse monétaire semble effectivement être le principal déterminant de l'inflation à long-terme.

L'hyperinflation.

En Allemagne, entre août 1922 et novembre 1923, le niveau des prix a été multiplié par dix milliards. Pire, en Hongrie, entre août 1945 et juillet 1946, ce chiffre a atteint 3,8 billions de milliards, soit 3.8×10^{27} . Plus récemment, le Zimbabwe a subi un épisode d'inflation massive en 2008 et le Venezuela en 2016.

Comme on le voit sur la figure 10.11, l'hyperinflation allemande s'est accompagnée

FIGURE 10.10 – Taux d'inflation vs taux de croissance annuel de la masse monétaire entre 1960 et 1990 dans 110 pays.



Source : G.T. McCandless and Weber, W.E. (1995), 'Some Monetary Facts', *Quarterly Review of the Federal Reserve Bank of Minneapolis*, 19(3), 2-11.

d'une forte augmentation de la masse monétaire, ce qui est cohérent avec la théorie quantitative de la monnaie. A l'issue de la première guerre mondiale, le gouvernement allemand, très affaibli financièrement, peinait à financer ses dépenses avec ses seules recettes fiscales. L'État a alors commencé à régler ses factures avec des billets fraîchement imprimés. Cette augmentation de la masse monétaire a alimenté l'inflation, ce qui a augmenté les dépenses (en termes nominaux) du gouvernement, qui a imprimé encore plus de billets pour les financer, etc. Rapidement, la situation est devenue incontrôlable. L'hyperinflation n'a cessé qu'avec la création d'une nouvelle devise.

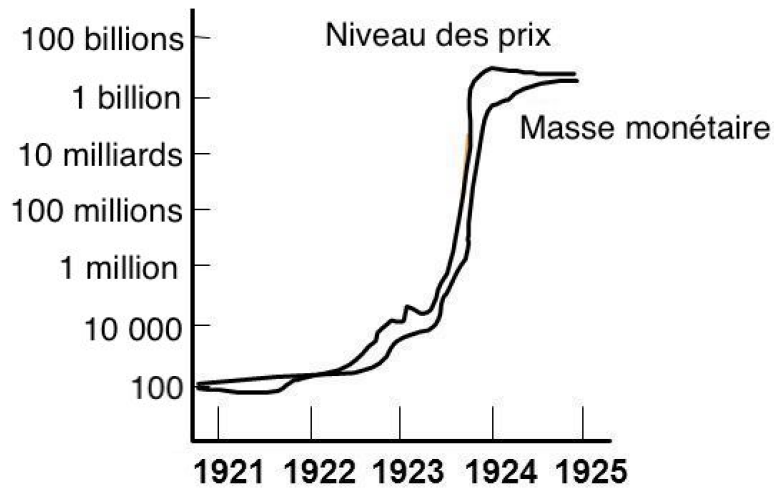
La recette de l'hyperinflation est simple et ne requiert que deux ingrédients :

- Un déficit budgétaire élevé ;
- L'absence d'épargnants disposés à prêter à l'État, qui n'a donc pas d'autre choix que d'imprimer de la monnaie pour financer ses dépenses.

Pour éviter de se retrouver dans une telle situation, il y a dans la plupart des pays une séparation claire entre le ministère des finances, responsable des comptes publics, et la banque centrale, responsable du niveau des prix. L'objectif de cette séparation est d'empêcher d'avoir recours à la politique monétaire pour financer des dépenses

publiques.

FIGURE 10.11 – Masse monétaire et niveau des prix (normalisés à 100 en janvier 1921) en Allemagne de 1921 à 1925.



Source : Hall, R. E. (2009), *Inflation : causes and effects*.

10.4.3 Le rôle de la monnaie à court terme : la théorie de la préférence pour la liquidité

Nous avons vu qu'à court terme, les prix et donc l'inflation π peuvent être considérés comme étant fixes ou rigides. Le taux d'intérêt réel $r = i - \pi$ est donc déterminé par le taux d'intérêt nominal i . Tâchons donc de caractériser l'impact de la masse monétaire sur le taux d'intérêt nominal, et donc réel. Pour cela on va s'appuyer sur la **théorie de la préférence pour la liquidité**.

La demande de monnaie. Les ménages peuvent détenir leur épargne sous forme de monnaie ou d'obligations. Chacun de ces deux actifs présentent des avantages et des inconvénients. La monnaie peut être utilisée pour réaliser des transactions, mais rapporte un intérêt nominal nul, puisqu'un billet de 100 euros aujourd'hui vaudra toujours 100 euros dans un an. Soulignons que la monnaie correspond soit à des pièces et à des billets, soit à des liquidités sur un compte bancaire qui ne rapportent rien, mais qui peuvent être utilisées pour effectuer des paiements par chèque, par virement ou avec une carte de débit. Les obligations rapportent quant à elles un intérêt nominal de i par an, puisque 100 euros investis dans des obligations donneront $100(1 + i)$ euros dans un an. En revanche, les obligations ne peuvent pas être utilisés pour effectuer des transactions. Vous ne pouvez pas acheter une baguette de pain, ni même une voiture, avec des obligations. Le règlement doit nécessairement se faire avec de la monnaie. Comment les ménages décident-ils donc de répartir leur épargne entre ces deux actifs ?

Un ménage souhaite détenir de la monnaie jusqu'à ce que le bénéfice marginal d'un euro supplémentaire soit égal au coût marginal de cet euro. Si un ménage dispose de peu

de monnaie, il doit constamment aller à la banque vendre des obligations afin d'avoir les liquidités nécessaires pour effectuer ses transactions. Dans ce cas, un euro supplémentaire de monnaie lui sera très utile. En revanche, si un ménage possède déjà beaucoup de monnaie, il est peu probable qu'un euro supplémentaire fasse une grosse différence dans sa capacité à régler ses achats. On suppose donc que *le bénéfice marginal de la monnaie est une fonction décroissante de la quantité de monnaie dont le ménage dispose.*

Le coût d'opportunité de la monnaie est égal au taux d'intérêt nominal : détenir un euro sous forme de monnaie, c'est renoncer à l'opportunité de recevoir un intérêt en détenant cet euro sous forme d'obligations. *Le coût marginal de la monnaie est donc égal au taux d'intérêt nominal.* Plus le taux d'intérêt est élevé, plus il est coûteux de posséder de la monnaie plutôt que des obligations. Par conséquent, la demande de monnaie est une fonction décroissante du taux d'intérêt nominal. Une hausse du taux d'intérêt nominal augmente le coût d'opportunité de la détention de monnaie : les ménages choisissent donc d'accroître le bénéfice marginal de la monnaie et d'en détenir moins.

En réalité, si les ménages souhaitent faire fructifier leur épargne, il peuvent acquérir bien d'autres actifs que des obligations, comme par exemple des actions ou des biens immobiliers. Ceci étant, ces autres actifs rapportent un intérêt et ne peuvent pas être utilisés pour effectuer des transactions. Notre analyse de la demande de monnaie demeure inchangée : inutile d'ajouter à notre théorie de la préférence pour la liquidité des variables qui complexifieraient la théorie sans avoir aucun impact sur les résultats qui nous intéressent.

L'offre de monnaie. La quantité de monnaie M dans l'économie, aussi appelée masse monétaire, est déterminée par la banque centrale, qui contrôle donc l'offre de monnaie. En pratique, comment est-ce que la banque centrale ajuste la masse monétaire ? En intervenant directement sur les marchés financiers où elle achète et vend des obligations : on parle d'opération d'*open market*. Pour augmenter la masse monétaire, la banque centrale achète des obligations avec de la monnaie nouvellement imprimée. Pour réduire la masse monétaire, il lui suffit de vendre ses obligations, ce qui retire de la circulation la monnaie ainsi récupérée. Pour un euro de monnaie en circulation, la banque centrale détient donc un euro d'obligation. La banque centrale reçoit des intérêts sur son portefeuille sous forme de monnaie : ces revenus sont transférés au ministère des finances, qui peut librement dépenser cette monnaie. Les recettes fiscales engendrées par la création monétaire sont appelées **seignuriage**.

Une banque centrale peut-elle faire faillite ?

On parle parfois du risque de faillite des banques centrales. Mais comment est-ce qu'une institution en charge d'imprimer la monnaie peut-elle faire faillite ? En réalité, il ne s'agit pas d'une faillite comme pour une entreprise.

Le rôle de la banque centrale est de contrôler la masse monétaire. Si elle effectue de mauvais placements, par exemple en achetant des obligations qui perdent de leur

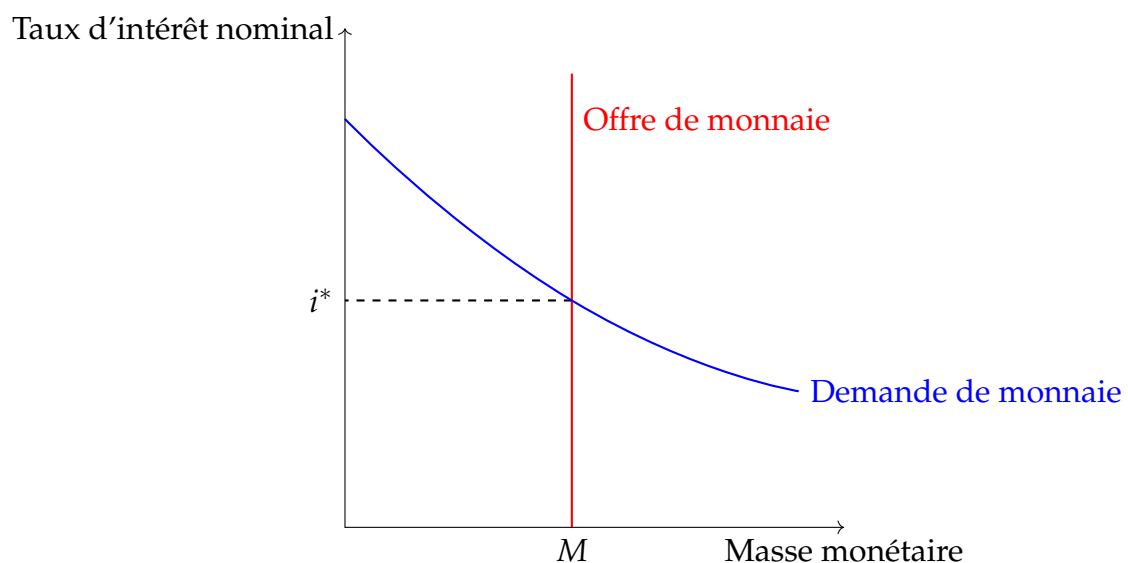
valeur, alors la quantité de monnaie dans l'économie devient supérieure à la taille de son portefeuille. La banque centrale perd donc sa capacité à diminuer la masse monétaire en vendant ses obligations. Dans ces circonstances, face à une accélération de l'inflation, la banque centrale se retrouve démunie.

La seule solution est de *recapitaliser* la banque centrale. Pour cela, le ministère des finances doit lever des impôts afin de financer l'achat d'obligations, qui sont transférées à la banque centrale, qui peut alors les vendre afin de réduire la masse monétaire. Cette opération revient à lever des impôts et à brûler la monnaie ainsi récupérée ! C'est le prix à payer pour diminuer la masse monétaire afin d'éviter une inflation trop élevée.

Durant la crise de la dette européenne de 2010 à 2013, l'Allemagne était très réticente à ce que la Banque Centrale Européenne achète des obligations grecques car elle craignait que la BCE perde alors le contrôle de la masse monétaire. La Grèce a fini par restructurer sa dette auprès de certains de ses créanciers, mais pas auprès de la BCE qui a ainsi évité d'enregistrer une perte !

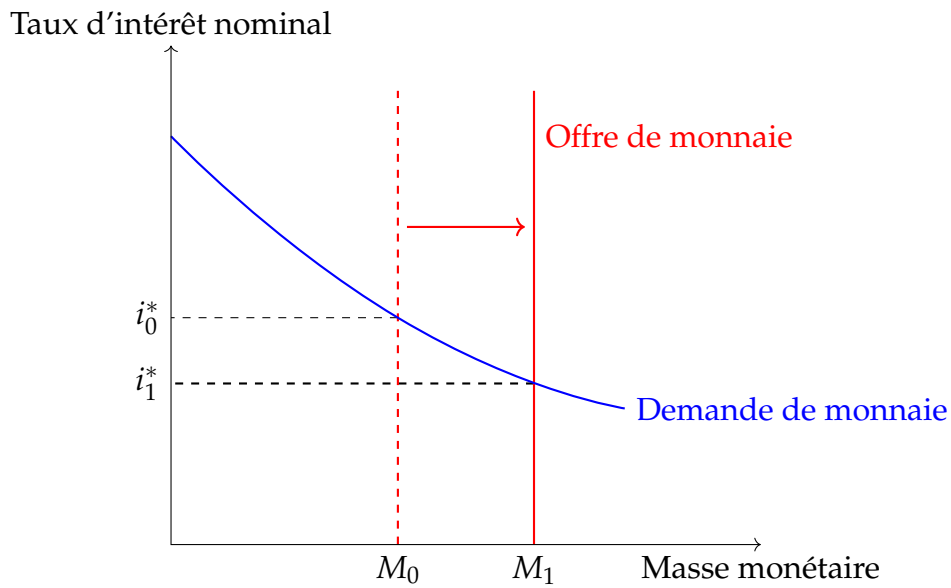
L'équilibre sur le marché de la monnaie. L'offre et la demande de monnaie sont représentées sur la figure 10.12. À l'équilibre, le taux d'intérêt nominal i^* égalise offre et demande de monnaie. Lorsque la banque centrale augmente la masse monétaire, comme sur la figure 10.13, le taux d'intérêt nominal baisse. Au contraire, une contraction monétaire provoque une augmentation du taux d'intérêt nominal.

FIGURE 10.12 – Offre et demande de monnaie.



Plus précisément, comment est-ce qu'une augmentation de la masse monétaire diminue le taux d'intérêt nominal ? Soit P le prix d'une obligation simple qui rapporte en tout et pour tout €1 dans un an. Ce prix étant la valeur actualisée d'un euro versé dans un an, le taux

FIGURE 10.13 – Impact d’une augmentation de la masse monétaire sur le taux d’intérêt nominal.



d’intérêt nominal correspondant i est tel que

$$P = \frac{1}{1 + i}$$

Le taux d’intérêt nominal vaut donc

$$i = \frac{1}{P} - 1$$

Lorsque la banque centrale augmente la masse monétaire, elle achète des obligations, ce qui augmente leur prix et diminue mécaniquement le taux d’intérêt nominal. Le taux d’intérêt nominal diminue jusqu’à ce que l’augmentation de la demande de monnaie des ménages soit égale à l’accroissement d’offre de monnaie de la banque centrale, ce qui rétablit l’équilibre sur le marché de la monnaie.

Soulignons qu’en contrôlant la masse monétaire, la banque centrale peut fixer le taux d’intérêt nominal au niveau de son choix. En pratique, les banques centrales conduisent leur politique monétaire en fixant le taux d’intérêt nominal plutôt que la masse monétaire. Il leur suffit pour cela d’ajuster la masse monétaire en fonction des fluctuations de la demande de monnaie, de manière à garder un taux nominal constant. Cela revient à fixer le prix des obligations. Par exemple, dans le cas de notre obligation simple, si la banque centrale est disposée à en acheter ou à en vendre une quantité infinie à 99 centimes, alors cela fixe le prix de l’obligation à 99 centimes, et donc le taux d’intérêt nominal à $1/0.99 - 1 \approx 1\%$. En effet, personne ne vendra cette obligation à moins de 99 centimes, sachant que la banque centrale est prête à l’acheter à ce prix, et personne ne l’achètera à plus de 99 centimes, sachant que la banque centrale la vend à ce prix. C’est bien le contrôle de la masse monétaire qui permet à la banque centrale de fixer le prix d’une obligation, et donc le taux d’intérêt nominal.

Impact d'un changement de la masse monétaire Nous disposons désormais de tous les ingrédients pour déterminer les conséquences d'un changement d'offre de monnaie de la banque centrale. Une hausse de la masse monétaire diminue le taux d'intérêt nominal i . Le niveau des prix et l'inflation π étant fixes ou rigides à court terme, le taux d'intérêt réel $r = i - \pi$ diminue. Cela stimule la demande de consommation des ménages et la demande d'investissement des entreprises. Face à cette hausse de la demande, les entreprises produisent davantage pour y répondre. Ainsi, une hausse de la masse monétaire stimule l'activité économique et augmente le PIB réel. Il s'agit là d'une *politique monétaire expansionniste*. De la même manière, une diminution de la masse monétaire réduit l'activité économique, ce qui correspond à une *politique monétaire restrictive*.

Ce mécanisme est loin d'être trivial. Il permet de comprendre comment une augmentation du nombre de billets imprimés par la banque centrale peut accroître, à court et moyen terme, la prospérité d'une économie. Autrement dit, un choc purement nominal peut avoir un impact réel sur l'économie! Déterminons maintenant comment la banque centrale peut utiliser ce mécanisme afin de stabiliser l'activité économique pour maintenir le plein emploi.

10.5 Politique monétaire optimale

Dans un premier temps, il est essentiel de déterminer le niveau optimal de l'activité économique. Nous verrons ensuite comment utiliser la politique monétaire afin de s'en rapprocher le plus possible.

Le niveau naturel de la production. Si tous les prix, les salaires et les taux d'intérêt étaient parfaitement flexibles, alors la production serait à un niveau efficace. Il s'agit là d'une conséquence du premier théorème du bien-être que nous avons vu au chapitre 3. Ce niveau théorique est appelé le **niveau naturel de la production**. En effet, dans cette situation, toutes les transactions mutuellement bénéfiques se réalisent. Certes, il peut y avoir des défaillances de marché, mais on considère habituellement que ce n'est pas le rôle de la politique monétaire que de chercher à y remédier. On peut donc faire abstraction de ces défaillances et considérer que le niveau naturel de la production est efficace.

Conformément à l'analyse macroéconomique keynésienne, nous savons qu'à court terme les prix sont fixes et le niveau de la production est déterminée par le niveau de la demande agrégée. Or la politique monétaire contrôle le taux d'intérêt nominal et donc la demande agrégée. Le cap de la politique monétaire est donc simple à établir.

La politique monétaire optimale consiste à maintenir la demande agrégée égale au niveau naturel de la production.

Lorsque la demande agrégée est inférieure au niveau naturel de production, l'économie est déprimée. L'insuffisance de la demande de biens et de services se traduit par une insuffisance de la demande de travail des entreprises. Ainsi, les dépressions économiques sont

typiquement caractérisées par un niveau de chômage élevé. Pour restaurer le plein emploi, la banque centrale doit augmenter la masse monétaire afin de réduire le taux d'intérêt nominal, et donc réel, ce qui stimule la demande agrégée.

A l'inverse, il est possible que la demande agrégée soit supérieure au niveau naturel de la production. En ce cas le PIB réel est trop élevé! Il s'agit d'une surchauffe de l'économie. La forte demande de biens entraîne une demande vigoureuse de travail des entreprises. A court terme, les salaires sont fixes et les employés travaillent plus qu'ils ne le souhaiteraient au regard de leurs rémunérations. Les prix et les salaires finissent par s'ajuster progressivement. La surchauffe de l'économie est typiquement une situation qui peut amorcer une spirale inflationniste des prix et des salaires : les travailleurs demandent des hausses de salaires que les entreprises répercutent sur le niveau des prix afin de maintenir leurs marges, ce qui incite les travailleurs à demander des salaires encore plus élevés, etc. Pour éviter ce scénario, la banque centrale doit réduire la masse monétaire afin d'augmenter le taux d'intérêt nominal, et donc réel, ce qui réduit la demande agrégée.

On considère généralement que le *plein emploi* est atteint dès lors que le PIB est à son niveau naturel. Le taux de chômage n'y est pas nul du fait des frictions qui caractérisent le processus de rencontre entre offre et demande de travail. Lorsque l'économie est déprimée, le chômage est trop élevé; mais lorsque l'économie est en surchauffe, le chômage est trop faible et les entreprises peinent à recruter pour répondre à la demande de leurs consommateurs.

Le taux d'intérêt réel qui est compatible avec le niveau naturel de la production est celui qui égalise l'offre et la demande de crédit, tel que nous l'avons vu au début du chapitre précédent. Il s'agit du **taux d'intérêt réel naturel**. Pour que la demande agrégée soit égale au niveau naturel de la production, le taux d'intérêt réel doit être égal au taux d'intérêt réel naturel. Par conséquent, si à court terme la banque centrale contrôle le taux d'intérêt nominal, et donc réel, elle ne peut pas s'éloigner du taux d'intérêt naturel sans déstabiliser l'activité économique. A moyen terme, le niveau du taux d'intérêt reste donc fondamentalement déterminé par l'offre et la demande de crédit, tel que nous l'avons vu au chapitre précédent.

Le ciblage de l'inflation. En pratique, le niveau naturel de production et le taux d'intérêt réel naturel sont très difficiles à mesurer. Comment savoir alors si la demande agrégée est trop forte ou trop faible? Il suffit de regarder le niveau d'inflation. Comme nous venons de le voir, lorsque la demande agrégée est trop forte, l'inflation a tendance à augmenter. À l'inverse, lorsque la demande est insuffisante, le niveau élevé du chômage exerce une pression à la baisse sur les salaires et, donc, sur les prix. L'inflation a alors tendance à diminuer.

L'évolution de l'inflation reflète donc l'écart entre le PIB réel (égal à la demande agrégée) et le niveau naturel de production. La politique monétaire optimale consiste donc à cibler un taux d'inflation constant. Dès lors que l'inflation augmente au dessus de sa cible, la banque centrale doit augmenter le taux d'intérêt nominal afin de réduire la demande agrégée. A contrario, lorsque l'inflation diminue en dessous de sa cible, la banque centrale doit réduire le taux d'intérêt nominal afin de stimuler la demande agrégée. La stabilisation de l'inflation au niveau de sa cible coïncide avec la stabilisation du PIB réel à son niveau naturel!

Aujourd'hui, la plupart des grandes banques centrales ont pour mandat de cibler un taux d'inflation constant. Pour la BCE, la Réserve Fédérale américaine, la Banque d'Angleterre et la Banque du Japon, cette cible est égale à 2%.²

On reproche parfois aux banques centrales de ne se soucier que de l'inflation, et pas assez de la croissance et de l'emploi. En réalité, la stabilisation de l'inflation coïncide avec la stabilisation de la croissance et de l'emploi à leurs niveaux efficaces. Si le chômage est durablement trop élevé malgré une inflation égale à la cible de la banque centrale, alors ce n'est pas dû à une insuffisance de la demande agrégée, mais à un dysfonctionnement du marché du travail contre lequel la politique monétaire ne peut rien.

Dans les années 1970, les banques centrales ont surestimé le niveau naturel de production, ce qui les a conduit à mener des politiques monétaires excessivement expansionnistes. Cela a abouti à la *stagflation*, combinaison d'une stagnation économique et d'un niveau d'inflation élevé. Ce n'est qu'à partir des années 1980 que les mécanismes que nous venons de voir furent suffisamment bien compris pour être intégrés à la conduite des politiques monétaires, ce qui conduisit progressivement à ce que le ciblage de l'inflation devienne la boussole des banques centrales. Cette nouvelle approche fut un franc succès : rarement la croissance, l'inflation et le chômage ne furent aussi stables, comme on peut le constater sur la figure 10.14. Les récessions étaient peu nombreuses et de courtes durée, l'inflation était proche de sa cible de 2%, tandis que le chômage était faible et peu volatil. On a appelé cette période de l'histoire macroéconomique, qui a débuté au milieu des années 1980, la « grande modération ». L'ironie est que cette « grande modération » s'est brutalement terminée en 2008, avec la pire crise économique depuis la Grande Dépression des années 1930.

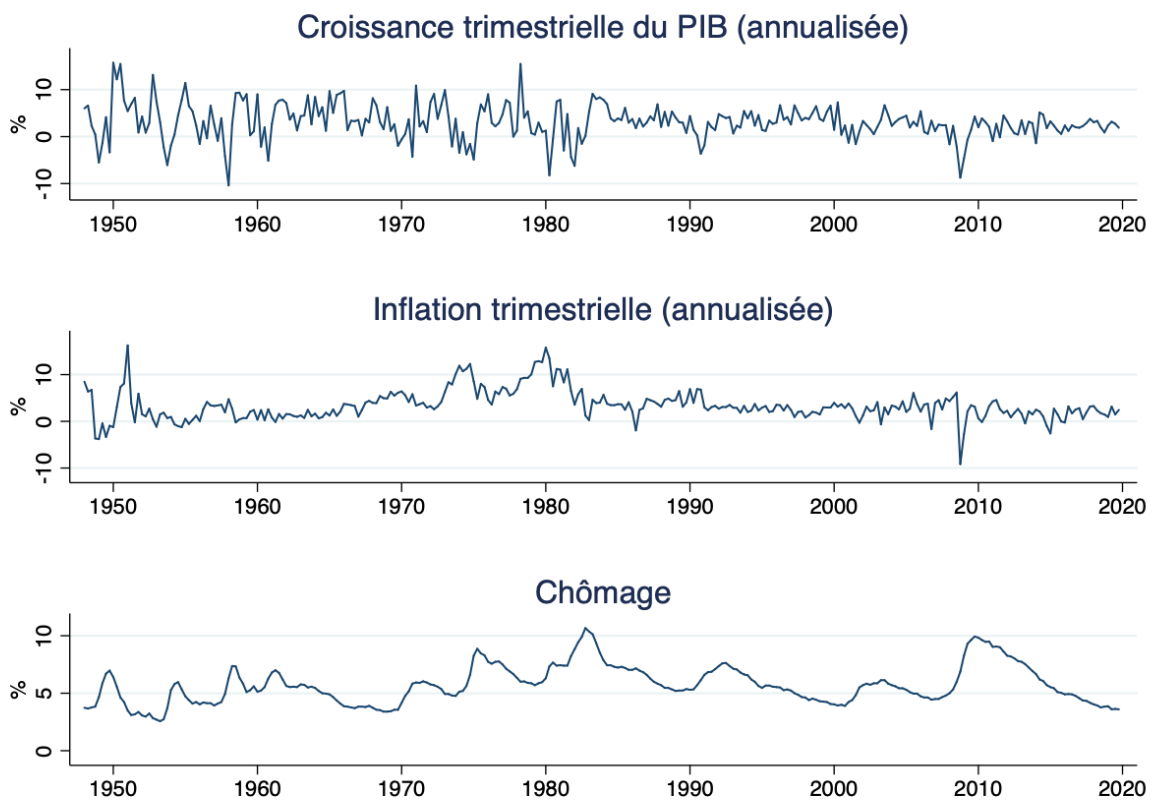
10.6 La Grande Récession de 2008

La trappe à liquidité. La Grande Récession de 2008 fut si violente qu'elle poussa la politique monétaire à bout, donnant alors un rôle à jouer à la politique budgétaire. L'éclatement de la bulle immobilière américaine et le resserrement du crédit ont provoqué une forte contraction de la demande agrégée. Les nombreux ménages qui vivaient à crédit ont dû se serrer la ceinture, tandis que les entreprises ont fortement diminué leurs investissements face à la crise à venir. La Réserve Fédérale a réagi en diminuant drastiquement le taux d'intérêt nominal, passé de 5,25% en juillet 2007 à 0% fin 2008. Malgré cette réaction rapide, l'inflation est restée durablement inférieure à sa cible de 2%, ce qui caractérise une insuffisance persistante de la demande agrégée. Un taux d'intérêt de 0% peut sembler très faible mais, en réalité, il était trop élevé au regard de l'ampleur de la dépression de la demande. Le Royaume-Uni et la zone Euro ont été confrontés à des situations similaires.

Pourquoi les banques centrales n'ont-elles pas baissé leurs taux d'intérêt davantage, à -1%, -2%, ou -3%? Tout simplement parce qu'il existe un actif financier qui rapporte un

2. Nous avons vu qu'au niveau naturel de production, l'inflation devait être constante, mais nous n'avons pas déterminé à quel niveau. En réalité, l'important est que l'inflation soit égale à la cible de la banque centrale, quelle que soit cette cible (1%, 2%, 4%...). Pour le comprendre, il faut faire intervenir le rôle des anticipations inflationnistes des ménages et des entreprises, ce qui va bien au delà de ce chapitre d'introduction à la macroéconomie.

FIGURE 10.14 – La grande modération aux États-Unis : du milieu des années 1980 à 2008, la croissance économique, l’inflation et le chômage furent bien moins volatiles que dans les décennies précédentes



Source : U.S. Bureau of Economic Analysis and U.S. Bureau of Labor Statistics.

rendement de 0% : la monnaie. Personne ne souhaite payer plus de 100 euros pour acheter une obligation qui rapportera 100 euros dans un an ; il est préférable de garder de la monnaie sous son matelas ! L’existence des billets implique donc une borne inférieure à 0% (en anglais, *Zero Lower Bound*) sur le taux d’intérêt nominal.³

A défaut de réduire le taux d’intérêt nominal, on pourrait espérer qu’un accroissement de la masse monétaire augmente l’inflation, ce qui réduirait le taux d’intérêt réel $r = i - \pi$ et stimulerait la demande agrégée. Malheureusement, il n’en est rien. En temps normal, la monnaie n’est utilisée que pour réaliser des transactions, conformément à la théorie de la préférence pour la liquidité. Mais, une fois le taux d’intérêt égal à 0%, la monnaie devient un support d’épargne attractif. Par conséquent, lorsque la banque centrale augmente la masse monétaire en achetant des obligations, au travers une opération d’*open market*, les épargnants qui vendent des obligations utilisent la monnaie qu’ils reçoivent comme support d’épargne. Monnaie et obligations sont désormais deux actifs qui rapportent tous deux 0%.

3. Dans certains pays, le taux d’intérêt nominal est parfois légèrement négatif. Cela est dû au coût de stockage de la monnaie, qui peut brûler ou être volée. Le rendement de la monnaie n’est donc pas nul, mais légèrement négatif. Il n’en reste pas moins que ce rendement de la monnaie fixe une borne inférieure en dessous duquel le taux d’intérêt nominal ne peut pas descendre. On parle parfois de *Effective Lower Bound* plutôt que de *Zero Lower Bound*.

L'accroissement de la masse monétaire perd donc sa capacité à stimuler l'inflation et l'activité économique, ce qu'on appelle la **trappe à liquidité**. En 2008, les banques centrales se sont ainsi retrouvées dans l'incapacité de relancer l'économie.

La relance budgétaire. Face à cette situation, comment éviter le scénario catastrophe de la Grande dépression des années 1930? Souvenez-vous, dans une économie fermée, la demande agrégée est égale à $C + I + G$. Une fois dans la trappe à liquidité, avec un taux d'intérêt nominal nul, les banques centrales ne peuvent relancer ni la consommation des ménages C , ni l'investissement des entreprises I . Il ne reste alors plus qu'une solution : accroître les dépenses gouvernementales G ! Pour compenser l'absence de demande du secteur privé, l'État doit dépenser davantage. Il s'agit d'une **politique budgétaire expansionniste**.

La hausse du chômage est de nature à amplifier la contraction de la demande du secteur privé, car les chômeurs sont obligés de se serrer la ceinture tandis que les employés, inquiets de la situation, ont tendance par précaution à épargner davantage. Les dépenses gouvernementales limitent la hausse du chômage, ce qui atténue la contraction de la demande. Ainsi, dans la trappe à liquidité où le taux d'intérêt nominal est fixe à 0%, un euro de dépenses gouvernementales augmente le PIB réel de plus d'un euro. C'est ce qu'on appelle le *multiplieur budgétaire*.

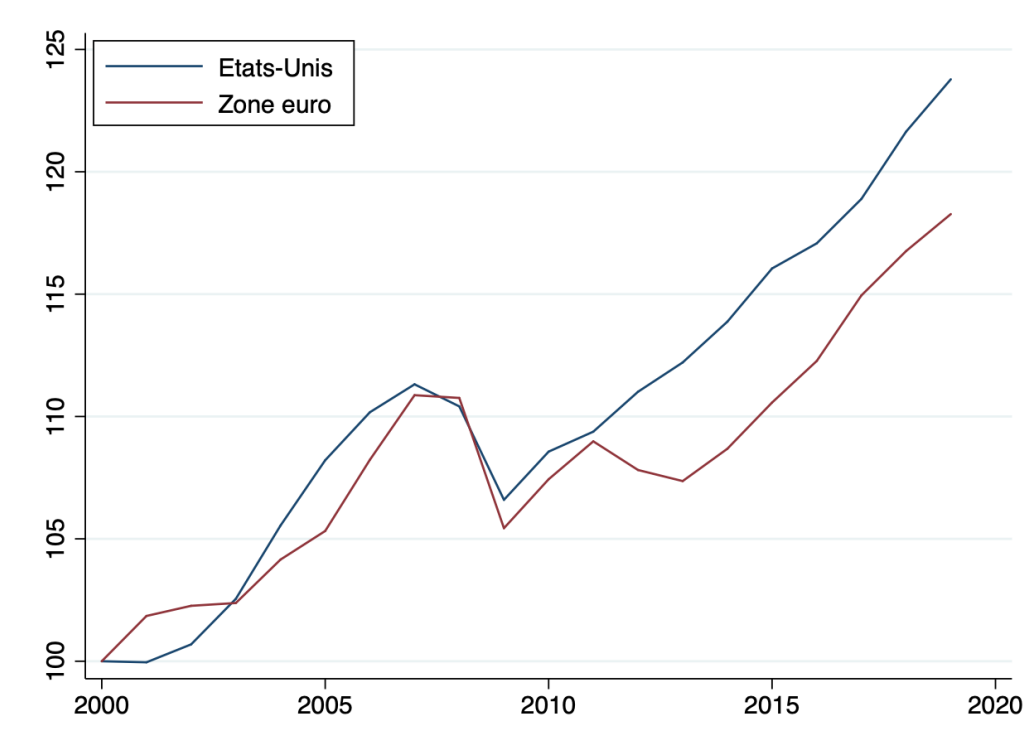
Dès son arrivée à la Maison Blanche en janvier 2009, Barack Obama a mis place un gigantesque plan de relance de 800 milliards de dollars : l'American Recovery and Reinvestment Act. Les impôts ont été temporairement baissés, les allocations chômage augmentées, les dépenses de santé et d'éducation accrues et des investissements ont été réalisés dans les infrastructures.⁴

Cette politique a engendré un déficit budgétaire de plus de 10% du PIB en 2009. Cela peut paraître totalement irresponsable. En réalité, c'était un des seuls moyens d'éviter un effondrement de l'économie et une répétition de la Grande Dépression des années 1930. En zone Euro, la relance a été plus timide, avant d'être rapidement remplacée par des mesures d'austérité. La dépression macroéconomique y a donc été plus durable qu'aux États-Unis d'où est partie la crise, comme le montre le figure 10.15.

Cette Grande Récession de 2008 nous rappelle qu'on ne gère pas une économie comme on gère une entreprise. Par temps de crise, un chef d'entreprise responsable réduit ses investissements, ses embauches et son endettement afin d'éviter la faillite et de surmonter la crise. L'impact de ces décisions sur le carnet de commande de l'entreprise est négligeable. En revanche, lorsque toutes les entreprises agissent ainsi, le chômage augmente et la demande se contracte, ce qui amplifie la crise. Le rôle du gouvernement est alors d'accroître son endettement et ses dépenses afin de compenser l'insuffisance de la demande privé par un accroissement de la demande publique, ce qui permet d'éviter un effondrement de l'économie.

4. Ce plan ne doit pas être confondu avec l'Emergency Economic Stabilization Act de 2008 qui a consisté à injecter 700 milliards de dollars dans le système financier afin d'éviter une série de faillites bancaires. Ce plan a stoppé la crise financière, mais n'a pas permis d'enrayer la crise économique.

FIGURE 10.15 – PIB par habitant (normalisé à 100 en 2000).



Source : U.S. Bureau of Economic Analysis et Eurostat.

La stagnation séculaire.

L'insuffisance de la demande est habituellement perçue comme étant un phénomène transitoire, qui se corrige naturellement au fur et à mesure que les prix et les salaires s'ajustent. Cette vision est correcte dans un monde où la banque centrale ajuste son taux d'intérêt nominal afin de hâter le retour au plein emploi. En revanche, dans la trappe à liquidité, où le taux d'intérêt nominal est scotché à 0%, la paradoxe de la flexibilité que nous avons vu précédemment implique que l'ajustement des prix et des salaires ne contribue pas nécessairement au retour au plein emploi. L'insuffisance de la demande peut alors être un phénomène persistant, qu'on appelle la **stagnation séculaire**. La Grande Dépression des années 1930 ne s'est terminée que par un gigantesque stimulus budgétaire : la Seconde Guerre mondiale.

L'exemple récent du Japon nous oblige à nouveau prendre au sérieux cette possibilité. En 1995, le Japon fut le premier pays riche à tomber dans une trappe à liquidité. Depuis, le taux d'intérêt nominal est toujours égal à 0%, la croissance économique est molle et, surtout, l'inflation est en moyenne de 0%, soit nettement moins que la cible de 2% de la Banque du Japon. Il semblerait que le vieillissement démographique ait largement contribué à l'affaiblissement de la demande, les ménages âgés ayant tendance à épargner pour leur retraite, puis pour leurs dépenses de santé, pour leurs enfants et leurs petits-enfants. En outre, les entreprises sont réticentes à investir dans une économie vieillissante où elles anticipent une demande modérée dans les décennies à venir.

A bien des égards, la zone Euro depuis 2008 se trouve dans une situation très similaire au Japon depuis 1995. Les solutions au problème existent, mais le chemin est étroit. Une possibilité consiste à augmenter la cible d'inflation, tout en mettant en oeuvre une politique budgétaire très expansionniste afin d'atteindre cette cible, ce qui permet de réamorcer la pompe de l'activité économique (Michau, 2021).

Bibliographie

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money*, MacMillan Press, London.

Michau, J.-B. (2021), 'Understanding secular stagnation', *Institut Des Politique Publiques Policy Briefs* 73, 1–7.

Exercices du chapitre 10

Exercice 1 : Le Produit Intérieur Brut

Décrire l'impact de chacune des possibilités listées ci-dessous sur le PIB. Pour les sept premières, tachez de décrire le problèmes de deux points de vue différents. [Indice : Souvenez-vous que le PIB peut être calculez à partir de la production, des revenus et des dépenses.]

1. Vous allez au restaurant.
2. Une femme épouse son chauffeur, qui n'est donc plus rémunéré pour les services qu'il continue à assurer.
3. Un commerçant vend pour 600 euros des ordinateurs qu'il a importé au prix de 400 euros.
4. Vous achetez une maison.
5. Vous louez une maison.
6. Vous vivez dans votre propre maison.
7. Vous entrez sur le marché du travail et trouvez un emploi.
8. Apple lance un nouveau modèle d'iPhone, qui est commercialisé au même prix et qui se vend aussi bien que le modèle précédent.
9. Facebook et Youtube deviennent accessibles dans votre pays.

Ces problèmes de mesure du PIB sont-ils une défaillance majeure de l'analyse macro-économique ?

Exercice 2 : La trappe à liquidité

1/ Décrire l'impact d'une augmentation de la masse monétaire M sur le taux d'intérêt nominal i . Représentez graphiquement cet effet. Comment réagit le PIB ?

2/ Supposons que la banque centrale augmente la masse monétaire jusqu'au niveau \bar{M} où le taux d'intérêt nominal atteint 0%. En supposant que cette intervention ne suffise pas à faire repartir l'activité économique, que se passe-t-il si la banque centrale augmente encore davantage la masse monétaire ? Quel impact cela a-t-il sur le PIB ? [Indice : Pour répondre à cette question, tracez la courbe de demande de monnaie lorsque $M > \bar{M}$ et décrivez précisément les mécanismes économiques sous-jacents.]

Exercice 3 : Instabilité de la demande de monnaie

1/ Quel est l'impact d'une hausse de la demande de monnaie sur l'économie? Décrivez précisément les mécanismes économiques à l'oeuvre.

2/ Comment la banque centrale peut-elle protéger l'économie d'une instabilité perpétuelle de la demande de monnaie?

Exercice 4 : Ciblage de l'inflation

A long terme, quels sont les coûts et les bénéfices d'un taux d'inflation élevé? Qu'en déduisez-vous concernant la cible d'inflation de la banque centrale?

Exercice 5 : Dépenses publiques

1/ Les dépenses publiques G sont soit financées par des impôts T , soit pas le déficit budgétaire $G - T$. Supposons que la consommation est une fonction linéaire du revenu disponible des ménages (défini comme leurs revenus net d'impôts). Interprétez le coefficient correspondant à la pente de la fonction de consommation. Quelles bornes peut-on imposer à ce coefficient?

2/ Quel est l'impact sur la production d'une augmentation d'un euro des dépenses publiques, financées par un accroissement du déficit budgétaire? Expliquez le mécanisme économique sous-jacent.

3/ Quel est l'impact sur la production d'une diminution d'impôts d'un euro, financée par un accroissement du déficit budgétaire? Comparez votre résultat à celui de la question précédente.

4/ Est-il possible de stimuler l'activité économique sans accroître le déficit budgétaire?

5/ En supposant que l'économie est initialement au niveau naturel de production, quel est l'impact d'une hausse des dépenses publiques sur l'inflation and sur les taux d'intérêt? En quoi est-ce que cela modifie votre réponse à la seconde question? En quoi la situation est-elle différente en trappe à liquidité?

Exercice 6 : Dépréciation

Soit Y le Produit Intérieur Brut (annuel), K le stock de capital, WL les revenus du travail, RK les revenus du capital, Π les profits économiques pures, et δ le taux de dépréciation du capital. Souvenez-vous que la production agrégée est toujours égale aux revenus agrégés :

$$Y = WL + RK + \Pi.$$

Une fraction importante du stock de capital appartenant aux entreprises, on peut difficilement distinguer les rendements du capital RK des profits économiques purs Π . On considère donc que les revenus du capital s'élèvent à $RK + \Pi$.

1/ Le Produit Intérieur Brut ne tient pas compte de la dépréciation du capital. Il surestime donc la création de richesse dans une année. Le Produit Intérieur Net est égal au PIB net de la dépréciation. Donnez une expression pour le PIN.

2/ Quel est la relation entre le PIN et les revenus nationaux? Expliquez.

3/ Empiriquement, à long terme, la part du travail WL/Y dans le PIB est égale à $2/3$ et la part du capital $(RK + \Pi) / Y$ à $1/3$. On sait aussi que la dépréciation s'élève à environ 14% du PIB. Définissez et calculez la part du travail et du capital dans le PIN.

4/ Empiriquement, le stock de capital s'élève à environ 4 années de PIB (la moitié correspondant au capital résidentiel et l'autre moitié au capital non résidentiel). Déduisez-en la taux de dépréciation annuel pour l'économie.