

EMMANUEL BUISSON-FENET
MARION NAVARRO

La microéconomie en pratique

CURSUS



3^e édition

ARMAND COLIN

EMMANUEL BUISSON-FENET
MARION NAVARRO

La microéconomie en pratique

3^e édition

ARMAND COLIN

Conception de couverture : Hokus Pokus créations

Images de couverture : Slavica/Getty Images – © Bloomberg/Getty Images

© Armand Colin, 2012, 2018 pour la présente édition

www.armand-colin.com

Armand Colin est une marque de
Dunod Éditeur, 11 rue Paul Bert, 92240 Malakoff

ISBN : 978-2-200-62261-9

Collection CURSUS Économie

Ouvrages publiés sous la direction d'Alain Beitone

- Alain Beitone, Christophe Rodrigues, *Économie monétaire*, 2017
- Alain Beitone, Lionel Lorrain, Christophe Rodrigues, *La dissertation de science économique*, 2016
- Marion Navarro, *Les inégalités de revenus*, 2016
- Didier Marteau, *Les marchés de capitaux*, 2^e édition, 2016
- Vincent Barou et Benjamin Ting, *Fluctuations et crises économiques*, 2015
- Mickaël Joubert et Lionel Lorrain, *Économie de la mondialisation*, 2015
- Éloi Laurent et Jacques Le Cacheux, *Économie de l'environnement et économie écologique*, 2^e édition, 2015
- Marc Bassoni et Alexandre Joux, *Introduction à l'économie des médias*, 2014
- Magali Chaudey, *Analyse économique de la firme*, 2014
- Antoine Bernard de Raymond et Pierre-Marie Chauvin, *Sociologie économique. Histoire et courants contemporains*, 2014
- Denis Anne et Yannick L'Horty, *Économie de l'emploi et du chômage*, 2013
- Jean-Luc Gaffard, *La Croissance économique*, 2011

Voir aussi

- *Dictionnaire de science économique*, Alain Beitone (dir.), 2016 (5^e édition)
- *Économie sociologie et histoire du monde contemporain*, Alain Beitone (dir.), 2018 (3^e édition)

Sommaire

[Couverture](#)

[Page de titre](#)

[Page de Copyright](#)

[Introduction](#)

[1 Les marchés concurrentiels](#)

[1. Le fonctionnement des marchés concurrentiels](#)

[1.1 Qu'est-ce qu'un marché concurrentiel ?](#)

[1.2 La loi de l'offre et de la demande](#)

[1.3 Gains à l'échange et exclusion du marché](#)

[2. Comment étudier un marché concurrentiel particulier ?](#)

[2.1 Les frontières d'un marché](#)

[2.2 L'interdépendance entre les marchés](#)

[3. La régulation d'un marché concurrentiel](#)

[3.1 La régulation par les prix](#)

[3.2 La régulation par les quantités](#)

[3.3 La régulation par les normes](#)

3.4 Comment évaluer les effets d'une intervention publique sur un marché ?

2 La formation de la demande des ménages

1. Le modèle de base du consommateur rationnel

1.1 Contrainte budgétaire et prix relatif des biens

1.2 Les préférences

1.3 La maximisation de la satisfaction

1.4 Les effets d'une variation du budget

1.5 Les effets d'une variation de prix

2. Prolongements et débats autour du modèle du consommateur rationnel

2.1 Consommer aujourd'hui ou consommer demain ?

2.2 La nouvelle théorie du consommateur

2.3 Des préférences endogènes et instables ?

2.4 Des choix rationnels ?

3. De la demande individuelle à la demande agrégée

3.1 La détermination de la courbe de demande

3.2 Élasticités et typologie des biens

3.3 Fiscalité et régulation de la demande de biens

3 Le comportement des firmes

1. Le choix de la combinaison productive

1.1 Les coûts de production

1.2 Le choix de la combinaison productive

2. L'offre de la firme sur un marché concurrentiel

3. Innovations, concurrence et politiques publiques

4. Les choix organisationnels des producteurs

4.1 Les coûts de transaction et le dilemme entre l'organisation et le marché

4.2 Coordination au sein de la firme et efficacité économique

4 Les imperfections de marché

1. Le pouvoir de marché

1.1 Les barrières à l'entrée et à la sortie des marchés

1.2 La différenciation des produits

2. Comportement des entreprises et structures de marché

2.1 Le comportement des monopoles

2.2 Le comportement des oligopoles

3. Les asymétries d'information

3.1 Le risque d'antisélection

3.2 Le risque d'aléa moral

5 Biens collectifs et externalités

1. Caractéristiques des biens et fonctionnement

du marché

1.1 Les différents types de biens économiques

1.2 La production de biens collectifs, de biens communs ou de biens de club peut-elle être assurée par le marché ?

2. Externalités et décisions économiques des agents

2.1 Les caractéristiques des externalités

2.2 La défaillance des marchés en présence d'externalités

3. Quelle régulation publique en cas de défaillance du marché ?

3.1 Quelles alternatives au recours au marché, en présence de défaillances ?

3.2 Quels outils pour réguler les comportements des agents sur les marchés, en présence de défaillances ?

6 Microéconomie du marché du travail

1. Les déterminants de l'offre de travail

1.1 L'arbitrage consommation/loisir

1.2 La théorie de la prospection

2. Les déterminants du salaire

2.1 Rémunération et qualité de la main-d'œuvre

2.2 Rémunération, type d'emploi occupé et pouvoir de négociation

3. L'explication microéconomique du chômage

3.1 La question du coût du travail

[3.2 Chômage frictionnel et rôle des institutions](#)

[7 Modifier l'équilibre du marché](#)

[1. L'équilibre général concurrentiel](#)

[1.1 Une allocation optimale des ressources](#)

[1.2 Un équilibre qui peut être très inégalitaire](#)

[2. Quels critères de justice ?](#)

[2.1 L'approche « bien-être »](#)

[2.2 Les limites de l'approche bien-être](#)

[3. Introduction à la taxation optimale des revenus](#)

[3.1 La démarche de la taxation optimale](#)

[3.2 Les principaux résultats en taxation optimale](#)

[Bibliographie](#)

[Index](#)

[Listes des documents et graphiques](#)

[Listes des encadrés](#)

Introduction

Le projet d'un ouvrage sur « la microéconomie en pratique » est né de notre expérience d'enseignants et de nos frustrations de lecteurs. L'enseignement de la microéconomie à des niveaux très divers nous a souvent confronté aux réactions critiques d'étudiants, qui ne sont pas avares de reproches envers cette discipline. Trois séries d'arguments critiques sont mobilisées régulièrement : la microéconomie serait d'abord déconnectée de la réalité, avec une utilisation parcimonieuse de données empiriques, au profit de « robinsonnades » ayant un rapport très lointain avec la vie économique d'aujourd'hui. Par ailleurs, elle masquerait une assez grande pauvreté conceptuelle sous la sophistication des techniques de calcul, donnant peu de sens à son enseignement. Enfin, elle serait largement idéologique, au service d'un projet libéral visant à démontrer l'efficacité générale du marché et à remettre en cause l'intervention de l'État.

Pour tous ceux qui ont l'occasion de fréquenter la production microéconomique contemporaine, ces critiques ont de quoi surprendre. En effet, les travaux empiriques se sont multipliés depuis quelques décennies, avec l'avènement de nouvelles techniques économétriques permettant d'aborder une infinité de problèmes concrets sous l'angle microéconomique. En outre, ce domaine a fait preuve, contrairement aux idées reçues, d'une grande richesse conceptuelle et de développements rapides dans des directions nouvelles, comme le montre par exemple le dynamisme de l'économie de l'environnement. Enfin, l'économie publique a très largement contribué à démontrer la nécessité d'une régulation étatique, si bien qu'il est erroné d'identifier la microéconomie moderne à une apologie libérale du tout marché. Le prix Nobel attribué en 2014 à Jean Tirole pour ses travaux d'économie industrielle sur la réglementation des marchés fournit une bonne illustration des avancées de la microéconomie et du souci d'applications pratiques, encore ignorées du grand public et parfois des étudiants.

Si le fossé s'est creusé entre la perception de la microéconomie par un trop grand nombre d'étudiants et la fécondité de cette approche, c'est selon nous en partie lié aux supports d'enseignement utilisés, qui font parfois la part trop

belle à une microéconomie très fortement mathématisée, et pauvre à l'inverse en applications pratiques. Le présent manuel voudrait apporter un élément de réponse. La « microéconomie en pratique » propose une découverte des grands modèles de la microéconomie et du raisonnement propre à cette discipline, en s'appuyant sur quatre principes.

- Chaque chapitre puise ses exemples dans un domaine particulier, de façon à montrer au lecteur la fécondité empirique de la démarche microéconomique, mais aussi parfois ses limites. Ainsi, le chapitre 1 s'appuie sur des exemples concernant le marché du logement, le chapitre 2 se réfère aux dépenses culturelles, le chapitre 3 à la production agricole, etc. Nous faisons référence dans chaque cas à un grand nombre d'études empiriques entièrement actualisées pour la troisième édition, de manière à illustrer de façon précise et vivante les théories microéconomiques. Nous espérons que les étudiants s'approprient ainsi plus facilement des notions parfois complexes, et seront au final plus facilement convaincus de la capacité de ces théories à éclairer le monde contemporain.
- La microéconomie, comme toute approche scientifique à visée explicative, repose sur la construction de modèles et leur confrontation avec la réalité. Par définition, cette activité suppose une certaine dose de simplification, un modèle n'éclairant jamais l'ensemble des contours du phénomène étudié. Si la microéconomie recourt intensément à la formalisation mathématique, cette dernière n'est pas nécessaire pour comprendre les concepts élaborés par les économistes. Au contraire, elle peut constituer un obstacle pour les étudiants, si elle se substitue à la réflexion sur la portée des notions utilisées. Nous avons donc pris le parti d'éviter systématiquement le recours à la formalisation, pour privilégier une approche notionnelle. Les étudiants auront tout le loisir d'apprendre par ailleurs à utiliser des outils mathématiques, qu'ils pourront acquérir aisément une fois saisis les fondements de la microéconomie. C'est également la raison pour laquelle nous avons privilégié l'approche graphique, qui permet de raisonner avec rigueur, sans nécessiter pour autant de recours aux mathématiques. Le présent manuel ne comprend pas moins de 55 graphiques qui sont autant d'occasions de s'entraîner au raisonnement microéconomique.

- Contrairement à certaines idées reçues, la microéconomie n'est pas par essence hostile à l'État. Elle montre en revanche sous quelles conditions, en pratique, l'efficacité des marchés passe par une régulation publique adéquate. C'est pourquoi nous avons fait le choix d'aborder l'intervention publique dans une économie concurrentielle dès le premier chapitre, sous la forme d'une présentation des grands principes de la régulation d'un marché, et de l'incidence fiscale. Chaque chapitre consacre un développement aux mécanismes d'intervention publique, par des subventions ou des taxes, des mécanismes de contrôle des quantités échangées, ou encore par la réglementation des marchés. Nous postulons que les étudiants s'approprient d'autant plus facilement le raisonnement microéconomique qu'ils saisissent dès le début de leur formation comment on modélise le rôle de l'État au sein d'une économie de marché. En effet, l'État est présent sur l'écrasante majorité des marchés, ne serait-ce que *via* la taxation des consommateurs et des producteurs, et il est très irréaliste de présenter un marché sans aucune intervention publique.
- Enfin, ce manuel souhaite montrer que l'approche microéconomique offre une perspective intéressante pour comprendre de nombreux phénomènes concrets et ne se limite pas au modèle néoclassique de base développé par Walras il y a près de 140 ans. Par ailleurs, il nous a semblé important de confronter cette approche à d'autres travaux issus de traditions différentes, qui nourrissent le débat aujourd'hui parmi les économistes. Ainsi, dans chaque chapitre, nous nous sommes efforcés de donner des exemples de travaux contemporains, dont les hypothèses et les critiques du modèle néo-classique ont enrichi les débats microéconomiques. De E. Ostrom à R. Coase, en passant par les différents institutionnalismes, les exemples ne manquent pas. Aucune activité scientifique ne peut se développer sans confrontation entre différents paradigmes, et nous avons pour cette raison mobilisé ponctuellement des travaux provenant d'économistes plus « hétérodoxes », ou de spécialistes d'autres sciences sociales, quand ils permettaient d'éclairer les débats microéconomiques contemporains.

Chapitre 1

Les marchés concurrentiels

L'exemple du marché du logement

1. LE FONCTIONNEMENT DES MARCHÉS CONCURRENTIELS

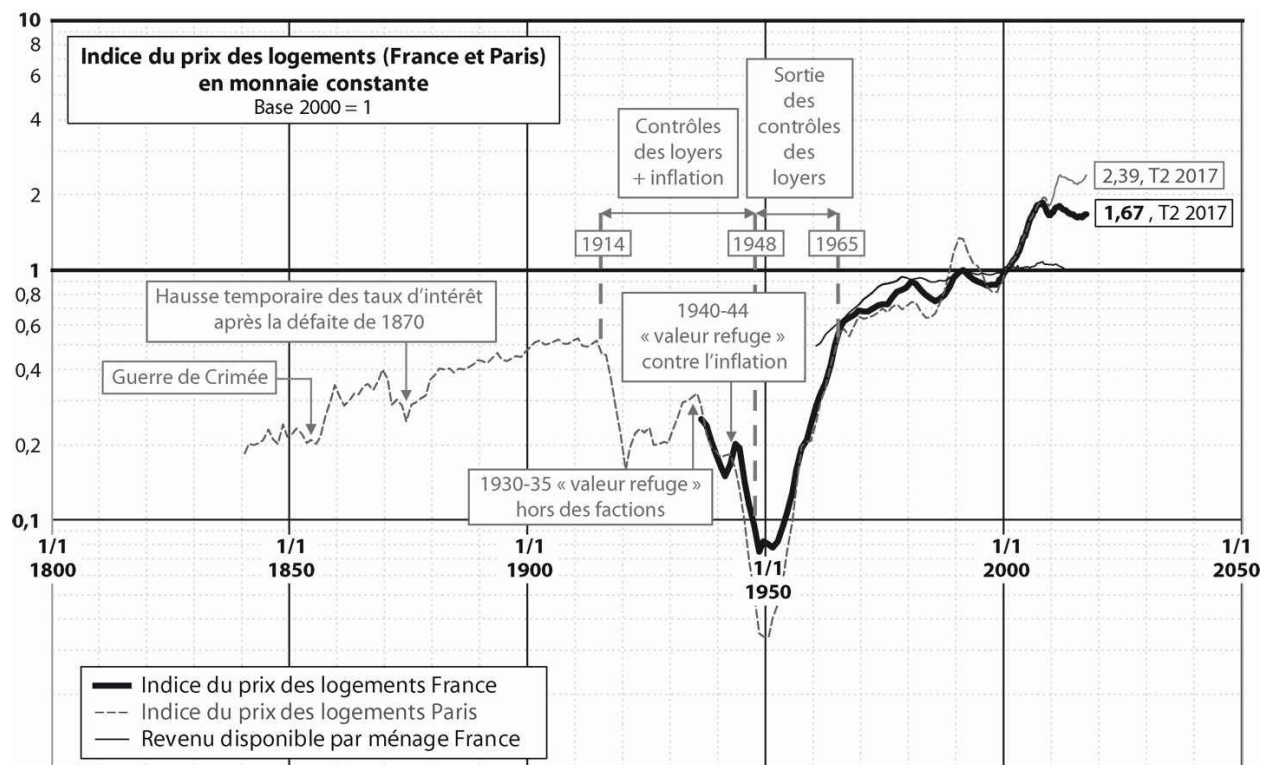
2. COMMENT ÉTUDIER UN MARCHÉ CONCURRENTIEL PARTICULIER ?

3. LA RÉGULATION D'UN MARCHÉ CONCURRENTIEL

La question du logement est au cœur des préoccupations actuelles : envolée des prix de l'immobilier, insuffisance de l'offre de logements, augmentation du nombre de mal logés et de sans-abri, etc.

Les différents dispositifs d'aide au logement n'ont pas réussi à enrayer l'envolée des prix d'achat de logements anciens, comme on le voit sur le **document 1.1**. En 2014, 23,6 % de la consommation finale effective¹ des ménages est allouée au service de logement. La part des dépenses consacrées au logement a sensiblement augmenté depuis les années 1980, où elle était de 2 points inférieure à son niveau de 2014. Les symptômes d'un mauvais fonctionnement du marché et d'une réglementation publique inadaptée sont nombreux. Quels sont les mécanismes économiques qui sont à l'origine de cette situation ? Comment l'État peut-il intervenir pour sortir de ce que l'on appelle souvent aujourd'hui « la crise du logement » ?

Document 1.1 : Évolution du prix des logements et du revenu disponible des ménages



Lecture : entre 2000 et 2017, le prix des logements en France a augmenté de 67 % (139 % à Paris) alors que le revenu disponible est resté quasiment constant. Les sommes étant indiquées en euros constants, cela signifie que l'on raisonne à prix constants

Source : CGEDD d'après l'INSEE

L'objectif du présent chapitre n'est pas d'étudier le fonctionnement du marché du logement dans toute sa complexité, mais de montrer comment un modèle théorique particulier (ici principalement le modèle de concurrence pure et parfaite) permet de comprendre certains mécanismes qui sont à l'œuvre sur ce marché, comme sur d'autres. Bien évidemment, le marché concret du logement est fort éloigné de celui du modèle de concurrence pure et parfaite. Le décalage existant entre tout modèle et la réalité n'est pas une limite en soi. La pertinence d'un modèle se juge avant tout à sa capacité à expliquer le réel. Aucun modèle ne suffit à lui seul à l'étude complète d'un phénomène concret et il existe d'autres modèles théoriques qui peuvent être mobilisés de façon très pertinente pour étudier certains aspects du fonctionnement du marché du logement. L'étude complète et exhaustive d'un marché concret nécessite de tenir compte d'une multitude de facteurs, de ses spécificités, et d'intégrer les apports de différents modèles et aussi d'autres disciplines que l'économie. Pour autant, si l'on veut comprendre les

mécanismes qui jouent sur le marché concret du logement et saisir les effets de l'intervention publique sur ce dernier, il faut démêler les différentes logiques qui sont à l'œuvre et l'éclairage apporté par le modèle de concurrence pure et parfaite semble fécond.

1. Le fonctionnement des marchés concurrentiels

1.1 Qu'est-ce qu'un marché concurrentiel ?

- Une approche empirique de la notion de concurrence

Un **marché concurrentiel** désigne un marché sur lequel la pression concurrentielle est forte. Les marchés concurrentiels sont en général des marchés sur lesquels il y a de nombreux offreurs et demandeurs. Les producteurs sont ainsi en compétition pour séduire les consommateurs qui, de leur côté, font « jouer la concurrence » pour bénéficier des conditions d'achat les plus favorables. Certains marchés concurrentiels peuvent toutefois ne comporter que peu d'offeurs qui sont soumis à la menace d'entrée de concurrents potentiels : c'est une **concurrence potentielle** (voir le [chapitre 4](#)). Dans l'approche néoclassique, on relie l'existence d'une forte concurrence au fait qu'aucun acteur n'est en mesure d'influencer par son comportement individuel le prix auquel s'échangent les biens : on dit qu'aucun acteur n'a de **pouvoir de marché**. La notion de marché concurrentiel renvoie à une réalité empirique : à un moment donné, sur un marché donné, la pression exercée par la concurrence est plus ou moins forte, et il est possible de la mesurer empiriquement. Il existe différents instruments de mesure qui permettent d'évaluer empiriquement le degré de concurrence d'un marché. Un des critères les plus connus est l'indice d'Herfindahl-Hirschmann (IHH), qui mesure la concentration de l'offre sur un marché en sommant le carré des parts de marché de toutes les entreprises du secteur considéré. L'IHH est notamment utilisé par les autorités de concurrence pour analyser et réguler le fonctionnement des marchés. Une valeur élevée de l'IHH témoigne d'une situation où il y a peu d'offeurs et où la compétition est donc moins intense. Le degré de concurrence qui existe sur un marché résulte de multiples facteurs, juridiques notamment. En France, depuis 1953 et la naissance de la politique de la concurrence, l'État régule la concurrence sur les différents marchés, et depuis le Traité de Rome en 1957, la Commission Européenne a développé la politique de la concurrence à l'échelle européenne. De ce fait, le degré de concurrence observé sur chaque

marché ne résulte pas uniquement du comportement des agents privés, mais en large partie de l'action publique.

Le marché du logement est un marché concurrentiel au sens où il y a à la fois beaucoup d'offreurs et de demandeurs. Si les acteurs disposent de pouvoir de négociation dans la fixation du prix de vente ou de location d'un bien, ils sont tout de même contraints de suivre les tendances du marché. Si les prix sont à la baisse, les vendeurs seront contraints de s'adapter s'ils veulent pouvoir vendre leur bien. La rencontre de l'offre et de la demande s'opère de façon décentralisée. Le secteur du logement social ne fonctionne pas lui selon un processus marchand mais en fonction de règles administratives, et doit être traité séparément.

• Le modèle de concurrence pure et parfaite

Au-delà de cette approche empirique de la concurrence, il est nécessaire de définir théoriquement ce concept afin de développer des modèles permettant de comprendre le fonctionnement des marchés concurrentiels, le modèle de concurrence pure et parfaite est le fruit des travaux d'économistes dits « néo-classiques », et propose une représentation simplifiée d'une économie où la concurrence s'exerce uniquement par les prix. Aucun agent ne peut jouer sur d'autres facteurs que le prix pour s'imposer (comme sa taille sur le marché, une différence de qualité du produit, un accès à une ressource privilégiée, etc.). Ce modèle permet donc de représenter ce qui se passerait dans le « cas limite » où la concurrence s'exercerait uniquement *via* les prix auxquels les agents sont disposés à acheter ou à vendre des biens en tout point identiques. Il permet de prédire comment les prix devraient s'adapter aux variations des quantités offertes ou demandées.

La **concurrence pure et parfaite** est définie par cinq conditions :

- une condition **d'atomicité** : il y a sur le marché de nombreux acheteurs et vendeurs de telle sorte qu'aucun acteur ne peut influencer à lui seul le prix d'équilibre (on dit que les agents sont **preneurs de prix**) ;
- une condition de **fluidité** : tout agent peut quitter ou entrer sur le marché à tout instant, il y a libre entrée et sortie du marché ;
- une condition d'**homogénéité** des produits : les produits échangés sur un marché sont strictement identiques ;

- une condition de **transparence** : l'information est **parfaite** et gratuite et tous les acteurs connaissent ainsi parfaitement la qualité des produits échangés, leur prix, etc. ;
- une condition de **mobilité parfaite des facteurs de production** : à tout instant, les travailleurs ou le capital utilisés pour la production d'un bien donné peuvent être mobilisés pour produire un autre bien destiné à un autre marché.

Ces conditions théoriques sont très exigeantes et au niveau empirique les marchés qui s'en approchent sont très rares. Le marché du logement n'est pas au sens strict un marché de concurrence pure et parfaite. Nous verrons pour autant que ce modèle est tout de même très utile pour comprendre certaines logiques à l'œuvre sur ce marché. En soi, il n'est pas problématique qu'un modèle théorique ne soit pas le reflet parfait de la réalité : par définition, tout modèle offre une représentation simplifiée et néglige certaines caractéristiques observables du phénomène étudié. Les microéconomistes considèrent que le seul critère valide pour juger de la qualité d'un modèle est sa capacité à expliquer des phénomènes concrets, en proposant un cadre rigoureux pour faire des hypothèses et en tirer des déductions cohérentes. De ce point de vue, le modèle de concurrence pure et parfaite a une certaine portée heuristique, comme on va le montrer dans ce chapitre : il permet par exemple, sous l'hypothèse d'une variation de la demande sur un marché, d'en déduire les variations de prix qui en découlent. On peut ensuite confronter les observations empiriques aux prédictions du modèle, et voir dans quelle mesure il permet effectivement d'éclairer les phénomènes étudiés. Comme de nombreuses disciplines, la microéconomie se fonde sur une démarche qui privilégie la construction de modèles permettant des prédictions rigoureuses et pouvant faire l'objet d'évaluations quantitatives précises, en acceptant la part d'irréalisme qui en découle².

Le modèle de concurrence pure et parfaite ne permet pas de comprendre l'ensemble des logiques à l'œuvre sur les marchés et de nouveaux modèles théoriques fondés sur d'autres hypothèses ont ainsi été développés : il s'agit principalement des modèles de concurrence imparfaite. Certains visent à mieux comprendre le fonctionnement de marchés concrets qui ne sont pas ou peu concurrentiels, et dont la logique de fonctionnement est mal expliquée par le modèle de concurrence pure et parfaite, d'autres peuvent être utilisés pour saisir l'effet des asymétries d'information, que l'on retrouve aussi sur

des marchés concurrentiels. Ces modèles sont présentés dans le chapitre 4.

Nous allons à présent voir comment le modèle de concurrence pure et parfaite peut être mobilisé pour expliquer certains mécanismes et éclairer des « faits stylisés » caractérisant le marché du logement.

1.2 La loi de l'offre et de la demande³

Comme nous l'avons vu, le prix des logements ne fait que s'accroître ainsi que les sommes consacrées aux dépenses de logement. Si l'on s'intéresse à l'évolution du poids des loyers dans les revenus (loyer moyen payé sur le revenu moyen des locataires pour la location de logements vides hors HLM et loi 1948), on constate que ce poids est passé de 12 % à 26 % entre 1970 et 2013.⁴ Dans quelle mesure le modèle de concurrence pure et parfaite peut-il être mobilisé pour expliquer les variations de prix constatées sur le marché du logement ?

• La détermination du prix d'équilibre

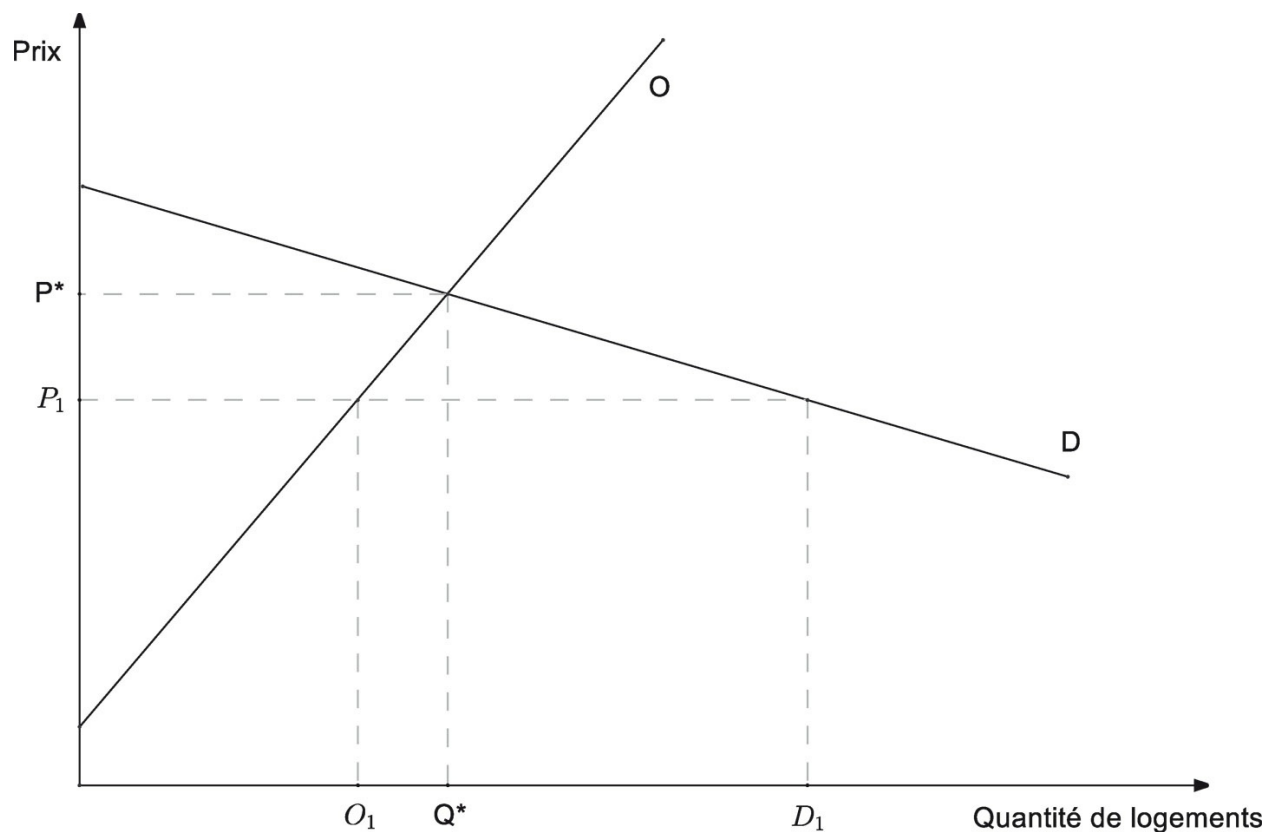
Dans le modèle de concurrence pure et parfaite, les prix dépendent uniquement de la confrontation de l'offre et de la demande. Comment représenter graphiquement l'offre et la demande d'un bien ?

Une **courbe d'offre** associe pour chaque niveau de prix la quantité offerte d'un bien. Appliquée au cas du logement, la courbe donne pour chaque niveau potentiel de prix des logements, la quantité totale de biens immobiliers que les propriétaires acceptent de mettre à la vente. Une **courbe de demande** représente pour chaque niveau de prix la quantité demandée d'un bien. Appliquée au cas du logement, la courbe donne pour chaque niveau potentiel de prix des logements, la quantité totale de biens immobiliers que les acheteurs sont prêts à acquérir. Il s'agit donc de courbes d'offre et de demande agrégées : elles représentent les quantités offertes et demandées par l'ensemble des acteurs du marché pour un prix donné.

Les courbes d'offre et de demande se coupent en un point unique : il n'y a qu'un **prix d'équilibre** tel que la quantité globale offerte sur le marché est égale à la quantité globale qui est demandée. Le modèle de base prédit que le prix de marché va se fixer à ce niveau. Deux explications peuvent être données pour justifier ce résultat.

Graphique 1.1 : Les courbes d'offre et de demande





Une première approche s'inscrit dans la tradition marshalienne⁵ avec une vision dynamique du fonctionnement des marchés. On suppose que le prix ne se fixe pas spontanément au prix d'équilibre mais que des mécanismes économiques y conduisent progressivement. Supposons qu'au départ, le prix P_1 des logements est inférieur au prix d'équilibre (noté P^* sur le [graphique 1.1](#)). À ce prix, la quantité offerte de logements notée O_1 est inférieure à la quantité demandée D_1 de logements. Les propriétaires qui avaient mis en vente leur logement réalisent, en voyant cet afflux d'acheteurs, qu'ils avaient fixés un prix trop bas pour leur bien et augmentent donc leur prix de vente. Supposons que leur réaction soit très forte et qu'ils fixent un prix supérieur au prix d'équilibre : cette fois, la quantité demandée est inférieure à la quantité offerte. Voyant que les logements se vendent mal, les propriétaires vont de ce fait abaisser leur prix. La situation ne sera à l'équilibre que lorsque le prix de marché sera égal au prix d'équilibre pour lequel la quantité offerte de logements est égale à la quantité demandée. En situation d'information parfaite et en supposant une parfaite flexibilité des prix, le processus d'ajustement a lieu rapidement et les mécanismes de marché conduisent à une situation d'équilibre. Les travaux récents de

microéconomie insistent sur l'existence d'une **viscosité des prix** qui ne s'ajustent qu'avec lenteur du fait par exemple de l'existence de **coûts de catalogue**, c'est-à-dire de coûts liés aux opérations concrètes de changement des prix (édition de nouveaux catalogues par exemple). Cette viscosité des prix qui découle des stratégies des acteurs conduit à ce que les prix soient rarement au niveau d'équilibre sur les marchés et donc à des phénomènes de **rationnement** de l'offre ou de la demande, c'est-à-dire à une situation où les quantités offertes et demandées sont différentes l'une de l'autre.

Une seconde approche s'inscrit dans la tradition walrasienne⁶ en supposant l'existence d'un agent économique (surnommé le **commissaire-priseur walrasien**) qui centralise les informations sur l'offre et la demande et fixe le prix en fonction. Pour chaque niveau de prix, il recense les quantités offertes et demandées et le prix va être fixé de façon à ce que les quantités offertes soient égales aux quantités demandées. Dans la réalité, une telle centralisation de l'information s'observe rarement, hormis sur certains marchés très centralisés comme le sont par exemple les marchés financiers organisés.

Par ailleurs, du fait de l'hypothèse d'information parfaite, tous les biens sont censés être échangés au même prix, le prix d'équilibre : c'est **la loi du prix unique**. Si l'information est parfaite, quel acheteur accepterait en effet en toute connaissance de cause de payer un logement plus cher qu'un autre en tout point identique ? Il n'y a donc pas d'échange hors du prix d'équilibre et ce prix fixé par les mécanismes de marché s'impose à tous les acteurs, ils sont **preneurs de prix**. En situation de concurrence pure et parfaite, on évoque la **fonction paramétrique des prix**, c'est-à-dire que les prix sont des paramètres auxquels les acteurs ajustent leurs comportements, sans pouvoir les changer.

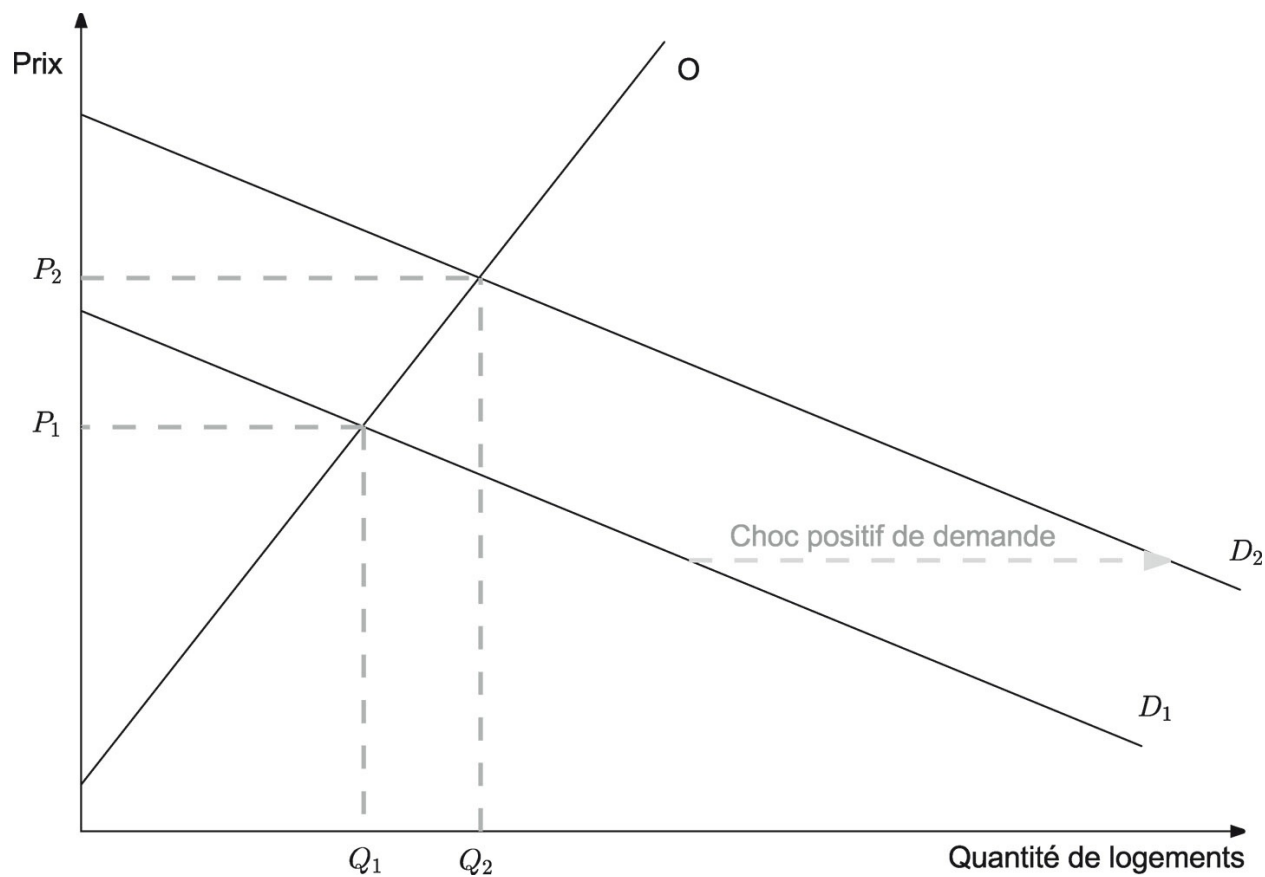
• L'impact des chocs d'offre et de demande sur le prix d'équilibre

Essayons de voir à présent comment ce modèle peut être mobilisé pour expliquer la hausse des prix des logements en France. Pour des raisons que nous ne détaillerons pas ici, les Français investissent de plus en plus dans la pierre et la demande de logements a ainsi fortement augmenté à long terme. Pour un niveau de prix donné, il y a aujourd'hui davantage de personnes qui veulent acheter un logement. En termes économiques, on parle d'un **choc de demande** positif dans ce genre de situation. Comme nous pouvons le voir sur

le [graphique 1.2](#), cela se traduit par un déplacement de la courbe de demande vers la droite : les ménages sont prêts à demander une plus grande quantité de logements pour chaque niveau de prix proposé.

Nous voyons que le choc de demande a fait augmenter le prix d'équilibre, qui passe de P_1 à P_2 sur le graphique, ce qui est conforme à l'évolution des prix observée sur le marché du logement ancien. Si le [graphique 1.2](#) illustre les conséquences d'un choc de demande positif, il est aisé de représenter l'évolution inverse, un choc de demande négatif qui serait dû à une restriction de la demande. De façon similaire, les **chocs d'offre** entraînent un déplacement de la courbe d'offre vers la gauche quand l'offre se réduit, et vers la droite quand l'offre augmente.

Graphique 1.2 : Un choc de demande positif



Il faut bien différencier les déplacements le long de la courbe qui représentent pour différents niveaux de prix les quantités que les agents sont disposés à offrir ou à demander dans les conditions actuelles (variation des quantités offertes ou demandées liée à une variation du prix), et les déplacements de la courbe d'offre ou de demande qui sont la conséquence d'un changement de comportement (variation de l'offre ou de la demande pour un prix donné). On se déplace le long de la courbe de demande quand on souhaite savoir par exemple comment les ménages adaptent la quantité demandée aux variations de prix, mais on déplace leur courbe de demande si une modification de leur revenu ou une augmentation de la population entraîne un changement des quantités demandées à prix constant.

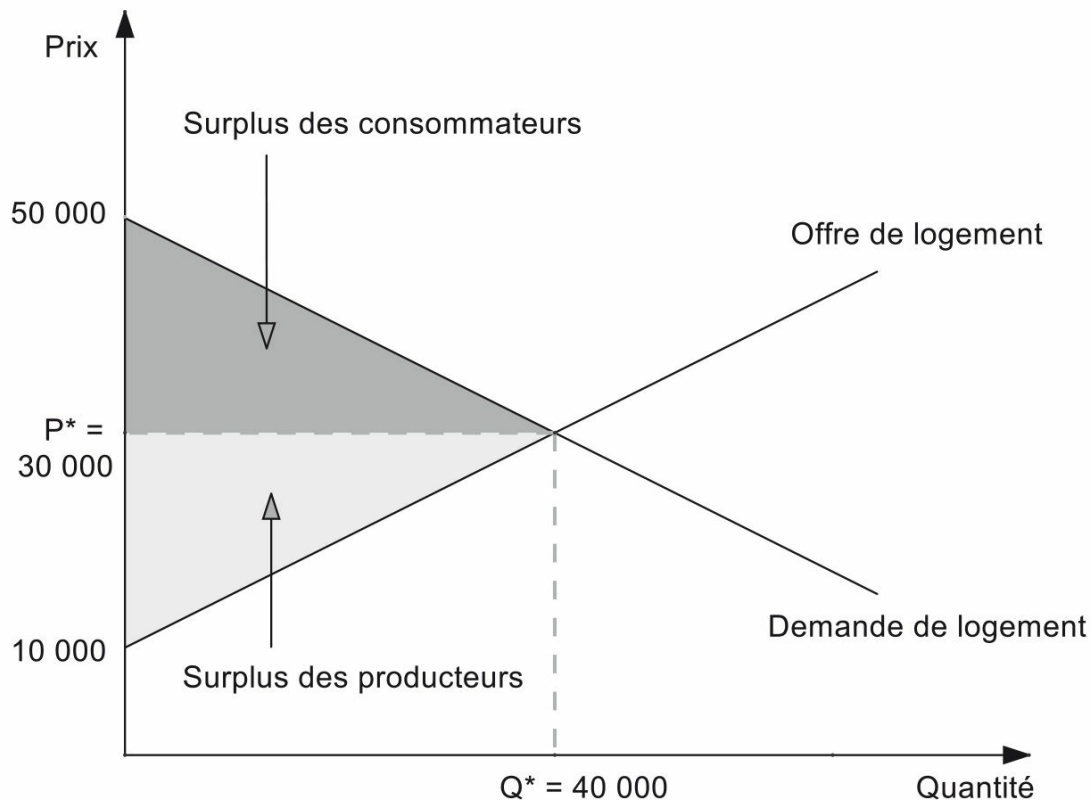
1.3 Gains à l'échange et exclusion du marché

- Le concept de surplus

Comment évaluer les gains et les pertes réalisés par les agents aux conditions d'équilibre du marché ? Sur un marché concurrentiel, tous les biens échangés se vendent au même prix et le prix d'un échange ne dépend donc pas des négociations entre un acheteur et un vendeur singuliers mais de l'offre totale et de la demande totale sur ce marché. Prenons l'exemple d'un marché du logement où des studios s'échangeraient à l'équilibre à un prix moyen de 30 000 € : dans ces conditions, les agents demandeurs, qui étaient prêts à acheter ces biens à un prix supérieur (40 000 € par exemple) réalisent une économie substantielle, puisqu'ils obtiennent leur logement moins cher que ce qu'ils étaient disposés à payer au maximum : l'équilibre concurrentiel leur a été profitable. Le concept de **surplus** a été développé pour évaluer précisément les gains réalisés dans les conditions d'équilibre du marché, par rapport aux prix que les agents offreurs demandaient au minimum pour vendre en fonction de leurs coûts de production, et par rapport au prix maximal que les agents demandeurs étaient disposés à payer en fonction de la satisfaction ressentie à la consommation du bien. La différence entre ce que les consommateurs étaient prêts à payer pour acquérir un bien (somme qui varie au niveau individuel, tous les consommateurs ne ressentant pas la même satisfaction à la consommation d'un bien) et la somme effectivement payée s'appelle le **surplus des consommateurs**. Comme on le voit sur le [graphique 1.3](#), le surplus est représenté par une surface et on peut en proposer une évaluation monétaire. La courbe de demande représente en effet les quantités que les consommateurs désirent acheter pour un prix donné et reflète les préférences des agents (voir le [chapitre 2](#)). Dans le cas présent, c'est un triangle dont la surface s'obtient en multipliant la quantité d'équilibre (en abscisse) par la différence de prix en ordonnée le tout divisé par deux ; il est égal à $40\,000 \times (50\,000 - 30\,000)/2 = 400$ millions d'euros pour les consommateurs. Cela signifie qu'aux conditions d'équilibre du marché, les consommateurs économisent ce montant pour obtenir un volume de logements qu'ils étaient disposés à payer plus cher. De même, la différence entre la somme effectivement reçue et les sommes minimales auxquelles les vendeurs étaient prêts à vendre (différentie là aussi en fonction des producteurs qui ne font pas nécessairement face aux mêmes coûts de production) s'appelle le **surplus des producteurs**. Il permet d'évaluer les gains réalisés par les offreurs dans les conditions d'équilibre du marché. Dans les conditions de la concurrence pure et parfaite, les quantités que les offreurs sont prêts à vendre pour un prix donné correspondent au volume de

production tel que le coût marginal est égal au prix (voir le [chapitre 3](#)). Ici, le surplus des producteurs est égal à $40\,000 \times (30\,000 - 10\,000)/2 = 400$ millions d'euros.

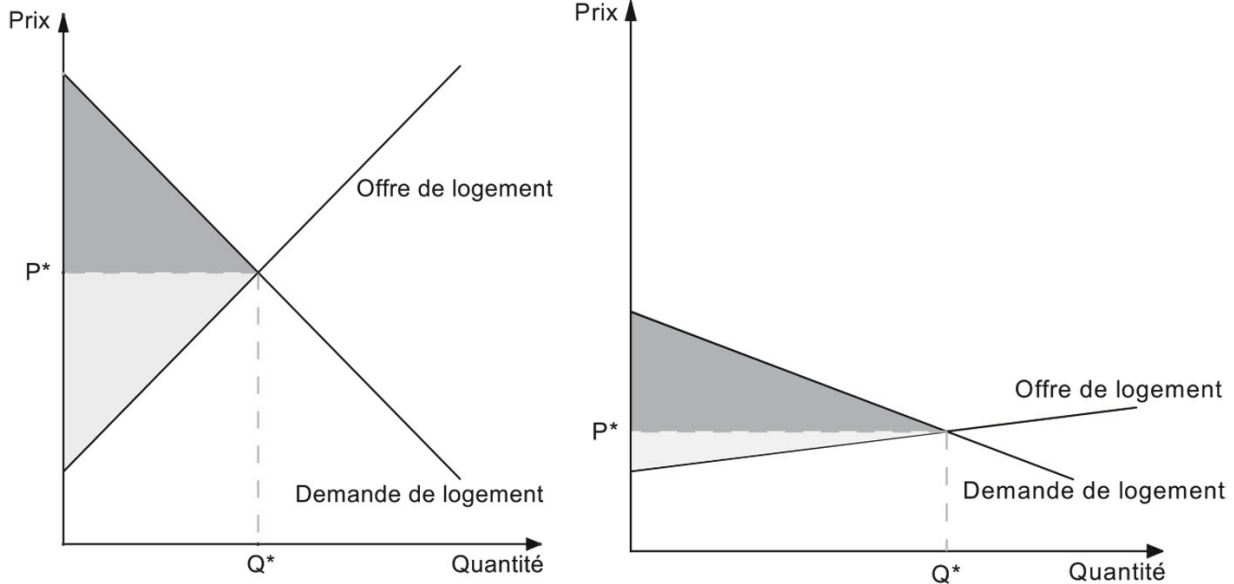
**Graphique 1.3 : Le surplus des consommateurs
et des producteurs**



Le **surplus total** représente les gains réalisés par l'ensemble des acteurs qui interviennent sur le marché. Dans ce cas, il renvoie à la somme du surplus des consommateurs et du surplus des producteurs. Il représente les gains à l'échange réalisés au prix d'équilibre par l'ensemble des agents présents, c'est-à-dire les montants qu'ils ont gagnés ou évités de dépenser aux conditions du marché, pour acquérir ou vendre un certain volume de biens. Dans le cas présenté ci-dessus, le surplus total est réparti de façon égale entre acheteurs et vendeurs. Il s'agit d'un cas particulier. L'ampleur des gains à l'échange et la répartition du surplus total entre les demandeurs et les offreurs dépend de la forme des courbes, comme on le voit sur le [graphique 1.4](#). La pente des courbes dépend elle-même de l'**élasticité-prix** de l'offre et de la demande (voir les chapitres [2](#) et [3](#)). L'élasticité-prix est un indicateur de la sensibilité des offreurs ou demandeurs à la variation des prix. L'élasticité-prix de la demande mesure ainsi la variation relative de la quantité demandée suite à une variation relative du prix et l'élasticité-prix de l'offre mesure la variation relative de la quantité offerte suite à une variation relative du prix.

Graphique 1.4 : Surplus total et élasticité de l'offre

et de la demande



Le calcul de la valeur du surplus total dans différentes situations de marché, par exemple avec intervention publique, permet de comparer les gains ou pertes réalisés par les agents. En situation de concurrence pure et parfaite, le surplus total est maximisé à l'équilibre et ce dernier correspond ainsi à une situation optimale au sens de Pareto (voir le [chapitre 7](#)) : il n'est pas possible d'accroître la satisfaction d'un agent sans dégrader celle d'un autre. Raisonnons par l'absurde pour expliquer ce résultat : si le prix est fixé à un niveau supérieur au prix d'équilibre, le surplus des producteurs sera plus élevé mais celui des consommateurs plus faible et on peut voir aisément sur un graphique que le gain des producteurs ne compense pas la perte des consommateurs. À l'inverse, si le prix est inférieur au prix d'équilibre, le surplus des consommateurs sera plus important mais cette augmentation ne compensera pas la perte au niveau de producteurs.

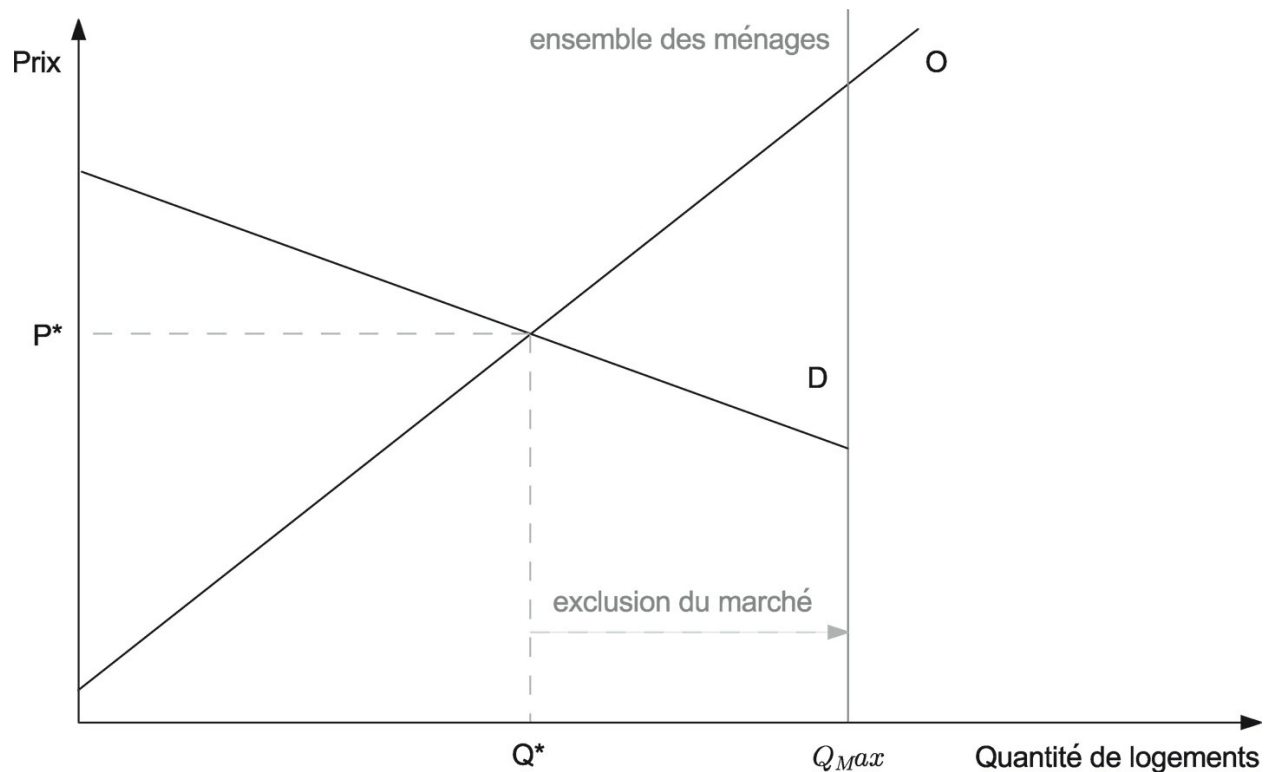
Cette situation découle du fait qu'à l'équilibre, les acheteurs qui accordent au bien une valeur supérieure à son prix décident de l'acquérir alors que les autres ne sont pas demandeurs. Côté offre, les producteurs dont les coûts sont inférieurs au prix vont être vendeurs tandis que ceux dont les coûts sont supérieurs ne participent pas au marché. L'allocation des ressources par le marché conduit ainsi à ce que les biens soient vendus aux acheteurs qui leur accordent la valeur la plus élevée par les producteurs qui produisent de la manière la plus efficiente, c'est-à-dire aux coûts de production les plus

faibles.

- **L'exclusion du marché**

Dans les conditions actuelles, les prix des logements sont trop élevés pour permettre à tous les ménages de se loger. Selon l'INSEE, 143 000 personnes sont sans-domicile en 2012, et plus de 2 millions de personnes vivent dans un logement privé de confort (Enquête Logement 2013). Le mal-logement s'explique par de multiples facteurs et le modèle de concurrence pure et parfaite en éclaire seulement certains aspects. Si à l'équilibre les quantités offertes sont égales aux quantités demandées, tous les agents ne sont pas nécessairement satisfaits puisque certains sont exclus du marché. Tous ceux qui voulaient vendre ou acheter à ce prix le peuvent, mais les ménages qui désiraient un logement à un prix inférieur ne trouvent pas à se loger. Ces ménages ne sont pas considérés comme des demandeurs à l'équilibre car au prix en question, ils ne peuvent pas acheter le bien. Le phénomène de l'exclusion par les prix ne doit donc pas être confondu avec le rationnement de la demande car les quantités offertes sont bien égales aux quantités demandées. Supposons qu'à la quantité $Q_{\max}$, sur le [graphique 1.5](#), tous les ménages ont un logement.

Graphique 1.5 : L'exclusion du marché



Par hypothèse chaque ménage peut au plus posséder un logement, en laissant de côté les résidences secondaires. Dans le cas présenté ci-dessous, une partie des ménages est exclue du marché à l'équilibre : le prix d'équilibre étant trop élevé pour eux, ils ne peuvent demander un logement à ce prix et se retirent donc du marché. L'**exclusion du marché** se mesure ici par la différence entre la quantité d'équilibre et la quantité maximale que les ménages peuvent demander, soit Q^* et Q_{max} . D'après le modèle de concurrence pure et parfaite, la concurrence ne bénéficie donc pas à la totalité des agents : elle permet d'obtenir un bien « au meilleur prix », mais elle exclut aussi une part des ménages de l'accès à ce bien : c'est l'exclusion par les prix. Il n'y a pas de rationnement car les quantités offertes et demandées sont égales mais il y a des personnes qui auraient aimé accéder au bien (mais pas au prix du marché) et qui ne le peuvent pas. Ils ne sont pas demandeurs au sens où ils ne sont pas prêts à payer le prix de marché pour accéder au bien. Puisque les logements rendent des services essentiels, dont les ménages ne peuvent durablement se passer, les États ou les collectivités locales sont amenés, dans la quasi-totalité des pays développés, à conduire des politiques de régulation du marché du logement, pour éviter une exclusion du marché trop importante. Les différents types d'intervention sont discutés dans la

troisième partie du présent chapitre.

2. Comment étudier un marché concurrentiel particulier ?

L'étude d'un marché particulier nécessite de définir de façon précise les limites du marché étudié et de tenir compte des interdépendances existantes entre les marchés.

Comment circonscrire le marché du logement ? Dans quelle mesure les choix des agents sur le marché du logement sont-ils liés à ceux qu'ils font sur d'autres marchés ?

2.1 Les frontières d'un marché

Par hypothèse, le modèle de concurrence pure et parfaite suppose une parfaite homogénéité des biens échangés sur un marché. De façon générale, un marché est défini en relation avec l'offre et la demande de biens qui partagent les mêmes propriétés. Dans les faits, les biens sont très rarement complètement similaires. Les biens immobiliers sont particulièrement hétérogènes : un logement a des spécificités en termes de confort et d'agrément, de localisation et d'environnement, de taille et même d'usage. Cela pose plusieurs problèmes à l'approche microéconomique du logement. Faut-il considérer qu'il existe un seul marché ou au contraire un très grand nombre, en fonction des caractéristiques des produits ? Sur quel territoire définir le marché du logement ?

Les biens immobiliers, entendus comme les locaux et bâtiments sur lesquels des agents disposent de droits de propriété, peuvent donner lieu à une multitude d'usages. Certains sont destinés à un usage professionnel, d'autres à un usage d'habitation. Pour simplifier l'analyse, nous nous limitons à cette dernière catégorie, qui regroupe à proprement parler les logements ; ce sont des biens durables qui fournissent un service à une ou plusieurs personnes, en leur offrant un lieu d'habitation. Deux grands types de marchés peuvent être distingués : le marché locatif qui regroupe l'ensemble des logements offerts à la location et le marché de l'acquisition de logements.

Ces deux types de marchés n'ont pas les mêmes systèmes de prix et n'obéissent pas aux mêmes règles. Sur le marché locatif par exemple, les agents s'entendent sur un loyer, fixé dans un contrat de location et le transfert de propriété entre acheteur et vendeur ne concerne que le droit d'*usus* (droit d'utiliser un bien), tandis que pour les biens disponibles sur le marché de l'acquisition de logements, il s'agit d'un transfert complet de propriété entre l'acquéreur et le vendeur. Nous ne détaillons pas ici le cas de la vente en viager qui correspond à un autre type de transfert de propriété. On peut donc définir deux « prix » différents, selon que l'on considère par exemple le prix de location au mètre carré, ou bien le prix d'acquisition au mètre carré. Mais il est aussi possible d'affiner l'analyse, et de considérer qu'il existe une grande diversité de marchés locatifs, et de marchés d'acquisition de logements. En fonction du lieu géographique, par exemple en centre-ville ou à la campagne, et du type de biens, selon qu'on a affaire à une maison de 100 m² ou bien à un studio de 20 m², il est possible de distinguer de façon plus précise les différents marchés du logement. On comprend bien en effet que l'offre et la demande de studios dans le centre-ville de Bordeaux ne peuvent pas être assimilées à l'offre et la demande de chalets à Courchevel. Le critère géographique occupe une place très importante dans la détermination du prix des biens immobiliers. Les personnes qui recherchent un logement souhaitent vivre dans une zone donnée, en fonction notamment de leur lieu de travail. Les logements sont associés à un territoire, qui leur donne une certaine valeur. La dimension territoriale est essentielle pour étudier le marché du logement : en France cohabitent ainsi des territoires où les prix sont très élevés et la demande de logements très importante comme dans les grands centres urbains, et d'autres où au contraire les prix sont faibles et l'offre excédentaire, comme dans les régions en déclin industriel et démographique.

On peut poursuivre à l'infini ce processus de segmentation des marchés en fonction des caractéristiques détaillées du logement (toilettes séparées de la salle de bain, double vitrage, ensoleillement, etc.). La segmentation du marché du logement est aussi à relier aux réglementations spécifiques qui encadrent son fonctionnement. Les enjeux sociaux liés au logement sont trop importants pour laisser le marché fonctionner librement. Le logement est par certains aspects un « **bien tutélaire** », c'est-à-dire un bien sur lequel l'État exerce une forte tutelle en encadrant les échanges (voir le [chapitre 5](#)). Les locataires sont par exemple protégés contre les expulsions durant la trêve

hivernale, l'augmentation des loyers est contrôlée, et les contrats de location font l'objet d'une réglementation contraignante. Cet encadrement public contribue à la segmentation du marché du logement, dans la mesure où l'État fixe des règles différentes selon les biens échangés : ainsi le fait que les frais notariés pour l'achat d'un logement neuf ne sont pas les mêmes que ceux d'un logement ancien contribue à différencier le marché du neuf et celui de l'ancien.

Pour définir les limites d'un marché, les économistes utilisent la notion de **marché pertinent**. Si les biens existants sont quasiment toujours différents les uns des autres, il est tout de même possible de considérer qu'il existe un marché du studio dans le centre-ville de Marseille ou un marché des livres au niveau national. Des biens différents s'échangent sur un même marché s'ils sont substituables aux yeux des consommateurs, c'est-à-dire si les acheteurs considèrent qu'ils peuvent arbitrer entre ces différentes possibilités pour satisfaire une même demande. Ainsi un studio à Rennes n'appartient pas au même marché qu'un studio à Bordeaux, les deux biens n'étant pas substituables aux yeux du consommateur. Deux studios situés à quelques rues d'écart appartiennent en revanche au même marché bien qu'ils ne soient pas exactement identiques (un peut être plus ensoleillé, plus grand, plus calme, etc.). Quand deux biens sont substituables, les demandeurs arbitrent entre les deux biens et les mettent en concurrence. Il est ainsi logique de considérer qu'ils appartiennent au même marché.

• Encadré 1.1 : La méthode des prix hédoniques, base de l'étude des caractéristiques d'un logement

Comment attribuer une valeur économique à chaque caractéristique propre à un logement, telle la présence d'un balcon, d'un parking ou d'une pièce supplémentaire ? Ces caractéristiques n'ont pas de prix « explicite » car elles ne font pas l'objet de transactions distinctes de celles portant sur le logement dans son ensemble. Personne n'achète uniquement un balcon. Résoudre ce problème est indispensable pour l'étude empirique du marché du logement, et notamment pour la construction d'indices des prix de l'immobilier afin de pouvoir distinguer l'évolution des prix à caractéristiques identiques de l'évolution de la qualité des logements. Pour ce faire les microéconomistes utilisent la méthode dite des « prix hédoniques ». Ils font l'hypothèse que les choix des agents sur le marché du logement permettent de révéler leurs préférences pour chaque caractéristique (voir le [chapitre 2](#)), ce qui permet de leur attribuer par ce biais un « prix implicite », soit un prix qui correspond à ce qu'ils seraient prêts à payer pour le supplément de satisfaction apporté par cette caractéristique en propre. Si la mise en œuvre est complexe, le principe est assez simple : on collecte un grand nombre de données sur des transactions effectives et pour chaque transaction, on évalue les différentes caractéristiques des logements. En utilisant ensuite des modèles économétriques, l'effet du changement d'une caractéristique sur le

prix de vente d'un bien est isolé ce qui permet de déduire le prix implicite que le consommateur est disposé à payer pour chaque caractéristique. Cette méthode a joué un grand rôle dans le développement d'études empiriques sur de nombreux marchés.

2.2 L'interdépendance entre les marchés

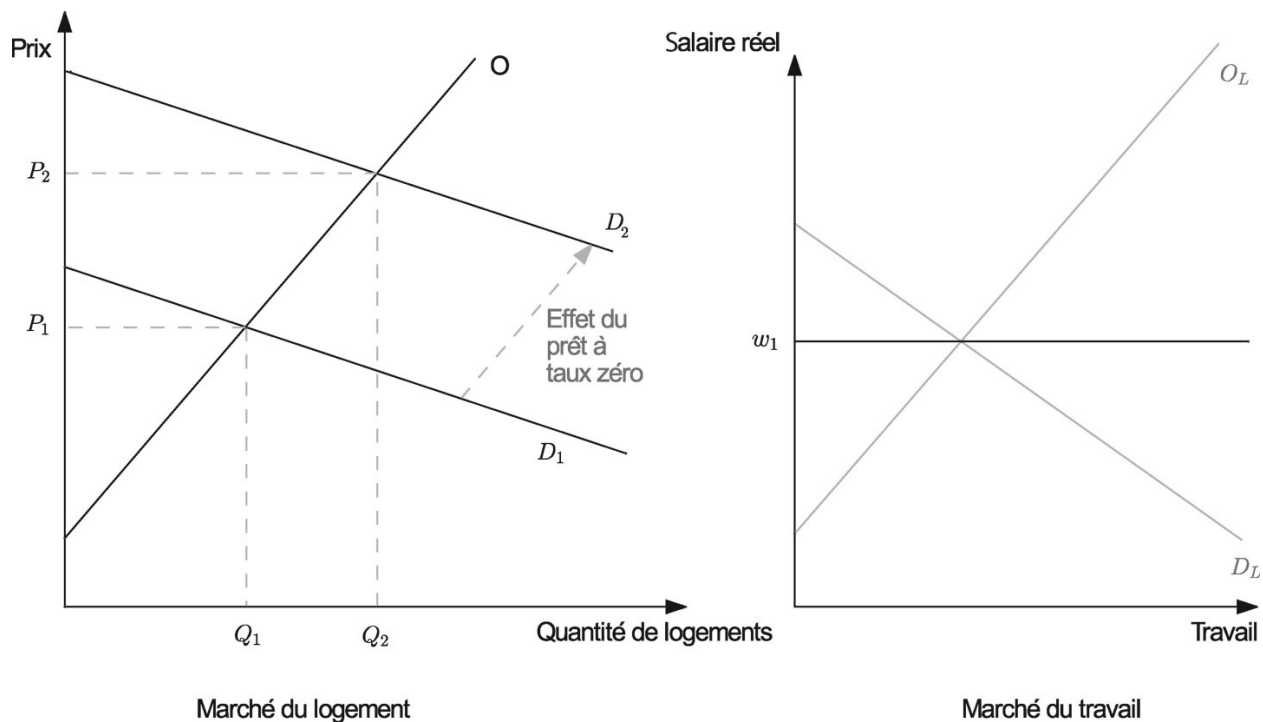
Pour décider d'acquérir un logement, un agent doit intégrer dans son raisonnement de nombreux paramètres, dont ses revenus actuels et futurs, le taux d'intérêt qu'il devra acquitter pour réaliser un crédit, etc. Ces différents paramètres sont déterminés sur d'autres marchés que celui du logement. Les différents marchés sont donc interdépendants : les changements de prix et de quantités d'équilibre sur un marché modifient ceux d'autres marchés, les agents acquéreurs sur un marché étant également clients sur d'autres, ils tiennent compte dans leur décision d'achat de l'ensemble des prix. Comment tenir compte de ces phénomènes d'interdépendances quand on étudie un marché particulier ?

La solution la plus simple est de considérer que le marché du logement est le seul à s'ajuster, *via* des mouvements de prix et de quantité d'équilibre, tandis que les différents prix sur les autres marchés sont des paramètres fixes, durant une période donnée. On considère qu'à court terme, les mouvements de l'offre et de la demande modifient l'équilibre sur le marché du logement, mais pas celui des autres : c'est ce qu'on appelle une analyse en **équilibre partiel**.

Pour bien comprendre le raisonnement en équilibre partiel, prenons l'exemple des effets d'une mesure d'aide au crédit destinée aux ménages accédant à la propriété. De nombreux dispositifs existent, notamment en France, pour favoriser l'accès à la propriété, le plus important est actuellement le « prêt à taux zéro » qui résulte de la fusion de différents dispositifs d'aide depuis janvier 2011 : c'est un prêt dont les intérêts sont pris en charge par l'État. Il est réservé aux primo-accédants à la propriété, sous condition de ressources et ne concerne qu'une partie de la somme nécessaire à l'achat du bien. Quels sont les effets d'un tel dispositif sur le marché du logement ? On devrait observer une augmentation de la demande avec un déplacement de la courbe vers la droite, les quantités demandées pour un prix donné étant plus importantes. À court terme, l'offre de logement ne peut guère s'ajuster en raison des délais de construction par exemple et une augmentation du prix du logement ne va ainsi conduire qu'à une faible

augmentation des quantités offertes, avec le déplacement le long de la courbe d'offre dont les caractéristiques n'ont pas été modifiées. On dit que l'offre est relativement rigide (courbe d'offre plutôt verticale sur le [graphique 1.6](#)). Pour connaître précisément la variation de la demande (de D_1 à D_2 sur le [graphique 1.6](#)), il faut tenir compte des prix pratiqués sur les autres marchés : le revenu des ménages dépend du taux de salaire déterminé sur le marché du travail, du niveau des loyers, de leur pouvoir d'achat et donc des prix des autres biens de consommation, etc. Pour simplifier la situation, on fait donc l'hypothèse en équilibre partiel que ces différents prix ne sont pas modifiés par les ajustements sur le marché du logement : quel que soit l'effet du prêt à taux zéro sur ce marché, l'équilibre sur les autres reste identique. Ainsi, le taux de salaire sur le marché du travail reste au niveau w_1 , comme on le voit sur le [graphique 1.6](#).

Graphique 1.6 : L'effet d'une aide en équilibre partiel



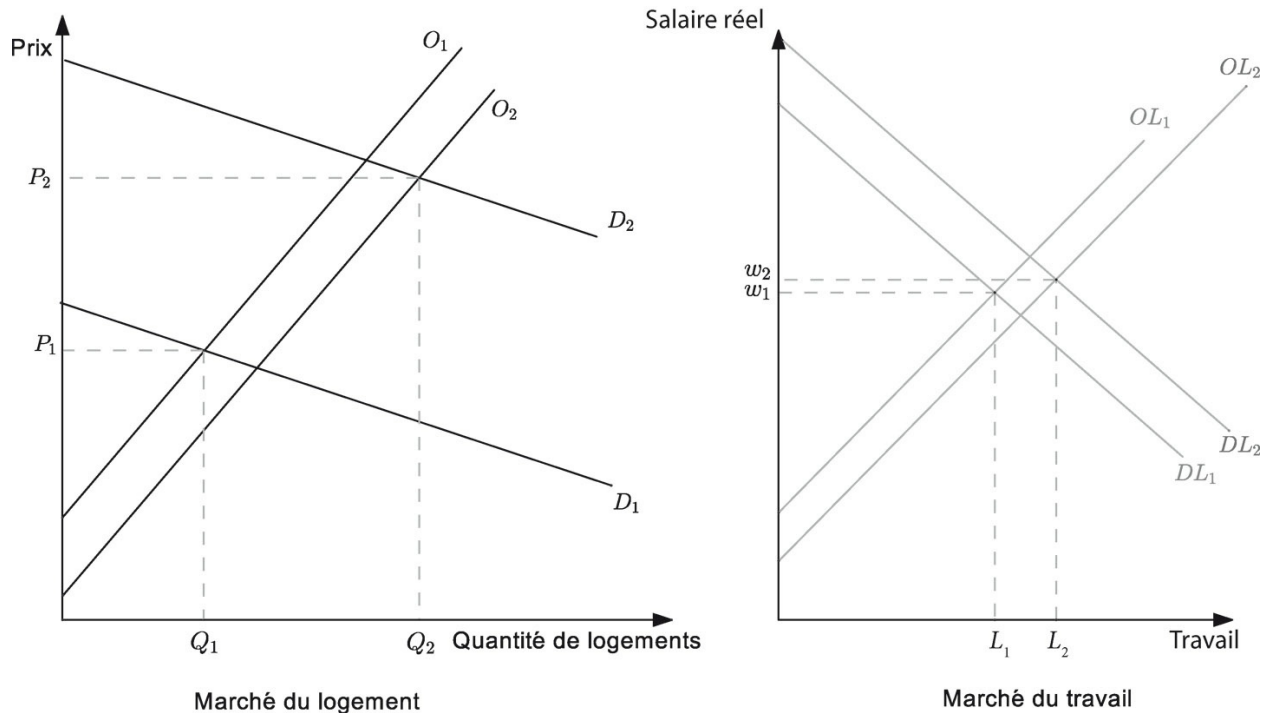
Le raisonnement en équilibre partiel permet donc de faire apparaître un résultat important : le « prêt à taux zéro » devrait augmenter le volume de logements auxquels les ménages accèdent, mais cela a un coût pour les ménages, la hausse sensible du prix d'équilibre de P_1 à P_2 , due à la rigidité de

l'offre sur ce marché, la courbe d'offre étant assez pentue. « Toutes choses égales par ailleurs », c'est-à-dire en considérant comme fixes les autres paramètres intervenant dans la décision des ménages, le « prêt à taux zéro » risque donc d'entraîner une hausse importante du prix des logements, car le marché ne peut pas s'ajuster autrement à court terme que par une hausse des prix, le nombre des biens immobiliers disponibles restant limité. C'est pourquoi ce type de mesure bénéficie paradoxalement en premier lieu à ceux qui sont déjà propriétaires de leur logement, et voient la valeur de leur patrimoine augmenter, alors que les primo-accédants vont devoir payer plus cher leur logement.

L'analyse en équilibre partiel reste simplificatrice, car en réalité l'effet de la mesure envisagée ne se limite pas au seul marché du logement : tous les prix sont susceptibles de s'adapter, en raison des mécanismes d'interdépendance entre les marchés.

Par exemple, les ménages peuvent décider d'augmenter leur offre de travail si le prix des logements augmente, afin d'accroître leur capacité d'emprunt : l'offre de travail passe de OL_1 à OL_2 sur le [graphique 1.7](#). De même, on peut faire l'hypothèse que la demande de travail augmente de DL_1 à DL_2 si les entreprises du bâtiment anticipent une reprise d'activité du secteur et embauchent pour y faire face, ce qui se traduit à son tour par un déplacement de l'offre de logement de O_1 à O_2 avec les nouvelles constructions. Ce processus d'ajustement progressif ne se limite pas aux marchés du logement et du travail, mais concerne l'ensemble des marchés avec lesquels ils sont reliés : le marché du crédit et celui des biens d'équipement destinés aux bâtiments vont s'ajuster également. Ces processus d'interdépendance sont très complexes, et peuvent au final renforcer ou au contraire réduire l'impact initial de la mesure sur le marché du logement. Pour étudier plus complètement l'effet du « prêt à taux zéro », il faut raisonner en **équilibre général**, c'est-à-dire construire une « maquette » de l'économie dans laquelle on détermine simultanément les prix et les quantités d'équilibre sur l'ensemble des marchés. Cela suppose que la demande et l'offre de chaque bien dépendent de l'ensemble des prix. Ainsi la demande de logement des ménages dépend à la fois du prix des logements, du taux de salaire, du prix des biens d'équipement, etc., chaque prix étant considéré comme une variable et non comme un paramètre fixe, contrairement au raisonnement en équilibre partiel.

Graphique 1.7 : L'effet d'une aide en équilibre général

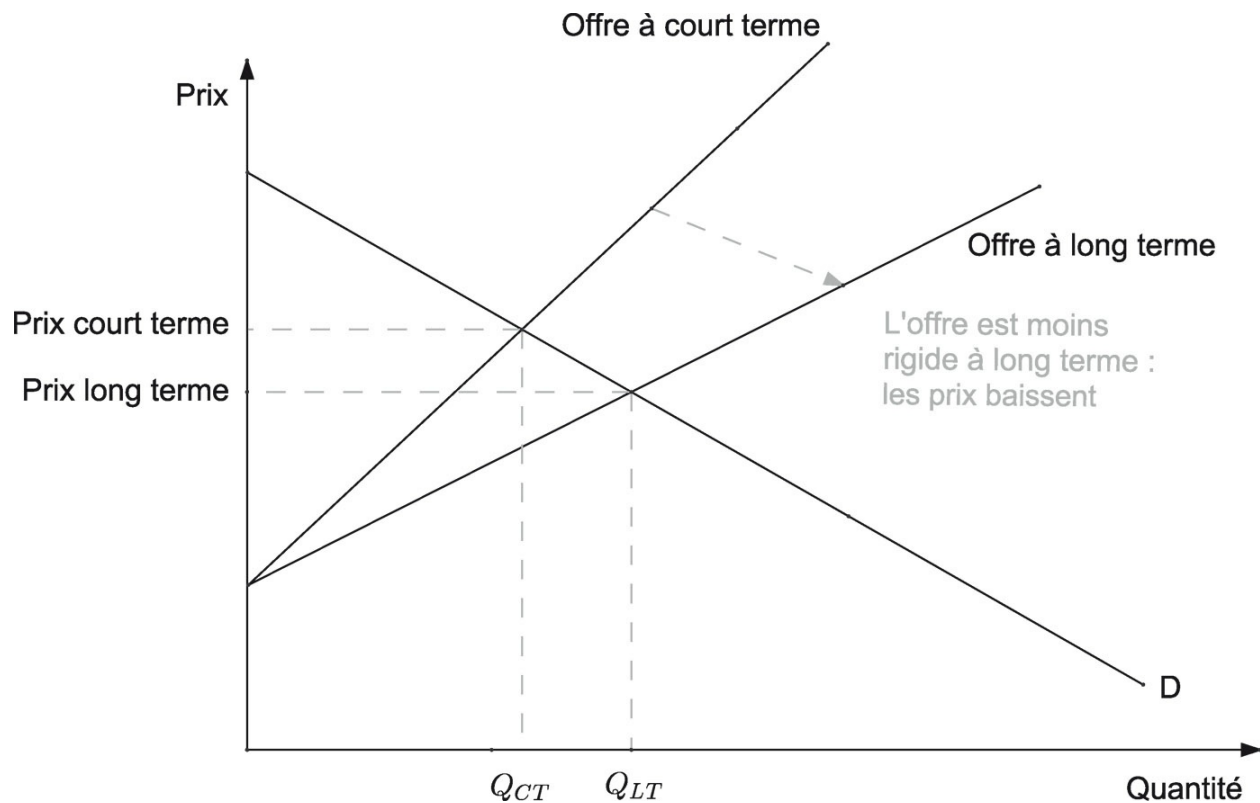


L'approche en équilibre général peut sembler plus rigoureuse, mais elle est aussi beaucoup plus lourde à mettre en place quand il s'agit de mener des études économétriques. Les modèles d'équilibre général sont en général très complexes, il est difficile de formaliser les relations d'interdépendance entre l'ensemble des marchés et le système d'équations n'a pas toujours de solution. Le choix entre le raisonnement en équilibre partiel et en équilibre général dépend en pratique des données disponibles, et du degré de simplification souhaité. Les économistes choisissent souvent des modèles en équilibre partiel, plus faciles à mettre en œuvre, mais dont les résultats doivent toujours être interprétés avec prudence car l'ensemble des interactions entre les marchés ne sont pas prises en compte.

Notre exemple pose un autre problème : il est fort possible que la hausse des prix entraînée par la mesure initiale, à savoir le « prêt à taux zéro », ait des conséquences différentes à long terme et à court terme. Tout dépend en fait de l'ajustement de l'offre et de la demande sur les différents marchés, car ils ne se font pas au même rythme et génèrent des situations de tension sur les prix ou de **rationnement**.

Graphique 1.8 : Équilibre à court terme et à long terme

sur le marché du logement



À long terme, l'offre de logement est probablement plus flexible, à mesure que de nouveaux projets immobiliers arrivent à échéance. En règle générale, l'offre n'a pas la même caractéristique à court ou à long terme, ce qui signifie que les courbes d'offre de court et de long terme n'auront pas la même forme, comme nous pouvons le voir dans le [graphique 1.8](#). On fait souvent l'hypothèse, comme c'est le cas sur ce graphique, que la pente de la courbe d'offre à long terme est moins forte qu'à court terme, car les agents sont plus réactifs aux prix sur la longue durée qu'à court terme, où il existe des rigidités. En lien avec cette évolution de l'offre avec le temps, les prix vont s'ajuster progressivement. L'ajustement de l'offre à la demande au cours du temps est particulièrement complexe sur le marché de l'immobilier, les décisions des agents offreurs et demandeurs n'étant pas prises avec le même « **horizon temporel** », c'est-à-dire qu'ils ne raisonnent pas tous sur les mêmes périodes. Il existe par exemple des phénomènes spéculatifs, car certains agents achètent des biens immobiliers non pour faire usage pendant plusieurs décennies des services qu'ils offrent, mais dans l'espoir de les vendre rapidement en profitant d'une hausse de prix. Ces processus spéculatifs favorisent fortement la volatilité des prix sur le marché du logement, et nécessitent une régulation publique pour limiter les crises

immobilières, comme certains pays en ont connu au moment de la crise des *subprimes* après 2007.

• Encadré 1.2 : Investir dans l'immobilier ?

Les biens immobiliers sont à la fois des biens de consommation et des biens d'investissement. En effet, le logement est d'une part un bien de consommation, car il fournit un service à ses occupants en leur offrant un toit et des éléments de confort, service consommé au cours d'une période donnée. Ce service peut être obtenu en échange d'un loyer, ou bien il est rendu par le logement acquis par l'agent lui-même. D'autre part, le logement est un bien patrimonial, c'est-à-dire un actif durable, dont l'acquisition résulte d'une stratégie d'accumulation du patrimoine. L'investissement repose sur une logique économique très différente de la consommation d'un service : le logement acquis augmente la valeur du patrimoine d'un propriétaire, alors que la consommation d'un service n'enrichit en rien celui qui le consomme. Un ménage qui achète un bien immobilier peut le faire dans différentes perspectives : l'occuper lui-même ou le louer (investissement locatif). Comment comprendre la décision d'un ménage qui achète un bien pour le louer ? Pourquoi investir dans la pierre plutôt que dans d'autres actifs (comme les actifs financiers) ? La microéconomie permet de comprendre les arbitrages entre une ou plusieurs alternatives. On appelle **coût d'opportunité** d'une décision le coût entraîné par la renonciation à la meilleure opportunité alternative : le propriétaire qui choisit d'investir dans un logement locatif renonce de fait à ce que lui aurait rapporté son capital s'il avait été placé sur le marché financier. Tout arbitrage suppose donc de tenir compte du coût d'opportunité des autres solutions.

Pour financer un investissement, il est souvent nécessaire d'emprunter une partie du capital requis pour acheter l'actif désiré. L'investisseur doit ainsi tenir compte du coût du crédit, plus ou moins important selon le niveau des taux d'intérêts. Dans le cas d'un investissement locatif, le logement rapporte ensuite des flux de revenus (les loyers), mais génère aussi des coûts à la charge du propriétaire (entretien, fiscalité, etc.). Au moment de la décision d'achat, l'investisseur tient aussi compte de la valeur de revente probable de son bien, en anticipant au mieux ses variations de prix. Il doit aussi intégrer l'inflation, son effet sur les prix et les revenus, d'autant que la décision d'investir l'engage à long terme. Tous ces éléments lui permettent de calculer le rendement d'un investissement locatif, c'est-à-dire le rapport entre les revenus annuels qu'il génère et le capital investi. Un investisseur rationnel devrait investir dans l'immobilier plutôt que sur le marché des actions si le rendement des placements financiers est inférieur au rendement d'un investissement locatif. Les observations empiriques montrent que les ménages ne se comportent pas toujours de manière aussi rationnelle. Selon J. Friggit², le rendement de l'investissement locatif est devenu très faible au cours des années 2000 alors que les prix d'acquisition s'envolaient. Des investisseurs rationnels auraient dû se tourner vers des placements financiers. Mais il n'en est rien en raison de la myopie des investisseurs, qui évaluent mal les risques liés au placement dans la pierre, notamment les fortes variations du prix des biens immobiliers à long terme, alors que les agents ont tendance à surévaluer les risques liés aux actifs financiers (la crainte des krachs boursiers par exemple). Il s'agit d'un biais systématique en faveur de l'investissement immobilier, c'est-à-dire d'un écart de comportement régulièrement observé par rapport à la décision la plus rationnelle (voir le [chapitre 2](#)). La myopie des acteurs joue un grand rôle dans les crises immobilières, qui interviennent d'autant plus facilement dans un contexte d'euphorie collective, où les agents sous-estiment les risques de retournement des prix immobiliers.

3. La régulation d'un marché concurrentiel

Dans les conditions de la concurrence pure et parfaite et sous réserve que les biens aient certaines caractéristiques (biens rivaux et excluables) et que les droits de propriété soient bien définis (voir [chapitre 5](#)), le libre jeu de l'offre et de la demande permet d'allouer efficacement les ressources. Les surplus du consommateur et du producteur sont maximaux, aucun côté du marché n'est rationné puisque pour un prix donné les quantités offertes sont égales aux quantités demandées, les prix relatifs reflètent les préférences des consommateurs et sont les plus bas possibles, égaux aux coûts marginaux de production. Les mécanismes qui sont à l'origine de cette situation sont détaillés dans les chapitres 2 et 3. Dans ces conditions, d'un point de vue théorique, on peut se demander pourquoi l'État voudrait modifier un tel équilibre. Pourtant, la puissance publique ne manque pas de raisons d'intervenir. Elle peut vouloir mettre en place un impôt sur les produits (comme la TVA) afin de trouver les ressources budgétaires dont elle a besoin pour financer son action dans d'autres domaines, même si le marché fonctionne efficacement. Elle peut aussi décider d'intervenir pour des considérations de justice sociale. Rien ne garantit en effet que l'allocation des ressources à l'équilibre soit juste, même si elle est efficace. Par exemple un prix d'équilibre élevé peut exclure du marché une partie des demandeurs, et être la cause du mal-logement d'une part trop importante de la population. Nous détaillerons dans le chapitre 7 comment les économistes ont conceptualisé les liens entre justice sociale et efficacité économique et la façon dont ont été intégrés des critères de justice sociale dans l'évaluation des situations économiques.

Au-delà des justifications théoriques de l'intervention publique sur les marchés de concurrence pure et parfaite, l'État peut intervenir sur des marchés concrets car leur fonctionnement est précisément caractérisé par diverses imperfections, et l'allocation des ressources n'a plus de raison d'être optimale dans ces conditions. La présence d'**externalités** ou d'**asymétries d'informations** rend l'équilibre sur le marché sous optimal, et une politique publique appropriée peut améliorer la situation (voir les chapitres [4](#) et [5](#)).

La situation du marché du logement est jugée très problématique par

beaucoup. À la fois injuste socialement (prix exorbitants, surtout pour l'achat d'un logement) et économiquement inefficace (existence de logements vacants alors que de nombreuses personnes peinent à se loger, manque d'entretien des logements existants, zones surpeuplées et zones désertées), les politiques publiques qui ont été menées n'ont pas réussi à changer la situation. Comment comprendre ces échecs ? L'État doit-il favoriser l'offre de logements en encourageant la construction ou aider les demandeurs à financer leurs achats ? Doit-on limiter administrativement la hausse des prix immobiliers ?

Le modèle de concurrence pure et parfaite peut être mobilisé pour comprendre certains des échecs des politiques passées en matière de logement et mettre à jour certains mécanismes liés à l'intervention de l'État. Il permet également d'anticiper les effets pervers potentiels de nouvelles politiques. Bien évidemment, l'action de l'État ne peut pas être évaluée uniquement à l'aune de ce modèle, mais nous verrons en quoi il éclaire certains mécanismes couramment observés.

L'État dispose de différents instruments pour modifier l'équilibre sur un marché :

- il peut jouer sur les prix par des systèmes de **taxes** et **subventions**, comme les aides au logement, mettre en place des **prix plafond** ou **prix plancher**, comme les loyers bloqués en France avant 1948 ;
- il peut jouer sur les quantités en instaurant des **quotas** ou en produisant lui-même, par exemple la loi SRU (loi relative à la solidarité et au renouvellement urbain) impose à certaines communes 20 % de logements sociaux sur leur territoire ;
- l'État peut enfin jouer sur la **réglementation** en modifiant les **normes** encadrant les transactions, par exemple en instaurant des normes environnementales.

Quels sont les avantages et inconvénients de ces différents instruments ? Comment mesurer l'impact d'une politique publique ? Dans la suite de ce chapitre, nous allons présenter différents types d'intervention de l'État et analyser la façon dont l'équilibre est modifié par la politique mise en œuvre. Il ne s'agit pas d'évaluer concrètement l'impact de telle ou telle politique mais de mettre en évidence l'existence de mécanismes très généraux. Nous raisonnerons en équilibre partiel, en ne tenant pas compte des répercussions

d'une politique sur les autres marchés afin de simplifier l'analyse.

3.1 La régulation par les prix

L'État peut modifier l'**allocation des ressources** en jouant sur les prix. L'allocation des ressources renvoie à la façon dont les différentes ressources disponibles sont affectées à des usages alternatifs. Prenons l'exemple du marché du logement locatif. Si, en France, la hausse des loyers est beaucoup moins forte que celle des prix d'achat des logements, le niveau des loyers reste élevé et difficilement supportable pour les ménages modestes, particulièrement dans les grandes villes. De quelle façon l'État peut-il intervenir pour que les loyers pèsent moins sur le budget des ménages ?

Trois solutions s'offrent à lui : subventionner les agents, les taxer, ou fixer des limites aux variations des prix (prix maximal ou minimal).

• Subventionner les agents

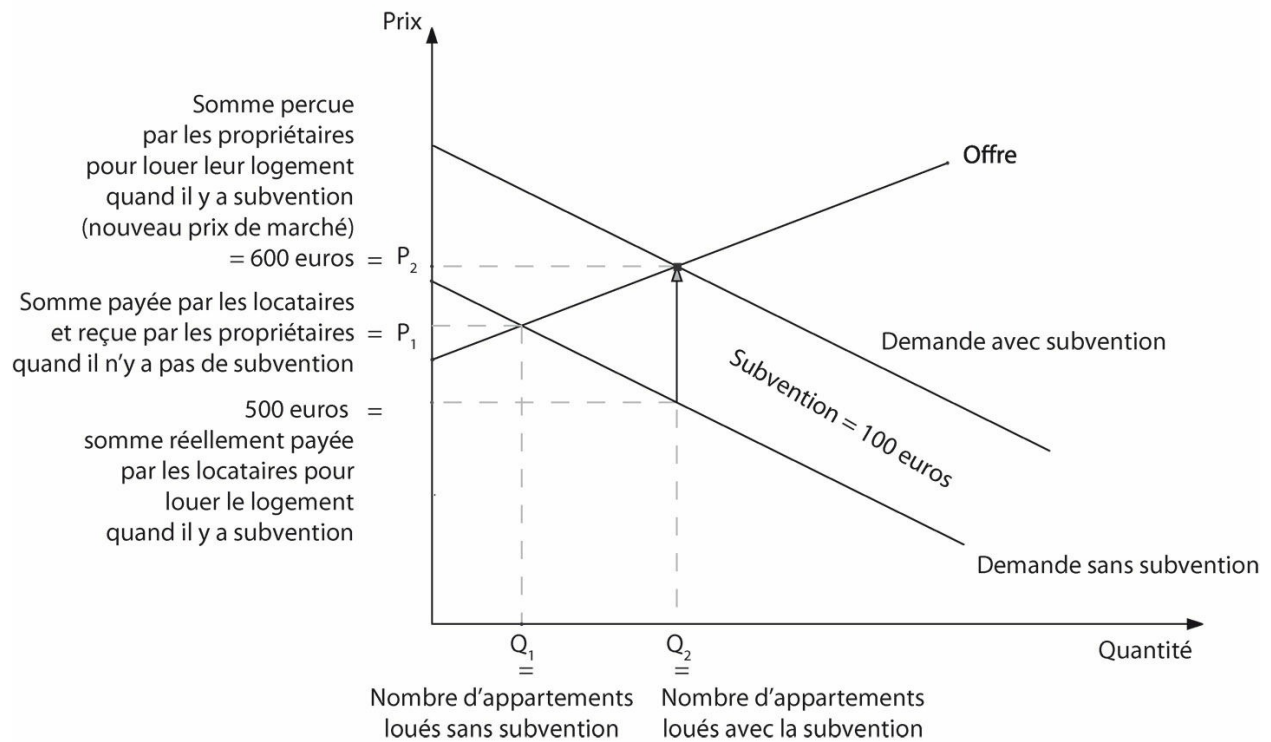
L'État peut tout d'abord verser des **subventions** aux ménages locataires, notamment les ménages à faibles revenus. L'idée est de ne pas jouer directement sur le prix du marché mais de donner une certaine somme à chaque ménage qui loue un appartement afin que le poids du loyer dans le budget soit plus faible. L'allocation d'aide au logement ou les aides personnalisées aux logements, comme les fameuses APL qui aident tant d'étudiants, relèvent de cette politique. Ce type d'initiative paraît pertinent mais en réalité l'État contribue à augmenter la demande en versant une subvention aux locataires, car certains ménages auparavant exclus du marché peuvent à présent y entrer et la courbe de demande va se déplacer ce qui augmente le prix des loyers en retour pour l'ensemble des ménages.

Supposons que l'État subventionne les personnes qui louent des studios à Paris pour compenser la cherté des loyers. Considérons deux cas simplifiés :

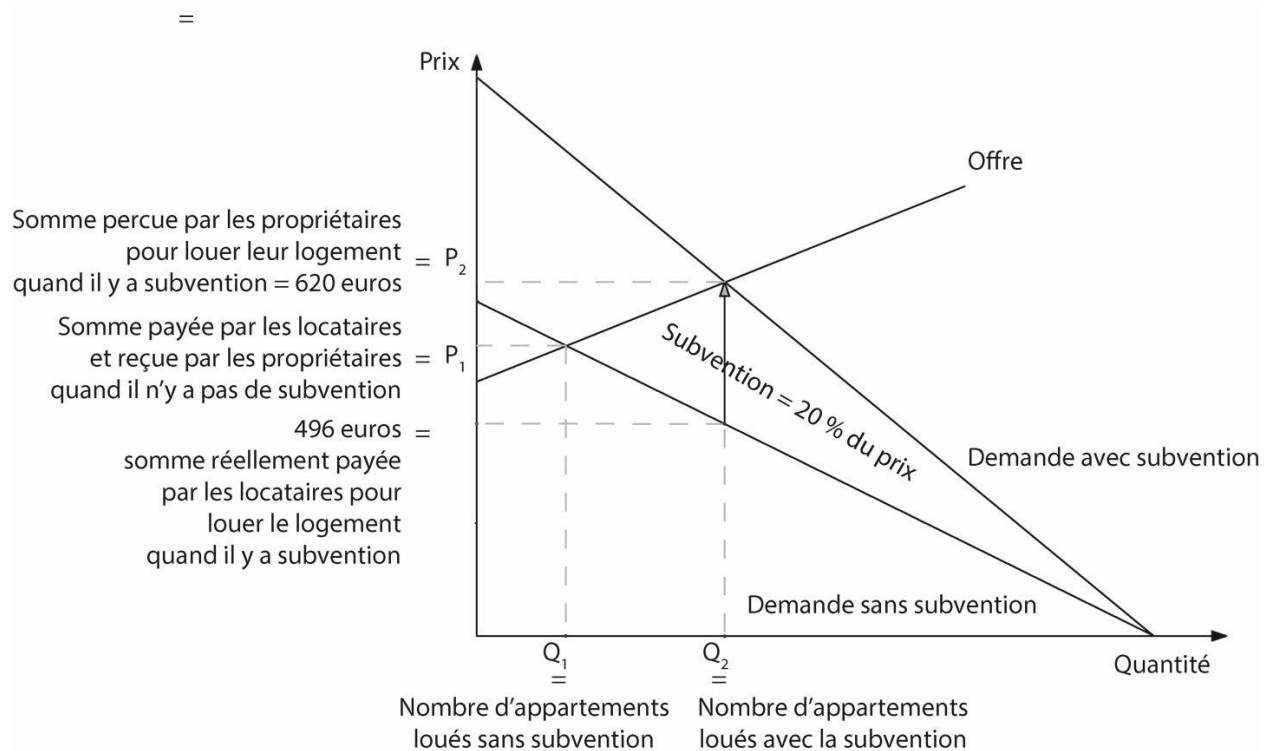
1) Une aide d'un montant fixe pour tous les ménages, par exemple 100 euros par mois à tous ceux qui louent un studio à Paris. Graphiquement, une aide ou un impôt de ce type entraîne un déplacement de la courbe, qui conserve la même pente en étant simplement « décalée » du montant de l'aide ou de l'impôt. Dans notre exemple, les ménages sont disposés à payer 100 euros supplémentaires pour chaque quantité demandée ([graphique 1.9](#)).

2) Une aide proportionnelle pour tous les ménages, par exemple une aide qui représente 20 % du loyer payé. Graphiquement, une aide ou un impôt proportionnel se traduit en revanche par un changement de pente de la courbe, car l'effet de la mesure s'exerce proportionnellement sur la demande, comme on le voit sur le [graphique 1.10](#).

Graphique 1.9 : Les effets d'une subvention d'un montant fixe de la demande



Graphique 1.10 : Les effets d'une subvention proportionnelle de la demande



On observe que les résultats de la politique de subvention sont ambigus : plus de logements sont loués mais le prix des loyers a augmenté. Certes, le coût réellement supporté par les locataires est inférieur à la situation sans subvention mais une partie de l'aide a bien profité aux propriétaires du fait de la hausse des loyers. En fonction de la forme des courbes d'offre et de demande, une grande partie de l'aide peut au final profiter aux propriétaires, ce qui n'était pas le but poursuivi. Des études ont ainsi montré que les APL n'ont pas vraiment contribué à faire baisser la part du logement dans le budget des étudiants vivant à Paris mais sont à l'origine d'une hausse des loyers du fait d'une offre de logement très rigide.

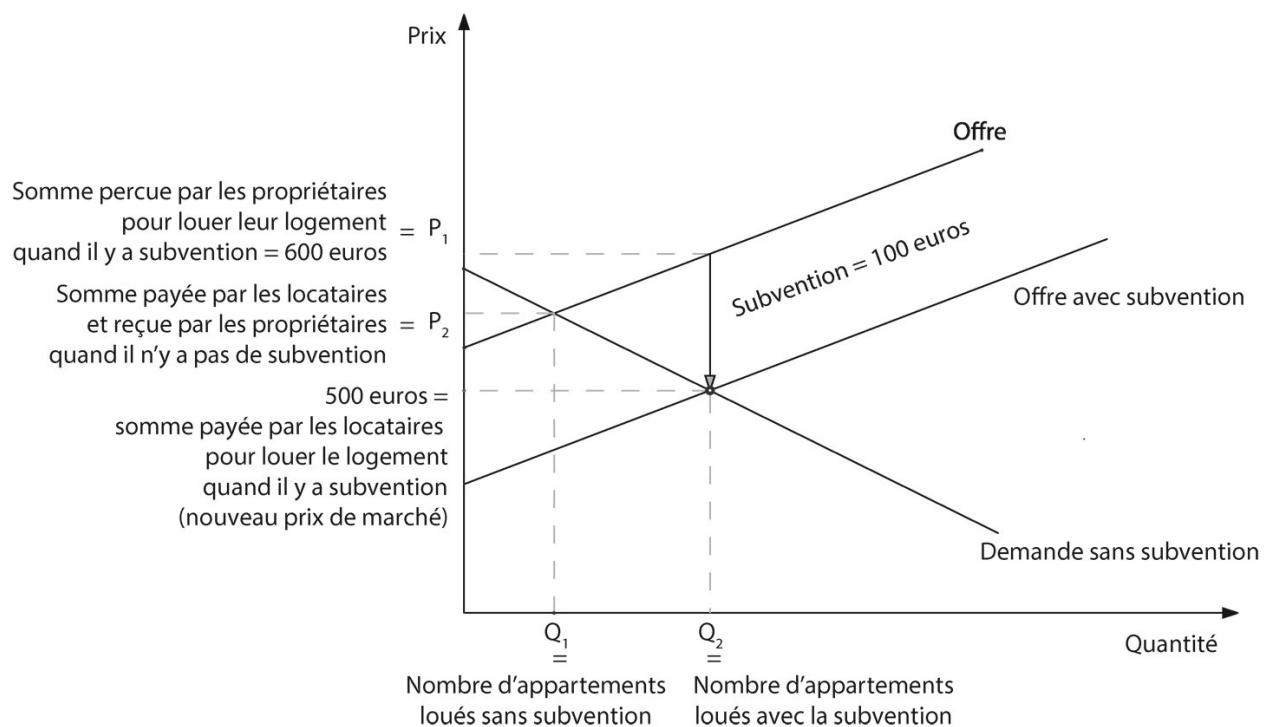
• Encadré 1.3 : Les effets des aides au logement

G. Fack (2005), A. Laferrère, A. et D. Le Blanc (2002) ont étudié la réforme d'extension des aides qui a eu lieu au début des années 1990 en France ; ils ont montré que les aides au logement sur les loyers avaient pour conséquence d'en faire augmenter les prix. Selon G. Fack (2010) « les estimations suggèrent qu'entre 50 % et 80 % des allocations logement perçues par les ménages à bas revenus suite à la réforme ont été absorbées par les augmentations de leurs loyers. L'étude suggère aussi que la réforme des aides, qui a permis aux étudiants de bénéficier de subventions pour se loger, a contribué à l'arrivée massive des étudiants sur le marché du logement, concourant ainsi à alimenter l'effet inflationniste des aides [...]. Ces résultats appellent deux commentaires. D'une part, la hausse des aides au logement a pu entraîner une amélioration de l'habitat, qui explique en partie la hausse des prix. Cependant, d'après les mesures sur la qualité des logements disponibles dans les enquêtes, cet effet semble trop faible pour suffire à expliquer la hausse des loyers. D'autre part, l'effet inflationniste à court et moyen terme peut s'atténuer dans le long terme, si la hausse des loyers améliore la rentabilité des investissements immobiliers et entraîne la construction de nouveaux logements. »

C'est ce type de considération qui est mis en avant par le gouvernement Philippe pour justifier la baisse des APL de 5 euros en 2017. L'idée est qu'*in fine* ce ne sont pas les locataires qui paieront ces 5 euros mais les propriétaires à travers des baisses de loyer. S'il est douteux de penser que 100 % de la baisse se reportera sur le niveau des prix et s'il est certain qu'un ajustement des prix à la baisse prendra du temps, la mesure devrait avoir un effet, certes minime, sur les prix des loyers, en limitant la demande sur le marché, ce qui bénéficiera à l'ensemble des locataires et pas seulement aux destinataires des APL. Cela ne règle pour autant pas le problème de l'accès au logement car la baisse des loyers découle ici d'une baisse de la demande du fait d'une diminution de la capacité à payer et non d'un moindre besoin de logement.

Pour éviter les effets inflationnistes des subventions de la demande, une solution pourrait être de coupler à ce type d'aides d'autres mesures de subvention qui favorisent l'offre de logement locatif. En France de nombreux dispositifs existent pour encourager les investissements locatifs comme les dispositifs Robien, Scellier, Pinel, etc. Une subvention versée à tous les offreurs sur un marché pour chaque transaction a néanmoins les mêmes effets qu'une subvention d'un même montant versée à tous les demandeurs (voir le [graphique 1.11](#)).

Graphique 1.11 : Les effets d'une subvention d'un montant fixe de l'offre

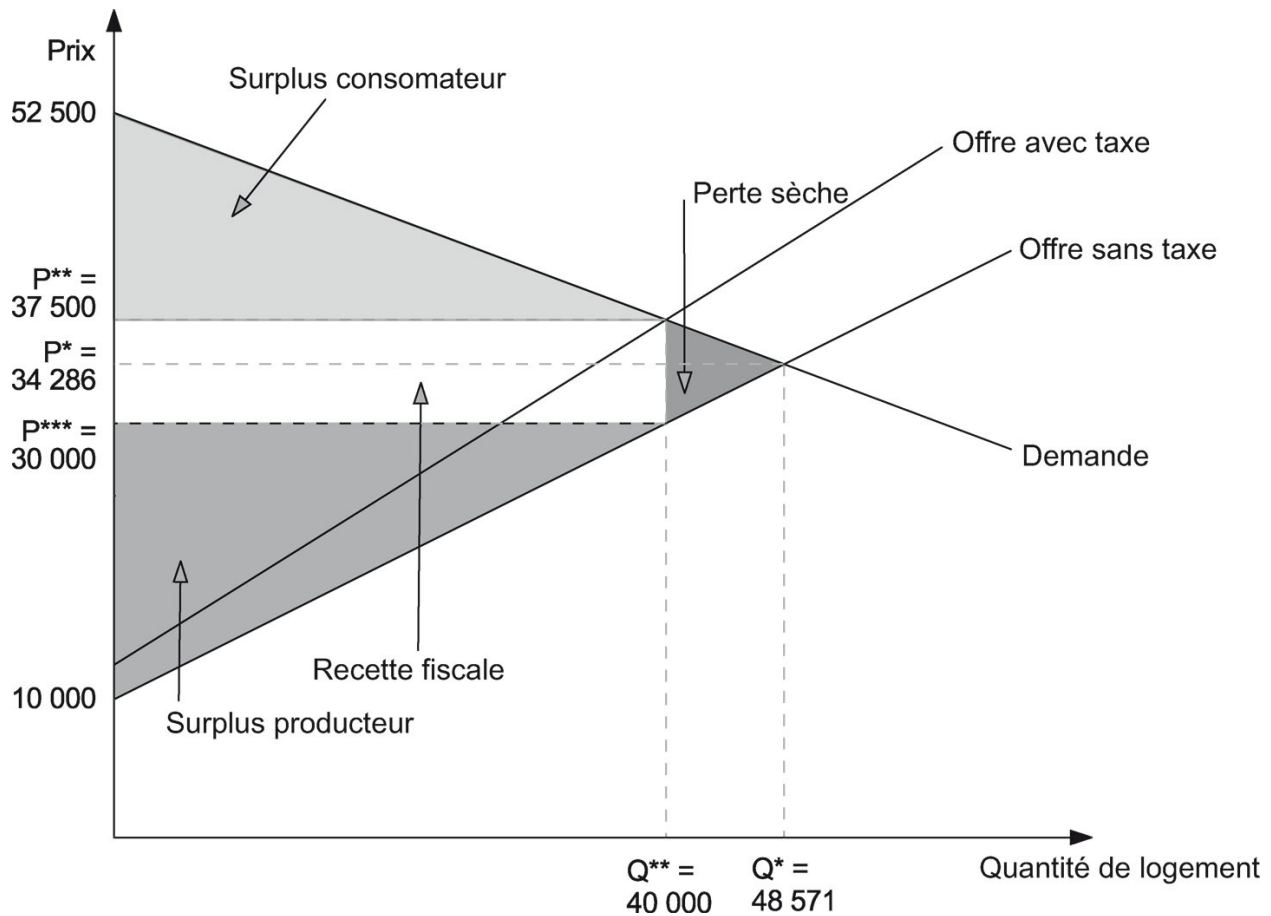


Le prix de marché diffère du cas d'une subvention des locataires de 100 euros mais les sommes réellement perçues et acquittées une fois les subventions prises en compte sont les mêmes dans les deux situations, comme nous pouvons le voir sur les graphiques précédents. En revanche, les subventions à l'offre ou à la demande n'ont pas les mêmes effets si elles ne concernent pas l'ensemble des transactions. Par exemple, une allocation logement versée sous condition de ressources ne concerne pas toutes les transactions mais seulement celles opérées par un ménage à bas revenus.

De fortes subventions peuvent faire baisser le coût des loyers et accroître les quantités offertes, cependant cette solution n'est pas pour autant idéale. Des politiques de subvention de l'offre et de la demande sont mises en œuvre en France et pourtant le mal logement reste un problème majeur. Les politiques d'incitations fiscales à l'investissement locatif ont conduit à des constructions excessives dans des zones où la situation était relativement satisfaisante sans rien changer dans certaines grandes villes où la crise du logement bat son plein. Pour les rendre plus efficaces, il faut donc différencier les mesures prises dans les « zones tendues » où l'écart entre offre et demande se fait le plus sentir, comme les centres urbains, et les autres territoires où la pression de la demande est moins forte.

Par ailleurs, il faut octroyer des subventions considérables pour que l'impact sur le marché soit important ce qui représente un effort budgétaire conséquent. De telles politiques impliquent ainsi de lever des ressources importantes pour les financer ce qui engendrera des distorsions. Les taxes modifient en effet l'allocation à l'équilibre et génèrent souvent des pertes d'efficacité.

Graphique 1.12 : Les effets d'une taxe proportionnelle



• Les effets de la taxation

Voyons comment l'effet d'une taxe peut être modélisé. Il s'agit en fait de l'effet inverse d'une subvention. De nombreuses taxes concernent le secteur du logement. La taxe foncière par exemple est acquittée par les propriétaires de logements et diminue la rentabilité d'une location. La taxe d'habitation est acquittée par les occupants d'un logement, qu'ils soient locataires ou propriétaires, et renchérit le coût d'une location. De quelle façon une taxe modifie-t-elle l'équilibre du marché ? Considérons le cas d'une taxe payée par les propriétaires au moment de la vente de leur bien et proportionnelle au prix de vente. Le [graphique 1.12](#) montre l'effet d'une taxe proportionnelle de 20 % : quand un propriétaire vend son bien, il ne conserve que 80 % du prix de vente et verse 20 % à l'État. La pente de l'offre ainsi taxée est plus forte car elle est égale à 1,2 fois celle de l'offre non taxée, car on ajoute 20 % au prix initial.

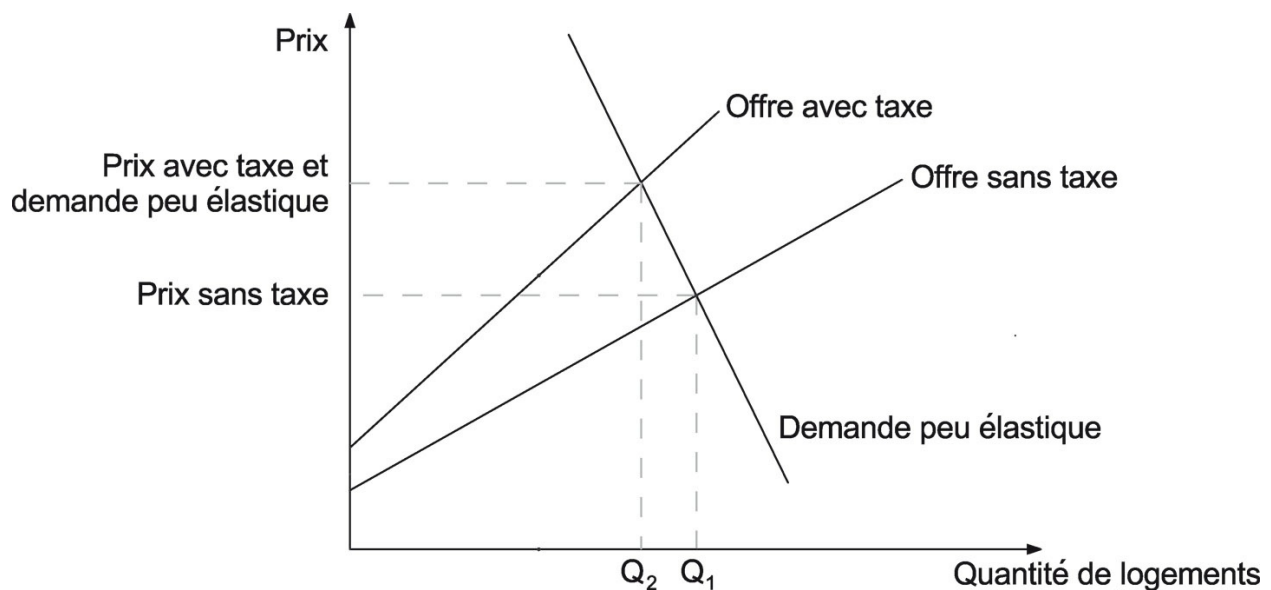
La **taxe** entraîne une baisse de la quantité offerte à l'équilibre sur le

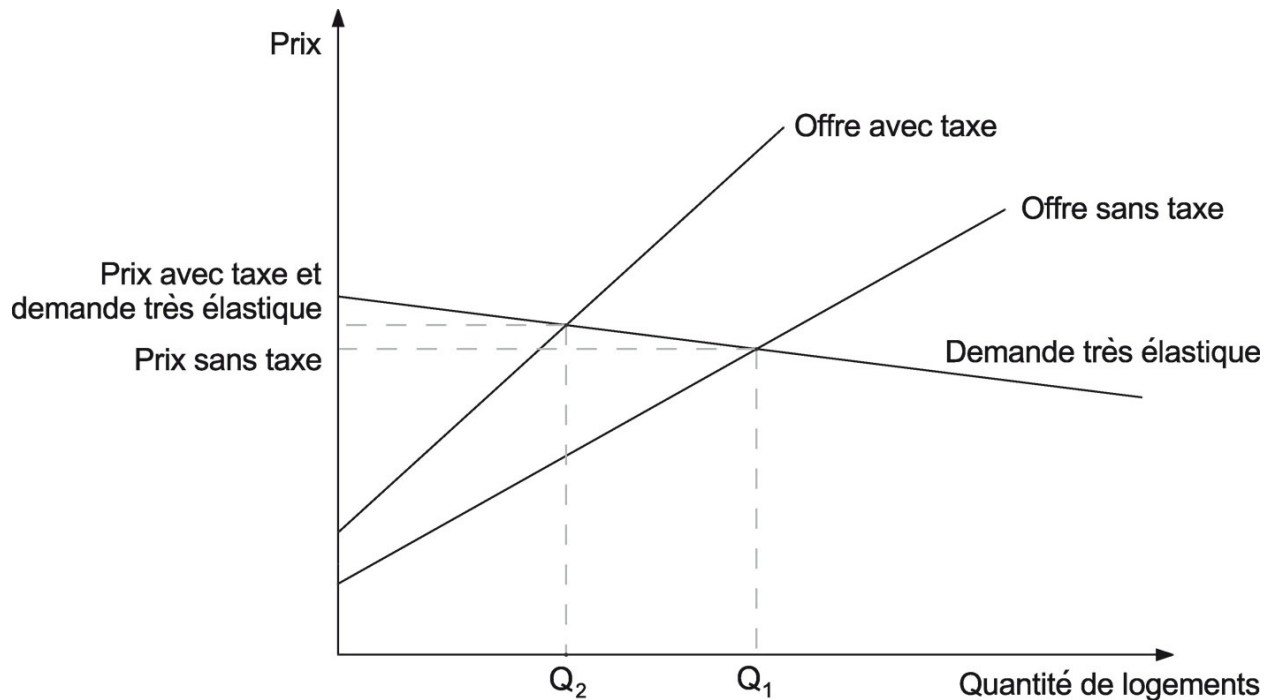
marché, qui passe de Q^* à Q^{**} comme on le voit sur le graphique : compte tenu de la forme des courbes d'offre et de demande, le nouveau prix d'équilibre P^{**} est plus élevé que l'ancien P^* , et la quantité échangée plus faible. P^{**} représente la somme que doivent dépenser les acheteurs pour acquérir le bien. Les propriétaires perçoivent eux la somme P^{***} car il faut déduire la taxe du prix payé par l'acheteur. En renchérissant les prix, la taxe a réduit le nombre de ménages qui peuvent se loger et a augmenté le prix moyen des achats sur le marché. Les surplus du consommateur et du producteur sont plus faibles qu'en l'absence d'intervention de l'État. On parle de **perte sèche** pour désigner la perte nette de surplus, c'est-à-dire les gains à l'échange auxquels les agents doivent renoncer en raison de la taxe. Avant intervention de l'État, le surplus total valait 1 032 133 750 euros ($48\,571 \times (52\,500 - 10\,000)/2$). La mise en place d'une taxe va réduire le surplus des offreurs et des demandeurs. Le surplus des consommateurs vaut à présent 300 millions d'euros (contre 442 336 097 euros avant intervention de l'État) et le surplus des producteurs 400 millions (contre 589 797 653 euros avant intervention de l'État). L'État a bénéficié de recettes fiscales égales à 300 millions d'euros ($40\,000 \times (37\,500 - 30\,000)$). Le surplus total est égal à la somme des surplus du consommateur, du producteur et de l'État, soit 1 milliard d'euros. La perte sèche est donc ici de 32 133 750 euros par rapport au surplus total avant la taxation. Une perte sèche importante ne signifie absolument pas que l'impôt soit inutile ou inefficace, car les sommes collectées peuvent être utilisées pour financer des subventions ou des services publics qui font accroître le surplus des consommateurs ou des producteurs dans d'autres domaines. C'est simplement un outil pour comparer différentes situations de marché, avec ou sans intervention publique, et mesurer les distorsions générées par la taxe sur ce marché précis. La taxe a conduit les agents à acheter plus cher un nombre moins grand de logements, et la notion de perte sèche propose une évaluation monétaire de cette « perte » pour la collectivité.

Bien que la taxe soit payée formellement par les propriétaires, nous voyons sur ce graphique qu'elle est au final supportée en large partie par les acheteurs. Sur 7 500 euros de taxe, 3 214 euros sont payés par les consommateurs (soit environ 43 % du montant de la taxe), car le prix de marché est passé de 34 286 euros à 37 500. Les propriétaires ne paient ainsi réellement que 4 286 euros. Les mécanismes de marché, en conduisant à des loyers plus élevés, font en partie supporter le poids de la taxe aux acheteurs.

On retrouve pour les taxes le même résultat que pour les subventions : il est équivalent, du point de vue des sommes perçues et acquittées, de taxer d'un montant identique toutes les ventes ou tous les achats. Il est ainsi crucial de distinguer l'entité qui paye administrativement une taxe (ou reçoit un transfert) des acteurs qui supportent économiquement le poids de la taxe (ou bénéficient du transfert). En effet, les taxes et subventions modifient l'équilibre du marché et affectent ainsi l'ensemble de ses acteurs. L'**incidence** désigne les répercussions d'une mesure fiscale (on parle d'**incidence fiscale**) ou encore d'une subvention sur les différents agents présents sur le marché concerné. Qui au final est amené à supporter économiquement la mesure fiscale, comment s'effectue le partage de l'effort entre les offreurs et les demandeurs ? Qui bénéficie réellement des subventions versées ?

Graphique 1.13 : Incidence fiscale et élasticité





La façon dont la taxe se reporte sur les différents agents dépend de leur sensibilité aux prix : une taxe sur l'offre est d'autant plus reportée sur la demande *via* une hausse des prix que la demande est peu élastique, comme on le voit sur le [graphique 1.13](#). En revanche, si la demande est très sensible aux prix, la taxe va être principalement supportée par les offreurs et le prix d'équilibre augmentera faiblement.

Nous voyons en effet sur le graphique qu'une même taxe sur l'offre a des effets très différents sur le prix d'équilibre en fonction de l'élasticité de la demande. Dans le cas d'une demande peu élastique, le prix de marché augmente fortement et ce sont les consommateurs qui supportent principalement le poids de la taxe. Dans le cas d'une demande très élastique, le prix de marché augmente peu et ce sont les offreurs qui supportent principalement la taxe. La notion d'incidence fiscale permet donc de comprendre pourquoi, selon les caractéristiques d'un marché, une taxe peut avoir des effets très variables sur les prix et la situation de chaque type d'agent. C'est un argument essentiel dans le débat public, car l'incidence d'une taxe est loin d'être systématiquement prise en compte lorsqu'il s'agit de décider entre plusieurs mesures fiscales.

• Contrôle des prix et rationnement

Pour éviter que le soutien à l'offre ou à la demande modifie le prix d'équilibre et conduise parfois à des loyers encore plus élevés qu'avant, certains ont avancé l'idée de plafonner les loyers. En France, les loyers étaient plafonnés entre 1914 et 1948 ; certains le sont restés ensuite mais la loi de 1948 a mis fin au contrôle généralisé des loyers. Aujourd'hui, les loyers ne sont fixés qu'en partie par des mécanismes de marché : en effet, au renouvellement du bail, les propriétaires ne peuvent pas augmenter le loyer comme ils l'entendent. L'État fixe une augmentation maximale. Le contrôle des prix est pratiqué sur de nombreux marchés, comme par exemple le marché du travail avec l'existence d'un salaire minimum. Quel effet a un contrôle direct des prix sur l'équilibre ?

L'instauration d'un **prix plafond** entraîne un déséquilibre entre les quantités offertes et demandées, si le prix considéré est inférieur au prix d'équilibre. Le nombre des locataires au prix plafond est plus élevé que le nombre de logements à louer à ce prix : il y a **rationnement** de la demande. L'ajustement ne pouvant plus se faire par les prix, il va se faire par les quantités : certains demandeurs ne trouveront pas à acheter la quantité désirée. Certes les loyers sont plus bas mais certains ménages ne trouvent pas de logement du fait d'une contraction de l'offre.

• Encadré 1.4 : le rationnement

Le rationnement désigne une situation où les quantités offertes ne sont pas égales aux quantités demandées. Le côté court du marché désigne le côté où les quantités sont les plus faibles. Quand les quantités offertes sont inférieures aux quantités demandées, on dit que la demande est rationnée. À l'inverse quand les quantités demandées sont inférieures aux quantités offertes, c'est l'offre qui est rationnée. Une situation de rationnement existe quand le prix est rigide à un niveau qui n'est pas égal au prix d'équilibre. Il ne s'agit pas d'un déséquilibre temporaire qui va disparaître avec la variation du prix. Il s'agit d'une situation qui découle de l'existence de mécanismes qui bloquent la flexibilité des prix. Cela peut être du fait d'un blocage des prix par l'État mais aussi du fait d'un comportement rationnel des acteurs. Il n'est en effet pas rationnel de modifier ses prix en permanence (existence de coûts de catalogue), ainsi les prix ne sont donc jamais parfaitement flexibles même quand aucun dispositif institutionnel ne limite leurs variations. Par ailleurs, les asymétries d'information peuvent conduire à ce que le prix se fixe à un niveau différent du niveau d'équilibre (voir [chapitre 4](#)).

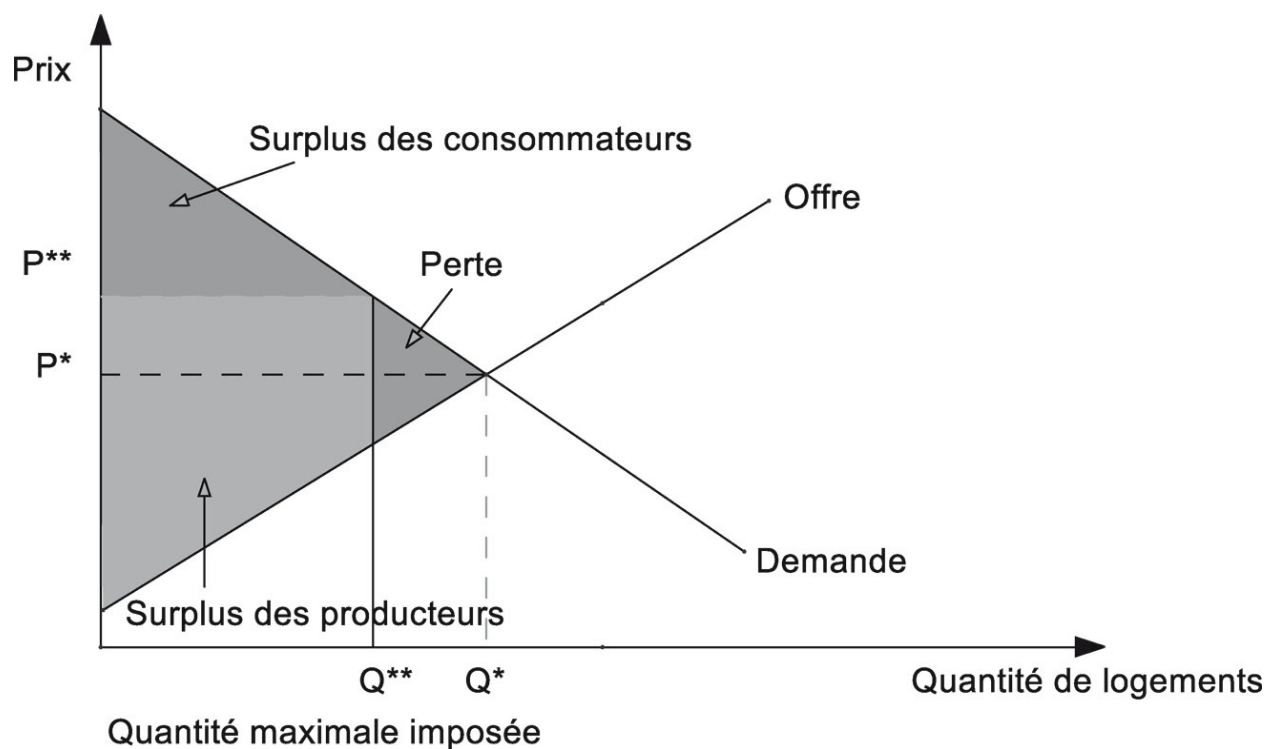
En situation de rationnement, l'**ajustement** ne se fait plus par les prix mais par les quantités. Ce ne sont plus les variations de prix qui déterminent l'allocation des ressources. Quand l'ajustement se fait par les prix, les biens et services sont alloués à ceux qui ont la capacité de payer leur prix et tous ceux qui ont cette capacité et désirent acheter le bien y ont accès. Il n'y a pas de rationnement car tous ceux qui veulent vendre ou acheter au prix de marché le peuvent. En situation de rationnement, au prix de marché, les quantités offertes ne sont plus cohérentes avec les quantités demandées et une autre procédure doit

être mise en place pour gérer l'allocation des ressources. En effet, les prix étant rigides, ce n'est plus le mécanisme des prix qui peut coordonner le comportement des acteurs.

Concrètement, un rationnement de la demande peut être géré par un système de file d'attente (premier arrivé, premier servi), un système de tirage au sort ou un système de ticket de rationnement, à titre d'exemples.

La baisse des loyers n'a pas conduit à ce que davantage de personnes accèdent au logement, bien au contraire. Par ailleurs, un des effets pervers de ce type de mesure est que les propriétaires n'ont aucun intérêt à entretenir le logement qu'ils louent. En effet, un appartement rénové ne pouvant pas être loué plus cher, le propriétaire n'a aucune incitation à investir dans une rénovation. Le contrôle des loyers va donc avoir pour conséquence de faire baisser la qualité moyenne des logements.

Graphique 1.14 : Les effets du contrôle des prix



Ces deux types d'effets mis en avant théoriquement se retrouvent au niveau empirique : en France, les loyers ont été bloqués entre 1914 et 1948, on a observé que la construction de nouveaux logements a été faible et les logements loués ont été très mal entretenus. L'État a mis fin au contrôle des loyers du fait de la vétusté et de la faiblesse du parc immobilier français en

1948.

Depuis 2012, il existe un encadrement de l'évolution des loyers lors de la relocation d'un logement ou lors du renouvellement du bail, dans les « zones tendues » (agglomérations qui connaissent une forte tension du marché locatif). Dans les communes concernées, le loyer ne peut donc pas être augmenté plus que l'indice de référence des loyers (IRL) fixé chaque année. Par ailleurs, depuis août 2015 à Paris et février 2017 à Lille, dans le cadre de la loi pour l'accès au logement et un urbanisme rénové (Alur) a été créé un nouveau dispositif, pérenne et complémentaire, d'encadrement du niveau des loyers. L'État ne fixe donc pas le prix des loyers mais leurs évolutions doit respecter certaines contraintes définies par l'État.

Aucune politique publique de régulation par les prix ne semble parfaite. Une solution pourrait être de coupler ces mesures avec des politiques de régulation par les quantités. En effet, si l'on considère le cas précédent, que se passerait-il si l'État produisait lui-même des logements afin d'augmenter l'offre ?

3.2 La régulation par les quantités

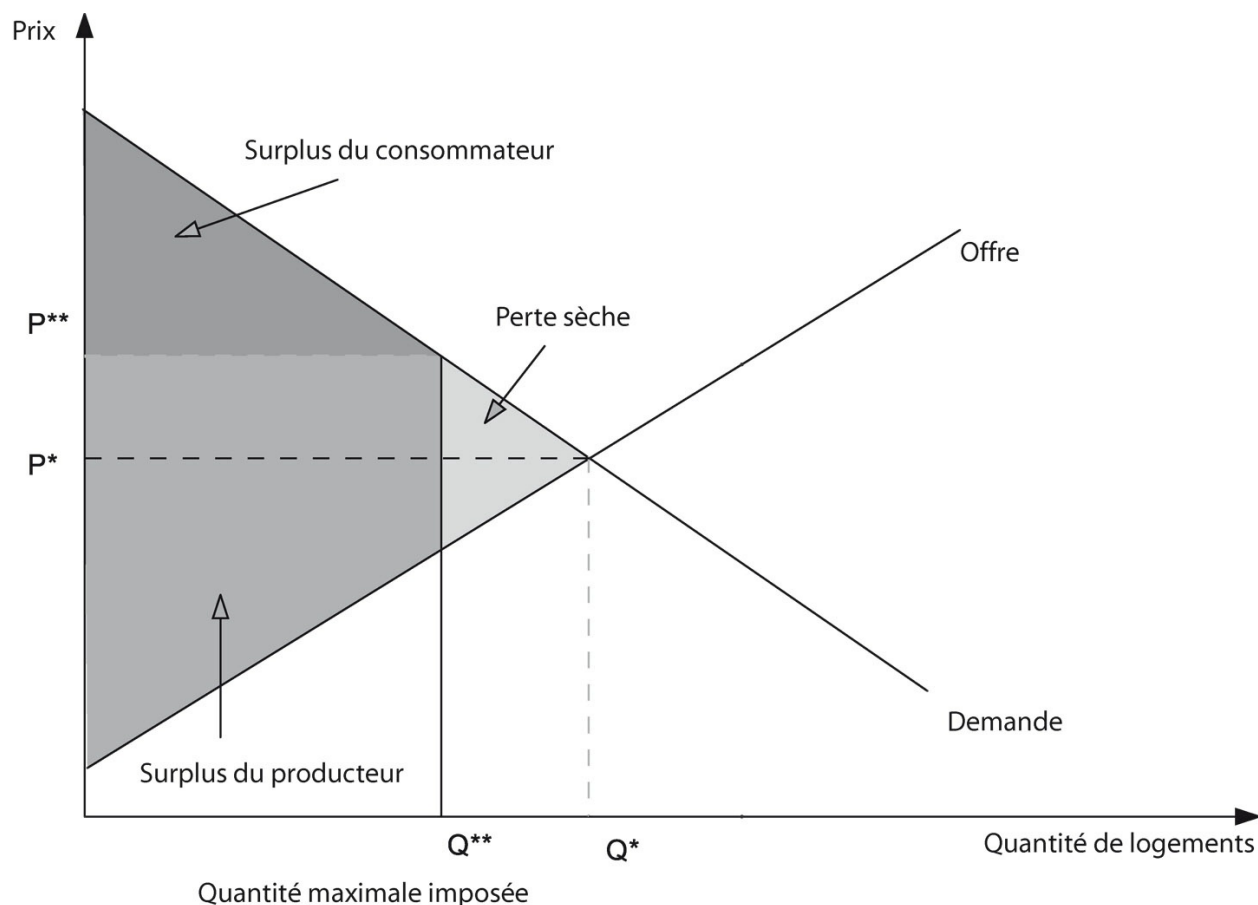
La régulation par les quantités suppose que l'État agit directement sur les quantités offertes ou demandées pour un niveau de prix donné et renvoie à deux grands types de politique : mise en place de quotas ou production publique.

• Les effets d'un quota sur le marché

Les **quotas** sont en général mis en œuvre pour soutenir un segment du marché et imposent une production maximale que les producteurs ne doivent pas dépasser ; ce fût le cas, par exemple, des quotas laitiers en matière de politique européenne agricole. Dans le cas du logement, on peut envisager une limitation volontaire de la construction de nouveaux logements. Supposons qu'une petite commune interdise toute nouvelle construction de logements pour préserver son « côté village » : elle impose ainsi une quantité de logement maximale. Ce type de contrainte sur l'offre de logement est assez courant sur les sites protégés ou les zones historiques où toute nouvelle construction est interdite. On parle de l'existence d'un **quota maximal** fixé

en dessous de la quantité qui aurait été offerte sans régulation.

Graphique 1.15 : Les effets des quotas



La mise en place d'un quota, comme on l'observe sur le [graphique 1.15](#), favorise la hausse du prix par rapport à une situation sans intervention de l'État et entraîne une perte de surplus des acheteurs : sans régulation, il y aurait plus de logements vendus, à un prix plus faible. Nous voyons que les propriétaires de logements déjà existants tirent parti de la mise en place d'un quota car leur bien voit sa valeur augmenter : ils bénéficient d'une rente de situation. La mise en place de quotas est donc problématique pour les nouveaux entrants sur le marché : ceux qui sont déjà propriétaires d'un bien ont tout intérêt à ce qu'un quota soit instauré, alors que ceux qui ne sont pas encore propriétaires sont pénalisés. Cette politique peut néanmoins se justifier au sein d'une collectivité si les habitants dans leur ensemble jugent que leur gain en bien-être lié à la préservation de l'esprit de village de leur commune (ce dont tout le monde bénéficie) est supérieur à la perte de bien-être pour les non propriétaires lié à l'augmentation du prix des logements.

Il est aussi possible d'envisager la mise en place d'un **quota minimal**. Dans le domaine du logement, il s'agit en général d'imposer l'existence d'un

certain nombre de logements sociaux sur un territoire donné. La loi SRU appliquée depuis 2000 impose ainsi l'existence d'au moins 20 % de logements sociaux sur certains territoires sous peine du paiement d'une amende. Si le quota minimal est supérieur à la quantité d'équilibre, il faut que l'État use de son pouvoir de contrainte pour forcer les individus à offrir des logements à la location. Certains évoquent dans cette optique l'idée d'une réquisition des logements vacants. Une telle possibilité existe dans la loi mais n'est pas appliquée. Cette solution serait en principe sans véritable effet car les propriétaires vendraient leur bien à des primo-acquéreurs souhaitant s'y installer devant la perspective d'une réquisition. L'État pourrait forcer les individus à louer leurs biens et ce à un tarif peu onéreux mais cette solution est difficilement envisageable en démocratie et serait peu efficace car les individus n'investiraient tout simplement plus dans le logement locatif. D'où le fait que les logements sociaux sont en général gérés directement par la puissance publique, pour réduire les tensions sur le marché.

• La production publique

Considérons à présent le cas d'une production publique de logements. Supposons que l'État mette en location les logements qu'il a produits. Quels vont être les effets d'une telle politique sur le prix des loyers ? Si l'État propose ses biens à des loyers décidés administrativement, cette offre supplémentaire ne peut être assimilée à une hausse de l'offre locative privée. L'offre publique va simplement absorber une partie de la demande de logements privés qui sera ainsi plus faible et va conduire à une baisse des prix du marché. Si l'État met en location ces biens au prix du marché – ce n'est en général pas le cas – cette politique se traduit par une hausse de l'offre totale de logements locatifs sur le marché et ainsi par une baisse des loyers. Plus la production publique est importante, plus l'effet est fort.

Jusqu'à présent, nous avons raisonné à court terme pour une forme de courbe donnée. Il faut tenir compte du fait que, sur le long terme, la forme des courbes peut changer selon les politiques menées. Prenons le cas évoqué ci-dessus. Imaginons que l'État décide de mener une politique active de construction de logements et que le prix des loyers baisse durablement. Il devient beaucoup moins avantageux de louer des biens et sur le long terme cette politique va déprimer l'offre privée. À court terme en revanche, l'offre privée de logement est peu élastique : un propriétaire ne peut que

difficilement le revendre rapidement ou l'utiliser autrement et il va accepter de louer son logement bien que les loyers soient moins hauts.

3.3 La régulation par les normes

Pour changer les comportements, l'État peut également modifier la **réglementation** qui encadre le fonctionnement du marché, ou encadrer les caractéristiques des biens qui sont échangés. Les marchés sont des institutions qui n'existent pas sans les règles qui les fondent. En imposant par exemple des normes sur le type de logement qu'on a le droit de louer (taille minimum, pas de peinture au plomb ni d'amiante, installation électrique aux normes, etc.), l'État exclut du marché un certain nombre de biens, notamment les moins chers, et va améliorer la qualité globale des logements mis à la location sur le marché. Cette hausse de la qualité moyenne des logements peut être néanmoins problématique si elle contribue à la hausse du prix moyen des logements et exclut ainsi du marché du logement un certain nombre de ménages qui n'ont pas les moyens de s'offrir des logements décents. Une solution pour pallier ce problème serait de subventionner les locataires à bas revenus pour qu'ils soient capables de payer des loyers plus élevés liés à des logements rénovés, ou que l'État construise lui-même des logements en bon état et qu'il accepte de les louer à des ménages à bas revenus pour un loyer modeste, en dessous du prix du marché. On voit qu'il est difficile de trouver le bon compromis entre l'amélioration de la qualité des logements grâce à la réglementation du marché, et son impact négatif sur le nombre de logements proposés, du fait des contraintes imposées aux propriétaires.

L'État peut aussi, par le recours à la règle, contraindre les ménages à intégrer les effets externes de leur comportement. En obligeant à ravalier les façades ou en imposant des normes environnementales sur les logements, l'État contribue ainsi à accroître le bien-être de tous les habitants. Considérons plus en détail l'obligation de l'entretien extérieur des logements : ce type d'entretien est coûteux et les propriétaires de logements sont souvent peu enclins à les assumer car la satisfaction qu'ils retirent d'une belle façade est inférieure au coût lié à son ravalement, surtout si le propriétaire loue le logement. Le rendement social du ravalement des façades est en revanche plus grand : si toutes les façades sont propres dans une ville, il sera beaucoup plus agréable de s'y promener, la ville attirera sans doute

plus de touristes, etc. On est dans un cas typique d'externalités où les rendements privés et sociaux divergent (voir le [chapitre 5](#)). Le marché est dans ce cas défaillant et il peut être intéressant pour la puissance publique d'imposer l'entretien des façades. Le cas des défaillances de marché sera détaillé ultérieurement mais on peut déjà noter que l'État dispose de différents instruments pour intervenir, dont le recours à la règle.

Considérons les effets d'une règle qui imposerait à tous les propriétaires d'installer des fenêtres à double vitrage.

Focalisons-nous sur le marché de la location. Une telle mesure revient à alourdir les coûts d'entretien du bien loué et diminue la rentabilité de la location. Cette mesure va donc conduire à diminuer l'offre de logement. Dans ces conditions, le prix des loyers augmente et moins de logements sont loués. En revanche, ils sont de meilleure qualité qu'avant. Pour éviter que l'amélioration de la qualité des logements ne se fasse au détriment des locataires, qui supportent au final en partie le coût des améliorations à travers la hausse de loyer, l'État peut par ailleurs subventionner la pose de double vitrage afin que le coût de la mise aux normes soit plus faible. Le financement de la subvention par l'impôt, est un moyen de faire supporter une partie du poids de la rénovation des logements par l'ensemble de la collectivité, qui bénéficie en retour de la meilleure isolation énergétique des logements et donc de la moindre consommation d'énergie et de la baisse de la pollution. Dans les pays développés, la puissance publique a très largement recours à ce type d'aide à la rénovation.

3.4 Comment évaluer les effets d'une intervention publique sur un marché ?

Pour choisir entre les différents types d'intervention publique sur un marché concurrentiel, il est indispensable d'évaluer précisément les effets de chaque mesure. Comment savoir, sans cela, quand il est préférable de mettre en place des normes contraignantes ou bien un système de taxes et de subventions pour réduire la consommation énergétique des logements, à titre d'exemple ?

Différentes méthodes sont généralement utilisées pour apprécier les effets d'une politique publique :

- L'analyse coûts-avantages compare les coûts liés à la mise en œuvre d'une politique aux gains engendrés par cette dernière. On va ainsi

comparer par exemple le coût d'une politique d'incitation fiscale à la construction de logement neuf aux bénéfices attendus d'une telle politique (nombre et qualité des logements produits, impact sur le prix des logements etc.). Le recours à l'économétrie permet de mesurer avec précision les modifications engendrées par la mise en œuvre de politiques publiques. Il est ainsi possible grâce à un tel outil d'évaluer par des méthodes expérimentales le nombre de logements neufs qui ont été construits grâce à la mise en œuvre d'une loi comme le dispositif Scellier. Ce calcul suppose un raisonnement « toutes choses étant égales par ailleurs » puisqu'il s'agit d'isoler les effets propres d'une politique. On distingue donc dans l'ensemble des logements neufs ceux qui ont été construits du fait de la mise en œuvre du dispositif de ceux qui auraient été construits de toute façon, même sans que le dispositif soit mis en place.

- D'autres approches raisonnent en termes de bien-être en s'appuyant sur l'utilité ressentie par les agents (voir le [chapitre 2](#) pour la définition du concept). Par exemple, l'analyse en termes de surplus, présentée ci-dessus, se base sur l'évaluation des gains et des pertes de surplus engendrés par l'action de l'État. Il s'agit d'une analyse en équilibre partiel qui ne tient pas compte des répercussions potentielles sur l'équilibre d'autres marchés. La mesure de la variation du surplus se fait en comparaison à une situation de concurrence pure et parfaite. Il est possible de complexifier cette approche en introduisant des fonctions de bien-être social qui pondèrent de façon particulière l'utilité de chaque agent. On va alors comparer la valeur du bien-être global avant et après l'intervention de l'État (voir le [chapitre 7](#)). Cette approche ne se contente pas de décrire d'un point de vue économique les conséquences de l'intervention de l'État mais propose une perspective normative afin de classer les différentes situations d'équilibre sur un marché donné, et de déterminer la politique optimale à mener pour le gouvernement.

Il convient de ne pas opposer ces méthodes : elles sont largement complémentaires. Par exemple pour mener une analyse en termes de coûts-avantages, il est utile d'avoir recours à l'analyse économétrique pour mesurer les effets d'une politique.

Notes

1. En comptabilité nationale, la consommation finale effective des ménages regroupe la consommation réalisée directement par les ménages et les dépenses de biens et services fournis par les administrations publiques aux ménages.

2. Pour approfondir la discussion sur la démarche modélisatrice en microéconomie, on pourra consulter B. Walliser, *Comment raisonnent les économistes ?* « Les fonctions des modèles », Odile Jacob, 2011 ou bien la première partie du manuel d'E. Wasmer, *Principes de microéconomie – Méthodes empiriques et théories modernes*, Pearson, 2014, intitulée « L'économie : une discipline construite sur les faits et leur analyse ».

3. C. Rougerie et J. Friggit, *INSEE Première*, n° 1297, mai 2010.

4. Source : CGEDD d'après INSEE, <http://www.cgedd.fr/prix-immobilier-presentation.pdf>

5. A. Marshall (1842-1924) est un des fondateurs de l'approche néo-classique. Professeur à l'université anglaise de Cambridge, il développa l'approche en équilibre partiel des marchés. Il est considéré comme un des fondateurs de l'économie industrielle.

6. L. Walras (1834-1910) est un économiste français, professeur à l'université de Lausanne, qui fut lui aussi un fondateur de l'approche néo-classique, et en particulier de la théorie de l'équilibre général.

7. J. Friggit, « Quelles perspectives pour le prix des logements après son envolée ? », in *Regards Croisés sur l'Économie*, « Pour sortir de la crise du logement », n° 9, mai 2011.

Chapitre 2

La formation de la demande des ménages

L'exemple des dépenses culturelles des ménages

1. LE MODÈLE DE BASE DU CONSOMMATEUR RATIONNEL

2. PROLONGEMENTS ET DÉBATS AUTOUR DU MODÈLE DU CONSOMMATEUR RATIONNEL

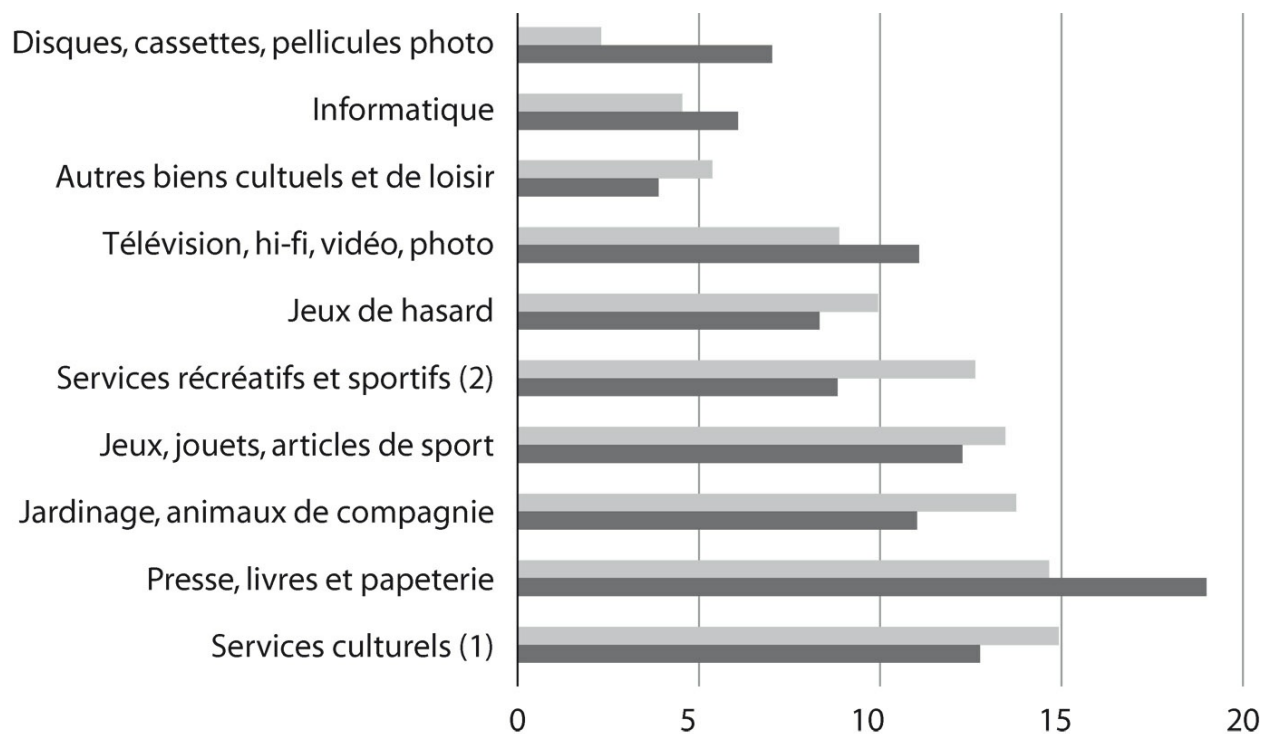
3. DE LA DEMANDE INDIVIDUELLE À LA DEMANDE AGRÉGÉE

Au cours des dernières décennies, la consommation des ménages a fortement augmenté en volume, suivant tendanciellement la croissance du revenu réel par habitant. Mais cette augmentation cache des disparités importantes selon la nature des dépenses de consommation : si la part de l'alimentation dans le revenu diminue en moyenne à long terme, celle du logement a tendance au contraire à augmenter (voir le [chapitre 1](#)). Les biens culturels sont associés à des activités socialement valorisées comme aller au spectacle, lire un livre, écouter de la musique, etc. Quand le revenu des ménages augmente, ils peuvent diversifier leur consommation, et ont tendance à consacrer aux biens culturels une plus grande part de leurs revenus.

La dépense consacrée à la culture, aux loisirs et aux communications représente 8,8 % de la consommation effective dans l'UE en 2014. Outre son montant, la structure de cette dépense a évolué dans le temps. Nous pouvons voir dans le **document 2.1** qu'il existe de fortes variations d'un poste de dépenses culturelles à l'autre : on observe par exemple la baisse rapide de la part des dépenses pour des disques, cassettes, pellicules photos et à l'inverse l'augmentation de la part des dépenses pour des services culturels au sein des dépenses culturelles et de loisirs.

Comment expliquer ces évolutions des dépenses culturelles ? Certains produits technologiques ont vu leur prix diminuer fortement, d'autres ont subi la concurrence de nouveaux biens au sein d'un budget des ménages limité.

**Document 2.1 : Part au sein des dépenses culturelles et de loisirs
en 2000 (gris foncé) et 2016 (gris clair)**



(1) : cinéma, spectacles vivants, musées, abonnements audiovisuels (y c. redevance TV), développements de tirage de photos, etc.

(2) : sport, location de matériel sportif, fêtes foraines, parcs d'attractions, voyages à forfait, week-ends, etc.

Champ : France.

Source : Insee, comptes nationaux – base 2010.

Les modifications des choix des consommateurs résultent-elles de la variation des contraintes économiques pesant sur eux ou d'une évolution de leurs goûts ? Pour répondre à ces questions, il faut se doter d'outils capables de représenter les principaux déterminants de la demande des ménages. Les microéconomistes se focalisent sur un petit nombre de variables, principalement le revenu des ménages, leurs goûts pour les différents biens et les prix relatifs des biens, c'est-à-dire les rapports entre les prix des différents biens. Dans quelle mesure le modèle microéconomique de base du consommateur peut-il être utile pour comprendre les dépenses culturelles des ménages ?

Ces dernières peuvent donner le sentiment, en première analyse, de résister à l'investigation microéconomique : elles seraient trop complexes pour se plier à des modèles assez frustes et seraient davantage déterminées par des phénomènes socio-historiques et non strictement économiques. Pourtant, on

assiste à l'émergence d'un champ dynamique de recherche, suite au travail précurseur de William Baumol en 1966¹, qui a donné le « *coup de gong de l'économie de la culture* » selon l'expression de X. Greffe (2010). L'objectif de ce chapitre est de montrer d'une part que le modèle de base de la théorie du consommateur est déjà très utile pour comprendre les caractéristiques générales des comportements de consommation, y compris dans le cas spécifique des biens culturels, et d'autre part que la consommation de ces biens a des spécificités qui ont été l'objet de débats et de confrontations avec des modèles alternatifs, ce qui rend visible au final le chemin parcouru par la théorie du consommateur.

• Encadré 2.1 : Comment mesurer les dépenses des ménages ?

L'étude du budget des ménages est un objet de prédilection des enquêtes quantitatives, depuis les travaux précurseurs d'A. Quetelet (1796-1874) et E. Engel (1821-1896), qui ont initié les méthodes de la statistique publique dès le milieu du XIX^e siècle.

Différentes enquêtes de l'INSEE permettent d'aborder le budget des ménages ; la principale est l'enquête « budgets de famille ». Elle consiste à étudier les dépenses d'environ 10 000 ménages, tous les cinq ans, à partir d'une nomenclature de près de 900 postes budgétaires, chacun regroupant des dépenses de même nature. On calcule alors le coefficient budgétaire de chaque poste, c'est-à-dire sa part dans le budget des ménages. Cela permet d'étudier l'incidence des variations de prix, de revenus ou des variables sociodémographiques (âge, sexe, profession, etc.) sur la consommation des ménages.

Parmi les dépenses des ménages, celles consacrées aux biens culturels font l'objet depuis 1979 d'une analyse spécifique par la DEPS, le service d'études statistiques du Ministère de la culture. Les données les plus complètes sont issues des deux dernières enquêtes « budget de famille », en 2006 et 2011. Une nouvelle enquête a eu lieu entre septembre 2016 et septembre 2017, dont les données ne sont pas encore parues. L'évolution rapide des comportements dans le domaine de la culture et des loisirs, avec l'émergence des produits numériques, oblige à revoir régulièrement la typologie des biens culturels. C'est la raison pour laquelle on regroupe les dépenses cultures-médias dans un même ensemble : acheter un journal numérique ou un journal papier, consommer des films ou de l'information sur internet, se déplacer au musée, au cinéma ou acheter un DVD, sont des pratiques largement substituables. Elles relèvent donc toutes de l'étude des biens culturels et sont interdépendantes selon l'évolution de leurs prix relatifs par exemple, même si ces pratiques ne sont pas perçues de la même manière par les acteurs.

1. Le modèle de base du consommateur rationnel

1.1 Contrainte budgétaire et prix relatif des biens

Les ressources des ménages étant limitées, ils doivent arbitrer au mieux entre les différents postes de dépenses de consommation. La démarche microéconomique s'appuie sur un modèle de base très simple, dont le but est de distinguer précisément l'effet des différentes variables agissant sur les choix des consommateurs.

L'unité de base est un **consommateur** qui prend des décisions d'achat. Il « arbitre » entre différents usages de son revenu : il peut s'agir d'un individu ou éventuellement d'un ménage qui prend des décisions communes, en mettant de côté dans cet ouvrage les négociations entre les membres du ménage.

Le consommateur a des moyens limités par sa **contrainte budgétaire**, c'est-à-dire par le budget dont il dispose pour consommer une certaine quantité de biens, compte tenu des prix pratiqués. La notion de contrainte budgétaire renvoie au fait que l'agent arbitre dans les limites d'un budget. La somme dépensée pour sa consommation ne peut être supérieure au budget dont il dispose. La contrainte budgétaire signifie ainsi concrètement que la somme des quantités achetées de chaque bien multipliées par leur prix doit être inférieure ou égale au budget du consommateur. Le budget du consommateur sur une période donnée dépend de son revenu sur cette même période mais ne se confond pas avec lui, car il peut décider d'en épargner une partie en vue d'une consommation future ou emprunter pour pouvoir consommer davantage que son revenu de la période. Nous étudierons plus loin dans ce chapitre la façon dont le consommateur fixe son budget consommation pour chaque période (par exemple son budget mensuel) et considérons simplement pour le moment que le consommateur dispose à chaque période d'un budget destiné à la consommation. Les travaux empiriques de l'INSEE montrent que le budget sur lequel les ménages ont réellement la possibilité de faire des choix est plus réduit que leur revenu disponible à court terme : elle a introduit dans ce but en 2008 la distinction entre **revenu arbitral** et revenu non-arbitral, qui renvoie au fait qu'une

part du revenu disponible des ménages n'est pas l'objet d'un véritable choix car il est destiné à des dépenses « pré-engagées » difficilement renégociables à court terme (contrats d'abonnement liés au logement ou à la communication, loyers, contrats d'assurance et services financiers, etc.). Cette distinction est importante pour comprendre pourquoi certains ménages modestes ont en fait très peu de marges de manœuvre dans leurs arbitrages, compte tenu du niveau élevé de dépenses pré-engagées, tandis que les ménages plus aisés ont un revenu arbitrageable plus important qui permet d'ouvrir un espace de choix plus large, dont bénéficient les dépenses culturelles (MARESCA *et alii*, 2011).

Les **prix** des biens déterminent quant à eux les quantités de biens qu'un consommateur est capable d'acquérir, compte tenu de son budget, c'est-à-dire son **pouvoir d'achat**. Les choix de consommation dépendent des prix relatifs des biens. Par exemple, la décision d'aller au théâtre ne dépend pas uniquement du prix des places de théâtre, mais également des prix des autres biens. En effet, le consommateur tient compte du fait qu'en achetant une place de théâtre, il renonce dans le même temps à d'autres possibilités de consommation, son budget étant limité. Plus le prix relatif d'une pièce de théâtre est important, moins le consommateur pourra acheter d'autres biens ou services.

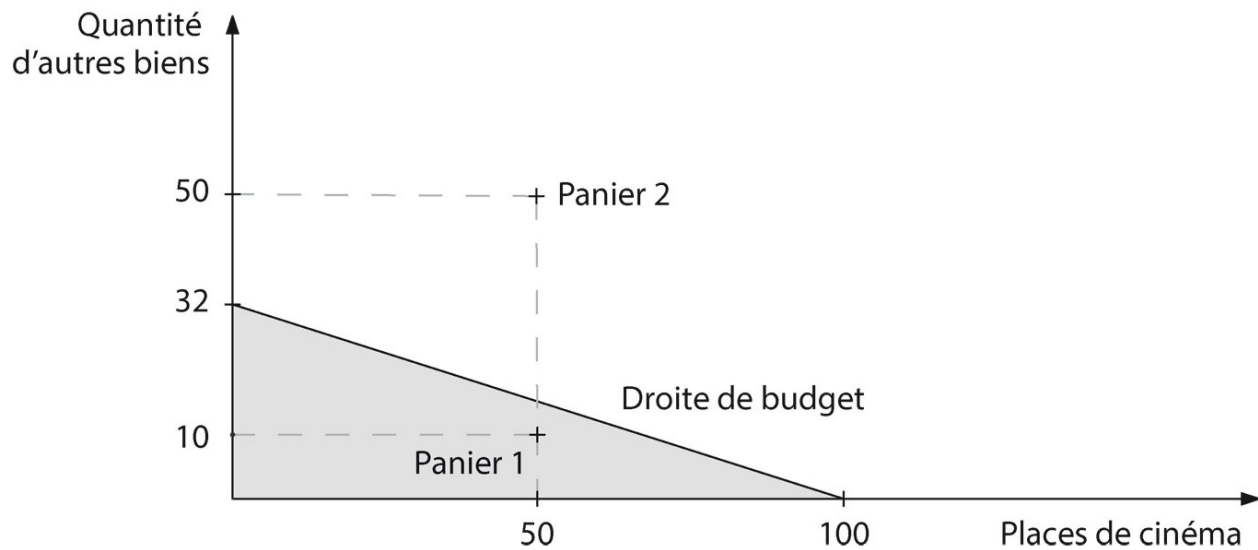
Chaque bien dispose de caractéristiques particulières (un film n'est jamais strictement équivalent à un autre, *idem* pour un jean, etc.) et le consommateur arbitre entre toutes ces caractéristiques, qu'il est supposé connaître. Pour simplifier le raisonnement, nous considérons dans le modèle de base de grandes catégories de biens comme par exemple les DVD, sans tenir compte des différences pouvant exister entre les biens qui composent cette catégorie, tous les DVD n'étant pas équivalents. En fait, le raisonnement présenté permet d'analyser d'une façon générale les choix du consommateur entre différentes possibilités d'achat, qu'elles se situent à un niveau très global (deux catégories de biens comme les DVD et le téléchargement légal de films) ou à l'échelle de deux produits bien précis (le choix entre le dernier prix Goncourt et le dernier prix Médicis).

La contrainte budgétaire qui pèse sur le consommateur peut se représenter graphiquement si l'on considère que le consommateur arbitre entre deux biens, les quantités consommées de chaque bien se trouvant sur un des axes. Considérons par exemple que le consommateur dispose d'un budget

(notée B) de 800 € annuels pour ses activités culturelles, et qu'il doit choisir entre les sorties au cinéma (leur quantité est notée C, et leur prix moyen p_C), et les autres biens culturels qu'il peut consommer (leur quantité est notée A, et leur prix moyen p_A). On pourrait tout aussi bien considérer qu'il arbitre entre le cinéma et le théâtre ou tout autre bien, cela ne modifie pas le raisonnement. Le modèle s'applique aussi à l'analyse des choix entre un grand nombre de biens mais dans ce cas la représentation graphique n'est simplement plus possible et c'est pourquoi nous considérons à chaque fois le cas de l'arbitrage entre deux biens, sans que cela n'ait d'incidence sur la validité de la démarche adoptée.

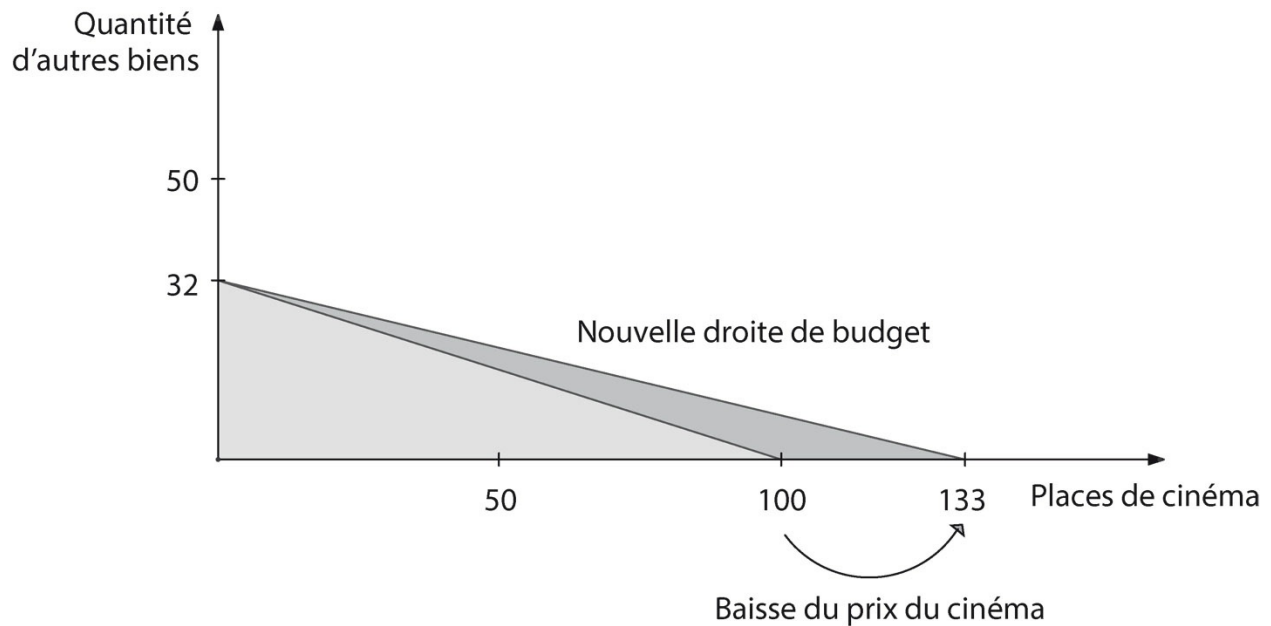
Si le prix des places de cinéma est de 8 € en moyenne, le consommateur peut au maximum acheter une quantité de 100 places de cinéma dans l'année ($800/8$). Si le prix moyen des autres biens culturels est de 25 €, alors il peut en consommer au maximum 32 ($800/25$). Le [graphique 2.1](#) présente sur les deux axes les quantités maximales qu'il peut consommer, et la droite qui relie les deux points s'appelle la **droite de budget** : la zone grisée en dessous de la droite correspond à toutes les combinaisons entre le cinéma et les autres biens qu'il peut s'offrir. On appelle **panier de biens** une combinaison entre une certaine quantité de chacun des biens consommés. Tous les paniers de biens se situant dans la zone grisée (ou sur la droite de budget) respectent la contrainte budgétaire du consommateur. Par exemple, le consommateur peut s'offrir le « panier 1 » (50, 10) composé de 50 sorties au cinéma dans l'année et 10 autres biens culturels, mais pas le « panier 2 » (50, 50).

Graphique 2.1 : La contrainte budgétaire



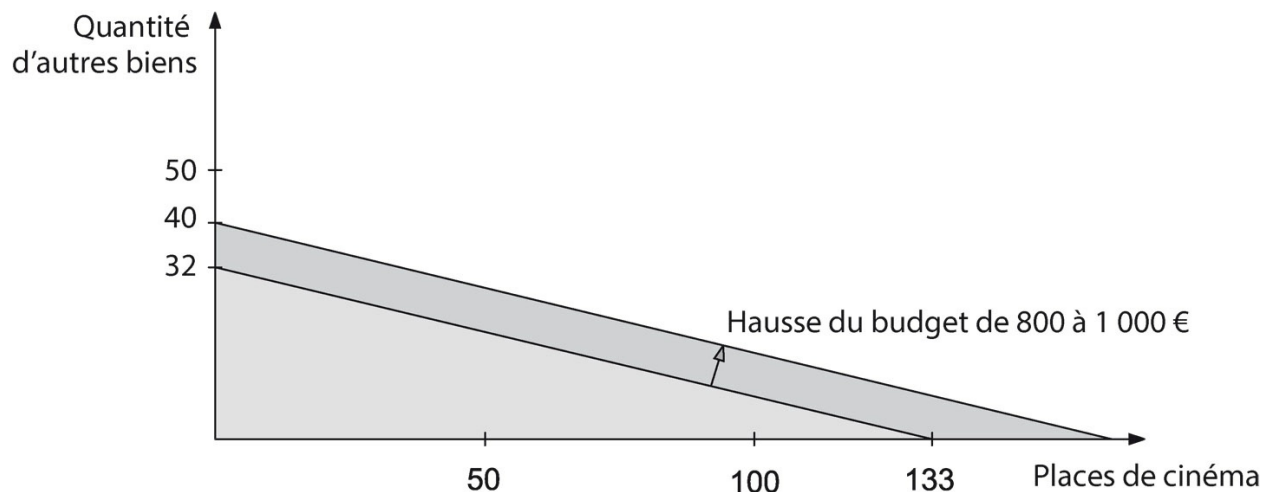
Grâce à la droite de budget, on voit très facilement ce qu'il advient lorsque le prix du cinéma diminue : le consommateur dispose d'un pouvoir d'achat plus important, comme le montre le [graphique 2.2](#). Dans l'hypothèse où le prix du cinéma passe à 6 €, la quantité maximale consommée est alors de 133 (soit $800/6$), et on voit que ses possibilités d'achat augmentent. C'est l'évolution inverse si le prix du cinéma diminue : le consommateur perd alors en pouvoir d'achat et les paniers de biens accessibles sont moins nombreux. On observerait le même type d'évolution si le prix des autres biens changeait. La pente de la droite de budget dépend ainsi des prix relatifs entre nos deux types de biens.

Graphique 2.2 : Effet d'une baisse du prix du cinéma



Supposons que le budget « culture » du consommateur augmente et s'établisse désormais à 1 000 €, le prix d'une place de cinéma se maintenant à 6 euros et le prix moyen des autres biens à 25 euros. Différents facteurs peuvent en être à l'origine : une hausse de ses revenus, ou bien le fait qu'il ait décidé d'accroître la part de son budget annuel consacré aux biens culturels. Dans ce cas, les quantités maximales de sorties au cinéma et des autres biens qu'il peut consommer sont respectivement de 166 (soit $1\,000/6$) et de 40 ($1\,000/25$).

Graphique 2.3 : Effet d'une hausse du budget

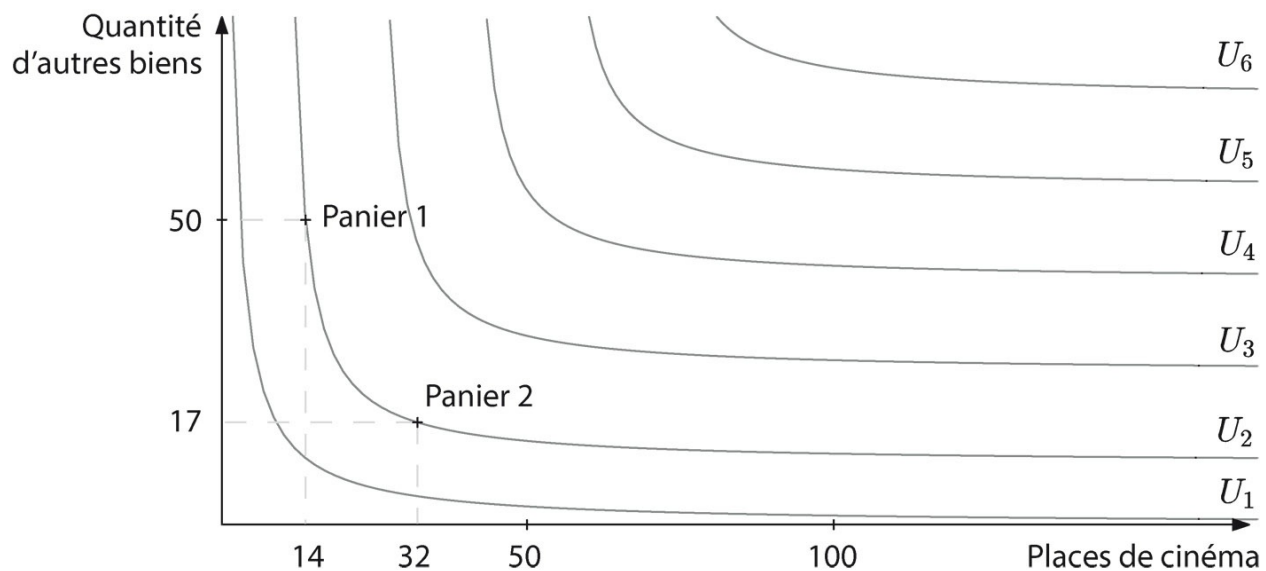


Le [graphique 2.3](#) montre que la droite de budget va se déplacer parallèlement vers la droite, car le consommateur peut s'offrir à la fois davantage d'un bien et de l'autre, alors que la pente de la droite de budget, qui dépend des prix relatifs, n'est pas modifiée. À l'inverse, une diminution des ressources déplace la droite de budget vers la gauche.

1.2 Les préférences

Les enquêtes empiriques soulignent que la consommation de biens culturels est caractérisée par des inégalités importantes, selon les caractéristiques sociodémographiques des consommateurs. Ces inégalités ne se réduisent pas uniquement à des écarts de revenu ou à l'effet de la structure des prix. Même quand ils sont gratuits, les biens culturels voient leur consommation varier fortement d'une catégorie à l'autre, en fonction des préférences des consommateurs. Selon une enquête du Credoc consacrée aux jeunes fréquentant les « nocturnes » gratuites de grands musées parisiens, seulement 6 % du public est employé ou ouvrier (BENHAMOU, THESMAR, 2011) alors qu'ils représentent environ la moitié de la population active. Sociologues et économistes portent des regards différents sur cette situation. Là où le sociologue cherche à comprendre comment les goûts de chacun sont socialement déterminés notamment par le milieu social, le microéconomiste considère ces goûts comme donnés et étudie la façon dont ils expliquent les choix de consommation. L'économiste évalue par exemple si les différences de dépenses culturelles entre jeunes consommateurs et personnes âgées viennent des écarts de revenus ou bien de différences de « goûts » entre les deux groupes. Une connaissance précise des déterminants de la consommation est essentielle pour mener des politiques culturelles efficaces : on ne lutte pas contre les inégalités d'accès à la culture de la même manière si elles sont dues à des écarts de revenus, à l'effet du prix de certains biens ou à des préférences différentes d'un groupe à l'autre. Ainsi, la probabilité de se rendre dans un musée ou de fréquenter un spectacle vivant dépend fortement de l'éducation d'après les enquêtes, car la poursuite d'études longues développe le goût pour ce type d'activités. Les consommateurs ayant des goûts différents, une baisse de prix ne suffira pas à démocratiser l'accès aux théâtres ou aux musées.

Graphique 2.4 : Les préférences du consommateur



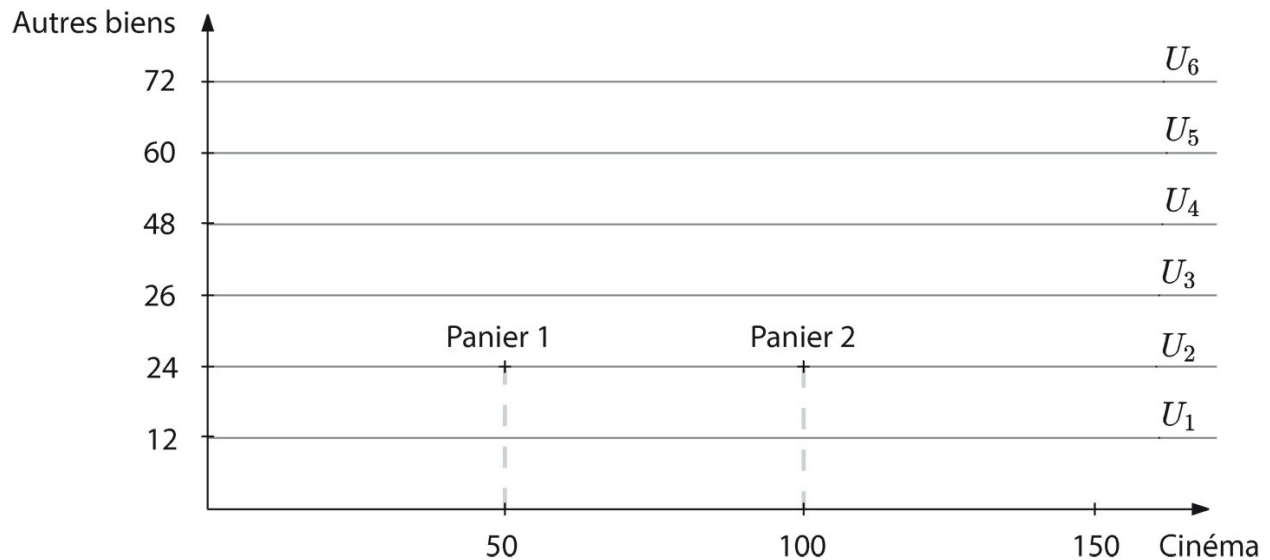
Le concept de **préférence** permet de représenter les goûts des consommateurs. Par hypothèse, les individus sont capables de hiérarchiser les différentes possibilités de consommation qui s'offrent à eux, de dire s'ils préfèrent une combinaison à une autre, et d'identifier celles qui lui semblent équivalentes. Pour reprendre l'exemple précédent du choix entre les sorties au cinéma et les autres dépenses culturelles, un individu doit ainsi être capable de dire s'il préfère aller beaucoup au cinéma et limiter ses autres dépenses culturelles ou l'inverse. Le [graphique 2.4](#) propose une représentation graphique de ces préférences.

Une **courbe d'indifférence**, comme la courbe U_2 regroupe tous les paniers de biens apportant le même niveau de satisfaction au consommateur. L'agent est ainsi indifférent entre consommer le panier 1 (14, 50) ou le panier 2 (32, 17). Les paniers situés sur la courbe d'indifférence U_2 sont préférés à ceux de la courbe U_1 . Plus une courbe d'indifférence s'éloigne de l'origine, plus elle représente un niveau élevé de satisfaction. Les courbes d'indifférence ne doivent pas se couper si l'individu fait des **choix cohérents**, c'est-à-dire qu'un même panier ne peut être à la fois sur une courbe, avec un niveau de satisfaction donnée, et sur une autre avec un autre niveau de satisfaction : un consommateur rationnel qui préfère le panier A au panier B et le panier B au panier C, préfère alors forcément le panier A au panier C. On dit que ses préférences sont **transitives**.

La forme des courbes est un instrument très pratique pour caractériser le goût d'un individu pour les biens considérés. Ainsi, le consommateur dont les

préférences sont représentées sur le [graphique 2.5](#) n'éprouve aucun intérêt pour le cinéma, puisque le fait de consommer 50 ou 100 sorties annuelles au cinéma (panier 1 ou 2) ne change rien à sa satisfaction, du moment qu'il a toujours accès à 24 « autres biens culturels » : il se situe toujours sur la courbe U_2 en étant indifférent entre les deux paniers.

Graphique 2.5 : Une absence de goût pour le cinéma



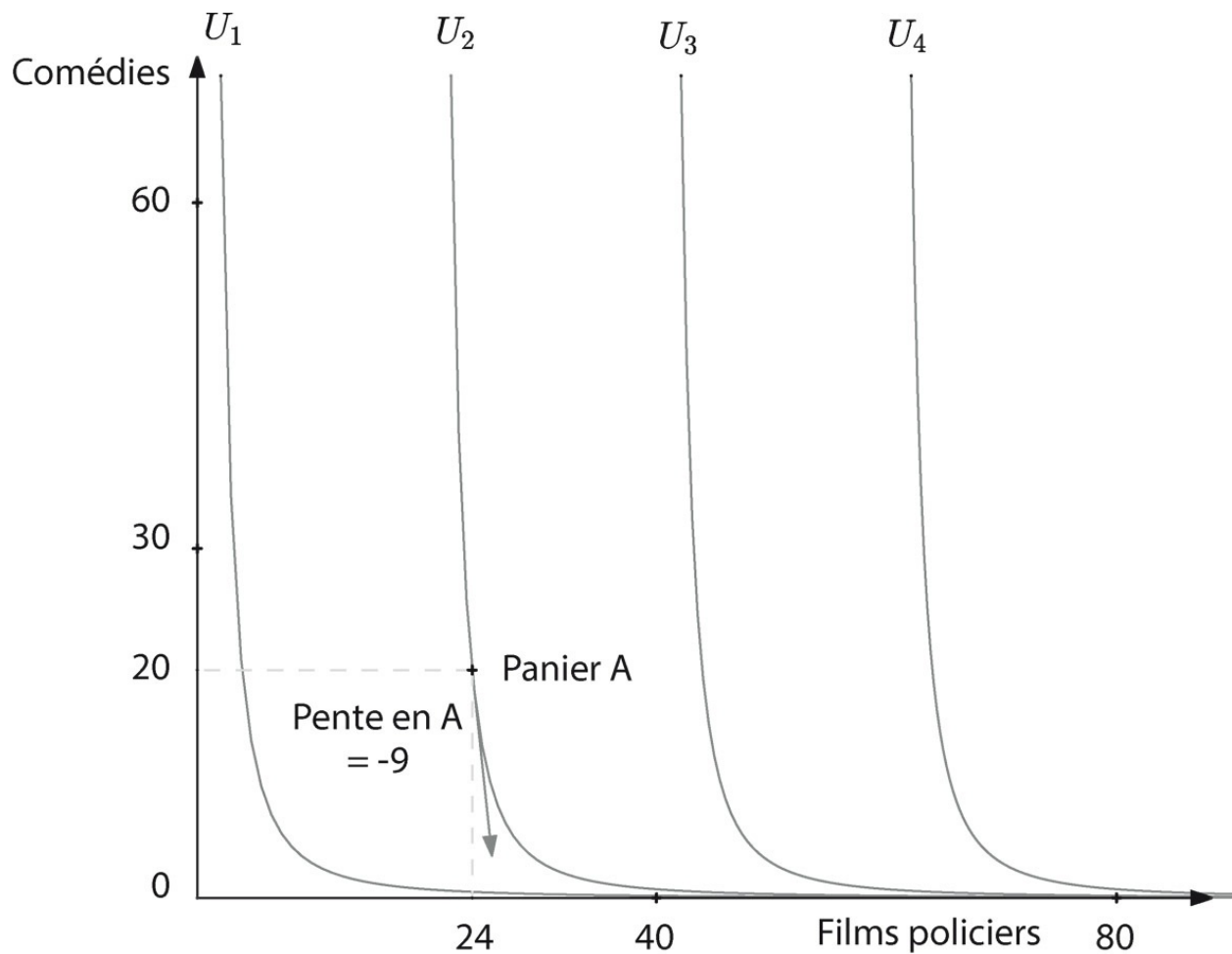
Des courbes d'indifférence horizontales ou verticales traduisent donc une absence complète de goût pour l'un des deux biens en jeu.

Dans le [graphique 2.4](#), les courbes d'indifférences sont strictement convexes : c'est le cas lorsque le consommateur « préfère les mélanges ». Le modèle de base postule qu'il a une préférence pour la diversité : considérons deux paniers de bien situés sur la même courbe d'indifférence, alors le consommateur préfère des paniers diversifiés s'il obtient toujours plus de satisfaction avec un panier de biens situé sur la droite reliant les deux paniers, donc composé d'un mélange des deux. Cette hypothèse n'est pas vérifiée dans le cas du consommateur dont les préférences sont représentées sur le [graphique 2.5](#).

Le modèle de base de la théorie microéconomique du consommateur suppose que ce dernier est insatiable : une quantité supplémentaire d'un bien, toutes choses étant égales par ailleurs, accroît toujours la satisfaction du consommateur. Il ne se lasse ainsi jamais complètement d'un bien, même si

l'utilité marginale décroît au fur et à mesure que la quantité consommée augmente.

Graphique 2.6 : Le taux marginal de substitution



La pente des courbes d'indifférence nous renseigne sur la manière dont le consommateur considère les deux biens, et est prêt à les substituer l'un à l'autre. Un amateur de cinéma peut par exemple apprécier fortement les films policiers, et accorder peu d'intérêt aux comédies. Partant d'un même panier de biens, une augmentation de 10 films policiers permet alors d'atteindre une courbe d'indifférence plus élevée qu'une augmentation de 10 comédies.

La très forte pente des courbes d'indifférence sur le [graphique 2.6](#) nous indique donc que l'individu accorde bien plus d'importance au bien en abscisse, les films policiers, qu'à l'autre bien. On appelle **taux marginal de substitution** le rapport entre la quantité de biens en ordonnée que l'agent est prêt à céder en échange d'une unité supplémentaire du bien en abscisse, sans perte de satisfaction, c'est-à-dire en restant toujours sur la même courbe d'indifférence. Le taux marginal de substitution varie en fait tout au long de la courbe d'indifférence et est égal à la pente de cette dernière. Si on se situe

au panier A, on voit que la pente de la courbe d'indifférence est de -9 , ce qui signifie que l'agent est prêt à céder 9 comédies contre un film policier supplémentaire, sans perdre de satisfaction : il accorde beaucoup plus d'importances aux films policiers qu'aux comédies. Si on se déplace sur la même courbe d'indifférence vers le bas, la pente (donc le taux marginal de substitution) devient plus faible, car l'individu est prêt à céder une moins grande quantité de comédies contre un film policier supplémentaire, dans la mesure où sa consommation de films policiers augmente et qu'il accorde moins d'importance à l'acquisition d'une unité supplémentaire de ce genre.

Les courbes d'indifférence ne s'expriment pas dans une unité particulière. Il suffit de faire l'hypothèse que les agents sont capables de « classer » les différents paniers de biens, sans associer à chacun un niveau de satisfaction ou un degré d'importance mesurable. Les indices associés sur le graphique à chaque courbe sont purement conventionnels ; c'est ce qu'on appelle une **conception ordinale** de l'utilité, à la différence d'une **conception cardinale** qui suppose que l'on peut mesurer l'utilité et lui attribuer une unité.

Les microéconomistes utilisent des **fonctions d'utilité** pour représenter sous forme mathématique la relation entre les quantités de biens consommées et la satisfaction obtenue par l'agent, autrement dit « l'utilité » qu'il leur accorde. La fonction d'utilité associe aux quantités de biens consommés une valeur chiffrée (par exemple une utilité égale à 5), mais cette valeur ne sert qu'à classer les différents paniers de biens : un panier de bien dont l'utilité est de 8 est préféré à un panier de bien dont l'utilité est de 4 mais cela ne veut pas dire qu'il lui apporte deux fois plus de satisfaction. La théorie néo-classique donne une grande place à la notion d'**utilité marginale**, qui désigne le supplément de satisfaction apporté par la consommation d'une unité additionnelle d'un bien. Par hypothèse, l'utilité marginale d'un bien est décroissante, c'est-à-dire que le supplément de satisfaction apporté par la consommation d'une unité supplémentaire d'un bien est de plus en plus faible à mesure que la quantité consommée augmente. La satisfaction globale augmente toujours avec la consommation d'une quantité supérieure de biens, mais le consommateur se lasse peu à peu du bien et le supplément de satisfaction par unité est de moins en moins élevé.

La notion de préférence reste très abstraite jusqu'à présent : on a simplement constitué des « types purs » comme le consommateur insatiable, ou le consommateur indifférent à un bien, mais comment connaître la forme

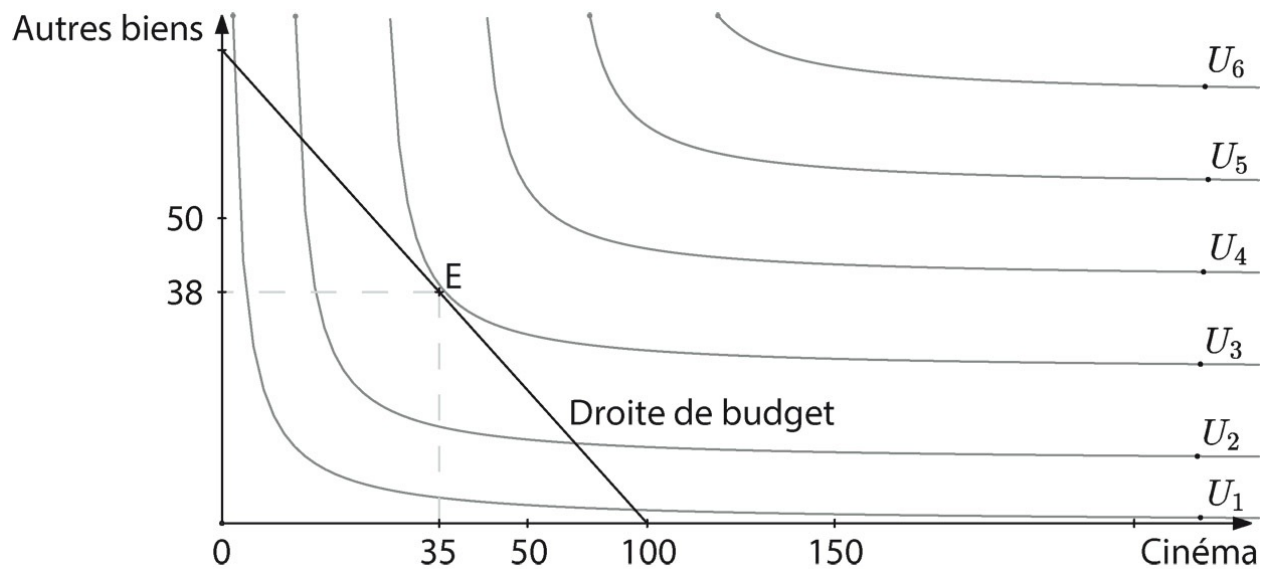
des courbes d'indifférence qui caractérisent en moyenne les amateurs de cinéma par exemple ? Une première solution développée par P. Samuelson part de l'hypothèse suivante : si la fonction d'utilité n'est pas directement observable, les choix des consommateurs nous révèlent en fait les préférences qu'ils ont exprimées. En connaissant la demande observée dans des contextes différents de prix ou de revenus, on peut estimer, grâce à la théorie des **préférences révélées**, la fonction d'utilité « révélée » par ces choix de consommation. Cette théorie a joué un rôle important en servant de base aux études empiriques sur la consommation, mais elle demeure contestée, car cette méthode suppose pour être valide, que la théorie du consommateur sur laquelle elle repose représente parfaitement les choix de consommation des individus, puisqu'on s'appuie sur la théorie pour retrouver les préférences des individus.

1.3 La maximisation de la satisfaction

Si le consommateur est rationnel, il cherche à tirer le meilleur parti de ses ressources pour augmenter son niveau de satisfaction. Le modèle de base repose sur l'idée que les agents cherchent à atteindre le niveau maximal d'utilité, compte tenu des prix pratiqués et de leur budget. La situation atteinte correspond alors à l'**équilibre du consommateur**. Il est possible de représenter graphiquement cet équilibre quand le consommateur doit choisir entre deux biens.

Nous voyons ainsi sur le [graphique 2.7](#), en faisant apparaître la droite de budget, que le consommateur maximise sa satisfaction au point E, car ce panier de bien lui permet d'atteindre le niveau de satisfaction le plus élevé compte tenu du prix des différents biens et de la somme dont il dispose. Remarquons qu'au point E la droite de budget a la même pente que la courbe d'indifférence U_3 : à l'équilibre, le point qui maximise la satisfaction du consommateur se situe nécessairement au point de tangence entre la droite de budget et la courbe U_3 pour être sur la courbe la plus élevée possible, celle qui touche en un seul point la droite de budget.

Graphique 2.7 : L'équilibre du consommateur



Par conséquent, la pente de la courbe d'indifférence au point E étant déterminée par le taux marginal de substitution, et la pente de la droite de budget par le rapport des prix, le taux marginal de substitution est donc égal, à l'équilibre, au rapport des prix des biens.

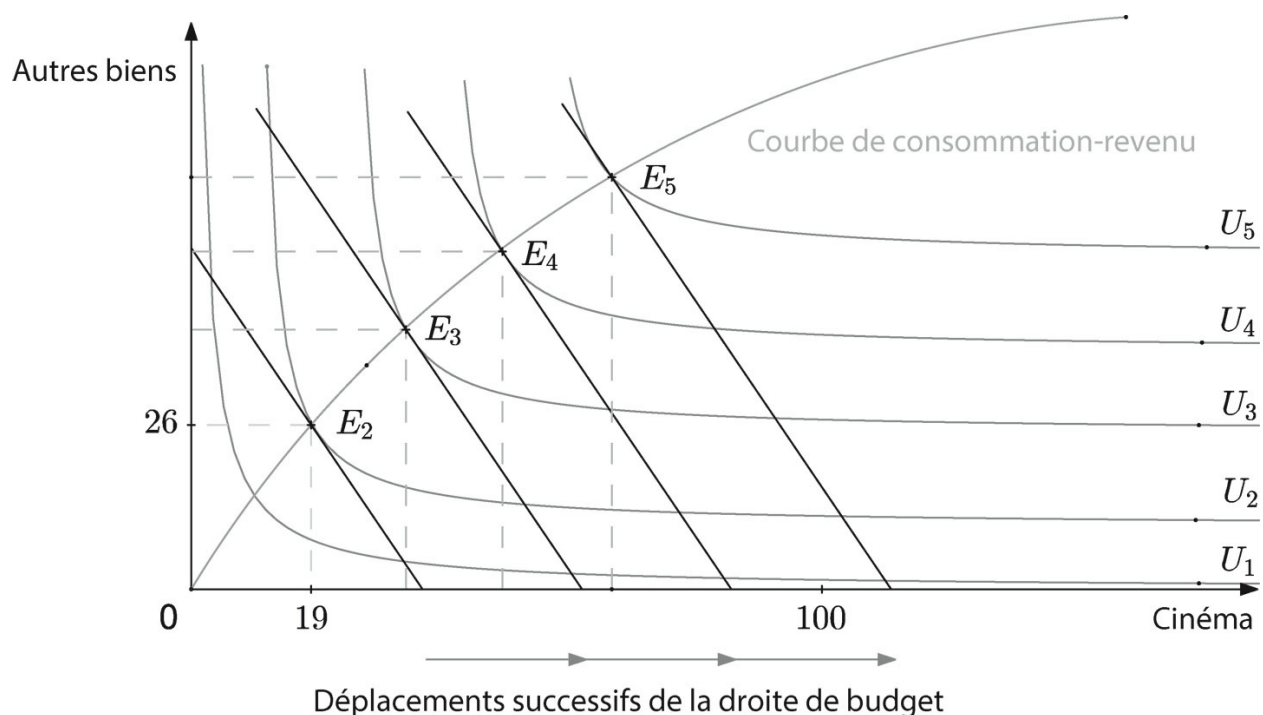
L'équilibre du consommateur peut être présenté sous une forme mathématique plus complexe si on représente les préférences par une fonction d'utilité dont on cherche la valeur maximale, sous contrainte du budget. Le recours à la formalisation est nécessaire pour déterminer l'équilibre du consommateur quand il existe de nombreux biens mais cela ne change rien à l'analyse. La **règle de consommation optimale** maximise la satisfaction d'un consommateur rationnel : à l'équilibre, les utilités marginales des différents biens pondérées par leurs prix sont égales entre elles. Si l'on reprend le [graphique 2.7](#) cela signifie qu'au point E l'utilité marginale de la 35^e sortie au cinéma divisée par son prix (8 € par exemple) est égale à l'utilité marginale du 23^e « autre bien culturel » (divisée par son prix moyen, soit ici 25 €). Pour le comprendre, il suffit de raisonner par l'absurde : si ce n'était pas le cas et que la 35^e sortie au cinéma rapportait plus de satisfaction supplémentaire par euro dépensé que les autres biens, alors il serait préférable de dépenser un peu plus en cinéma (et un peu moins dans les autres biens) pour en tirer parti, et donc le panier E ne serait pas le choix optimal puisqu'il existerait un panier de bien apportant plus de satisfaction au consommateur ; en conséquence cette loi est forcément vérifiée à l'équilibre. Un consommateur rationnel adapte ainsi ses choix aux

variations des prix et de l'utilité marginale des différents biens consommés. Par exemple, l'arrivée des téléviseurs à écran plat, au cours des années 2000, a réduit l'utilité marginale des téléviseurs cathodiques qui deviennent moins attractifs que les nouveaux modèles. Les producteurs n'ont pas eu d'autres choix que de réduire rapidement les prix relatifs des téléviseurs cathodiques pour s'adapter et continuer à en vendre, avant que la baisse des coûts de production des téléviseurs à écran plat ne chasse complètement du marché les anciens modèles.

1.4 Les effets d'une variation du budget

L'augmentation du budget modifie les choix du consommateur : il peut acheter davantage de biens, et le surcroît de pouvoir d'achat peut être réparti diversement entre les biens selon ses préférences. Dans le [graphique 2.8](#), l'individu arbitre entre les sorties au cinéma et les autres biens culturels et bénéficie de hausses successives de son budget, entraînant ainsi une translation vers la droite de sa droite de budget, la pente restant identique puisque'elle dépend des prix relatifs.

Graphique 2.8 : Courbe de consommation-revenu et hausse du budget



Dans ces conditions, le choix optimal passe du panier E_2 (qui correspondait à la courbe d'indifférence U_2), au panier E_3 , puis E_4 , puis enfin E_5 . Si on relie tous les paniers de biens optimaux on obtient la **courbe dite de consommation-revenu**, qui associe les différents paniers choisis par le consommateur aux variations de son revenu. Il serait plus juste de parler de courbe de consommation-budget, puisque c'est le budget consacré à la consommation qui augmente ici, pas forcément son revenu. La forme de la courbe de consommation-revenu nous renseigne sur les caractéristiques des biens : plus elle s'infléchit vers l'axe des abscisses, comme c'est le cas sur le graphique, plus cela signifie qu'il consacre une part croissante de son budget au cinéma. Pour le comprendre, raisonnons sur le cas limite où la courbe devient presque parallèle à l'axe des abscisses : cela signifie que le supplément de budget est presque exclusivement consacré au cinéma, les autres dépenses culturelles n'augmentent presque plus. Les biens dont la part dans le budget augmente quand le budget lui-même augmente sont appelés des **biens supérieurs** (voir la partie consacrée à la [typologie des biens](#) ci-dessous).

L'effet global de la variation du budget d'un bien sur la demande de ce bien peut s'apprécier par le calcul, au niveau individuel, de **l'élasticité-revenu de la demande** : c'est le rapport entre la variation relative de la demande d'un bien (par exemple +3 %) et la variation relative du budget de la personne (par exemple + 2 %), ce qui donne une élasticité-revenu de + 1,5 dans ce cas. Sur le [graphique 2.8](#), l'élasticité-revenu du cinéma est supérieure à un puisque la quantité demandée augmente plus vite que le revenu (voir le [chapitre 1](#)).

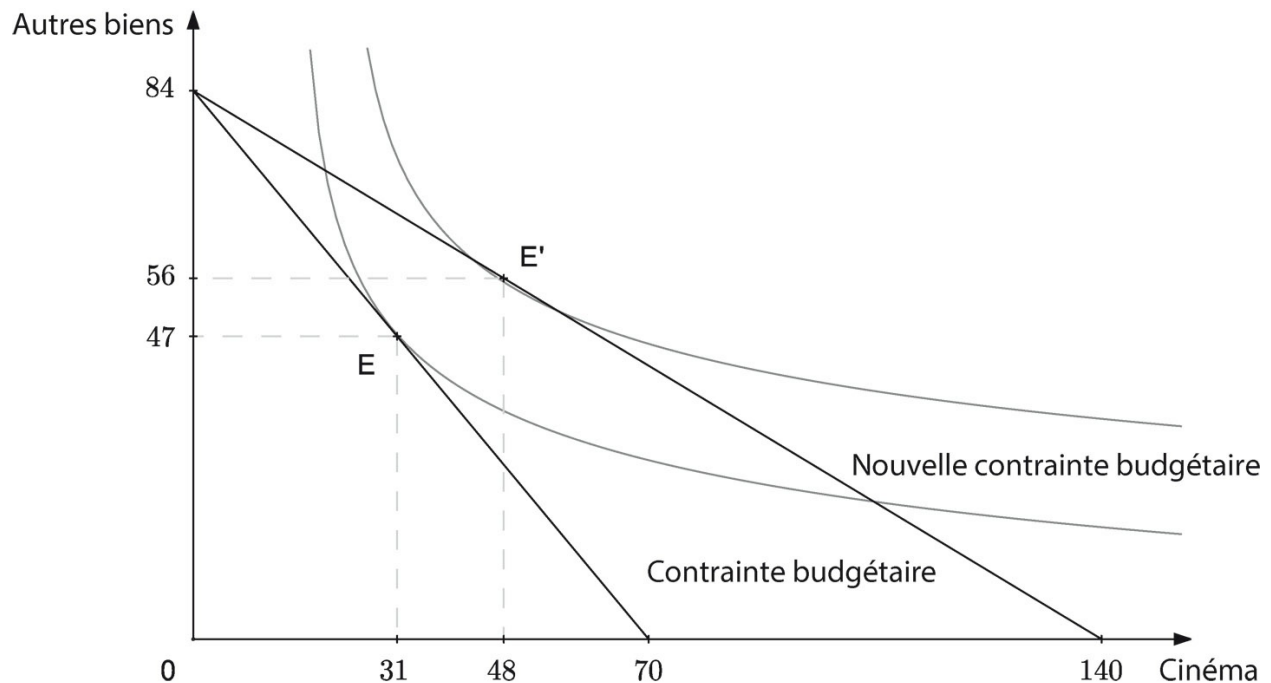
1.5 Les effets d'une variation de prix

Supposons que la municipalité d'une petite ville décide de subventionner le cinéma local en prenant en charge la moitié du prix du billet, pour le rendre plus attractif. Comment cette politique va-t-elle modifier les choix des consommateurs ?

Par hypothèse, les consommateurs arbitrent entre le cinéma et les autres biens culturels, pour un budget donné. Si le prix des places est divisé par deux, cela signifie qu'ils pourront s'offrir deux fois plus de sorties au cinéma, comme on le voit sur le [graphique 2.9](#), la nouvelle contrainte budgétaire

passé alors d'un maximum de 70 à un maximum de 140 sorties annuelles au cinéma, sur l'axe des abscisses. L'impact de la variation du prix du billet va dépendre des préférences des consommateurs. Dans le cas du [graphique 2.9](#), le nouvel équilibre, en passant de E à E', se caractérise par une hausse de la demande de cinéma, mais aussi par une hausse de la demande des autres biens.

Graphique 2.9 : L'effet d'une division par deux du prix des places de cinéma

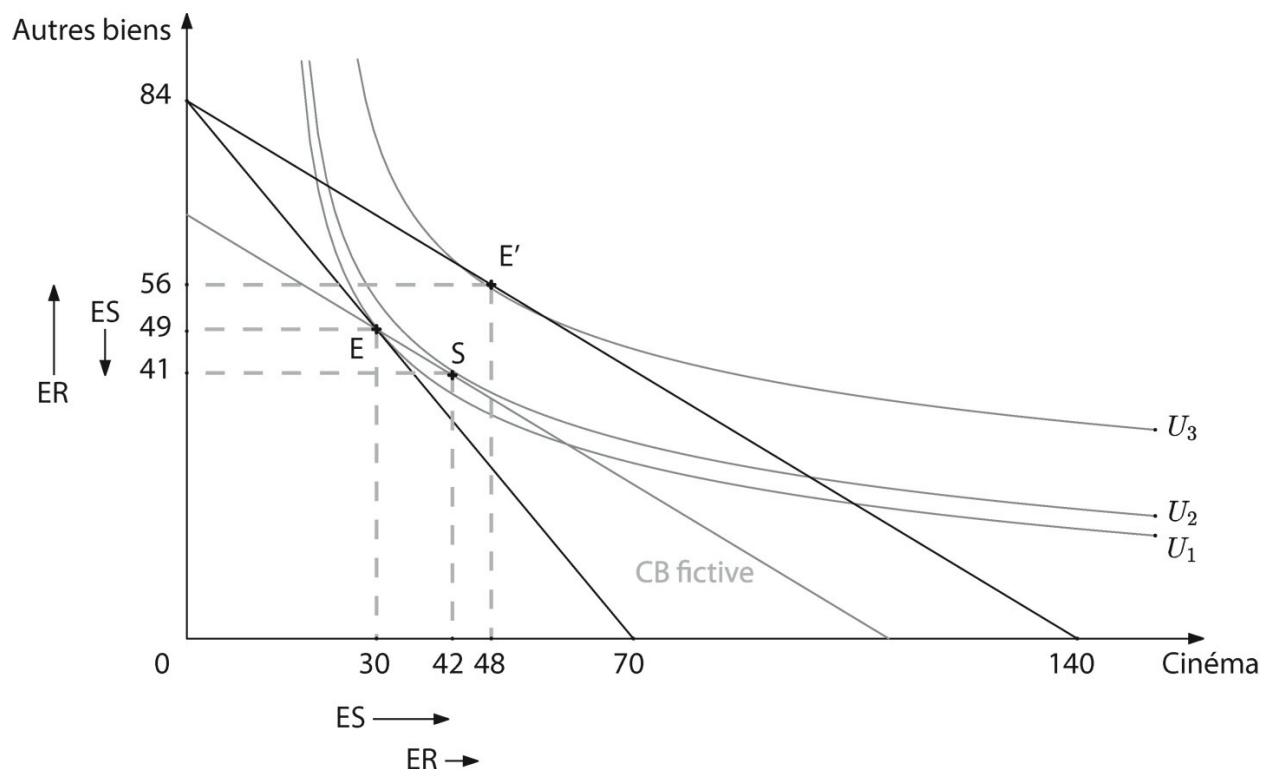


Pour en comprendre la signification, il faut distinguer les deux effets potentiels d'une variation du prix :

- La baisse du prix du cinéma modifie les prix relatifs des biens : le cinéma est devenu plus attractif en termes de prix ce qui incite à s'y rendre davantage. C'est **l'effet de substitution**, qui agit toujours en sens inverse des prix : une hausse des prix entraîne un effet de substitution négatif pour le bien qui subit la hausse, et inversement pour les autres biens.
- La baisse du prix du cinéma accroît le pouvoir d'achat de l'agent, car ce dernier achète un même panier de biens moins cher qu'avant, les places de cinéma lui coûtant deux fois moins cher. Il peut utiliser ce

surcroît de pouvoir d'achat pour aller davantage au cinéma et/ou pour consommer d'autres biens. Il s'agit de l'**effet revenu**, qui est indéterminé : selon les préférences de l'agent et les caractéristiques des biens, il entraîne une hausse ou une baisse de la consommation de cinéma et des autres biens.

Graphique 2.10 : Effet substitution (ES) et effet revenu (ER)



Pour distinguer graphiquement l'effet revenu et l'effet de substitution suite à un changement du prix d'un bien, on fait apparaître² une troisième contrainte budgétaire fictive comme on le voit sur le [graphique 2.10](#), qui remplit deux conditions : elle est parallèle à la nouvelle contrainte budgétaire qui passe par E', autrement dit sa pente dépend des nouveaux prix relatifs, après la baisse des prix du cinéma. En revanche, elle permet toujours d'acheter le panier de bien initial E, puisqu'elle passe par ce dernier, ce qui traduit le fait que le consommateur a conservé le « pouvoir d'acheter » ledit panier.

Cette contrainte budgétaire « fictive » se base sur les nouveaux prix relatifs, mais maintient aussi le pouvoir d'achat initial. Si on fait apparaître le panier S correspondant au choix optimal selon cette contrainte budgétaire « fictive », on peut estimer que le déplacement de E à S met en évidence l'effet de substitution, puisque le pouvoir d'achat est maintenu alors que les prix relatifs changent, tandis que l'effet revenu apparaît entre S et E', quand les contraintes budgétaires ont la même pente et les mêmes prix relatifs, mais se déplacent en fonction du pouvoir d'achat de l'agent. Sur le graphique, les effets revenus (ER) et effets de substitution (ES) n'évoluent donc pas dans le même sens. Pour le cinéma, ils se cumulent car ce consommateur utilise une

partie des gains de pouvoir d'achat liés à la baisse des prix du cinéma pour y aller davantage. En fonction des préférences, les deux effets peuvent agir en sens inverse, et au final, dans notre exemple, l'effet revenu l'emporte sur l'effet de substitution pour les autres biens culturels : le pouvoir d'achat gagné est utilisé pour en consommer davantage.

En conséquence, on peut déterminer la demande individuelle de chaque bien en faisant varier les prix relatifs. L'effet global de la variation du prix d'un bien sur sa demande dépend, au niveau individuel, de l'**élasticité-prix de la demande** qui se calcule en faisant le rapport entre la variation relative de la demande d'un bien (par exemple + 3 %) et la variation relative de son prix (par exemple – 2 %), ce qui donne une élasticité-prix de – 1,5.

On mesure aussi l'effet de la variation du prix d'un bien sur la demande d'un autre bien. L'acquisition d'un livre par exemple, ne dépend pas seulement de son prix, mais aussi du prix d'autres biens entre lesquels le consommateur doit choisir : le prix d'autres loisirs comme le cinéma, par exemple, ou l'achat de logiciels peut avoir une incidence sur leur consommation. L'**élasticité-prix croisée de la demande** se calcule en rapportant la variation relative de la demande d'un bien à la variation relative du prix d'un autre bien, et nous indique si la consommation de ce bien est sensible à d'autres prix.

2. Prolongements et débats autour du modèle du consommateur rationnel

2.1 Consommer aujourd'hui ou consommer demain ?

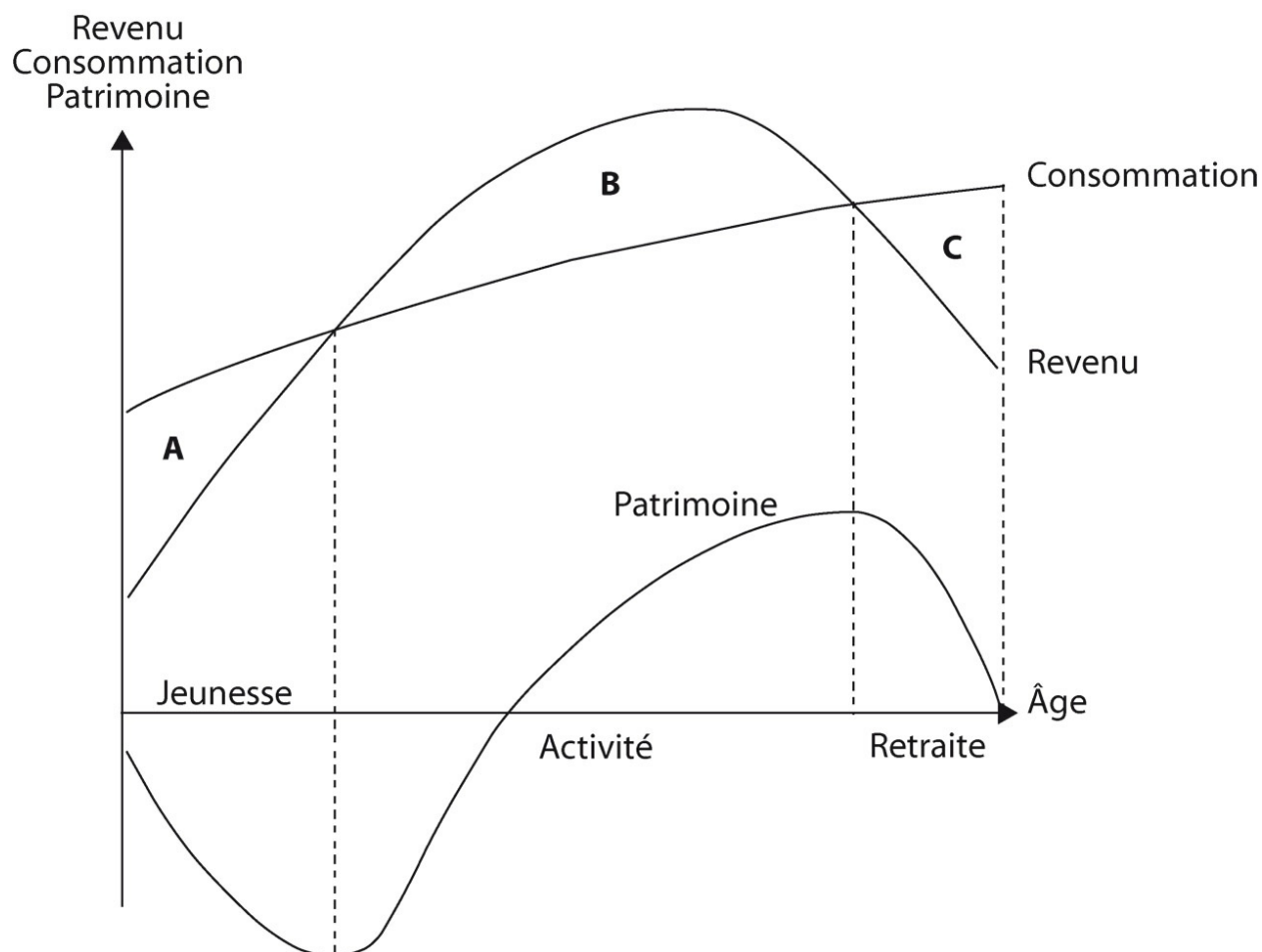
Certains biens culturels ont un coût élevé et nécessitent soit de réduire dans un premier temps sa consommation pour **épargner** et pouvoir acheter ces biens dans un second temps, soit d'emprunter la somme manquante pour pouvoir acquérir le bien tout de suite, ce qui implique une réduction de sa consommation future afin de rembourser le prêt et les intérêts associés. Par exemple, l'achat d'une télévision, la préparation d'un voyage pour visiter le patrimoine culturel d'un pays lointain nécessitent des dépenses importantes, donc de les financer sur une période longue. Le budget dont dispose un individu pour sa consommation à chaque période peut ainsi différer de son revenu.

Supposons qu'un agent ait un revenu de 20 000 € cette année, et s'attende à un revenu identique l'an prochain : il peut décider de consommer davantage que son revenu courant cette année, en remboursant l'année suivante, ou faire l'inverse au contraire, selon sa préférence pour la consommation présente par rapport à la satisfaction apportée par une consommation future. Un agent rationnel maximise donc sa satisfaction sur plusieurs périodes en faisant des choix intertemporels de consommation, en fonction du revenu attendu à chaque période, de l'évolution des prix, et des taux d'intérêt qui agissent sur la décision d'emprunter ou d'épargner. Quand on raisonne sur plusieurs périodes, on considère que le consommateur fait face à une **contrainte budgétaire intertemporelle**. Il détermine le budget qu'il va consacrer à la consommation pour chaque période en tenant compte de la conséquence de ses choix sur les périodes suivantes. Ainsi, le consommateur intègre le fait que s'il épargne, il pourra consommer plus dans le futur car son épargne pourra être placée.

Le choix entre consommation et épargne varie tout au long de l'existence, et le **modèle du cycle de vie** proposé par F. Modigliani (1918-2003) est fondé sur l'idée que les individus cherchent à lisser leur consommation au

cours de leur existence, alors que leurs revenus varient fortement selon l'âge. Afin de connaître une progression régulière de leur niveau de vie, ils doivent emprunter et épargner différemment selon les périodes. Sur le **document 2.2** on voit que dans la période de la jeunesse, la consommation désirée est supérieure au revenu (zone A sur le graphique), et les ménages doivent emprunter ce qui diminue leur **patrimoine net**, ce dernier étant la différence entre les actifs détenus par les ménages (produits financiers, biens immobiliers, biens de valeur, etc.) et leurs dettes. Les ménages peuvent ensuite reconstruire une épargne positive (zone B) au cours de leur période d'activité, ce qui permet de rembourser les dettes puis de se constituer un patrimoine, qui peut être utilisé au cours de la retraite pour compenser la perte de revenu (zone C). Les trois courbes du revenu, de la consommation et du patrimoine net sont donc liées, et le taux d'épargne, entendu comme le rapport entre l'épargne et le revenu, évolue tout au long de l'existence.

Document 2.2 : Revenu, consommation et patrimoine dans le modèle du cycle de vie



Source : d'après P. Villieu (2008), *Macroéconomie – Consommation et épargne*, La Découverte.

Notons que si la valeur du patrimoine augmente en raison d'une hausse du prix d'un actif, cela offre de nouvelles possibilités de consommation aux ménages qui peuvent utiliser cet enrichissement pour consommer davantage. Certains produits culturels ont eux-mêmes une valeur patrimoniale : les tableaux et les objets d'art peuvent prendre de la valeur au cours du temps, ils ne sont pas « consommés » lors de leur acquisition. Les ménages ont donc la possibilité d'acquérir des biens culturels dans un but patrimonial, mais les études empiriques montrent que de façon générale, la rentabilité de l'achat d'œuvres d'art est plus faible que celle des actifs financiers. Il faut donc supposer que les individus achètent des œuvres d'art pour en tirer une satisfaction plus esthétique que financière, ou qu'ils évaluent mal l'évolution des prix des actifs (biais de perception de l'inflation par exemple).

Le modèle du cycle de vie permet de comprendre pourquoi l'âge moyen

d'une population a un impact sur son comportement d'épargne, mais il reste très sommaire car il ne tient pas compte des aléas qui peuvent affecter le revenu au cours du temps, et de l'adaptation des ménages. Ces derniers sont censés connaître parfaitement l'évolution de leurs revenus. **M. Friedman** a jeté les bases d'une nouvelle théorie du consommateur, qui donne une place importante aux anticipations des ménages : on ne dépense pas de la même manière un revenu perçu comme « transitoire » et ce qu'il appelle le **revenu permanent**, c'est-à-dire la part du revenu sur lequel le ménage pense pouvoir compter de manière durable. M. Friedman considère que le revenu permanent détermine de façon stable la consommation des agents, et dépend non seulement des revenus d'activité, mais de l'ensemble de la richesse sur laquelle ils peuvent compter (patrimoine, capital humain, etc.). Cette approche et les débats avec les keynésiens qui en résultent ont joué un grand rôle en macroéconomie. Sur le plan microéconomique, cela permet d'insister sur l'importance des anticipations, et de la perception par les ménages de la stabilité de leur revenu dans la décision de consommer.

2.2 La nouvelle théorie du consommateur

Dans les années 1960, trois articles s'inscrivant dans des perspectives proches vont proposer une nouvelle théorie du consommateur. Ils sont écrits respectivement par G. Becker (1965), R. Muth (1966) et K. Lancaster (1966). Les trois articles entendent compléter voire refondre la théorie néo-classique du consommateur en proposant d'aborder autrement la question du choix des consommateurs. Tous ces travaux conservent l'hypothèse d'un acteur rationnel et optimisateur mais le consommateur est présenté comme un producteur de sa propre satisfaction, à partir des différents « intrants » dont il a besoin pour cette production, à savoir les biens et services disponibles sur les marchés et le temps nécessaire à leur transformation pour produire l'usage qui génère de la satisfaction. Le consommateur n'achète pas des biens car ils ont un intérêt en eux-mêmes mais pour leurs caractéristiques. Les biens achetés permettent de produire des biens ou services qui ont un certain usage aux yeux du consommateur. Les caractéristiques des biens correspondent ainsi aux services que les biens rendent aux consommateurs. Les biens ne sont pas les objets directs de l'utilité, ils se combinent pour produire de l'utilité. Le consommateur s'intéresse à la somme des caractéristiques du panier de biens qu'il achète. Becker introduit le temps comme un intrant dont

le coût est le taux de salaire auquel le consommateur renonce en ne travaillant pas du fait de l'utilisation de ce temps. Ce taux de salaire dépend du capital humain du consommateur.

Ainsi si un consommateur a pour but de consommer un bon repas il existe plusieurs façons d'arriver à ce résultat et le consommateur va acheter les biens qui lui permettent de combler cette attente de la façon la plus satisfaisante possible. Il peut acheter un plat préparé au supermarché, aller au restaurant, se faire livrer ou se cuisiner lui-même un repas en achetant différents produits de base. Ces différentes options n'ont pas les mêmes coûts et ne demandent pas toutes la même quantité de temps. En fonction des préférences du consommateur, du prix relatif des biens et notamment du coût d'opportunité de son temps, il va préférer une option ou une autre.

Appliqué à la culture, ce modèle permet de comprendre pourquoi certains agents ayant une forte préférence pour les activités culturelles limitent pourtant leur pratique. Plus le capital humain est élevé, plus le coût d'opportunité du temps passé aux activités culturelles augmente (voir le [chapitre 6](#)) : un manager de haut niveau peut se passionner pour l'opéra, mais il ne peut y consacrer trop de temps sans « perdre » beaucoup en opportunités de gagner un revenu élevé par son travail, si l'on suppose que la rémunération augmente avec le capital humain.

Cette approche permet de comprendre des changements dans la demande de biens et services sans avoir à les renvoyer à des « changements de préférences » exogènes, non explicables du point de vue économique. Si un nouveau produit est consommé par exemple, c'est peut-être tout simplement qu'il a des caractéristiques permettant d'atteindre de façon plus satisfaisante l'objectif du consommateur, dont les préférences n'ont pas changé. Le rôle des prix relatifs, du capital humain et des revenus est remis au centre de l'analyse.

Lancaster, s'inscrivant ici dans la lignée des travaux de E. Chamberlin (1953), insiste sur la non homogénéité des biens vendus sur un marché. Ils diffèrent par leur qualité, la qualité d'un bien ou d'un service renvoyant à l'ensemble de ses caractéristiques constitutives. Plus l'offre de biens est variée, plus le consommateur pourra atteindre un niveau de satisfaction élevé, compte tenu de sa contrainte budgétaire, car il pourra réaliser des combinaisons plus adaptées à ses attentes à moindre coût. Il pourra trouver plus aisément des biens ayant les caractéristiques désirées. Ce modèle permet

ainsi de mettre en évidence l'attachement des consommateurs à la diversité elle-même, dans le domaine de la culture comme dans d'autres : une situation où l'on peut choisir dans une large gamme de biens augmente les chances du consommateur d'en trouver un conforme à ces attentes. Les consommateurs préfèrent généralement, quand les produits sont différenciés, avoir accès à de vastes librairies ou à des cinémas multiplex pour pouvoir trouver un produit conforme aux caractéristiques qu'ils privilégient. Pour Lancaster, les imperfections de l'information dont dispose le consommateur sur les caractéristiques des biens risquent de le priver d'un bien plus adapté à ses attentes, par méconnaissance. Dans ce contexte la publicité joue un rôle important, non pour changer ses préférences mais pour lui faire connaître les biens disponibles sur le marché et leur spécificité.

• Encadré 2.2 : Une approche sociologique de l'incertitude sur la qualité des produits

Les sociologues de la culture s'intéressent beaucoup aujourd'hui aux dispositifs et aux institutions qui facilitent les choix des consommateurs (guides, systèmes de fidélisation, etc.). C'est un domaine porteur de débats féconds, et parfois critiques, entre les deux disciplines. L. Karpik met ainsi en cause le modèle néo-classique de K. Lancaster dans *L'économie des singularités* (2007), en considérant qu'il ne permet pas de penser la consommation de ce qu'il appelle des biens singuliers, c'est-à-dire des biens qui sont « incommensurables » les uns avec les autres, et ne peuvent pas se réduire à une addition de caractéristiques clairement mesurables. Les vins ou les œuvres artistiques sont des biens singuliers : chacun est unique et leur qualité est très difficilement mesurable objectivement. Selon L. Karpik, la consommation de ces biens repose alors sur des « dispositifs de jugement » construits socialement, qui permettent de coordonner les agents sur le marché en partageant des manières communes de juger les produits. Il montre toute l'importance de la construction des marchés par des guides qui reposent sur une convention partagée sur ce qu'est un « bon vin » ou une « grande œuvre ». Le travail de L. Karpik a donné lieu à des échanges croisés entre économistes et sociologues dans un numéro de la *Revue française de sociologie* qui lui était consacré (STEINER, 2008). L'économiste J. Gautié considère en particulier que la théorie des biens singuliers reste de portée moins générale que l'approche microéconomique qui présente l'avantage de pouvoir assouplir progressivement, et une à une, une série d'hypothèses qui permettent à la fois de s'intéresser aux marchés standards où les biens sont homogènes et aux spécificités de marchés imparfaits.

On peut définir plusieurs types de biens selon l'information dont on dispose sur leurs caractéristiques.

Les **biens d'expérience** (Nelson, 1970) sont des biens pour lesquels l'information sur la qualité du bien est acquise en consommant/utilisant le bien et les **biens de recherche** (Nelson, 1970) des biens pour lesquels

l'information sur le bien est obtenue en analysant le bien avant l'achat. Les **biens de croyance** (Darby, Karni, 1973) sont des biens pour lesquels l'information sur la qualité des biens ne peut être obtenue ni avant ni après l'achat du bien.

Les biens culturels sont souvent des biens d'expérience. Comment savoir à l'avance si un film particulier, un livre ou un spectacle procurera une réelle satisfaction ? Le modèle de base ne répond pas à cette interrogation, il postule en quelque sorte que le consommateur a « vu d'avance » les films entre lesquels il doit choisir.

Quand les biens sont différenciés et qu'il est difficile d'évaluer leur qualité respective. Le consommateur pour limiter le risque d'être déçu face à la profusion de produits dont il a du mal à évaluer la qualité peut recourir à différentes **institutions**, en charge de certifier la qualité ou de sélectionner les produits à la place du consommateur. C'est particulièrement le cas dans le domaine des biens culturels : guides de cinéma, chronique critique d'un journal spécifique, etc. Les travaux empiriques insistent sur l'importance de ces mécanismes de certification qui structurent les marchés de biens culturels, et sur l'adoption d'un comportement propre à réduire les coûts de recherche d'information. La fidélité à un lieu, à une marque ou un distributeur peut être un bon moyen de réduire l'incertitude, comme le montre le travail d'O. Guillon (2009) sur la fidélité du public du Parc de la Villette, où se trouvent différentes expositions, temporaires ou permanentes : l'assiduité à un lieu est un moyen de prolonger la satisfaction des expériences antérieures en sélectionnant une « visite au Parc » sans avoir nécessairement à s'informer sur les qualités propres des biens culturels disponibles au moment de la visite.

2.3 Des préférences endogènes et instables ?

Jusqu'à présent, nous supposons que les préférences d'un acteur étaient données et n'étaient pas influencées par la consommation des acteurs ou par celle de leurs pairs.

Or, on constate que le goût pour un bien ou un service peut évoluer avec sa consommation et en fonction de la consommation des autres. Quand un cinéophile se rend plusieurs fois par semaine au cinéma, il ne voit pas forcément son envie d'aller au cinéma décroître pour autant. L'utilité marginale n'est pas toujours décroissante et peut même augmenter avec

l'évolution des préférences des individus. Une plus grande culture cinématographique permet d'apprécier encore davantage la projection d'un film. La satisfaction liée à la consommation augmente à mesure que l'on accumule de l'expérience, fruit des consommations antérieures. Les biens culturels font souvent partie de cette catégorie de biens à utilité marginale croissante : les amateurs éclairés, qui consomment déjà beaucoup, tirent davantage de plaisir que les novices d'un nouveau bien. Mais comment alors expliquer que les individus ne se jettent pas à corps perdu dans cette consommation, qui peut être à l'origine d'un processus addictif dans ces conditions, l'usage du bien augmentant le désir de le consommer ? En prenant en compte notamment le fait que leur consommation prend du temps et qu'il existe un coût d'opportunité du temps.

Ce raisonnement peut également s'appliquer à des consommations addictives comme le tabac et les autres produits à l'origine d'addictions, pour lesquels on observe également l'absence de phénomène de satiété et donc la non-décroissance de l'utilité marginale. Cette situation s'observe aussi dans le cas de collectionneurs qui ne se lassent pas d'avoir toujours plus du bien dont ils font la collection.

Ainsi un consommateur peut prendre de plus en plus de plaisir à aller au théâtre à mesure qu'il se familiarise avec le genre. Ses préférences évoluent en fonction de sa consommation. Dans le modèle standard du consommateur rationnel, le plaisir associé à la consommation des différents types de bien n'est pas censé évoluer à mesure que les quantités consommées changent. Les préférences sont supposées être stables. Le modèle envisage qu'elles puissent changer sur le long terme mais ce changement est censé provenir de facteurs exogènes à l'acte de consommation.

Les études empiriques mettent en avant l'instabilité des préférences pour les biens culturels. Une enquête de L. Lévy-Garboua et C. Montmarquette³ sur le public du théâtre confirme l'effet de l'expérience passée sur la demande courante, mais le goût pour le théâtre ne s'accroît pas systématiquement, chaque nouvelle expérience permet en fait au consommateur de « découvrir » un peu plus précisément ce qu'il apprécie ou non, et ce processus d'apprentissage modifie sa demande par rapport à ce bien, mais aussi par rapport à ses substituts. Les différents publics ne réagissent pas de la même manière, les consommateurs occasionnels préférant par exemple les spectacles les moins risqués. Une expérience

d'évaluation des goûts musicaux plus récente (2011) portant sur un groupe de 120 participants permet de confirmer que le jugement sur la qualité des morceaux dépend de l'expérience cumulée au cours du processus « d'apprentissage par la consommation », plutôt que des goûts affirmés au début de l'expérience. Cela implique que les goûts des consommateurs ne préexistent pas aux choix de consommation, mais se transforment progressivement à mesure que le consommateur fait de nouvelles expériences.

Par ailleurs, les préférences des individus peuvent s'avérer plus complexes et parfois même contradictoires. Pour l'illustrer, on notera que certaines pratiques culturelles sont marquées par un net décalage entre ce que déclarent les agents sur la qualité des biens, ou sur l'importance qu'il leur attribue, et ce que révèlent leurs pratiques réelles : ainsi les consommateurs ont une opinion sur la qualité des programmes de télévision très différente de ce que montrent leurs pratiques réelles. Ils ont en effet tendance à regarder des programmes pour lesquels ils expriment publiquement un jugement négatif et n'assument pas leur attrait pour certaines émissions, films, etc. (MOURAU, VIDAL, 2009). Pour l'expliquer, T. Schelling considère que les individus sont l'objet d'un conflit intérieur entre différents systèmes de préférences, qui ne sont pas unifiés, par exemple des préférences liées au parcours personnel et des préférences partagées avec les membres d'un groupe d'appartenance, des préférences liées à leurs goûts impulsifs de court terme et à leurs goûts « raisonnés » de long terme, etc. Cette « théorie des moi multiples » éclaire le fait que les préférences déclarées sont loin de correspondre aux préférences révélées par les pratiques. Cette situation est assez fréquente pour les biens culturels, quand se jouent les contradictions entre les choix de pratiques jugées légitimes aux yeux des autres et les choix individuels.

2.4 Des choix rationnels ?

Dans le modèle de base, les choix du consommateur sont supposés rationnels : compte tenu de leurs préférences, il tire le meilleur parti possible des informations et des ressources dont il dispose. Or rien ne garantit qu'il en soit ainsi : les agents font-ils toujours des choix cohérents ? Réagissent-ils de façon rationnelle aux changements de leur environnement économique ?

Pour répondre à ces questions, le mieux est de tester, par une démarche

expérimentale, le degré de rationalité des comportements, et plus généralement de repérer expérimentalement les régularités dans le comportement du consommateur. Le développement de l'**économie comportementale** va dans ce sens : le terme désigne un vaste champ d'études visant à observer les comportements des agents économiques afin d'étudier les raisons d'éventuelles déviations entre les comportements observés et les prédictions des modèles théoriques, en s'appuyant sur un raisonnement expérimental identique à celui des sciences de la nature. Elle peut s'appuyer sur différentes méthodes :

- Des « **expériences de laboratoire** » pour tester une hypothèse dans un environnement contrôlé afin de raisonner « toutes choses égales par ailleurs », en demandant par exemple à un public réuni dans une même salle de procéder à des choix de manière anonyme, en contrepartie généralement d'une compensation monétaire. On parle aussi d'économie expérimentale. Ce terme est généralement réservé aux expérimentations de laboratoire, mais la démarche expérimentale est utilisée dans d'autres contextes comme nous pouvons le voir dans les points suivants.
- Les « **expériences contrôlées** » se déroulent dans un environnement naturel (non créé artificiellement pour les besoins de l'expérience), et consistent à modifier une règle ou un mécanisme pour en observer les effets économiques. On compare un « groupe test » soumis au changement de règle dont on étudie l'impact, et un « groupe de contrôle » qui ressemble en tout point au premier, sauf qu'il n'est pas affecté par le changement considéré. Les expériences contrôlées reposent sur une méthodologie statistique très proche de celle que l'on utilise en biologie ou en médecine, par exemple pour tester les effets d'un médicament⁴.
- Les « **expériences naturelles** » se déroulent également dans un environnement naturel, mais s'appuient sur l'examen d'une situation accidentelle, non voulue par le chercheur, qui permet de « simuler » l'effet d'une variable dans des conditions quasi-expérimentales. Par exemple, la comparaison de la consommation culturelle dans les villes où l'on ouvre un cinéma et dans des villes équivalentes mais sans cet événement permet de raisonner « toutes choses égales par ailleurs » sur les effets de l'offre culturelle sur les comportements, à condition de

contrôler rigoureusement que les deux types de villes ont les mêmes caractéristiques dans l'échantillon.

Cependant, les travaux d'économie comportementale montrent que les agents ne se comportent pas toujours de façon rationnelle au cours de leur cycle de vie : ils peuvent par exemple faire preuve d'**incohérence temporelle**, c'est-à-dire ne pas agir demain comme ils prévoient de le faire aujourd'hui. Par exemple, les agents ont tendance à reporter au lendemain la décision d'épargner, alors qu'ils regretteront ce choix dans le futur, on parle d'effet procrastination. Dans ces conditions, l'absence de régulation du marché conduit à une mauvaise allocation des ressources, car les agents consomment davantage qu'ils ne le feraient s'ils parvenaient à maintenir des choix cohérents. Cette myopie des acteurs est souvent mise en avant pour justifier l'intervention tutélaire de l'État. Par exemple, l'assurance automobile au tiers est rendue obligatoire par l'État, de même que l'assurance du logement. Certains économistes proposent de substituer des incitations à la fonction tutélaire de l'État. **R. Thaler** et **C. Sunstein** suggèrent ainsi, à partir d'études expérimentales, de mettre en place des « coups de pouce » (ou **nudges** en anglais) de nature à aider les agents à faire de meilleurs choix. Les agents ayant souvent du mal à changer leurs habitudes mêmes quand les conditions se sont modifiées (ce que l'on appelle le biais de statu quo), les auteurs conseillent ainsi par exemple de mettre en place des options de « choix par défaut » en faveur des plans d'épargne-retraite. Dans ces conditions, sauf si un individu exprime explicitement son refus pour cette option, on considère par défaut qu'il souhaite souscrire à un plan d'épargne retraite. Les études expérimentales montrent en effet que la décision de « décocher » un choix que l'on sait favorable dans une liste d'options n'est pas équivalente au choix symétrique de « cocher » cette option. Ces auteurs défendent donc un **paternalisme libertarien**, c'est-à-dire une attitude incitative de la part des autorités publiques, qui aide les individus à faire de meilleurs choix sans se substituer à eux.

Les recherches en économie comportementale ont joué un rôle majeur dans l'évolution de la microéconomie ces dernières décennies, et ont valu le Prix Nobel d'économie à R. Thaler en 2017. Elles se sont particulièrement intéressées à la **rationalité** des choix de consommation. Un agent est rationnel s'il est capable d'évaluer les différentes options qui s'offrent à lui, et de faire ensuite des choix cohérents par rapport aux fins poursuivies et aux

ressources dont il dispose. Les travaux expérimentaux montrent qu'en fait la rationalité des agents est limitée et contextuelle, c'est-à-dire que les agents ne sont pas toujours en situation de faire « les meilleurs choix », et suivent des procédures de décision variables en fonction du contexte. Ils font parfois preuve de **biais cognitifs**, c'est-à-dire de tendances régulièrement observées à s'écarter du comportement attendu de la part d'un agent rationnel. Voici quelques exemples adaptés à la demande de biens culturels :

- L'**effet de dotation** intervient quand un agent accorde plus de valeur à la détention d'un objet qu'à son acquisition alors qu'un agent rationnel devrait considérer que l'utilité marginale de l'acquisition d'un bien supplémentaire est équivalente à la désutilité marginale de sa session. Par exemple, être disposé à payer moins cher pour acquérir un CD d'occasion que le prix exigé pour s'en séparer. En présence d'effets de dotation, les préférences du consommateur ne sont plus systématiquement transitives, autrement dit les courbes d'indifférences peuvent « se croiser », ce qui nous éloigne du modèle standard vu précédemment. L'effet de dotation intervient dans de nombreux domaines comme le montrent les travaux de D. Kahneman (prix Nobel en 2002). Par exemple, les opérateurs de services multimédia (internet, téléphone mobile, abonnement télévisé...) peuvent faire payer plus cher les « anciens clients » qui ne souhaitent pas changer leurs habitudes et sont donc enclins à accepter un tarif plus élevé que les « nouveaux clients », qui doivent être attirés par des tarifs plus avantageux, car ils ne sont pas encore « dotés » de ce bien et n'ont donc pas d'attachement particulier pour lui.
- Les agents n'évaluent pas toujours correctement le prix qu'ils seraient prêts à payer pour obtenir un bien, mais ils s'efforcent de maintenir cohérents leurs choix, sur une base très conventionnelle : c'est ce qu'on appelle la « **cohérence arbitraire** ». Par exemple, des agents sont prêts à payer un spectacle particulier à des tarifs très différents selon les circonstances de leur choix, et les comparaisons qu'ils peuvent faire avec d'autres prix, qui servent de point de comparaison arbitraire. Une expérience de vente aux enchères menée par D. Ariely (2008) montre ce processus à l'œuvre : on vend différents biens aux enchères, en demandant au préalable aux agents d'inscrire comme prix de départ les deux derniers chiffres de leur numéro de sécurité sociale ; ils sont ensuite libres de fixer leur prix maximal d'achat et celui qui a indiqué

le prix le plus élevé emporte l'enchère. L'expérience montre que le prix de départ, purement arbitraire, sert bien de point d'ancrage aux prix demandés ensuite, la corrélation statistique entre les deux étant très forte. Le **biais d'ancrage** qui en résulte se vérifie dans de nombreuses circonstances, et les firmes en jouent car il est souvent très avantageux de parvenir à imposer un tarif comme « normal » du point de vue des consommateurs. Au cours des années 2000, les ventes de musique en ligne sur des sites comme celui d'Apple ont été l'occasion par exemple d'un nouvel « ancrage » des prix sur une norme différente du prix du CD. Si les consommateurs sont prêts au fond à accepter des prix très différents pour un même type de bien, la régulation publique des marchés est donc nécessaire pour éviter que les firmes n'en jouent systématiquement.

- Les travaux d'économie comportementale mettent en avant l'existence de nombreux autres biais cognitifs (biais de représentativité, aversion aux pertes, etc.). Ils ne condamnent pas pour autant le modèle standard du consommateur, mais poussent à adopter une attitude « agnostique », selon ce que nous apprend l'étude empirique d'un marché particulier, pour décider s'il est pertinent de l'utiliser ou s'il faut au contraire rechercher d'autres modèles.

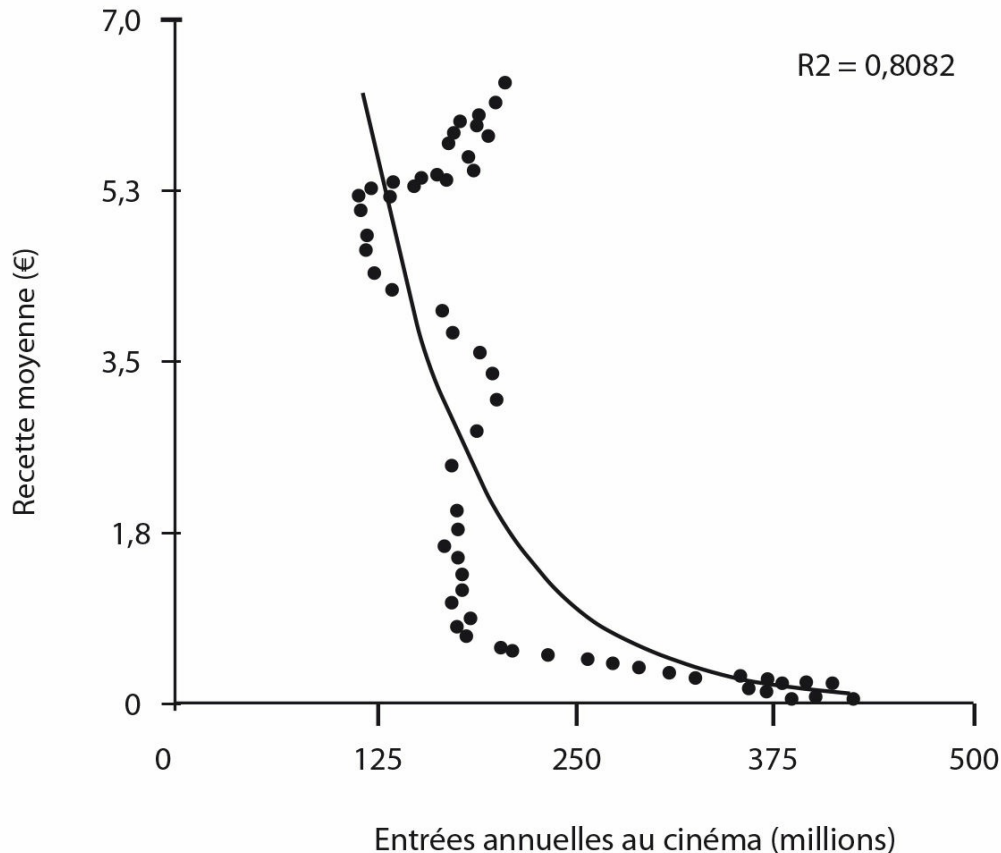
3. De la demande individuelle à la demande agrégée

3.1 La détermination de la courbe de demande

La **demande agrégée**, qui s'exprime sur un marché pour un prix donné, se déduit des demandes individuelles. Pour chaque niveau de prix, compte tenu du prix des autres biens et de son budget, chaque consommateur exprime une certaine demande. En sommant les quantités demandées par chacun, on obtient la demande agrégée pour un niveau de prix donné.

Le **document 2.3**, représente la courbe de demande de places de cinéma avec en abscisses les quantités demandées et en ordonnées le prix moyen des places (qui correspond ici à la recette moyenne du producteur par place vendue). Les points figurent les données annuelles en France entre 1947 et 2010. La courbe a été obtenue par une méthode de régression simple, en faisant l'hypothèse que la seule variable agissant sur la demande de cinéma est le prix des places : il s'agit alors de trouver la courbe qui passe « au plus près » du nuage de points. Le travail du statisticien consiste à tester différents modèles de régression pour sélectionner le mieux ajusté aux données. On peut estimer la « qualité » de la courbe obtenue (c'est-à-dire son adéquation aux données étudiées) par divers indicateurs statistiques, le plus courant étant le coefficient de détermination R^2 , qui varie entre 0 (ajustement nul aux données) et 1 (régression parfaitement ajustée aux données).

Document 2.3 : Estimation statistique de la courbe de demande de cinéma



Source : Données du Centre National du Cinéma, période 1947-2010, calculs des auteurs.

Une méthode de régression de ce type reste très sommaire, elle porte sur des données agrégées, en faisant l'hypothèse que la demande de cinéma répond à un même « modèle » durant plusieurs décennies. Les enquêtes économétriques portant sur des données individuelles sont beaucoup plus fiables, car elles s'appuient sur les différences de comportement entre individus à un moment précis où ils font face au même contexte. Elles permettent de distinguer l'effet propre de chaque variable sur le comportement et pas seulement l'effet du prix du bien ; or de nombreuses variables distinctes du seul prix peuvent agir sur la demande : évolution du revenu des ménages, présence de biens substituables, évolution des goûts des consommateurs, etc.⁵

3.2 Élasticités et typologie des biens

• Courbe de demande et élasticité-prix

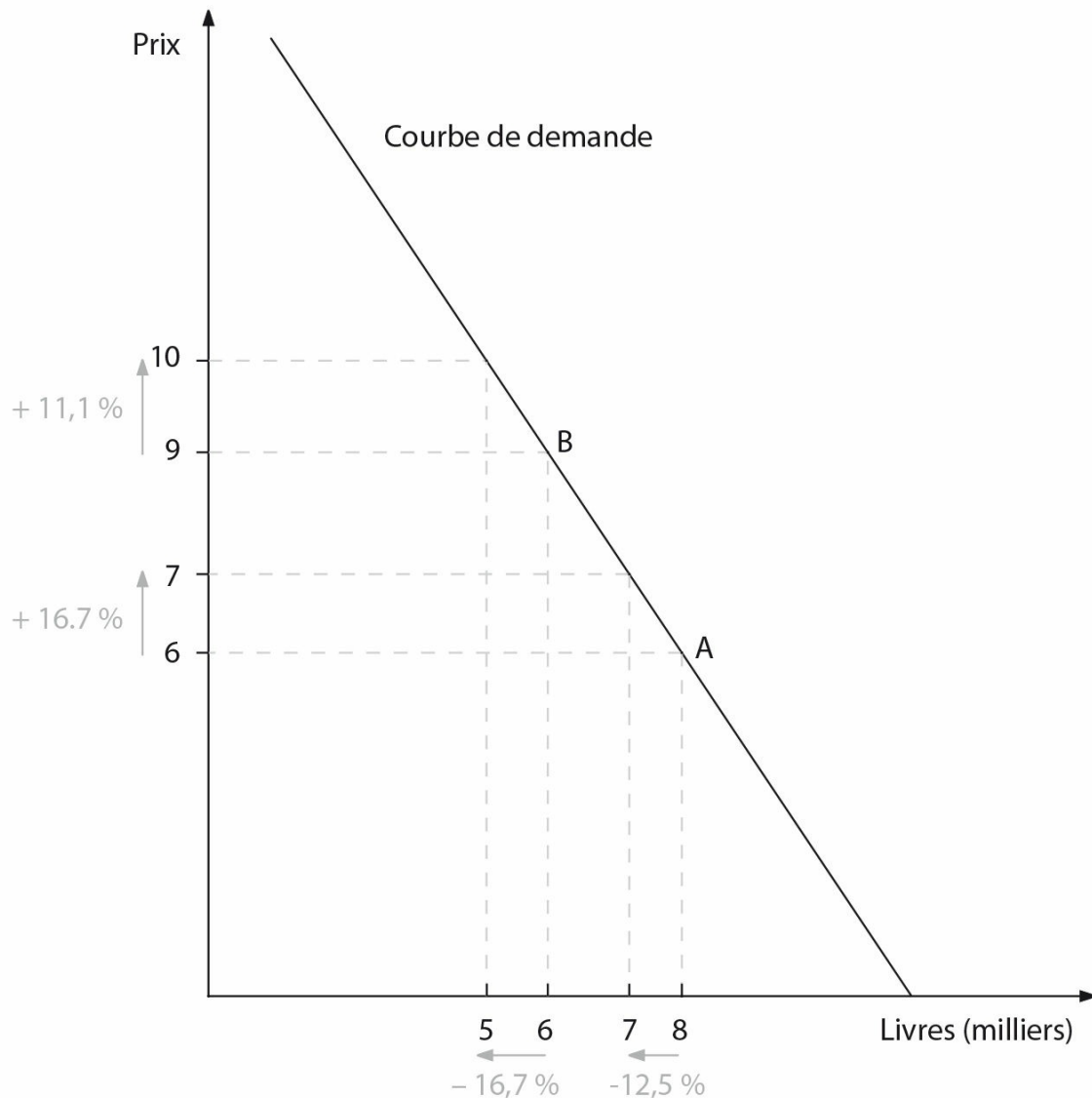
Pour connaître au mieux l'effet spécifique des prix sur la demande agrégée, les économètres utilisent des modèles plus complexes évaluant l'effet spécifique de chaque variable, de façon à raisonner « toutes choses égales par ailleurs », en neutralisant l'influence des autres variables retenues. L'utilisation de ce type de modèles permet d'estimer au mieux l'**élasticité-prix de la demande** des agents au niveau agrégé, c'est-à-dire la sensibilité de la demande aux variations relatives du prix du bien (voir le [chapitre 1](#)). Dans une étude sur la demande de cinéma, S. Ragot montre par exemple, à partir d'un modèle complet de ce marché intégrant les différentes variables susceptibles d'influer sur la consommation de cinéma que l'élasticité-prix est de $-1,34$ à court terme et de $-1,66$ à long terme (RAGOT, 2010). Cela signifie qu'une hausse de 1 % du prix des places de cinéma entraîne à court terme une baisse de 1,34 % de la demande de cinéma, et permet de caractériser précisément la façon dont la demande réagit aux prix, une donnée centrale si l'on s'interroge par exemple sur l'efficacité d'une politique culturelle de soutien au cinéma. La distinction entre élasticité-prix de court terme et de long terme est également importante, car sur de nombreux marchés les comportements sont plus stables en courte période, ce que montre la valeur plus élevée de l'élasticité à long terme : une hausse des prix du cinéma a des effets plus importants sur la durée, car les consommateurs s'équipent et se tournent progressivement vers d'autres biens culturels, comme le suggère la progression de la VOD (la vente de films en ligne).

La valeur de l'élasticité-prix nous renseigne sur la forme de la courbe de demande agrégée d'un bien. Si l'élasticité-prix est très faible et proche de zéro, la demande est inélastique et la courbe a une pente importante puisque par définition, une forte variation du prix entraîne une faible variation de la demande. À l'inverse, une demande fortement élastique aux prix se traduit par une courbe presque horizontale, pour laquelle une faible variation de prix entraîne une variation importante de la quantité demandée.

La valeur de l'élasticité change tout au long de la courbe de demande, sauf cas particulier des courbes de demande à élasticité constante. Cela est lié au fait que l'on calcule l'élasticité à partir d'une valeur initiale et qu'une même variation d'une unité ne donnera pas lieu au même taux de variation selon cette valeur initiale. Sur le [graphique 2.11](#), on voit que la valeur de l'élasticité de la demande de livres est différente selon qu'on se situe

initialement au point A, avec une élasticité de $-12,5/+16,7 = -0,75$, et au point B, avec une élasticité de $-16,7/+11,1 = -1,5$. Ce point est important quand on commente la valeur empirique d'une élasticité-prix : il faut se souvenir qu'elle a été calculée « au voisinage » de la quantité d'équilibre dont on est parti, rien ne garantit qu'elle soit identique pour des quantités très différentes.

Graphique 2.11 : L'élasticité-prix varie le long de la courbe de demande



Pour lire correctement la courbe de demande agrégée, la notion de **disposition à payer** est essentielle. La courbe du [graphique 2.11](#) indique par exemple au point B que les consommateurs seraient prêts à payer au maximum 9 euros pour acquérir une quantité de 6 000 livres. La courbe de demande ne donne donc pas le prix que les consommateurs souhaitent payer, mais le prix maximal qu'ils sont disposés à payer pour une quantité donnée. Pour estimer la disposition à payer, on peut soit interroger directement un échantillon de consommateur en testant leur disposition à payer divers biens⁶, soit utiliser une méthode d'évaluation économétrique, comme on l'a vu dans la section précédente.

• Les différentes élasticités de la demande

Les différents types d'élasticité de la demande évoqués au niveau individuel peuvent être calculés pour la demande agrégée, comme on le voit dans le **document 2.4**.

Document 2.4 : Les différentes élasticités de la demande

Type d'élasticité	Définition	Exemples
Élasticité-prix de la demande	$\frac{\text{Taux de variation de la demande}}{\text{Taux de variation du prix du bien}}$	Élasticités comprises entre – 0,2 et – 0,5 pour le spectacle vivant dans les pays développés
Élasticité-prix croisée de la demande	$\frac{\text{Taux de variation de la demande}}{\text{Taux de variation du prix d'un autre bien}}$	Demande de livre par rapport au prix relatif du disque : élasticité de 0,6 en France en longue période
Élasticité-revenu de la demande	$\frac{\text{Taux de variation de la demande}}{\text{Taux de variation du revenu}}$	Élasticité-revenu du spectacle vivant de 1,52 et de 0,61 pour la presse en 2006
Élasticité-prix de la demande de court terme	$\frac{\text{Taux de variation de la demande à court terme}}{\text{Taux de variation du prix du bien à court terme}}$	Élasticité-prix de court terme de la demande de cinéma : – 1,34
Élasticité-prix de la demande de long terme	$\frac{\text{Taux de variation de la demande à long terme}}{\text{Taux de variation du prix du bien à long terme}}$	Élasticité-prix de long terme de la demande de cinéma : – 1,66

Remarque : l'élasticité-revenu de la demande calculée au niveau agrégée s'appuie

généralement sur le revenu disponible des ménages, comme c'est le cas dans ce tableau, et non sur le budget des ménages.

Source des données empiriques : S. Ragot (2010), F. Benhamou (2008), B. Maresca *et al.*, (2011), M. Perona (Mémoire DEA de sciences économiques, 2004, *Essai de mesure de l'élasticité-prix de la demande de livres et de disques en France, 1960-2003*).

• Typologie des biens selon la valeur de leur élasticité

La détermination empirique des différentes valeurs de l'élasticité pour un bien donné est déterminante pour connaître ses caractéristiques, et prédire la réaction du consommateur à son égard : c'est un peu le « tableau périodique des éléments » du microéconomiste. Ainsi l'**élasticité-prix** d'un bien permet de préciser s'il a tendance à suivre le comportement « typique » de la plupart des biens, dont la consommation diminue quand le prix augmente avec une élasticité-prix négative dans ce cas. Au contraire, si l'élasticité-prix est positive, cela signale un **bien atypique**, plus rarement vérifié : le prix de certains biens sert de « signal » de leur valeur aux yeux des consommateurs (effet de snobisme), par exemple le prix élevé des œuvres d'un artiste peut en augmenter la demande car cela signifie qu'il est « à la mode ». On se trouve alors en présence d'une **externalité de réseau**, c'est-à-dire d'une situation où la demande d'un bien par un individu dépend de la demande des autres consommateurs envers ce bien.

L'élasticité-revenu permet de classer les biens selon l'évolution de leur consommation quand le revenu augmente. On s'appuie généralement sur le revenu disponible du ménage, c'est-à-dire sur le revenu dont il dispose réellement pour consommer et épargner, compte tenu des taxes et impôts versés, et des allocations reçues. Comme on le voit dans le **document 2.5**, si l'élasticité-revenu est négative, on a affaire à un **bien inférieur**, dont la quantité consommée diminue en moyenne dans le budget lorsque le revenu augmente, car les agents lui préfèrent des biens de meilleure qualité, par exemple un modèle « premier prix » d'appareil photo par rapport au reste de la gamme. C'est rarement le cas des biens culturels, mais plutôt réservé à des biens de consommation de base. Les **biens normaux** ont une élasticité-revenu comprise entre 0 et 1 : la quantité consommée augmente avec le revenu, mais la part dans le budget diminue puisque la demande augmente moins vite que le revenu. Par exemple, la presse a une élasticité-revenu inférieure à 1, car un ménage qui s'enrichit augmente bien sa consommation de presse mais pas dans les mêmes proportions que son revenu. Cela

correspond à un grand nombre de biens de consommation courante. Enfin les **biens supérieurs** ou « biens de luxe » ont une élasticité supérieure à 1 : ils sont très sensibles à la hausse du revenu, les quantités consommées augmentant plus que proportionnellement par rapport à la hausse du revenu. Cela correspond souvent à des biens non indispensables mais très prisés, vers lesquels on se tourne dès que le niveau de vie le permet. En règle générale, les biens culturels sont des « biens de luxe », dont la part dans le budget des ménages augmente quand ils s'enrichissent.

L'élasticité-prix croisée permet quant à elle de déterminer la nature des relations entre deux biens. Si elle est négative, cela signifie que les biens sont **complémentaires** puisque la hausse du prix de l'un réduit la demande de l'autre : les biens sont consommés ensemble de préférence, à l'image des DVD et de leur lecteur. Si l'élasticité-prix croisée est positive, cela signifie au contraire que les biens sont **substituables**, puisque la hausse du prix de l'un augmente la demande de l'autre, car le consommateur préfère lui substituer ce dernier : une hausse du prix du théâtre entraîne par exemple une hausse de la demande de cinéma.

Document 2.5 : Les caractéristiques des biens selon leurs élasticités

Valeur de l'élasticité	Caractéristiques des biens	Biens concernés
élasticité-prix de la demande < 0	bien typique : si l'élasticité prix est négative, la demande du bien diminue quand le prix augmente	la plupart des biens
élasticité-prix de la demande > 0	bien atypique : la demande du bien augmente quand le prix augmente	certain produits de luxe, avec un « effet de snobisme »
élasticité-prix croisée de la demande > 0	biens substituables : la demande du bien augmente quand le prix d'un bien substitut augmente	les livres et les DVD ; le cinéma et le théâtre
élasticité-prix croisée	biens complémentaires : la demande du bien diminue	les logiciels et les ordinateurs, les DVD

de la demande < 0	quand le prix du bien qui lui est complémentaire augmente	et les lecteurs de DVD
élasticité-revenu de la demande < 0	bien inférieur : la demande du bien diminue quand le revenu augmente	Les produits « premier prix » de la grande distribution, les produits de faible qualité
élasticité-revenu de la demande comprise entre 0 et 1	bien normal : la demande du bien augmente quand le revenu augmente, mais moins vite que le revenu	la presse, l'équipement informatique et télévisuel
élasticité-revenu de la demande > 1	bien supérieur : la demande du bien augmente plus vite que le revenu	loisirs, transports aériens, le théâtre, l'opéra, le cinéma

3.3 Fiscalité et régulation de la demande de biens

La quasi-totalité des marchés de biens font l'objet de taxes à la consommation qui augmentent leur prix et les biens culturels n'échappent pas à la règle. Les biens peuvent aussi être subventionnés, ce qui a l'effet symétriquement inverse (voir le [chapitre 1](#)). La plupart des taxes ou subventions concernent les producteurs et affectent donc l'offre des biens (modification de la courbe d'offre, la courbe de demande restant elle inchangée) Les subventions prennent la forme d'aides directes à la production, très fréquentes dans le spectacle vivant, d'allègement de la TVA par rapport au taux normal, comme c'est le cas par exemple de la TVA sur les livres. Il s'agit parfois de dispositifs plus complexes, tel le soutien des intermittents du spectacle par un régime avantageux d'allocations-chômage, qui repose sur les relations d'interdépendance entre le marché du travail et le marché du spectacle vivant.

Les biens culturels font l'objet d'une fiscalité spécifique, pour diverses raisons. Les biens culturels sont à l'origine d'externalités positives, qui justifient de les subventionner ou de réduire la taxation car en l'absence d'intervention publique le marché ne tire pas efficacement parti de ces effets (voir le [chapitre 5](#)) : par exemple, la production d'œuvres cinématographiques contribue au rayonnement culturel d'un pays ce qui bénéficie à la collectivité. D'autre part, la politique publique soutient la diversité culturelle, c'est-à-dire une production très diversifiée considérée

comme préférable collectivement, en usant des incitations fiscales appropriées pour éviter la tendance à l'uniformisation des produits culturels. Cela nécessite des mécanismes spécifiques de régulation pour éviter qu'un petit nombre de producteurs dominent ces marchés, en particulier lorsqu'il existe d'importantes économies d'échelle comme dans les industries culturelles (livre, cinéma, disque, etc.). Par exemple, le prix unique du livre contribue à une politique de rééquilibrage des territoires par le maintien de librairies dans les villes moyennes et de librairies spécialisées (PERONAT, 2010).

Enfin, la production de biens culturels est soutenue dans de nombreux pays pour éviter sa disparition en raison de prix trop élevés dus au mécanisme de la « fatalité des coûts » identifié par W. Baumol et W. Bowen. Selon eux, les activités culturelles ne réalisent pas de gains de productivité et subissent donc une augmentation relative inéluctable de leurs coûts, par rapport au reste de l'économie qui connaît des gains de productivité réguliers : c'est ce qu'on appelle la **loi de Baumol**. Les prix relatifs des biens culturels se dégradent nécessairement par rapport aux autres : il faut toujours le même nombre d'heures de travail pour produire un costume d'opéra, mais beaucoup moins pour produire les biens industriels, donc la production artistique coûte de plus en plus cher par rapport aux autres, si le taux de salaire des travailleurs dans le secteur de la culture reste le même que dans les autres secteurs.

Pour ces différentes raisons, la production de spectacle vivant est une activité particulièrement subventionnée : selon une enquête de la DEPS sur les compagnies de spectacle vivant entre 2007 et 2009, les recettes liées à la vente des spectacles représentent 45,1 % de leurs ressources, tandis que la part des subventions publiques directes est de 36,7 %, sans compter les autres formes d'aide. Le prix payé par les spectateurs peut donc être fort éloigné du coût réel. L'action d'une subvention sur l'offre est donc décisive pour diminuer le prix d'équilibre, mais elle pose problème si la demande est relativement rigide par rapport aux prix (voir le [chapitre 1](#)). Dans ces conditions, les subventions aux activités culturelles peuvent avoir des effets anti-redistributifs, car elles permettent à un public au revenu plus élevé que la moyenne de fréquenter à moindre coût des salles de spectacle, sans pour autant attirer massivement un nouveau public. Pour répondre à ces critiques, une solution consiste à différencier les tarifs selon les publics et leur consentement à payer. Dans leur rapport sur la valorisation du patrimoine culturel (2011), F. Benhamou et D. Thesmar notent qu'une large partie du

public des musées est très peu sensible aux prix, alors que ces derniers sont souvent subventionnés à hauteur de 50 % comme le Louvre. Ils proposent donc d'introduire une discrimination par les prix, en augmentant fortement les tarifs des visiteurs non communautaires qui accordent peu d'importance au prix des musées dans le budget total de leur voyage, et d'adopter une politique tarifaire beaucoup plus flexible pour les autres, de façon à attirer de nouveaux publics.

Notes

1. W. Baumol et W. G. Bowen, *Performing Arts : The Economic Dilemma*, New York, The Twentieth Century Fund, 1966.
2. La méthode présentée ici, dite de « Slutsky-Samuelson », n'est pas la seule disponible pour distinguer effet revenu et effet de substitution : une solution alternative, dite « méthode de Hicks », consiste à maintenir le niveau de satisfaction initial plutôt que le pouvoir d'achat correspondant au panier initial, en appliquant les nouveaux prix relatifs de façon à faire apparaître une nouvelle contrainte budgétaire qui est tangente à la courbe d'indifférence de départ. Les deux méthodes ne donnent pas des résultats similaires, sans qu'il y ait d'argument majeur pour trancher entre les deux.
3. Les deux expériences sont présentées dans les articles suivants : L. Lévy-Garboua et C. Montmarquette, « A microeconomic study of theatre demand », *Journal of Cultural Economics*, vol. 20, n° 1, 1996 ; L. Lévy-Garboua *et al.*, « Learning and addiction in the cultivation of musical tastes : an experiment », Working paper, AFSE Congress, 2011.
4. Pour une présentation détaillée de ces méthodes, voir A. Jatteau, *Les expérimentations aléatoires en économie*, La Découverte, coll. « Repères », 2013.
5. Pour approfondir sur les méthodes empiriques de détermination de l'offre ou de la demande, voir L. Behagel, *Lire l'économétrie*, La Découverte, coll. « Repères », 2012.
6. À titre d'exemple, les enquêtes sur la disposition à payer des biens culturels comme les musées montrent que les consommateurs sont souvent disposés à payer bien plus cher que le coût réel du billet, et lui donnent une valeur y compris s'ils ne s'y rendent jamais, car ils accordent de l'importance à son existence même. Voir F. Benhamou et D. Thesmar (2011, p. 33 et suiv.).

Chapitre 3

Le comportement des firmes

L'exemple des entreprises agricoles

1. LE CHOIX DE LA COMBINAISON PRODUCTIVE

2. L'OFFRE DE LA FIRME SUR UN MARCHÉ CONCURRENTIEL

3. INNOVATIONS, CONCURRENCE ET POLITIQUES PUBLIQUES

4. LES CHOIX ORGANISATIONNELS DES PRODUCTEURS

En France métropolitaine, entre 2010 et 2013, le nombre d'exploitations agricoles baisse de 8 %. Cette diminution se poursuit à un rythme annuel moyen proche de celui de la décennie précédente. En revanche, la superficie agricole moyenne des exploitations métropolitaines s'agrandit de 8 %, atteignant 61 hectares en 2013 contre 56 en 2010. Le nombre d'exploitations de grande dimension économique progresse de 9 % par rapport à 2010. Avec 176 000 exploitations elles sont désormais les plus nombreuses que les autres catégories. *A contrario*, le nombre d'exploitations de dimension petite ou moyenne diminue, respectivement – 20 % et – 12 % par rapport à 2010². On assiste donc à un mouvement assez rapide de concentration des terres et d'augmentation de la taille des exploitations, qui ont tendance également à se spécialiser dans une seule production, la part des polycultures diminuant. À travers ces évolutions, il s'agit d'accroître le rendement des terres qui dépend de la taille des exploitations, de la technologie utilisée et du cours des matières premières, souvent très volatils. Les professions agricoles connaissent des crises récurrentes du fait des fortes variations des prix, telle la crise du lait en 2008, ou la crise des prix agricoles en Europe en 2014, à cause de l'embargo russe suite au conflit ukrainien. Une chute des prix peut ainsi conduire certains agriculteurs à la misère alors qu'une hausse brutale de ces derniers peut entraîner la famine des populations dans certaines régions du monde.

Les marchés agricoles sont les prototypes même des marchés concurrentiels, sur lesquels les agents subissent les variations de prix sans avoir de prise sur ces derniers : ils réunissent en effet un très grand nombre de producteurs qui sont **preneurs de prix** (*price takers*), c'est-à-dire que la production décidée par l'un d'eux n'a aucune influence sur le prix d'équilibre du marché. De nombreux produits ont une qualité standardisée, comme les

céréales par exemple, et de ce fait il existe de vastes marchés de biens relativement homogènes : on se trouve alors dans une situation qui s'approche du modèle de **concurrence pure et parfaite** (voir le [chapitre 1](#)).

L'objet de ce chapitre est de présenter la façon dont la microéconomie a modélisé le comportement des entreprises en situation de concurrence et de voir comment ces modèles peuvent être utilisés pour comprendre certains aspects du comportement des entreprises, en privilégiant l'exemple des entreprises agricoles.

1. Le choix de la combinaison productive

Si les exploitants agricoles ne peuvent pas influencer le prix de vente de leur production, ils décident en revanche des quantités produites et de la façon dont ils mettent en œuvre la production. Leur but est de minimiser les coûts de production pour atteindre un volume de production donné. La connaissance des coûts est donc déterminante pour étudier l'offre de biens sur un marché concurrentiel.

1.1 Les coûts de production

• Définition

Les **coûts de production** désignent l'ensemble des dépenses que doit effectuer un producteur pour réaliser une certaine quantité d'un bien ou d'un service. Ils incluent les coûts générés par les facteurs de production (paiement des salaires, achat de machines, etc.) et l'achat de biens de consommation intermédiaire, terme qui s'applique aux biens non durables utilisés au cours du processus de production, comme le fourrage pour un éleveur de bétail. Les coûts sont dits **coûts explicites** quand ils entraînent des dépenses effectives, comme l'achat d'une machine agricole par exemple. Les économistes considèrent qu'un agent rationnel ne tient pas seulement compte des dépenses effectivement réalisées, mais intègre aussi dans le calcul des coûts de production les **coûts implicites**, c'est-à-dire les ressources potentielles auxquelles il renonce en faisant un choix productif. Le coût de production comprend donc le **coût d'opportunité** (voir le [chapitre 1](#)) entraîné par la décision de produire : un agriculteur qui possède un terrain renonce ainsi à le louer s'il décide de l'utiliser pour produire du blé. Le coût implicite de l'utilisation de sa terre correspond dans ce cas au revenu locatif auquel l'agriculteur aurait pu prétendre.

Le **document 3.1** fournit une illustration du calcul du coût moyen de production par kilo de viande bovine. Le **coût moyen** est le rapport entre le coût total et la quantité produite, c'est donc le coût par unité réalisée.

Les « coûts de production en charges réelles » indiqués dans le tableau correspondent aux coûts explicites engagés par les éleveurs, le fourrage a bien été acheté par exemple dans le but de nourrir le bétail. L'éleveur subit également des coûts implicites, dans la mesure où il a renoncé pour s'occuper de ses bêtes au revenu qu'il pourrait gagner en proposant son travail sur le marché. Il faut donc évaluer le coût d'opportunité entraîné par cette perte potentielle de revenu, et aussi le coût d'opportunité de ses capitaux propres et de ses terres qui auraient pu être utilisés à d'autres usages. Dans le modèle de concurrence pure et parfaite, on fait l'hypothèse que les facteurs de production sont rémunérés à leur prix de marché (voir [chapitre 6](#)). L'hypothèse retenue dans le **document 3.1** est celle d'un coût horaire du travail de l'agriculteur de 1,5 SMIC, pour estimer par ce biais le coût implicite de son travail : s'il a besoin de travailler 50 heures par semaine pour élever 100 têtes de bétail, il faut donc associer à cette quantité produite un coût implicite de $50 \times 11,25$ soit 562,5 € (avec un SMIC horaire net de 7,5 € heure, donc 11,25 € pour 1,5 SMIC). La production de 100 unités lui coûte bien la renonciation à ce revenu hebdomadaire.

Document 3.1 : Coûts moyens de viande bovine en 2008

Aliments concentrés achetés et produits	0,7
+ Charges sur les surfaces fourragères	0,4
+ Frais vétérinaires et autres frais d'élevage	0,5
+ Charges de mécanisation	1,2
+ Charges de bâtiments	0,3
+ Frais généraux	0,8
+ Frais financiers	0,2
= Total coût de production en charges réelles	4,1
<i>Charges supplétives (rémunération des facteurs auto-fournis) :</i>	
+ Rémunérations forfaitaires des terres en propriété (base retenue : fermages locaux)	0,0
+ Rémunérations forfaitaires des capitaux propres (base retenue : taux	0,2

d'intérêt de 1,5 %)	
+ Rémunérations forfaitaires du travail non salarié (base retenue : 1,5 fois SMIC)	1,0
= Total coût de production y compris rémunération des facteurs auto-fournis	5,3
Produits bovins viandes (ventes – achats d'animaux + variation stock animaux)	3,2
+ Aides	1,5
= Total produit	4,7

Données : valeur moyenne en euros par kilo de carcasse (jeune Charolais).

Source : FranceAgrimer (2010), « Prix et marge dans la filière bovine », Rapport préliminaire, données issues de l'Institut de l'élevage.

Si les entreprises achètent la totalité des facteurs de production dont elles ont besoin sur le marché, les coûts de production sont tous des coûts explicites, mais si en revanche une partie des facteurs de production est fournie par l'entrepreneur, comme c'est généralement le cas des agriculteurs, alors l'intégration des coûts implicites est indispensable.

Le **coût marginal** de production désigne la variation du coût de production entraînée par la réalisation d'une unité supplémentaire. Combien doit dépenser en plus un éleveur qui désire produire une tête de bétail supplémentaire ? Un coût marginal de 4 € le kilo pour la 100^e tête de bétail signifie ainsi que le coût au kilo d'une bête supplémentaire est de 4 € quand il augmente son cheptel de 99 à 100 unités³.

On distingue généralement les **coûts fixes**, comme par exemple les charges liées aux bâtiments ou les frais généraux, des **coûts variables** qui dépendent du volume de production : pour élever deux fois plus de bovins, il faut deux fois plus de fourrage ou de soins vétérinaires. Le **coût total** est égal à la somme des coûts fixes et des coûts variables. La distinction entre coûts variables et coûts fixes n'a de sens qu'à court terme : le producteur doit faire face à des contraintes qui l'empêchent d'ajuster immédiatement une partie de son outil de production. Par exemple, sur une période donnée, il produit avec un volume donné de surface cultivable, des bâtiments d'une certaine

taille, etc. Mais à long terme ces rigidités n'existent plus, il peut construire des bâtiments, acheter et vendre de la terre si bien qu'il n'existe plus de coûts fixes.

Certaines dépenses sont dites récupérables s'il est possible pour le producteur de revendre les actifs achetés : un agriculteur peut ainsi revendre une machine agricole s'il n'en a plus l'utilité. D'autres dépenses sont en revanche irréversibles : on parle de **coûts irrécupérables**. Imaginons par exemple qu'un agriculteur ait dépensé 1 000 € pour rechercher un nouveau fournisseur d'engrais : il ne peut pas « récupérer » cette dépense en revendant à un autre cette recherche d'information. Qu'il décide de produire ou non, elle restera à sa charge. Les coûts irrécupérables ne sont intégrés dans le calcul des coûts de production qu'au moment où ils sont engagés. Ainsi, toutes les dépenses irrécupérables effectuées dans le passé ne sont pas prises en compte dans les coûts de productions alors que le coût d'un bien d'équipement acheté dans le passé l'est car le producteur a la possibilité de le revendre et renonce donc à cette somme en le gardant.

• La forme des coûts de production

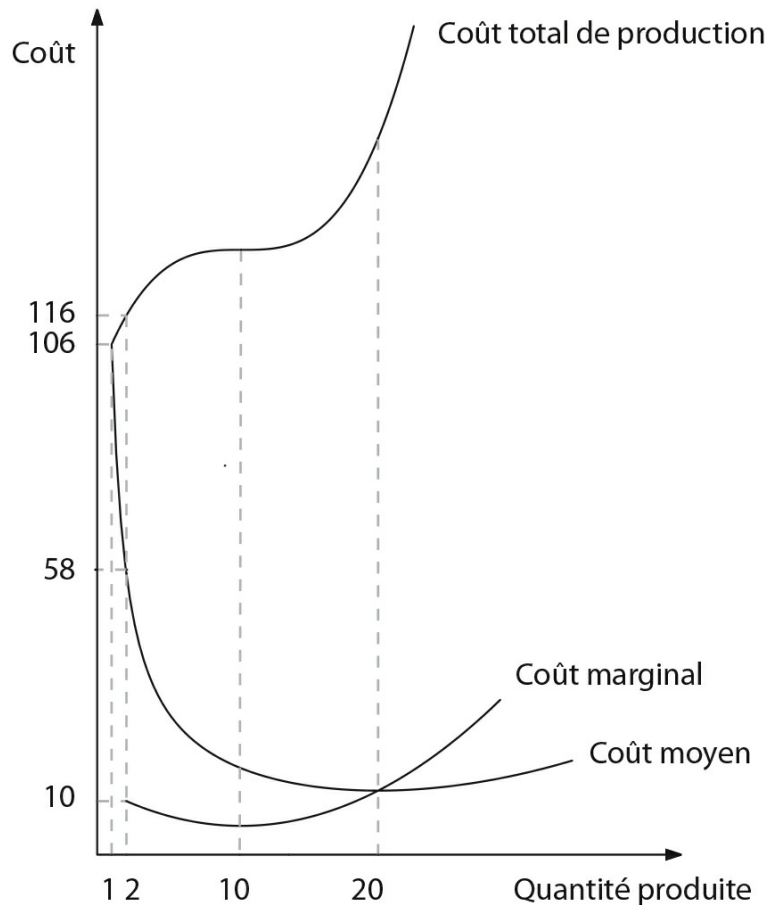
Si les producteurs cherchent le niveau de profit maximal, ils ont intérêt à minimiser les coûts de production afin que le coût moyen soit le plus faible possible. En situation de concurrence parfaite, le prix est déterminé sur le marché. Pour un prix de vente donné, plus le coût d'une unité produite est faible, plus le profit de l'entreprise est élevé.

Sur une période donnée, les producteurs disposent d'un certain savoir-faire et de certaines machines qu'ils ne peuvent pas facilement faire évoluer. Dans ces conditions, ils déterminent la somme minimale qu'ils doivent dépenser pour chaque niveau de production. On parle d'un **coût total de court terme**, car à long terme les producteurs peuvent changer leur équipement et adopter de nouvelles technologies de production. À partir des coûts totaux, il est possible de déduire la valeur du coût moyen et du coût marginal pour chaque niveau de production quand les producteurs s'organisent au mieux pour limiter leurs coûts.

Considérons un autre exemple, celui d'un petit agriculteur produisant des tomates, vendues par plateaux, ayant la structure de coûts hebdomadaires représentée sur le [graphique 3.1](#).

Le coût total du premier plateau est de 106 euros alors que celui de deux unités est de 116 euros. La production de la seconde unité a ainsi obligé le producteur à dépenser 10 euros de plus. Le coût marginal du deuxième plateau est donc de 10 euros. Le coût moyen d'un plateau s'obtient en divisant le coût total par la quantité produite. La **courbe de coût marginal** et la **courbe de coût moyen** s'obtiennent ainsi à partir de la courbe de coût total. Ainsi le coût moyen du premier plateau coûte 106 euros alors que le coût moyen de deux plateaux est de 58 euros ($116/2$). Les coûts fixes de production sont, dans cet exemple, très faibles, ce qui est le cas pour les petits producteurs qui n'investissent pas dans la construction de serres ou dans l'achat de machines. Dans certains secteurs, les coûts fixes sont très importants et la production de la première unité nécessite des dépenses très élevées. Le coût marginal est donc décroissant jusqu'à 10 plateaux puis devient croissant. Pour de faibles niveaux de production, ici inférieurs à 10, le producteur gagne donc en efficacité à mesure que le volume de la production augmente. Notre agriculteur utilise ainsi plus d'engrais et d'eau pour produire davantage sur sa parcelle de terrain, il doit dépenser des sommes supplémentaires en travail et en engrais quand il veut produire davantage. Cependant, le surcoût lié à chaque unité peut être décroissant car il organise plus efficacement ses déplacements sur son terrain, met en place un système d'irrigation plus adapté, etc. À partir d'un certain niveau de production toutefois, le coût marginal devient croissant. C'est lié à un effet de saturation. Si l'agriculteur dispose d'une surface fixe de terrain, à partir d'un moment augmenter la production devient de plus en plus coûteux car on ne peut pas espérer produire toujours davantage sur une surface donnée : ajouter de l'engrais et de l'eau ne sert plus à rien au-delà d'une certaine quantité et peut même s'avérer nuisible.

Graphique 3.1 : Coût total, coût moyen et coût marginal de court terme

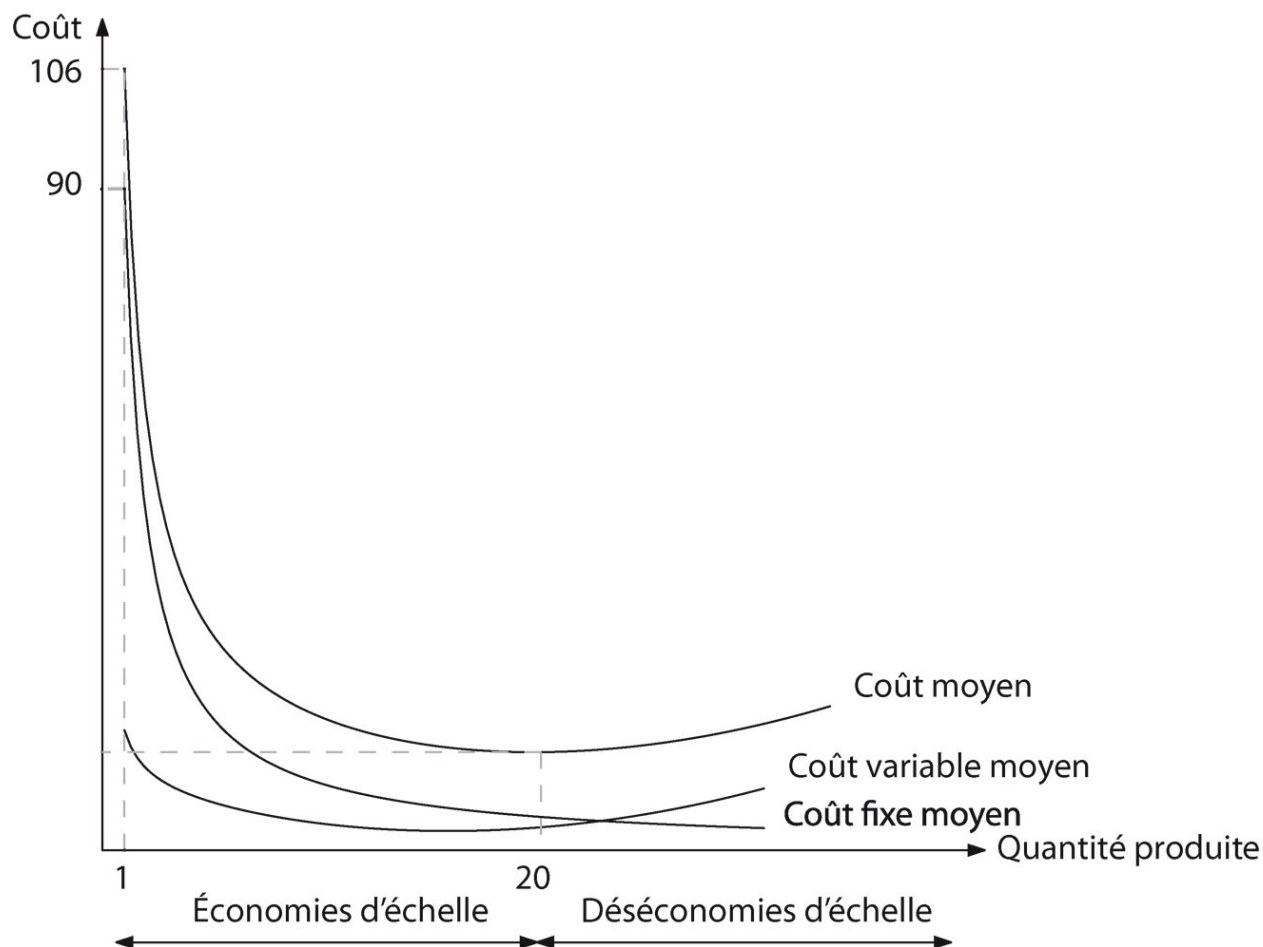


Le processus observé ici est très général, il repose sur ce que les économistes appellent la **loi des rendements factoriels décroissants** (ou rendements marginaux décroissants) : l'augmentation d'un seul facteur de production, comme le travail par exemple, la quantité des autres facteurs restant constante, entraîne une hausse de la production de moins en moins importante, au-delà d'un certain seuil. Autrement dit, la **productivité marginale du travail**, c'est-à-dire le surcroît de production entraîné par une unité supplémentaire de travail, a tendance à augmenter de moins en moins vite, car le travail utilisé sur une exploitation agricole, avec un volume de terre et de bâtiments donné, perd progressivement de son efficacité. La productivité d'un facteur ne dépend pas de ses seules caractéristiques mais aussi de la façon dont il est associé aux autres facteurs. Cette loi permet d'expliquer qu'à court terme le coût marginal soit croissant au-dessus d'un volume de production donné, les biens d'équipements et la terre dont disposent un producteur étant à court terme en quantités fixes, seuls certains facteurs sont variables, comme le travail.

Quand le coût marginal diminue, le coût moyen diminue lui aussi. Ce dernier continue à décroître tant que la production d'une unité supplémentaire coûte moins que le coût moyen des unités précédentes. En effet, si par exemple le coût marginal de la 10^e unité est de 5 € tandis que le coût moyen de 9 unités est de 20 €, la production d'une 11^e unité fait baisser le coût moyen. À partir du moment où le coût marginal de production devient supérieur au coût moyen de production, ce dernier va devenir croissant avec le volume de production puisque chaque unité supplémentaire coûte plus que la moyenne des coûts des unités précédentes. Voilà pourquoi la courbe de coût marginal coupe forcément la courbe de coût moyen en son minimum, ici pour une production de 20 plateaux de tomates.

Quand le coût moyen est décroissant en fonction du volume de production, on dit que l'entreprise réalise des **économies d'échelle**. À l'inverse, si le coût moyen augmente avec la quantité produite, on parle de **déséconomies d'échelle**.

Graphique 3.2 : Les courbes de coût moyen, coût fixe moyen et coût variable moyen



Le coût moyen total est égal à la somme des coûts fixes moyens et des coûts variables moyens. La **courbe de coût moyen** correspondant à la situation de l'agriculteur se décompose entre la courbe de **coût variable moyen** et la courbe de **coût fixe moyen** comme on le voit sur le [graphique 3.2](#).

La courbe de **coût fixe moyen** est forcément décroissante : quand la quantité produite augmente, les coûts fixes sont répartis entre une plus grande quantité de biens. Nous considérons ici un coût fixe de 90 euros, ce qui est très faible et peut correspondre à l'achat d'un matériel minimal et la location hebdomadaire d'une petite parcelle de terrain par un petit producteur.

La forme de la **courbe de coût variable moyen** dépend de l'évolution des coûts variables avec la quantité produite. Si l'agriculteur avait pour seul coût variable le temps de cueillette des tomates, alors le coût variable moyen serait constant car le temps de cueillette nécessaire est proportionnel au volume de tomates réalisé, et il serait donc représenté par une droite horizontale. Mais la

courbe de coût variable moyen a généralement une « forme en U », comme sur le [graphique 3.2](#), car le coût variable moyen change avec la quantité produite. Il décroît dans un premier temps, à mesure que l'on produit dans des conditions plus efficaces, et le coût variable moyen croit au contraire au-delà d'un certain niveau de production quand il devient de plus en plus difficile d'augmenter la production. Le producteur peut ainsi rationaliser sa façon d'arroser et de planter quand il augmente son volume de production ce qui conduit à des gains d'efficacité. En revanche, au-delà d'un certain seuil, il devient de plus en plus difficile de produire plus pour une surface de terrain donné et sans investissement dans des machines. Le graphique montre qu'au-dessous de 20 plateaux, l'éleveur réalise des économies d'échelle car son coût moyen de production est décroissant. Ces dernières découlent à la fois de l'existence de coûts fixes et de la baisse des coûts variables moyens.

La théorie du producteur développé dans le cadre du modèle de concurrence pure et parfaite ne s'intéresse qu'aux situations où il existe des déséconomies d'échelle au-delà d'un certain seuil. En effet, si les entreprises bénéficiaient d'économies d'échelle pour tous les volumes de production envisagés, le marché tendrait à se concentrer et il ne resterait qu'un petit nombre d'entreprises, voire même une seule (cas de **monopole naturel**) : le marché ne serait plus concurrentiel. Le chapitre 4 présente les modèles qui ont été développés pour rendre compte du comportement des firmes dans un cadre non concurrentiel.

1.2 Le choix de la combinaison productive

Nous venons de présenter les courbes de coûts qui indiquent pour chaque volume de production, le coût de production associé. Si nous avons expliqué la forme générale de ces courbes, il convient à présent de rentrer dans le détail de l'explication pour comprendre comment se détermine la valeur précise du coût de production pour chaque volume de production. Le producteur cherche à minimiser ses coûts et pour chaque volume de production, le coût indiqué dans la partie précédente correspond en réalité au coût minimum pour produire ce volume de biens ou services. Produire, c'est combiner des facteurs de production afin de réaliser un bien économique. La combinaison travail/capital mobilisée pour produire est appelée **combinaison productive**. Le but de cette sous-partie est d'expliquer comment le producteur détermine la combinaison productive qui lui permet de produire

un certain volume de production à moindre coût.

Le modèle néo-classique de base se focalise sur les problèmes de choix de la combinaison productive selon le prix et la productivité des facteurs. Il sert notamment à mettre en évidence les effets des changements de prix relatifs des facteurs de production sur les décisions des firmes, par exemple l'effet d'une hausse du coût du travail ou du prix du pétrole sur les entreprises agricoles, en faisant l'hypothèse que la firme ne connaît pas de problème d'organisation et tire toujours le meilleur parti des facteurs disponibles. Il y a de nombreux cas où l'organisation interne des firmes n'est pas aussi efficace, mais ces questions sont étudiées séparément (voir la [troisième partie](#)).

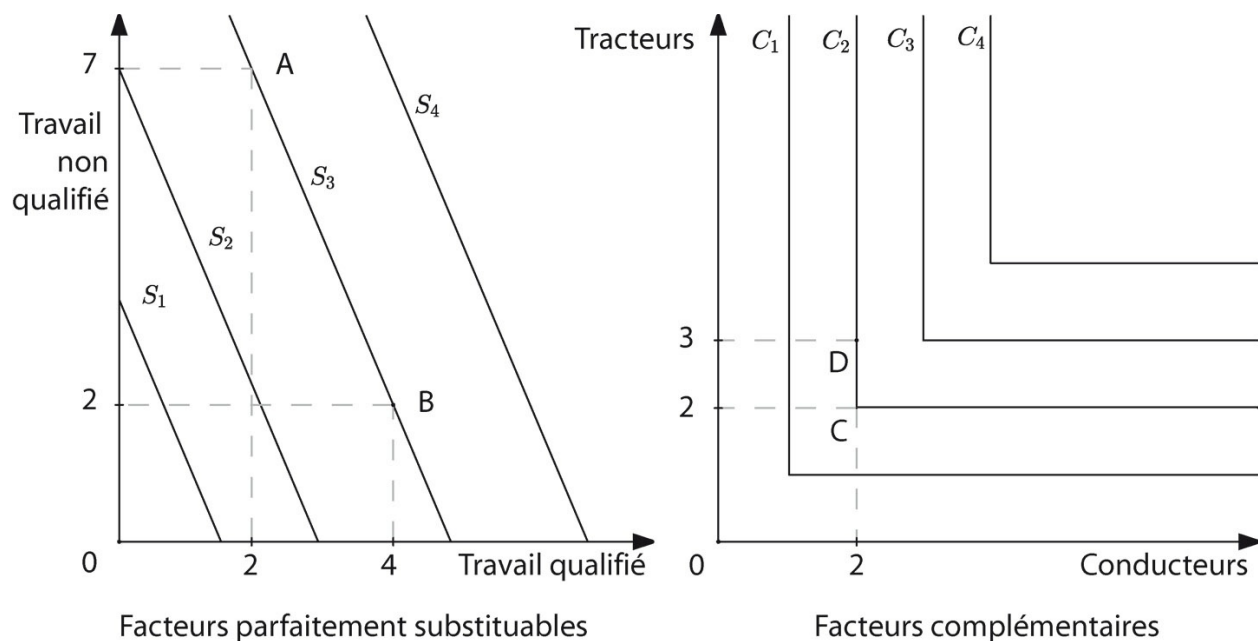
• La fonction de production

Il existe différentes façons de produire une certaine quantité de biens : un agriculteur peut décider de produire des légumes en serre ou en plein champ. On appelle **fonction de production** la relation entre les quantités de facteurs de production utilisés, par exemple le volume de travail (L) et de capital (K) et la quantité de biens réalisés avec ces facteurs (notée Q). On peut distinguer un grand nombre de **facteurs de production** : la production agricole nécessite de la terre, du travail non qualifié et du travail qualifié, divers types de biens capitaux comme les machines agricoles, les bâtiments ou les semences, etc. Mais le plus simple, pour raisonner graphiquement, c'est de retenir seulement deux facteurs, généralement le capital et le travail. Pour représenter la fonction de production, on met sur les deux axes les deux facteurs de production, comme on le voit sur le [graphique 3.3](#). Les courbes tracées représentent chacune un niveau de production et les points d'une même courbe indiquent les différentes combinaisons productives possibles pour atteindre ce niveau de production. Ces courbes s'appellent des **isoquantes (ou courbe d'isoproduit)**. Elles prennent des formes différentes selon l'état de la technologie.

Ce graphique montre deux types de fonctions de production. À droite, les facteurs de production sont **complémentaires** : ils doivent être associés dans une proportion fixe pour être efficaces, augmenter la quantité d'un seul facteur n'a pas d'effet sur la quantité produite. Pour produire la quantité C2, l'exploitant agricole peut recourir à un certain nombre de tracteurs et de conducteurs, mais seule la combinaison E est efficace (2 tracteurs et

2 conducteurs), car il ne sert à rien d'acheter un tracteur supplémentaire s'il n'y a pas aussi un conducteur supplémentaire pour le conduire. Ainsi la production est identique en D (2, 3) et en E. Les isoquantes sont alors en forme d'équerre, et seule la combinaison « en pointe » – comme E – est efficace. Ce cas de figure survient fréquemment quand les contraintes techniques font que les facteurs doivent être associés dans une proportion fixe.

Graphique 3.3 : Fonctions de production à facteurs substituables et complémentaires



À gauche, les facteurs de production sont **parfaitement substituables**. Deux facteurs sont substituables quand il est possible de les combiner dans des proportions diverses pour produire une même quantité de bien (par exemple S3). Supposons que pour vendanger une vigne on dispose de travailleurs qualifiés habitués à cette tâche et de travailleurs non qualifiés, qui débutent. On peut indifféremment utiliser les deux types de travailleurs, qui sont donc substituables, il suffit de faire en sorte que la différence de productivité entre les deux soit compensée par la quantité de facteurs utilisée. Nous pouvons lire sur le graphique qu'il est équivalent en termes de volume de production de disposer de 4 travailleurs qualifiés et de 2 travailleurs non qualifiés ou de 2 travailleurs qualifiés et de 7 travailleurs non qualifiés. Si les facteurs sont parfaitement substituables, la diminution d'un facteur est

compensée par l'augmentation de l'autre dans une proportion constante. La pente de l'isoquante nous indique la quantité de travail qualifié nécessaire pour remplacer une unité de travail non qualifié en moins. La valeur de cette pente est appelée le **taux marginal de substitution technique** (TMST). Il est de ce fait égal au rapport des productivités marginales des facteurs et est constant dans le cas de facteurs parfaitement substituables (dans ce cas l'isoquante est donc une droite). Le cas de facteurs parfaitement substituables est peu courant, et on fait généralement l'hypothèse que les facteurs de production sont partiellement substituables, ce qui donne des isoquantes convexes, comme sur le [graphique 3.4](#).

• Le choix de la combinaison productive et le prix des facteurs

Les coûts de production des exploitations agricoles sont affectés par le volume de facteurs de production utilisé, et par leur prix. Considérons d'abord le facteur travail, qui est acheté à son prix de marché. Pour simplifier l'analyse on retient une seule qualité de travail notée L ci-dessous, achetée à un **taux de salaire** unique noté w , qui correspond au prix d'une unité de travail (voir le [chapitre 6](#)). Pour le déterminer, il faut intégrer non seulement le salaire brut mais aussi les autres éléments qui sont à la charge de l'employeur et vont donc être pris en considération lorsqu'il s'agit d'embaucher, comme les impôts sur les salaires, les cotisations patronales, etc. On parle parfois de **taux de salaire superbrut** pour le distinguer du taux de salaire brut. Une étude de l'INSEE estime le coût horaire du travail en moyenne à 34 euros en France en 2012, pour fixer un ordre de grandeur qui masque de fortes variations selon le secteur et la qualification⁴.

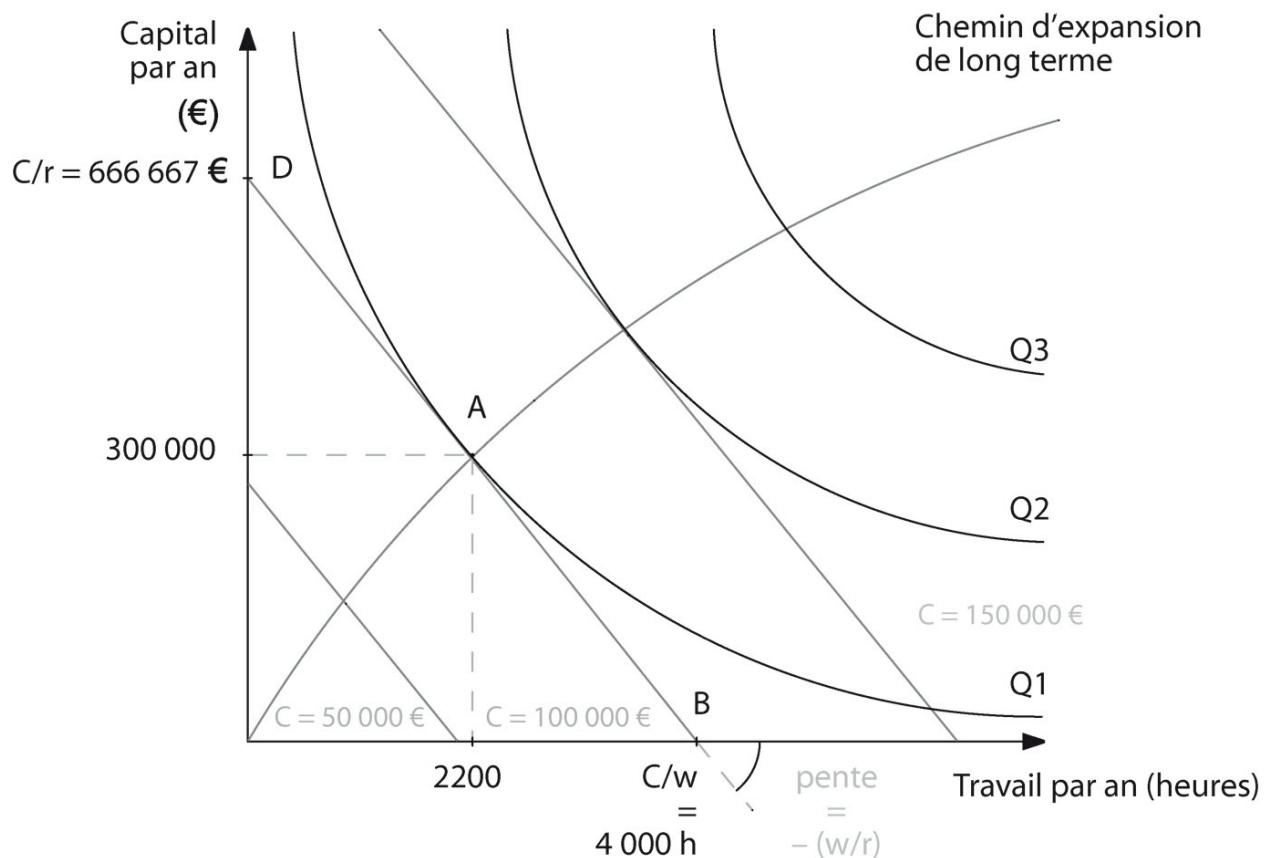
Le prix d'une unité de capital, noté r , est assez complexe à déterminer à première vue, tant le capital se présente sous des formes diversifiées, puisqu'il inclut l'ensemble des biens d'équipements achetés par le producteur (comme un tracteur). Par ailleurs, il faut tenir compte des éventuels intérêts que le producteur doit verser s'il a fait un prêt pour financer ses achats. Mais le raisonnement économique est au fond assez simple, car il ne repose pas sur l'estimation de sa valeur comptable mais sur sa valeur économique. La quantité de capital utilisée est exprimée sous forme monétaire du fait de l'impossibilité d'agréger des biens de nature différente et représente la valeur des biens d'équipement utilisés. Ce montant n'indique pas pour autant le coût

du capital car les machines peuvent être revendues à chaque fin de période et le producteur récupérerait ainsi une partie de la somme engagée initialement. Le coût lié à l'utilisation des machines se mesure en tenant compte de leur perte de valeur, appréciée grâce au **taux de dépréciation**. Il représente la décote à appliquer en cas de revente du capital. À ce coût, il faut ajouter la rémunération à laquelle le producteur renonce s'il avait placé la somme nécessaire à l'achat des machines sur les marchés financiers dans le cas où il avait constitué une épargne pour financer ses achats ou ajouter les intérêts à payer s'il n'avait pas épargné et doit emprunter. Dans les deux cas, le coût est identique et correspond au taux d'intérêt défini sur le marché des fonds prétables. En effet, les marchés étant supposés parfaits, les rendements sont les mêmes partout et le taux auquel le producteur doit emprunter sur les marchés est censé être identique au taux auquel ses placements seraient rémunérés s'il avait de l'épargne. Pour apprécier le prix total d'une unité de capital, il faut aussi tenir compte de nombreux paramètres non développés ici, comme le degré de risque des placements et la fiscalité sur le capital. Pour simplifier, on peut considérer que le prix d'une unité de capital correspond à la somme du taux de dépréciation et du taux d'intérêt, que l'on peut assimiler au rendement moyen du capital que l'on peut observer sur les marchés. Les intérêts sont versés au producteur ou au prêteur selon que le producteur emprunte ou non les fonds nécessaires à l'achat du capital. Le **coût d'usage du capital** correspond de ce fait au produit du prix réel des biens d'équipement par ce taux r qui est en fait un prix de location ou d'immobilisation.

Prenons l'exemple d'une exploitation agricole qui achète 300 000 euros de machines⁵. Ce capital noté K doit d'abord être entretenu, pour conserver sa valeur, et on peut lui appliquer un **taux de dépréciation** pour l'estimer : si un tracteur par exemple a une durée de vie de 15 ans en moyenne il se déprécie d'un quinzième chaque année. Ce capital a ensuite un coût d'opportunité, car l'agriculteur, en revendant son exploitation, pourrait investir ces 300 000 euros et en tirer le rendement moyen sur le marché du capital. Pour fixer les idées, une étude de l'INSEE estime entre 12 et 15 % le coût d'usage du capital en France, sur la période 1984-2002¹. Le profit tel que défini précédemment (différence entre la recette totale et le coût total de production) correspond ainsi à un sur-profit et non à la rémunération du capital qui se calcule à partir du taux r : s'il n'est pas nul, cela signifie que dans le secteur en question, il est possible de dégager des revenus supplémentaires aux

rémunérations moyennes observées pour le travail et le capital dans les autres secteurs (puisqu'elles déterminent le coût du travail et du capital). Afin de trouver la meilleure combinaison productive pour produire une quantité donnée, un producteur de céréales doit tenir compte des contraintes technologiques qui déterminent sa fonction de production, et du coût respectif du travail et du capital. Le [graphique 3.4](#) représente les isoquantes de sa fonction de production, qui associe un certain volume de capital utilisé au cours de l'année, et un volume annuel de travail. Les isoquantes sont convexes parce que l'exploitant a intérêt à utiliser un mélange de travail et de capital : on ne peut espérer faire de l'agriculture sans l'un ou sans l'autre.

**Graphique 3.4 : le choix de la combinaison productive
d'un producteur de céréales**



Le coût d'une combinaison productive s'obtient en multipliant chaque volume de facteur utilisé par son prix. S'il n'y a que deux facteurs de production (travail et capital), il se calcule donc ainsi : coût = $wL + rK$. La **droite d'isocoût** relie toutes les combinaisons productives présentant le même coût de production. Pour chaque niveau de production, il reste à trouver la combinaison productive qui permet d'atteindre ce niveau à moindre coût en cherchant celle qui se trouve sur la droite d'isocoût la plus proche de l'origine, donc avec le coût le plus faible. Pour chaque niveau de production, on recherche la somme minimale à dépenser pour la produire. Cette somme dépend du coût des facteurs et de la technologie du producteur (qui dicte les quantités nécessaires de facteurs). Ces informations sont en fait nécessaires pour construire la courbe de coût total présenté en début de chapitre.

La pente de la droite d'isocoût est égale au rapport des prix des facteurs : si le taux de salaire augmente le coût du capital restant inchangé, la pente devient plus forte et l'exploitant peut acheter moins de travail. Dans l'hypothèse où le taux de salaire serait de 25 euros, et le coût d'usage du capital de 15 %, le producteur peut au plus acheter pour ce coût $100\,000/25$

= 4 000 heures de travail (ce qui nous donne le point B sur la droite d'isocoût, et $100\,000/0,15 = 666\,667$ euros de capital au maximum (point D).

La combinaison A est celle qui permet de produire la quantité Q_1 à moindre coût. En ce point, on constate que la droite d'isocoût et l'isoquante ont la même pente. Or la pente de la première est égale au rapport des prix des facteurs – (w/r) et la pente de l'isoquante est négative et donnée par le rapport des productivités marginales. Ainsi, quand le producteur choisit la combinaison productive la moins onéreuse possible pour atteindre un certain niveau de production, le rapport des productivités marginales est égal au rapport des prix des facteurs. On peut le comprendre en raisonnant par l'absurde. Supposons que le rapport de la productivité marginale du travail à celle du capital soit égal à 4, alors que le rapport des prix est égal à 2. Le producteur a alors intérêt à changer sa combinaison et à utiliser une unité de plus de travail et une de moins de capital, puisque l'écart des productivités est supérieur à l'écart des prix, il produira davantage par euro dépensé, du coup il ne s'agit pas d'une combinaison optimale.

Les conséquences de cette règle sont importantes, car cela signifie que dans le modèle de concurrence pure et parfaite les **facteurs de production** sont rémunérés à leur **productivité marginale**. L'intérêt de ce modèle est de prédire les effets d'un changement de prix d'un facteur sur les choix productifs. À long terme, le coût du travail a tendance à augmenter plus vite que le coût du capital, dans ces conditions les producteurs modifient leurs choix productifs et utilisent relativement plus de capital que de travail. Graphiquement, la pente des courbes d'isocoût augmente, et la combinaison optimale se déplace en direction du capital sur chaque isoquante. Les applications empiriques sont nombreuses dans le secteur agricole. Un article de J. Cavailhès *et al.* (2011) se propose par exemple d'étudier les conséquences du développement de l'habitat péri-urbain sur les choix productifs. On assiste en effet à une extension rapide de l'habitat à la périphérie des centres urbains, qui accroît le prix du capital foncier : les terres coûtent plus cher à proximité des villes que loin d'elles, car il y a une pression pour en modifier l'usage afin d'y construire. Par exemple, en région toulousaine, le prix des terres diminue de 0,2 à 0,3 % par kilomètre de distance avec les centres urbains. Le modèle prédit que dans ces conditions il devient plus intéressant d'utiliser relativement plus de travail quand le prix relatif du capital augmente, et les observations le confirment. On passe de

0,5 travailleur par hectare à 10 km du centre urbain à 0,2 travailleur par hectare à 50 km. Les agriculteurs utilisent une combinaison bien plus intensive en travail quand le coût de la terre s'élève à proximité des centres urbains.

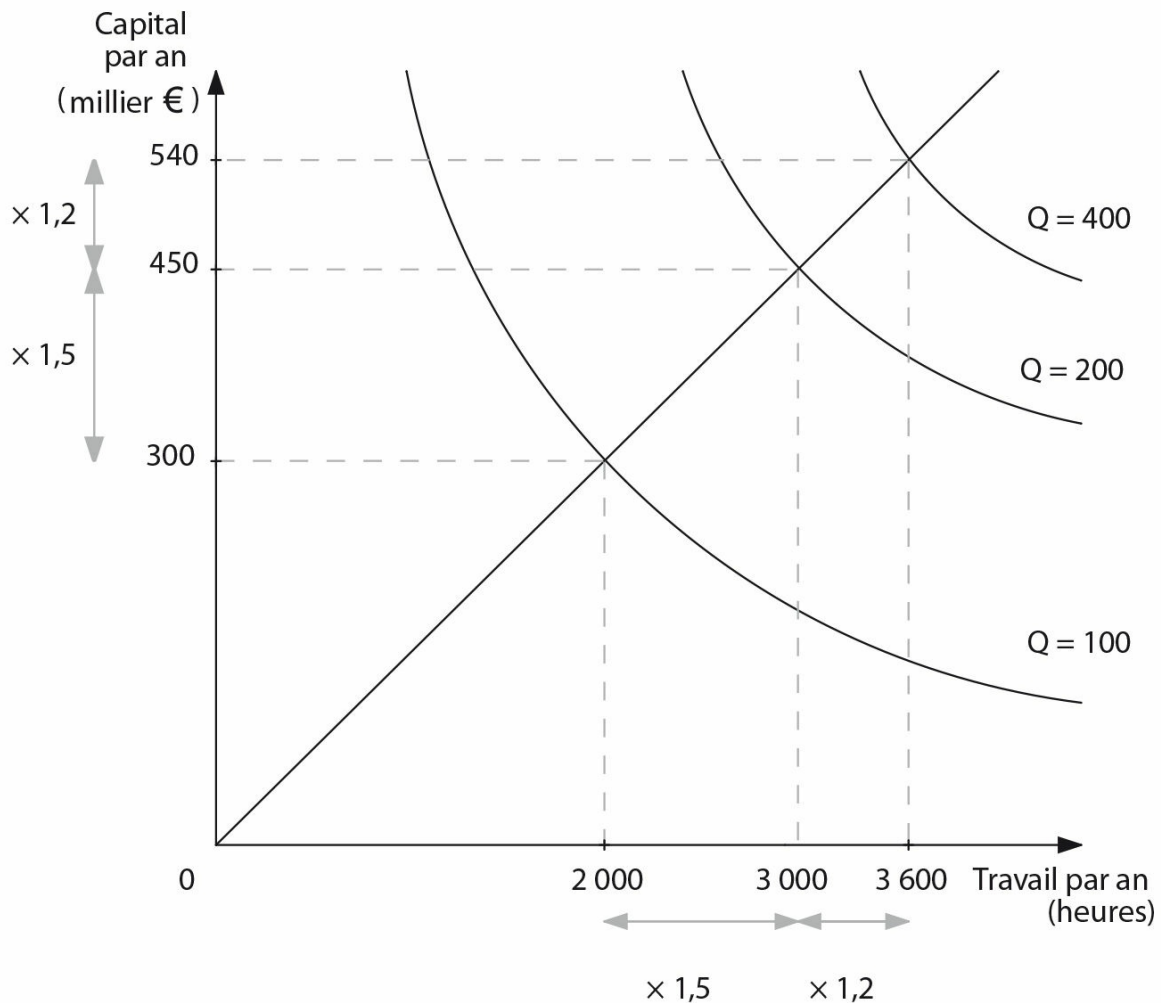
À long terme, quand une firme se développe elle modifie progressivement ses choix productifs en suivant son **chemin d'expansion de long terme**, c'est-à-dire la courbe qui relie les combinaisons optimales quand les coûts envisagés augmentent, comme on le voit sur le [graphique 3.4](#).

• Rendements d'échelle et rendements factoriels

La notion de rendement peut avoir des sens différents en science économique, et il est important de ne pas confondre rendements d'échelle et rendements factoriels, pour comprendre comment évolue l'efficacité productive quand on produit davantage.

Le **rendement factoriel** (ou rendement marginal) et la productivité marginale d'un facteur sont des termes synonymes : il s'agit du supplément de production entraîné par une variation d'une unité de la quantité d'un facteur. Comme indiqué précédemment, la productivité marginale d'un facteur finit toujours par décroître au-delà d'un certain seuil, car on ne peut produire toujours davantage en augmentant la quantité d'un seul facteur, les autres étant fixes. C'est la **loi des rendements décroissants**, qui a une place très particulière dans l'histoire des idées économiques. Elle a pour origine les observations, au XVIII^e siècle, sur le rendement factoriel des terres agricoles : les meilleures terres sont utilisées en premier, par conséquent le rendement factoriel diminue quand on utilise de nouvelles terres, moins faciles à exploiter. Elle sert alors de fondement aux théories de la population et de la croissance de l'école classique anglaise, puis sera un des fondements théoriques de la microéconomie néo-classique, à la fin du XIX^e siècle. Elle a une portée très générale aujourd'hui, puisqu'on considère que les exceptions à la loi des rendements décroissants sont peu nombreuses.

Graphique 3.5 : Les rendements d'échelle croissants



La notion de **rendements d'échelle** est très différente, car elle porte sur l'effet d'une hausse de l'ensemble des facteurs de production, dans la même proportion. Si les volumes de capital et de travail utilisés augmentent conjointement de $x\%$, et que la production augmente de plus de $x\%$, on dit que les rendements d'échelle sont croissants, comme c'est le cas sur le [graphique 3.5](#) qui présente l'exemple d'une entreprise de production laitière : une multiplication par 1,5, puis par 1,2 du volume de chaque facteur se traduit par un doublement de la production (qui passe de $Q = 100$ à $Q = 200$, puis de $Q = 200$ à $Q = 400$). Il s'agit donc spécifiquement d'un effet de taille : un élevage laitier de petite taille rencontre des pertes d'efficacité, augmenter l'échelle de la production permet de mieux utiliser travail et capital, par exemple en limitant les déplacements inutiles, en rationalisant l'usage des bâtiments, etc.

Graphiquement, les rendements d'échelle croissants se traduisent par des isoquantes qui se rapprochent quand la production augmente, et inversement si les rendements d'échelle sont décroissants. En effet, rappelons que sur les axes nous avons les facteurs de production, et sur chaque isoquante la quantité produite : si les isoquantes se rapprochent cela veut dire que la quantité produite augmente « plus vite » que les facteurs de production. Les rendements d'échelle sont donc croissants. À l'inverse, si les isoquantes ont tendance à s'éloigner de plus en plus, cela veut dire qu'en doublant par exemple les facteurs de production utilisés, on ne parvient pas à doubler la quantité produite (l'isoquante correspondant à une quantité double se trouve plus loin sur le graphique), donc les rendements d'échelle sont décroissants. S'il existe des rendements d'échelle croissants en passant d'une petite à une grande exploitation, ils ne se poursuivent généralement pas au-delà d'une certaine taille dans les activités agricoles, car il devient souvent de plus en plus coûteux de concentrer des milliers d'animaux en un seul lieu. Au-delà d'un certain seuil, les rendements d'échelle deviennent donc souvent décroissants dans ce secteur, c'est-à-dire qu'une hausse conjointe du capital et du travail se traduit par une augmentation moins que proportionnelle des quantités produites. Contrairement aux rendements factoriels, il n'y a pas de règle générale et il faut examiner au cas par cas si une firme connaît des rendements d'échelle croissants, constants ou décroissants pour différents niveaux de production.

Il ne faut pas confondre rendements d'échelle croissants et **économies d'échelle**, présentées en amont, même si les deux notions sont liées. Les rendements d'échelle renvoient à une amélioration de l'efficacité technique quand on augmente la taille de la production, la hausse de chaque facteur étant proportionnelle. Dans ces conditions, les rendements d'échelle entraînent forcément des économies d'échelle, qui se traduisent par une baisse du coût unitaire quand la production augmente, puisque les facteurs de production sont plus efficacement utilisés. Mais l'inverse n'est pas vrai : une laiterie peut faire des économies d'échelle alors même que ses rendements d'échelle sont constants, par exemple si la taille permet d'accroître son pouvoir de négociation auprès de ses fournisseurs, et de réduire ses coûts d'approvisionnement. Donc les rendements d'échelle croissants sont un facteur parmi d'autres d'économies d'échelle.

La **productivité globale des facteurs (PGF)** se distingue de la notion de rendement d'échelle. Elle mesure l'efficacité globale de la combinaison

productive. La PGF est souvent assimilée au progrès technique mais elle englobe d'autres facteurs tels que les infrastructures et les institutions. Elle renvoie aussi à l'effet de la variation de l'ensemble des facteurs sur la production mais se calcule différemment. Elle représente la part de l'augmentation du volume de production qui ne peut être expliquée par une augmentation de la quantité des facteurs de production utilisés.

2. L'offre de la firme sur un marché concurrentiel

Le modèle de base de la microéconomie du producteur en situation de concurrence part du principe que les producteurs cherchent à maximiser leur revenu. Le producteur est l'acteur qui met en œuvre la production. Il choisit le type de production qu'il veut réaliser, le volume de production, emploie les travailleurs et se procure les biens d'équipements nécessaires. Le profit désigne au sens strict la partie de la valeur générée par l'activité productive qui va au détenteur du capital. C'est donc un terme plus restrictif que celui de revenu du patrimoine qui englobe tous les revenus tirés par un individu en lien avec la possession d'un actif. Dans la théorie néoclassique le profit est défini comme la différence entre les recettes totales liées à la vente de la production et les coûts totaux de production. Cette définition a pu être considérée comme problématique car elle ne tient pas compte du revenu perçu par les personnes qui ont avancé les fonds nécessaires à la production (les intérêts) qui est ici perçu comme un coût. Il faut donc bien avoir en tête que même si le profit au sens néoclassique est nul, le prêteur (ou le producteur, s'il a mobilisé ses propres fonds) perçoit tout de même un revenu et le capital est donc bien rémunéré. Il y a donc bien du profit (mais pas au sens néo-classique), l'activité productive ne permet simplement pas de générer une somme supplémentaire par rapport à la rémunération du capital à son taux de référence, par exemple sur le marché des fonds prêtables. Quand le profit au sens néoclassique est nul, le capital rapporte le taux d'intérêt à celui qui avait les fonds nécessaires à la production pour acheter le capital, peu importe qu'il provienne d'un emprunt ou des fonds personnels de l'investisseur. C'est pour cette raison qu'on devrait plutôt parler de sur-profit pour qualifier la définition néo-classique du profit.

• Le profit maximal sur un marché

En situation de concurrence, le prix de vente se fixe sur le marché et aucun producteur n'a, au niveau individuel, la possibilité de modifier ce prix. C'est généralement le cas des marchés de matière première comme le blé ou le coton : le prix est déterminé au niveau mondial et les agriculteurs subissent le

prix du marché de façon passive, ce qui est à l'origine du mécontentement des agriculteurs lorsque les prix mondiaux viennent à s'effondrer.

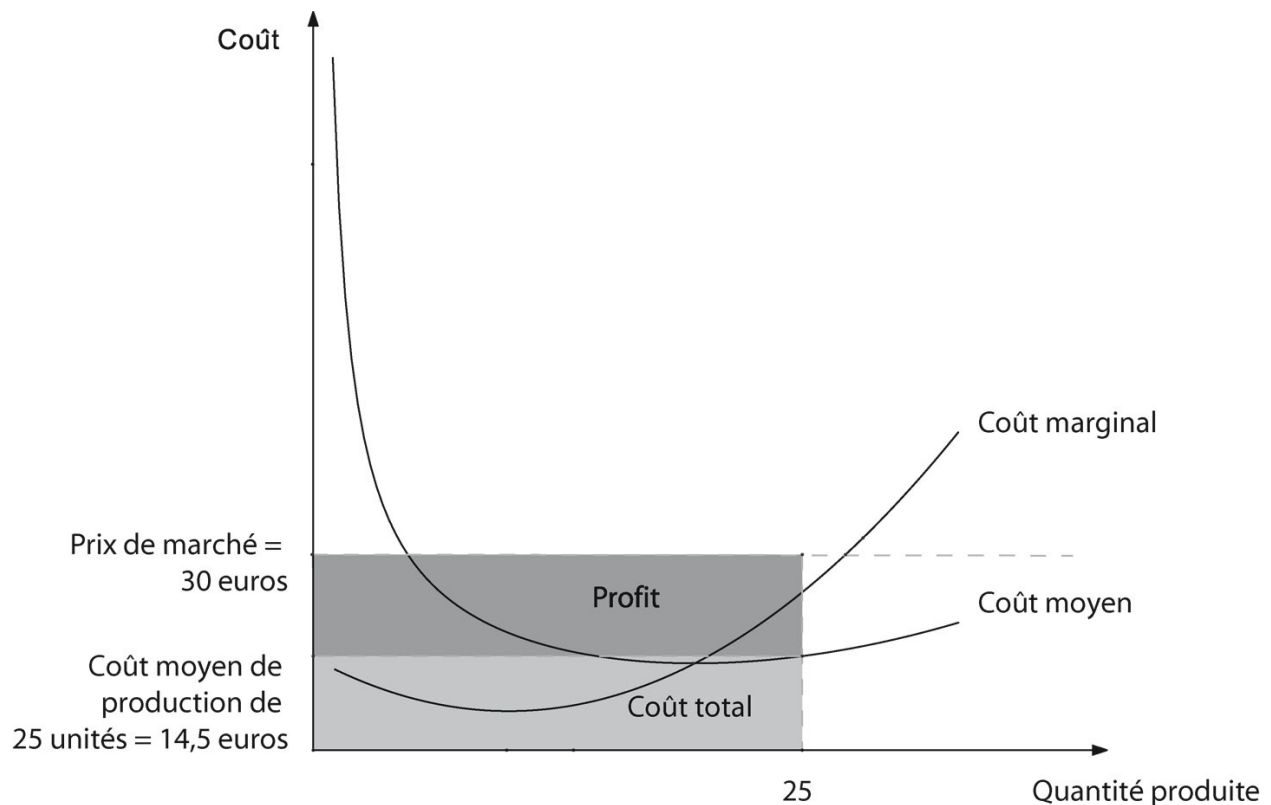
Compte tenu du prix qui se forme sur chaque marché, un producteur rationnel devrait calculer le profit maximal qu'il peut réaliser sur chaque produit et devrait choisir d'opérer sur le marché le plus lucratif. Si la technologie et le savoir-faire d'un producteur peuvent évoluer dans le temps, à court terme la structure de ses coûts est fixe. Le producteur ne peut donc que jouer sur les quantités offertes. Quelle est la quantité qui maximise son profit ?

La **recette totale** d'un producteur est égale à la quantité vendue multipliée par le prix de vente. La **recette moyenne** correspond à la somme reçue en moyenne par le producteur pour chaque unité vendue. Quand un producteur vend toutes ses unités au même prix (prix identique pour tous les clients), la recette moyenne est donc égale au prix du marché. La **recette marginale** désigne le surcroît de recette générée par la vente d'un bien supplémentaire. En situation de concurrence parfaite, la recette marginale est égale au prix : quelle que soit la quantité vendue le prix de vente reste identique et correspond donc au surcroît de recette entraînée par une unité supplémentaire. En situation de concurrence, le prix est donc égal à la recette moyenne et à la recette marginale.

Compte tenu du prix pratiqué sur le marché, le producteur peut en déduire le profit associé à chaque niveau de production.

Dans le cas représenté sur le [graphique 3.6](#), le prix de marché d'un plateau est de 30 euros. Pour chaque quantité produite, le producteur connaît la valeur de son coût moyen. S'il produit 25 unités, elles lui coûteront en moyenne 14,5 euros par unité à produire, le coût total pour 25 unités est donc égal à 362,5 euros ($25 \times 14,5$), ce qui correspond à l'aire gris clair sur le [graphique 3.6](#). Pour un prix de vente de 30 euros, la valeur de la recette totale s'obtient en multipliant la quantité produite par le prix : pour une production de 25 unités, la recette totale est donc égale à 750 euros. Le producteur dégage ainsi un profit égal à 387,5 euros ($750 - 362,5$) lorsqu'il produit 25 unités ce qui correspond à la surface gris foncé sur le graphique.

Graphique 3.6 : Le profit du producteur pour une quantité donnée



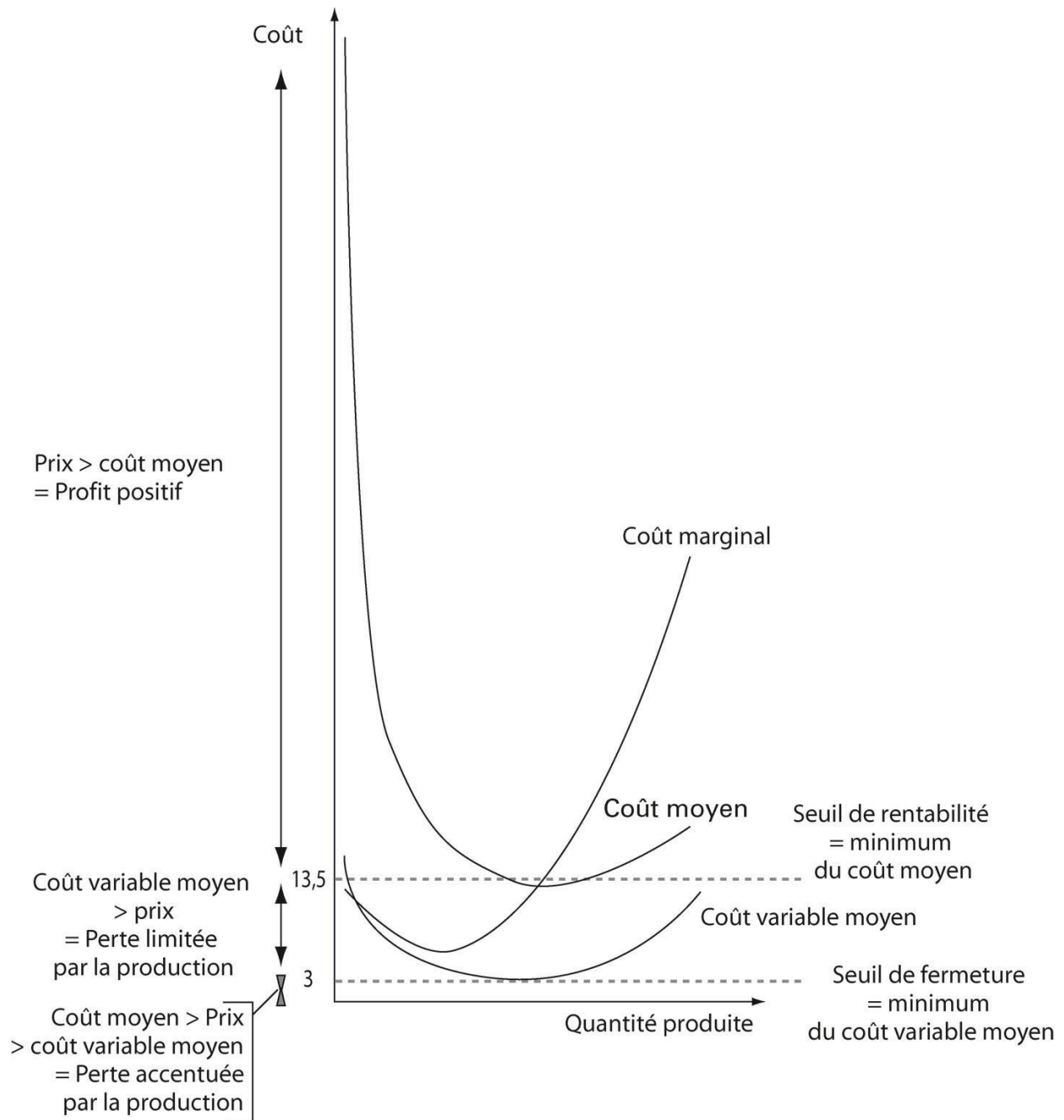
Pour qu'un producteur ne produise pas à perte, il faut que le prix de vente (donc la recette moyenne dégagée) soit supérieur au coût moyen de production. Si le prix de marché se fixe en dessous du minimum du coût moyen, il n'est pas rentable de produire : la production lui coûte plus qu'elle ne lui rapporte quel que soit le volume de production. La valeur du minimum du coût moyen s'appelle le **seuil de rentabilité**. En raison de l'existence de coûts fixes, le producteur a tout de même intérêt à produire à perte, dans certaines conditions. Supposons qu'un agriculteur ait loué une machine agricole pour l'année et qu'il ne puisse résilier son contrat de location. Qu'il produise ou non, il lui faudra payer ce coût. L'agriculteur avait anticipé un prix de vente plus élevé et pensait qu'il serait rentable de produire, celui-ci est plus faible que prévu mais certains achats ne peuvent être annulés. La valeur du coût fixe ne peut être modifiée à court terme et, de ce fait, ne joue pas sur sa décision de produire ou pas à court terme. Les coûts variables qui eux varient par définition en fonction du volume de production vont en revanche jouer dans la décision du producteur. Si le prix de vente est supérieur au coût variable moyen, cela signifie que chaque unité

supplémentaire produite engendre un surcoût de production inférieur au prix de vente. Il est rationnel pour l'agriculteur de produire cette unité bien qu'il produise à perte car il peut réduire cette perte par rapport à celle obtenue en ne produisant pas. Le producteur décide donc de cesser toute production quand le prix est inférieur au minimum du coût variable moyen car le fait de produire augmente ses pertes quelle que soit la quantité réalisée : il s'agit du **seuil de fermeture**, qui apparaît sur le [graphique 3.7](#).

Si le prix de vente se situe au-dessous de 3 euros le producteur va se retirer du marché. S'il est compris entre 3 euros et 13,5 le producteur continuera à court terme à produire mais réalisera des pertes.

Ce raisonnement permet de comprendre la situation des producteurs de tomates lors des crises de surproduction agricole, qui surviennent périodiquement. Durant l'été 2014, avec la conjonction de l'embargo russe et d'une situation météorologique exceptionnelle, les prix des fruits et légumes tombent très nettement en dessous du seuil de rentabilité, et les autorités publiques ont déclenché la procédure de « crise conjoncturelle », qui intervient lorsque les prix tombent à un niveau anormalement bas, pour soutenir les agriculteurs. Une telle situation ne peut pas perdurer, mais les producteurs ne ferment pas pour autant leurs exploitations dès que les prix tombent en dessous du seuil de rentabilité.

La valeur du profit peut être calculée pour chaque volume de production et il est possible de déterminer par comparaison le volume de production qui maximise le profit. Il n'est pourtant pas nécessaire de procéder ainsi. En effet, il existe une façon simple de déterminer la quantité optimale de production que nous allons à présent expliquer. Produire 25 unités (comme nous l'avons envisagé au [graphique 3.6](#)) n'est pas rationnel compte tenu de la structure de ses coûts de production. La production d'une unité supplémentaire lui rapporterait 30 euros alors qu'elle ne lui coûterait que 17 euros à produire (coût marginal de la 26^e unité), donc il est rentable de la réaliser. Tant que le coût de production d'une unité supplémentaire (coût marginal) est inférieur à la recette marginale (ici le prix de vente), il est rationnel pour le producteur de produire une unité supplémentaire. Dès que le coût marginal devient supérieur à la recette marginale, le producteur voit son profit diminuer en produisant davantage. Le volume de production qui maximise le profit du producteur se détermine donc en calculant la quantité qui permet d'**égaliser la recette marginale au coût marginal**.

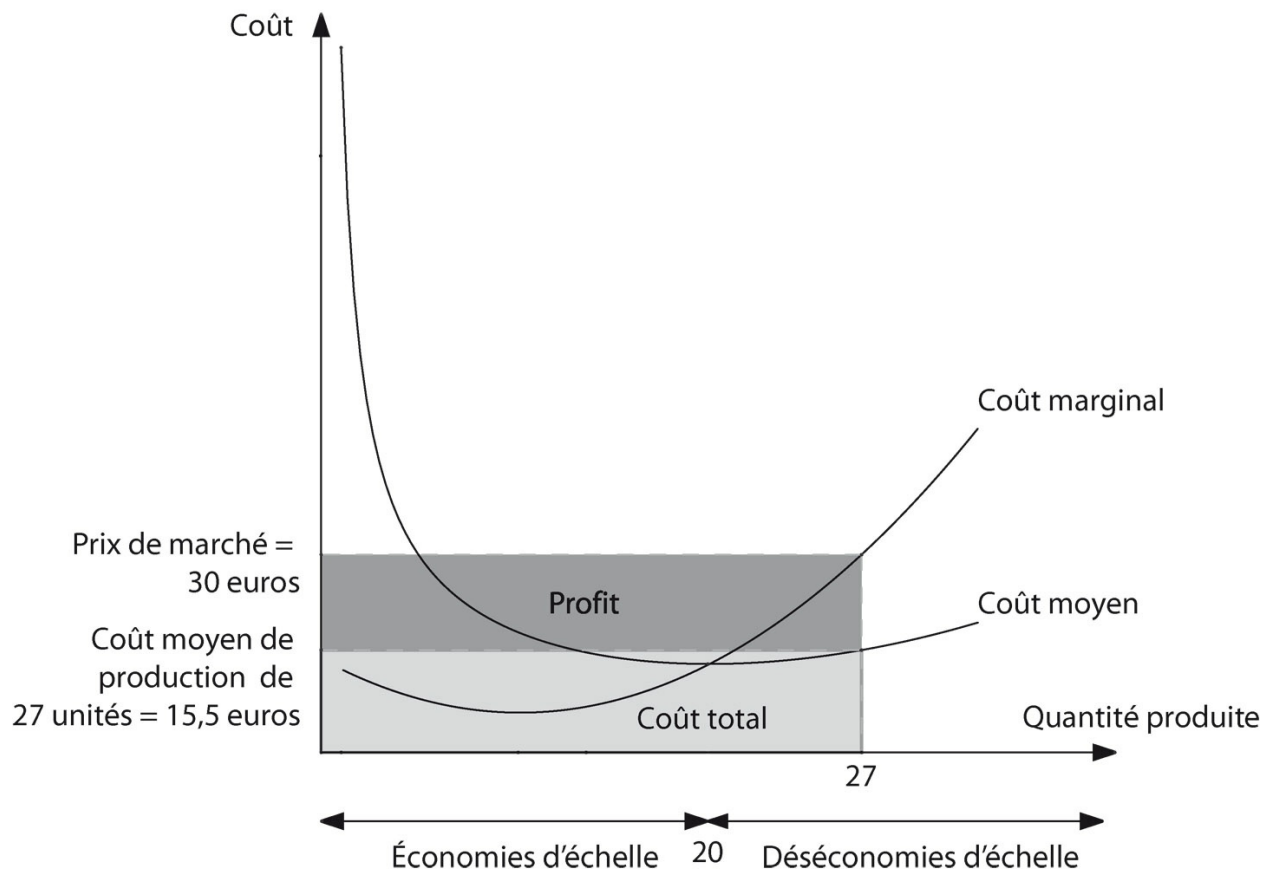


En réalisant 27 unités, le producteur réalise un profit égal à 391 euros ($27 \times 30 - 27 \times 15,5$). Compte tenu du prix du marché et de la structure de ses coûts, il ne peut espérer un profit supérieur sur ce marché.

Que se passe-t-il si le prix de marché diminue ? Logiquement, le

producteur adapte son offre pour que la recette marginale soit toujours égale au coût marginal de production. On comprend donc que la courbe de coût marginal est en fait la **courbe d'offre** de court terme du producteur, puisqu'elle permet de connaître les quantités offertes en fonction du prix de marché, du moment qu'il est supérieur au minimum du coût variable moyen. Dans les faits, les producteurs ne sont pas aussi rationnels que le stipule le modèle et produisent parfois des volumes de production qui ne sont pas optimaux. La valeur du profit est toutefois importante aux yeux des producteurs et ces derniers essaient en général d'adapter au mieux leur production pour tirer le profit le plus grand possible de leur activité.

Graphique 3.8 : La maximisation du profit



L'agrégation des décisions d'offre au niveau individuel permet de construire la **courbe d'offre agrégée** et c'est en fonction de cette dernière et de la courbe de demande que va se fixer le prix d'équilibre. À long terme, la courbe d'offre peut se déplacer. En effet, nous avons jusqu'ici considéré l'offre de court terme uniquement, quand le nombre de producteurs est donné sur un marché : on ne peut s'improviser exploitant agricole « du jour au lendemain », et ceux qui sont déjà présents sur le marché font face à des coûts fixes même s'ils décident d'avoir une production nulle. À court terme, un maraîcher dispose d'un certain équipement et décide simplement de produire plus ou moins de légumes en fonction de leurs prix et de la structure de ses coûts.

La courbe d'offre de long terme du secteur résulte d'un processus très différent et plus complexe. À long terme en effet, un producteur peut modifier le capital fixe dont il dispose, ou décider d'opérer sur un autre marché en produisant un autre légume que la tomate, s'il existe des possibilités plus rentables.

• L'offre d'un producteur à long terme

À long terme, le producteur peut investir et modifier en profondeur son organisation, il n'est donc plus pertinent de distinguer coûts fixes et coûts variables, puisque tous les facteurs de production sont modifiables. Par exemple, un exploitant agricole a l'occasion de vendre ou d'acheter des terres, construire un nouveau bâtiment, etc. À long terme, des investissements permettent de bénéficier d'économies d'échelle supérieures, ce qui réduit le coût moyen minimum de long terme, à un niveau inférieur à celui de court terme. La distinction entre seuil de fermeture et seuil de rentabilité n'a en fait plus de sens à long terme : une firme qui s'attend durablement à des pertes n'a pas intérêt à produire, donc la production à long terme est nulle quand le prix est inférieur au minimum du coût moyen puisque la firme fait durablement des pertes. Si le prix est supérieur à ce seuil, l'**offre de long terme** est donnée par la courbe de coût marginal de long terme, selon une logique similaire à celle observée à court terme.

En règle générale, la courbe d'offre de long terme est plus sensible au prix et donc plus horizontale que la courbe d'offre de court terme (voir le [chapitre 1](#)) : il est plus aisé de modifier le type de production effectué sur la durée (par exemple changer de spécialité agricole) et ainsi de quitter le marché si les perspectives de profit sont trop faibles.

• Le choix du type de bien produit

Les producteurs comparent à long terme les perspectives de profit sur les différents marchés afin de décider du type de bien qu'ils veulent produire. La décision de produire du blé ne va donc pas dépendre uniquement du profit attendu en produisant du blé, mais aussi du profit qu'il pourrait réaliser en produisant du maïs, du soja, etc. Les agents doivent donc intégrer dans leur calcul le **coût d'opportunité** entraîné par le choix de produire un bien plutôt qu'un autre.

Si un secteur apparaît plus profitable qu'un autre, de nouveaux producteurs vont décider d'opérer sur le marché ce qui contribue à faire baisser les prix, donc la rentabilité des entreprises opérant sur ce marché. Dans le modèle de concurrence pure et parfaite, on fait l'hypothèse qu'il n'existe pas de **barrières à l'entrée**, autrement dit que les agents sont libres d'entrer et de sortir sur le marché. Si de nombreux producteurs, au niveau mondial,

décident de produire du maïs car la production de ce type de biens semble être la plus rentable, le prix du maïs va diminuer, à terme une autre production peut devenir à son tour plus rentable, et le processus d'ajustement se poursuit. Au final, par ce mouvement d'entrée et sortie des entreprises, la valeur du profit sur les différents marchés s'égalise et diminue pour tendre vers zéro. En effet, tant qu'il existera une perspective de profit sur un marché, des nouvelles entreprises vont se créer et l'offre va augmenter. L'absence de barrières à l'entrée est rarement observée parfaitement, mais les marchés agricoles laissent une réelle latitude aux producteurs pour changer de production.

Sur un marché, la pente des courbes d'offre et de demande dépend de leurs élasticités respectives (voir le [chapitre 1](#)). L'élasticité-prix de la demande des produits agricoles est généralement assez faible, et donc leur demande rigide par rapport aux prix, car les consommateurs ont besoin de se nourrir, que les prix soient faibles ou élevés, et dans ces conditions, l'ajustement se fait par les prix plus que par les quantités. À court terme, les prix atteignent parfois des niveaux très élevés, suite par exemple à un choc climatique qui a détruit la production d'une partie des agriculteurs, et pousse les producteurs en place à augmenter leur prix. À long terme, l'offre va s'ajuster puisque le niveau élevé de la rentabilité attire de nouveaux producteurs. Mais ce processus d'ajustement nécessite un certain décalage dans le temps. La volatilité des prix met en danger le fonctionnement des marchés où la demande est rigide, comme les marchés agricoles, et peut être à l'origine de crises ou de famines. Elle nécessite fréquemment une régulation publique, par exemple la constitution de stocks pour limiter les tensions sur le marché. Elle dépend plus largement des **institutions publiques** mises en place car les famines ne naissent pas tant des aléas climatiques que de la passivité des pouvoirs politiques à lutter contre elles : selon l'économiste indien A. Sen (2003), un environnement institutionnel démocratique est nécessaire pour éviter les famines, en permettant un débat public sur la nécessité de réguler les marchés agricoles quand une crise émerge. Les facteurs institutionnels jouent ainsi un rôle décisif dans le bon fonctionnement des marchés.

• L'efficacité allocative d'un marché concurrentiel

Comme nous venons de le voir, à long terme, la rentabilité des entreprises opérant sur les marchés concurrentiels devrait être nulle. Un exploitant

agricole qui réalise un profit nul n'a pas de raison de cesser de produire, car l'absence de profit ne signifie pas que ses efforts ne sont pas rémunérés. Il est simplement en situation de payer le travail qu'il fournit et le capital qu'il apporte à leur prix de marché (ou de payer les intérêts s'il a emprunté les fonds pour acquérir le capital nécessaire), le paiement des facteurs étant intégré au calcul du coût de production. Il tire le même rendement de ses facteurs que s'il les utilisait dans d'autres activités, comme par exemple un placement sur le marché du capital, ou bien en menant une activité salariée.

Dans les conditions de la concurrence, un marché en concurrence pure et parfaite utilise au mieux les facteurs de production, puisqu'il permet de les rémunérer à leur prix de marché, et pas davantage. Il n'y a pas de phénomène de rente. Les producteurs moins efficaces que les autres sont évincés du marché. En effet, la baisse tendancielle du prix de vente des biens conduit à ce que seules les entreprises à faibles coûts de production ont intérêt à produire à ce prix, ce qui élimine peu à peu les entreprises disposant de structures de coûts non compétitives. Dans le modèle néo-classique de base, toutes les entreprises ont la même technologie mais on peut sans difficulté introduire de l'hétérogénéité dans les technologies de production avec des entreprises moins performantes que d'autres, ces dernières se faisant simplement éliminer à terme par le jeu de la concurrence ne pouvant être compétitive. Les producteurs restants sur le marché ne captent pas de revenus supplémentaires au minimum nécessaire pour rester sur le marché. Les consommateurs paient les biens à leur coût marginal, et ils ne peuvent espérer les payer moins cher car c'est le seuil minimal pour que tous les exploitants présents restent sur le marché. Si le prix descendait encore en dessous du prix d'équilibre de long terme, il serait inférieur au coût marginal et les exploitants réduiraient leur production. Un marché de concurrence pure et parfaite permet donc d'atteindre la meilleure **efficacité allocative** possible : la quantité produite ne peut augmenter avec les facteurs disponibles, sans réduire la production d'autres biens.

Pour cette raison, les autorités publiques cherchent généralement à renforcer la concurrence sur les marchés. Si les marchés observés ne sont jamais parfaitement concurrentiels, renforcer la concurrence permet généralement de réduire le profit des entreprises et de diminuer les prix des biens. Cependant, si les imperfections du marché sont importantes, laisser librement jouer le marché peut au contraire s'avérer collectivement inefficace (voir le [chapitre 4](#)). Les entreprises cherchent de leur côté à éviter la

concurrence : construire des barrières à l'entrée des marchés permet de se prémunir de l'arrivée de nouveaux concurrents par exemple. Les entreprises cherchent aussi à adopter des modes de production plus performants que celui des autres pour bénéficier de coûts plus faibles que leurs concurrentes et ainsi disposer d'un profit à court terme alors que les autres dégagent un profit nul. À long terme, les **innovations** technologiques se diffusent et les entreprises se livrent ainsi à une course perpétuelle à l'innovation, qu'elle porte sur le type de bien produit ou la technologie utilisée, pour disposer d'un profit positif. En situation de concurrence imparfaite, certains producteurs réussissent donc à dégager un profit confortable alors que d'autres sont au bord de la faillite, dans les mêmes conditions de prix. Dans le domaine agricole, une étude de l'INSEE (2007) montre que les écarts de revenus sont non seulement importants selon les spécialisations agricoles, mais surtout à l'intérieur d'une même spécialisation. Les écarts les plus importants se trouvent dans la viticulture et la production de fruits, là où les différences de qualité des facteurs sont les plus fortes (un coteau parfaitement ensoleillé, une terre qui permet d'obtenir un certain label de qualité) tandis qu'à l'inverse les écarts de revenus sont bien moindres chez les producteurs laitiers, où les conditions de production sont plus homogènes.

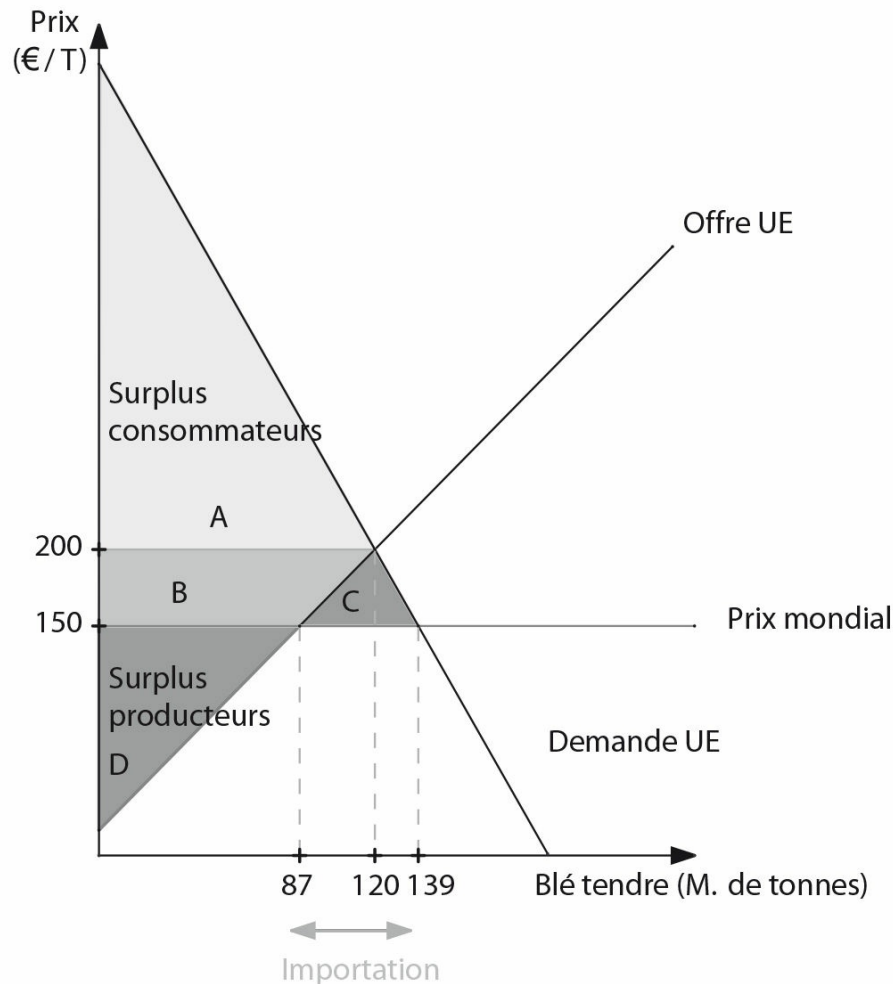
• La subvention de la production

Dans les pays développés, la production de biens agricoles est fréquemment subventionnée. La principale raison tient au fait que les coûts de production dans les pays développés sont en général plus élevés qu'ailleurs : l'élévation du niveau de vie augmente la rémunération des facteurs, dans ces conditions les producteurs des pays développés se feraient éliminer par le mécanisme de la concurrence s'ils vendaient aux prix du marché mondial.

Comment prédire les effets de l'ouverture à la concurrence internationale ? Considérons un exemple simplifié, à propos du marché européen. On suppose qu'en situation d'autarcie, en l'absence de commerce avec des pays non membres de l'Union Européenne, le prix du blé tendre en Europe se fixe à 200 euros la tonne, compte tenu de l'offre des producteurs européens et de la demande des consommateurs de la zone. En s'ouvrant au commerce mondial, le prix du blé diminue pour se fixer à 150 euros la tonne, car les producteurs étrangers ont la capacité de proposer des prix plus compétitifs que les producteurs européens. Les producteurs européens n'ayant pas,

individuellement, la possibilité d'influencer les prix de marché, ils se voient contraints de vendre leur production à ce prix, les consommateurs n'achetant pas leurs produits s'ils proposent des prix plus élevés que leurs concurrents pour le même produit.

Graphique 3.9 : L'équilibre du marché européen sans subventions



Nous pouvons voir sur le [graphique 3.9](#) qu'en situation de libre-échange, les agriculteurs produisent moins qu'en économie fermée et vendent leur production à un prix inférieur. En effet, en autarcie, ils produisaient 120 millions de tonnes de blé contre 87 millions en situation de libre-échange. Les consommateurs en revanche demandent davantage de biens que par le passé du fait de la baisse des prix (139 millions de tonnes en situation de libre-échange contre 120 millions en autarcie) et de la possibilité d'importer la production manquante (ici 52 millions de tonnes). La diminution des quantités produites et des prix se traduit par une perte de surplus du producteur donnée par la surface B (voir le [chapitre 1](#) pour l'interprétation du surplus). En revanche, les consommateurs sont gagnants ce qui se traduit par une hausse de leur surplus de A en économie fermée, à A + B + C en économie ouverte.

Ainsi, l'ouverture complète du marché agricole européen à la concurrence,

sans intervention publique, se traduirait par une forte hausse des importations pour répondre à la demande au prix mondial, par une baisse sensible du revenu des agriculteurs européens mais elle bénéficierait en revanche aux consommateurs.

La mise en place de la politique agricole commune (PAC) repose sur la volonté de maintenir une activité agricole substantielle en Europe, et de protéger les revenus des agriculteurs. La politique agricole commune (PAC) est donc un des piliers de la construction européenne, depuis les années 1960. Elle s'est progressivement transformée, en passant d'une politique de soutien direct aux prix à une politique d'aides directes aux agriculteurs. Pour comprendre les raisons de ces transformations, il faut souligner que les différents instruments disponibles n'ont pas le même impact pour les différents agents économiques (BUREAU, 2007). La PAC a longtemps reposé sur des **prix garantis**, c'est-à-dire des prix-planchers maintenus grâce à une subvention des pouvoirs publics, par hypothèse au-dessus du prix d'équilibre européen en l'absence d'intervention. À ce niveau de prix, l'offre européenne excède la demande européenne. Ce système d'aide est très coûteux pour les autorités, diminue le surplus des consommateurs qui voient les prix augmenter et implique que l'excès d'offre doit être retiré du marché intérieur, sous une forme ou sous une autre, pour stabiliser les prix au niveau du prix garanti. Selon les périodes, la PAC a conduit à la constitution de stocks d'excédents agricoles, qui ont posé des difficultés de gestion, ou bien des subventions réservées à l'exportation pour retirer des produits du marché européen.

Au total, l'économie subit des **pertes sèches** qui dépendent du mode de taxation pour financer les subventions publiques et notamment du niveau des taxes douanières. Un mécanisme de garantie des prix a donc un coût collectif pour les pays européens, et il a par ailleurs un impact négatif sur le bien-être des pays en développement qui ne sont plus compétitifs face aux prix garantis dont bénéficient les agriculteurs européens. C'est la raison pour laquelle l'Union européenne est passée progressivement à un système de subventions directes aux producteurs plutôt que de garantie des prix, système collectivement plus efficace d'après les évaluations dont on dispose.

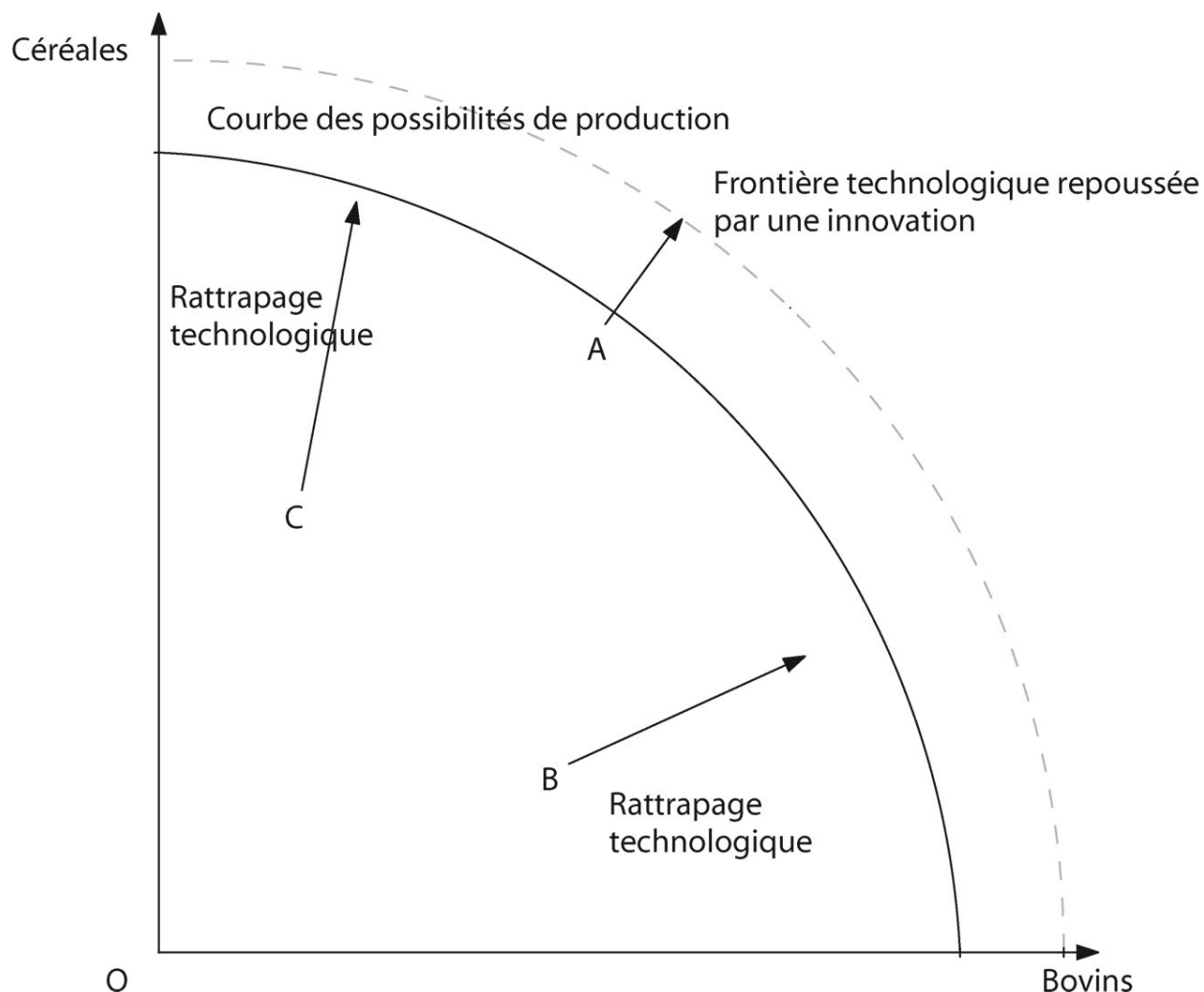
3. Innovations, concurrence et politiques publiques

À court terme, la technologie de production est donnée et les entreprises se contentent de déterminer pour chaque niveau possible de production, la combinaison productive la plus avantageuse en termes de coûts afin de maximiser leur profit. À long terme, des innovations technologiques sont possibles et il convient de s'interroger sur ce qui peut les favoriser. Les innovations sont sources de gains de productivité qui conduisent à des baisses de coûts de production mais pour qu'elles émergent des dépenses importantes sont nécessaires en amont pour financer la recherche nécessaire à leur conception. L'agriculture a connu de forts gains de productivité, et aujourd'hui il suffit d'un petit nombre d'exploitants pour nourrir la population française, alors qu'il y a quelques décennies une population active agricole beaucoup plus nombreuse produisait des quantités inférieures. L'accroissement de la productivité peut reposer sur une augmentation du volume de capital et de travail utilisé, ce qui permet de bénéficier d'économies d'échelle. . Mais à long terme, il n'est pas possible d'assurer une progression régulière des gains de productivité sans changer de technologie : une exploitation agricole aujourd'hui n'a pas seulement plus de capital et/ou de travail qu'une exploitation des années 1950, elle a incorporé de nombreuses **innovations** et a bouleversé les technologies utilisées. Une innovation vient modifier la technologie existante, que ce soit par la création d'un produit nouveau, d'un nouveau procédé de production ou d'une nouvelle organisation de la firme.

Par hypothèse, le modèle de concurrence pure et parfaite n'est pas adapté pour saisir les processus d'innovation, car la technologie des firmes est donnée par la fonction de production pour l'ensemble des firmes, et connue à l'avance. De façon générale, l'innovation réduit la concurrence quand elle permet à la firme d'acquérir un monopole provisoire sur un produit, par exemple *via* la protection d'un brevet, et les modèles de concurrence imparfaite sont plus appropriés ces situations (voir le [chapitre 4](#)). Cependant, l'innovation n'est pas absente des marchés concurrentiels, comme on le voit dans le secteur agricole, qui possède des caractéristiques spécifiques. D'une part, les innovations résultent, davantage que dans d'autres secteurs, de

l'effort de recherche public et de transferts technologiques de la part de l'industrie : la taille des exploitations agricoles est un obstacle à l'innovation, et une nouvelle semence où un nouveau processus d'exploitation des sols ont plus de chances d'être né dans un laboratoire de recherche ou dans un grand groupe de l'industrie agro-alimentaire. Selon les données de l'OCDE (2011), la recherche & développement dans le secteur agricole résulte d'abord des efforts de recherche publique, et est très inégalement répartie : aux États-Unis, où la productivité du travail dans l'agriculture est l'une des plus élevées au monde, l'effort de recherche public est aussi un des plus élevé, puisqu'il atteint plus de 4 % du PIB agricole.

Graphique 3.10 : Innovations et frontière des possibilités de production



Dans ces conditions, l'enjeu est de savoir s'il faut développer la concurrence sur les marchés agricoles pour accroître leur capacité productive à long terme, ou au contraire protéger ce secteur pour y diffuser de façon plus volontariste les innovations, *via* des politiques publiques. La question est au cœur des travaux de P. Aghion sur les facteurs de la croissance économique (2012), même s'ils ne portent pas spécifiquement sur les secteurs agricoles. Selon lui, il n'y a pas de réponse systématique à cette question, et tout dépend de la distance à la **frontière technologique**, c'est-à-dire du degré d'éloignement du secteur considéré par rapport aux performances du pays où les producteurs sont les plus efficaces, et utilisent donc la meilleure technologie du moment. Le [graphique 3.10](#) permet d'éclairer ce problème, en s'appuyant sur la notion de **frontière des possibilités de production**, qui représente graphiquement les limites de ce que l'on peut produire à un

moment donné au sein d'une économie avec un certain volume de facteurs dans deux types de produits agricoles, ici les céréales et les bovins.

Le raisonnement est limité à deux produits pour des raisons graphiques évidentes, mais il vaut pour une production plus variée. Les flèches indiquent l'évolution de la production des deux biens au sein de chaque pays, au cours d'une période donnée. Les pays émergents, comme B ou C, sont loin de la frontière des possibilités de production, cela veut dire qu'ils ne maîtrisent pas les technologies permettant un haut niveau de productivité des facteurs. Au cours de la période, la flèche indique qu'ils sont engagés dans un processus de rattrapage puisque la distance à la frontière se réduit, mais dans deux directions différentes : les exploitants du pays C se spécialise davantage dans les céréales, et ceux du pays B dans les bovins. Chaque pays tire donc parti de ses dotations en facteurs de production et de leur coût pour orienter sa production. Au cours de la période, on observe que la frontière technologique a été repoussée, puisque le pays A a développé les technologies permettant d'aller au-delà des possibilités de production initiales.

Selon P. Aghion, les **économies en rattrapage**, comme B ou C, ne doivent pas développer la concurrence, ni mener la même politique industrielle que les pays à la frontière technologique, qui forment des **économies d'innovation**. Dans une économie en rattrapage, il est préférable de mener des politiques volontaristes d'investissement public pour augmenter rapidement les capacités productives, de limiter la concurrence pour protéger les secteurs émergents et pour que se forment des entreprises de taille suffisante pour bénéficier d'économies d'échelle. Dans ces conditions, les producteurs pourront plus facilement rattraper leur retard technologique. C'est bien ce qu'on observe dans le secteur agricole européen au début des Trente Glorieuses, dans une agriculture largement en retard sur les États-Unis, qui restent encore aujourd'hui le pays leader en termes technologiques : la PAC et le protectionnisme douanier ont permis à l'agriculture européenne de se développer, tandis que se formaient de grands groupes dans l'industrie agro-alimentaire. Mais dans une économie d'innovation, proche de la frontière technologique, il faut au contraire développer la concurrence car ce sont les jeunes firmes, porteuses d'innovation, qui permettent de repousser la frontière technologique. L'innovation est nécessaire pour que les pays situés à la frontière technologique ne soient pas condamnés à une croissance faible. Sans l'existence de gains de productivité, la simple accumulation du capital par tête ne pourrait durablement soutenir la croissance car du fait de

l'existence de la loi des rendements marginaux décroissants, la productivité marginale du capital ne ferait que décroître et l'accumulation du capital conduirait à une augmentation de plus en plus faible du volume de production. Dans cette logique, les gouvernements des pays développés aujourd'hui ont intérêt à mener des « politiques de l'offre » permettant de développer la compétitivité de leurs entreprises. Les pays situés dans les économies de rattrapage ne connaissent pas ce problème car en accumulant du capital, ils intègrent aussi de nouvelles technologies qui n'étaient pas utilisées précédemment, ce qui n'est pas le cas des pays situés dans les économies d'innovations qui utilisaient déjà les machines les plus performantes. Ces considérations expliquent la réduction du protectionnisme dans le secteur agricole pour inciter à l'innovation, et la nécessité de mener une politique industrielle incitative, favorable à l'émergence de petites entreprises plus innovantes, par exemple dans le secteur très porteur de la « croissance verte » où se situent des gisements de croissance pour l'agriculture (biocarburants, technologies permettant de réduire le recours aux engrais, etc.). Selon P. Aghion, développer la concurrence sur les marchés et mettre en place des politiques publiques favorisant l'innovation vont de pair : il ne faut pas choisir entre l'État et la concurrence, mais trouver les synergies entre les deux.

4. Les choix organisationnels des producteurs

Jusqu'à présent, on s'est focalisé sur l'effet des prix sur les comportements des producteurs. Ils adaptent leur production aux prix des facteurs et aux prix des biens déterminés sur des marchés concurrentiels. Si le modèle de concurrence pure et parfaite est bien adapté pour comprendre l'effet des prix sur les comportements des firmes, il n'explique pas pourquoi certaines sont plus efficaces que d'autres, dans les mêmes conditions de marché, pas plus qu'il ne permet de comprendre pourquoi les firmes adoptent des organisations très différentes. Ces questions relèvent davantage de l'économie des organisations, qui s'est développée autour d'hypothèses absentes du modèle standard : la première est que le passage par le marché peut entraîner des coûts de transaction, qui modifient le comportement des producteurs ; la deuxième est que l'efficacité des firmes ne dépend pas seulement des prix, mais aussi de leurs choix organisationnels.

4.1 Les coûts de transaction et le dilemme entre l'organisation et le marché

Les agriculteurs exploitants sont à première vue le prototype des entreprises individuelles, où le propriétaire du capital est en même temps celui qui fournit le facteur travail. Quand un agriculteur a besoin de main-d'œuvre supplémentaire, il peut recourir à des membres de sa famille salariés ou non, qui fournissent toujours la majorité du travail sur les exploitations, ou au travail d'appoint de saisonniers. Pourtant, on constate un développement régulier du salariat dans les exploitations agricoles. En France métropolitaine, en 2015, le travail agricole, mesuré en unité de travail annuel (UTA), est assuré pour 18 % par des salariés permanents. Dans le même temps, les sociétés agricoles se développent au détriment des entreprises individuelles⁶. Une société se distingue d'une entreprise individuelle par son statut juridique : comme dans l'industrie, elle peut regrouper plusieurs associés dans le capital, elle limite la responsabilité des associés en cas de faillite et favorise l'embauche. Selon l'INSEE, la part des sociétés parmi les exploitations agricoles est passée progressivement de 6,4 % en 1988 à 29,9 %

en 2010, et elles représentent 57 % des surfaces cultivées par les exploitations professionnelles en 2007. Comme expliquer que des sociétés se développent dans le secteur agricole, plutôt que de recourir au marché pour acheter les services ou le travail saisonnier dont elles ont besoin ?

Une première hypothèse serait que les exploitations agricoles connaissent des économies d'échelle telle qu'une croissance de la taille des firmes est la solution adaptée. Ce phénomène est très fréquent dans l'industrie, où la taille est un facteur souvent décisif, mais dans ce cas les entreprises se concentrent et on ne se situe plus sur des marchés concurrentiels caractérisés par une atomicité de l'offre (voir le [chapitre 4](#)). Sur les marchés agricoles, les économies d'échelle restent limitées, au-delà d'un certain seuil, et dans ce cas de figure le modèle de concurrence pure et parfaite ne nous offre pas de réponse, car il est fondé sur l'idée que le recours au marché d'entreprises individuelles pour acheter les facteurs dont elles ont besoin est la solution appropriée. R. Coase (1910-2013) a proposé une explication à ce problème dans un article de 1937⁷ qui joue un rôle majeur dans le développement de l'économie des organisations : si les firmes ne passent pas systématiquement par le marché pour trouver les services productifs et les facteurs dont elles ont besoin, c'est que le recours au marché à un coût, et par conséquent l'embauche de salariés permanents et le développement de la société permet de limiter ces coûts, appelés **coûts de transaction** (CHAUDEY, 2014). Ils regroupent tous les coûts nécessaires pour faire aboutir une transaction sur un marché, que ce soit les coûts de recherche d'information et d'élaboration des contrats ex ante, ou les coûts ex post de surveillance et de contrôle de la bonne exécution des contrats, ou de renégociation en cas de rupture du contrat. Par exemple, la simple embauche d'un saisonnier suppose des coûts pour prospecter et rechercher la personne qui convient, et pour ensuite vérifier qu'il accomplit bien la tâche demandée. Si l'exploitant a régulièrement besoin de rechercher des saisonniers, il est plus efficace d'embaucher une personne, pour limiter les coûts de transaction. En l'absence de coûts de transaction, il serait préférable de recourir systématiquement au marché pour trouver les services productifs dont l'entreprise a besoin, mais plus ces coûts sont élevés et plus il est préférable de « faire en interne » ce que l'on pourrait faire réaliser par un autre agent sur le marché.

Pour R. Coase, il existe donc deux modes de coordination entre les agents. La **coordination par le marché** repose sur les réactions des agents aux

variations de prix, résultant des comportements d'autres agents ; elle n'est pas gratuite contrairement à ce que postule le modèle de base néo-classique, à cause des coûts de transaction. Inversement, la **coordination par la hiérarchie**, au sein d'une firme, repose sur la relation de pouvoir entre le salarié et son employeur, qui peut ordonner l'exécution d'une tâche mais subit aussi des coûts pour assurer le bon fonctionnement de son organisation. Le développement des firmes dépend alors d'un arbitrage entre ces deux types de coûts, selon la nature de leur marché.

Les travaux d'O. Williamson ont prolongé la théorie de R. Coase, à partir des années 1960, au sein de ce qu'on appelle la **nouvelle économie institutionnelle**. O. Williamson considère que les agents font preuve de **rationalité limitée**, c'est-à-dire qu'ils ne sont généralement pas capables de trouver les solutions optimales aux objectifs poursuivis comme on le postule dans le modèle de concurrence pure et parfaite, en raison de capacités de calcul limitées, d'un environnement incertain ou trop complexe. Ils adoptent alors des procédures de décisions différentes de celles d'un agent optimisateur, disposant de toutes les capacités de calcul et des informations nécessaires, et capable de définir des **contrats complets** dans ses transactions marchandes, c'est-à-dire des contrats dont toutes les variables sont vérifiables, sans incertitude sur leur réalisation. Si les producteurs ne peuvent signer que des **contrats incomplets**, ils prennent le risque que le contrat ne soit pas réalisé comme prévu. Un contrat de service pour monter un site web est par exemple incomplet s'il n'est pas possible de préciser certaines qualités du site et si le webmaster peut proposer un site de médiocre qualité tout en respectant les clauses du contrat. Certains contractants peuvent profiter de l'incomplétude des contrats pour faire preuve d'**opportunisme**, selon le terme de O. Williamson : ils tirent parti de la situation, par stratégie et par ruse, pour minimiser leurs efforts au détriment de l'intérêt du contractant. C'est le cas d'un webmaster proposant à un exploitant agricole décidé à faire de la publicité pour ses produits un site web inadapté à ses besoins, faute des compétences nécessaires pour réduire l'incomplétude du contrat. Les coûts de transactions peuvent alors être élevés, et réduire l'intérêt de l'achat d'un tel service.

Le producteur doit donc choisir le bon type de contrat, adapté à son environnement et aux prestations dont il a besoin : il peut décider de faire faire une tâche à un prestataire extérieur, quand les coûts de transactions ne sont pas rédhibitoires, ou décider au contraire de faire en interne cette tâche,

quitte à embaucher un salarié supplémentaire, pour réduire les coûts de transaction. Tout dépend de sa situation, et notamment du degré de **spécificité des actifs** selon O. Williamson : des **actifs sont spécifiques** lorsqu'ils perdent de la valeur dans un usage alternatif par rapport à leur usage actuel. Ainsi, un tracteur est un actif peu spécifique car il est aisé à utiliser sur une autre exploitation, sans coûts supplémentaires, mais un site web a une plus grande spécificité car si un contractant fait défaut, il sera plus difficile à utiliser. Les actifs peuvent être corporels (une machine par exemple) ou incorporels (l'expérience cumulée d'un salarié). La théorie des coûts de transaction développée par O. Williamson indique donc que les agents choisissent des contrats marchands avec des prestataires extérieurs lorsqu'ils ont affaire à des actifs peu spécifiques, ce qui limite notamment les risques d'opportunisme, et qu'au contraire ils ont intérêt à faire faire en interne certaines tâches, lorsque les actifs sont très spécifiques.

Le secteur agricole connaît le même mouvement d'**externalisation des services** que l'industrie : les exploitants agricoles ont tendance à se spécialiser sur leur « cœur de métier » et à faire réaliser les autres tâches par des sociétés de service, qui bénéficient d'économies d'échelle en se spécialisant dans un domaine particulier. Ce recours aux prestataires de service est plus important pour des tâches mettant en œuvre des actifs peu spécifiques car il s'agit de services très standards, par exemple la location de machines agricoles, mais il peut également concerner des tâches complexes et difficiles à réaliser en interne (soins vétérinaires, contrôle qualité, etc.). Le développement des nouvelles technologies, auxquels les agriculteurs ont largement recours aujourd'hui, contribue à réduire les coûts de transaction en réduisant la recherche d'information sur les partenaires, en mutualisant les expériences, etc.

Pour partager certaines ressources et assurer l'écoulement de leurs produits, les exploitants agricoles peuvent mettre en place des formes d'organisation qui ne se résument pas au dilemme entre le marché et la hiérarchie, présenté ici. Une des spécificités du secteur agricole, c'est la présence de nombreuses coopératives, les trois quarts des exploitants étant adhérents à au moins l'une d'entre elles⁸. Une **coopérative** est un regroupement d'exploitants qui apportent le capital et gèrent en commun tout ou partie de leur production, qui peut par exemple être revendue par la coopérative à des conditions plus favorables que si chacun vend de son côté ses produits. Les coopératives font partie de l'économie sociale et solidaire,

héritières du syndicalisme agricole. Elles n'ont pas vocation à être un centre de profits, et si elles dégagent des marges elles sont redistribuées à leurs membres. Le monde rural connaît depuis très longtemps l'existence d'**institutions** spécifiques pour gérer des ressources communes (voir le [chapitre 5](#)), qui ont été mises en lumière par les travaux d'E. Ostrom, Prix Nobel d'économie en 2009 conjointement avec O. Williamson. Elle insiste en particulier sur le fait que les institutions permettant la mise en commun de ressources sont des lieux de négociation, qui permettent de les gouverner selon des formes très variées. Les modes de coordination ne sont pas figés et des formes complexes existent : ainsi la **coopération**, telle qu'elle est présentée ici, propose une alternative au dilemme classique entre la coordination par le marché et la coordination par la hiérarchie au sein de la firme.

4.2 Coordination au sein de la firme et efficacité économique

• La théorie des incitations

À quelles conditions les facteurs de production utilisés au sein d'une organisation productive sont-ils utilisés de façon efficiente ? Dans le modèle de concurrence pure et parfaite, ce problème est mis de côté puisque l'usage optimal des facteurs de production est considéré comme allant de soi. Or un salarié agricole peut fournir plus ou moins d'effort, une terre louée à un exploitant par son propriétaire peut être médiocrement mise en valeur, etc. S'il existe des asymétries d'information entre les agents, qui ne peuvent pas vérifier les efforts réalisés par les autres (voir le [chapitre 4](#)), et si les contrats sont incomplets, alors il faut trouver le moyen de mettre en place des incitations adéquates pour coordonner les efforts de chacun.

La **théorie des incitations** porte sur les situations où des imperfections de l'information, au sein d'une organisation, rendent impossibles la mise en place de contrats complets pour rémunérer un **agent** en fonction de sa contribution. Par exemple, il n'est pas possible de rémunérer un salarié à sa productivité marginale, car celle-ci est difficilement observable. Il faut alors mettre en place les institutions appropriées pour inciter les agents à faire les efforts souhaités. Le modèle s'applique aux différentes **relations d'agence** entre un « principal » qui est le donneur d'ordre et l'agent chargé d'exécuter

un contrat, par exemple la relation entre l'employeur et le salarié, entre une entreprise cliente et son fournisseur, les actionnaires et le manager d'une société, etc. La théorie des incitations s'appuie également sur la théorie des jeux (voir le [chapitre 4](#)) pour comprendre les stratégies des agents.

La rémunération du propriétaire des terres constitue une bonne application du modèle de l'agence. En France en 2010, 75,9 % des surfaces agricoles utiles (SAU) sont louées à des propriétaires fonciers (qui peuvent être d'autres agriculteurs), principalement par des **contrats de fermage**, c'est-à-dire par une location contre une rémunération fixe. On est bien en présence d'une relation principal/agent entre les propriétaires et les exploitants. Le fermage s'est fortement développé, puisqu'il représentait seulement la moitié de la SAU en 1980 (COURLEUX, 2011). La théorie des incitations permet de comprendre pourquoi, dans les pays développés, le contrat de fermage s'est imposé au détriment du métayage, qui repose sur le partage de la récolte entre le propriétaire et l'agriculteur. En effet, le contrat de fermage incite l'agriculteur à faire le maximum d'efforts puisqu'une fois qu'il a payé le propriétaire la totalité des gains additionnels lui reviennent. Au contraire, le métayer n'a pas intérêt aux mêmes efforts car il ne reçoit qu'une partie des gains additionnels, et si les rendements sont décroissants, ses efforts cessent plus rapidement d'être rentables. Toutefois, ce système se rencontre encore dans certains pays émergents, car il permet de mieux partager les risques dans des contextes de faible productivité agricole où le métayer ne serait pas certain de produire assez pour payer le propriétaire.

• La diversité institutionnelle des firmes

Certaines approches récentes de la firme insistent davantage sur la diversité des firmes, en fonction de leur environnement institutionnel. Les deux approches évoquées ci-dessous sont plus éloignées de la théorie microéconomique standard, dans la mesure où elles insistent sur le caractère situé des comportements des agents selon le contexte institutionnel (CHAUDEY, 2014).

La **théorie évolutionniste** de la firme a été initiée par **R. Nelson** et **S. Winter**⁹ dans les années 1980, elle repose sur l'idée que les firmes sont des lieux d'apprentissage où les individus développent des compétences spécifiques à leur organisation, pour traiter les problèmes auxquels ils font

face. Selon eux, les individus n'ont ni le temps ni les outils nécessaires pour réaliser des calculs d'optimisation, dans un environnement de travail souvent très instable et incertain. Ils s'appuient alors sur des **routines**, c'est-à-dire des schémas de comportement pré-établis, qui permettent d'orienter rapidement l'action. Par exemple, un éleveur a mis en place des routines pour s'occuper quotidiennement de son bétail, il ne s'agit pas forcément des solutions optimales mais elles sont issues de son expérience et permettent de ne pas perdre de temps à se décider. Dans un environnement concurrentiel et très fluctuant, les entreprises qui s'appuient sur les routines les plus efficaces s'adaptent mieux que les autres : les auteurs font la comparaison avec un gène qui est sélectionné au long du processus d'évolution parce qu'il favorise la survie des entités porteuses. L'approche de Nelson et Winter insiste donc sur la diversité des cultures d'entreprise, en fonction des routines qu'elles ont développées et des compétences cumulées, cette théorie est souvent mobilisée pour comprendre pourquoi certaines firmes parviennent à se montrer plus innovantes que d'autres. Les innovations ne sont pas absentes du monde agricole, et les réussites et les échecs des producteurs qui se lancent dans « l'agriculture bio » fournissent un bon exemple de ces processus d'adaptation.

L'économie des conventions, développée en France dans les années 1980 insiste quant à elle sur une autre caractéristique institutionnelle : pour que des individus puissent coopérer, ils doivent se mettre d'accord sur les principes de qualification du monde qui les environne : comment définir ce qui est efficace ou non, juste ou injuste, adapté ou inadapté ? Suite aux travaux du sociologue L. Boltanski et de l'économiste L. Thevenot¹⁰, cette théorie s'oppose à l'idée selon laquelle les individus partageraient systématiquement la même manière d'évaluer ce qu'ils doivent faire. Au contraire, ils montrent qu'ils peuvent s'appuyer sur des « conventions » différentes, qui éclairent la variété que l'on observe dans l'organisation des firmes. Dans un article consacré aux producteurs de camembert en Normandie, P. Boisard et M.T. Letablier¹¹ montrent la fécondité de cette approche. Les producteurs se divisent entre l'industrie artisanale du camembert au lait cru, et les camemberts industriels au lait pasteurisé. Ils montrent que l'organisation du travail et la gestion de la main-d'œuvre renvoient à des conventions opposées : la production artisanale relève d'une « convention domestique » qui valorise l'expérience partagée et l'appartenance à un même groupe, et à l'inverse la « convention industrielle », qui valorise le respect de normes

standardisées et de procédures impersonnelles fixées par des règlements, domine dans la production de camemberts industriels. Par exemple, le recrutement d'une ouvrière chez un producteur artisanal valorisera l'appartenance à une famille bien connue, tandis que le recrutement dans l'industrie valorisera les qualifications objectivées par des diplômes. Dans un même secteur d'activité, on peut donc trouver des entreprises ayant adopté des organisations opposées, car elles ne sont pas appuyées sur les mêmes conventions.

Notes

1. X. Boutinet et S. Quantin, « Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises (1984-2002) », *Économie & Statistique*, n° 413, 2008.
2. Une exploitation agricole est définie par la production d'un bien agricole (végétal ou animal), une gestion indépendante et une surface minimale d'exploitation. La distinction entre les exploitations repose depuis 2010 sur la « production brute standard » (PBS) qui donne la production potentielle d'une exploitation en appliquant des coefficients de production standardisés à sa surface, son cheptel, la nature de ses produits, etc. Une petite exploitation a une production brute standard inférieure à 25 000 €, une exploitation moyenne entre 25 000 et 100 000 €, et une grande exploitation au-delà.
3. Pour simplifier l'analyse, les coûts sont donnés ici en kilos de viande produite, tandis que les quantités concernent le nombre de vaches élevées, ce n'est donc pas la même unité qui a été retenue. Il n'est pas possible de contrôler la production de viande au kilo près, mais il suffit de prendre en compte le poids moyen d'une carcasse pour passer des têtes de bétail à la production au kilo.
4. A. Luciani, « Le coût horaire de la main d'œuvre en France », *Insee Première*, n° 1567, 2015.
5. Le patrimoine professionnel moyen d'un agriculteur est de 664 100 euros en 2015 (patrimoine brut). Source INSEE, enquête patrimoine.
6. <https://www.insee.fr/fr/statistiques/2569448?sommaire=2587886>
7. R. Coase, « The nature of the firm », *Economica*, vol. 4, 1937, traduit dans le recueil de textes suivant : R. Coase, *L'entreprise, le marché et le droit*, Éditions d'organisation, 1988. L'article est disponible à l'adresse http://www.persee.fr/web/revues/home/prescript/article/rfec_0769-0479_1987_num_2_1_1132.
8. Coop de France, « Poids économique et social 2010 de la coopération agricole et agroalimentaire française », novembre 2010, cité in Dedieu (2011).
9. R. Nelson et S. Winter, *An evolutionary theory of Economic change*, Cambridge, 1982.
10. L. Boltanski et L. Thevenot, *De la justification*, Gallimard, 1991.
11. P. Boisard et M. T. Letablier, « Le camembert : normand ou normé. Deux

modèles de production dans l'industrie fromagère », *Cahiers du Centre d'Étude de l'Emploi*, n° 30, PUF, 1987. Voir F. Eymard-Duvernay (2004).

Chapitre 4

Les imperfections de marché

L'exemple du secteur bancaire

1. LE POUVOIR DE MARCHÉ

2. COMPORTEMENT DES ENTREPRISES ET STRUCTURES DE MARCHÉ

3. LES ASYMÉTRIES D'INFORMATION

Le secteur bancaire regroupe l'ensemble des institutions financières dotées du pouvoir de création monétaire. C'est un secteur où les imperfections de la concurrence sont très nombreuses, en raison des propriétés très spécifiques des activités bancaires et du contrôle exercé par l'État sur ce marché, comme le montre un rapide historique. Entre la fin de la Seconde Guerre mondiale et le début des années 1980, en France, l'État a joué un rôle prépondérant dans la régulation de ce secteur : nationalisations de nombreuses banques et encadrement très contraignant de leurs activités les obligeant à se spécialiser dans certaines opérations, ce qui a contribué au cloisonnement des banques selon leur activité. La loi bancaire de 1984 a révolutionné le secteur en le déréglementant très largement et en donnant ainsi un poids plus fort aux mécanismes de marché. Cette évolution de la législation a facilité la concurrence entre les banques. En réaction, les banques sont entrées dans une logique de concentration. Ainsi, fin 2016 les groupes bancaires français détiennent près de 8 milliards d'actifs, dont près de 83 % sont contrôlés par les six plus grands groupes¹. Les fusions entre banques ont permis à ces dernières de réaliser des économies d'échelle, de se constituer un **pouvoir de marché** – elles sont capables d'agir sur les prix – et de préserver ainsi leur profit. En situation de forte concurrence, celui-ci tend en effet à devenir nul, comme l'explique la théorie du producteur en concurrence pure et parfaite (voir le [chapitre 3](#)). Aujourd'hui, le secteur bancaire est un secteur assez concentré comme en témoigne son **indice d'Herfindahl-Hirschmann (IHH)** dont la valeur est en progression en France avant la crise de 2008, passant de 509 en 1999 à 698 en 2008, avant de décroître après la crise bancaire jusqu'à une valeur de 572 en 2016² (voir le [chapitre 1](#) pour la définition de l'indice). Les disparités sont grandes en Europe, de façon générale la concentration est beaucoup plus élevée dans les petits pays, où seuls quelques réseaux bancaires sont de taille suffisante pour se maintenir, alors que la concurrence

est plus vive pour conquérir des parts de marchés dans les pays les plus peuplés : l'IHH atteint en 2016 la valeur de 2 097 aux Pays-Bas, et même 2 406 en Estonie, alors que l'indice le plus faible est de 277 en Allemagne, pays le plus peuplé d'Europe. Dans le secteur bancaire, certains acteurs très importants sont capables d'influencer par leur comportement le prix de vente des services proposés.

Par ailleurs, une des caractéristiques du secteur bancaire est l'ampleur de l'asymétrie d'information, c'est-à-dire des situations dans lesquelles les co-contractants n'ont pas accès au même niveau d'information. Par exemple, les clients qui déposent des fonds auprès d'une banque évaluent de façon imprécise sa solvabilité et inversement les banques qui octroient des crédits ne disposent que d'informations partielles sur la situation de leurs débiteurs.

Pour comprendre l'ensemble des mécanismes qui structurent ce secteur, il est nécessaire de mobiliser d'autres modèles que celui de concurrence pure et parfaite, présenté dans les chapitres précédents. Par définition, aucun marché réel ne répond jamais parfaitement aux critères d'un modèle et il n'est pas pertinent de discréditer un modèle au motif qu'il ne correspond pas exactement à la réalité. Certains modèles sont toutefois plus adaptés que d'autres pour comprendre le fonctionnement des marchés où la concurrence entre entreprises est faible et où l'asymétrie d'information est forte entre acheteurs et vendeurs, ils permettent de mieux saisir la diversité du fonctionnement des marchés. On distingue les modèles intégrant l'hypothèse d'information imparfaite de ceux qui étudient le pouvoir de marché des entreprises. Ils montrent que dans des situations de **concurrence imparfaite**, le marché ne conduit pas à la formation d'un équilibre qui maximise le bien-être social : il existe des pertes d'efficacité qui nécessitent une intervention de l'État. L'objet de ce chapitre est de présenter ces différentes approches qui peuvent être mobilisées pour comprendre les comportements des banques et la relation qu'elles entretiennent avec leurs clients.

1. Le pouvoir de marché

En concurrence pure et parfaite, les agents sont preneurs de prix. Aucune entreprise ne dispose d'un pouvoir de marché. De ce fait, à l'équilibre, le prix de marché se fixe à un niveau tel qu'il est égal au minimum du coût moyen de production et que le profit des entreprises est nul (voir le [chapitre 3](#)). En situation de concurrence le marché est ouvert par hypothèse et si un secteur est lucratif de nouvelles entreprises entrent sur le marché, le prix de vente baisse alors et le profit des entreprises diminue. Si l'on veut comprendre pourquoi certaines activités comme le secteur bancaire permettent aux entreprises de dégager un profit élevé et durable, il est impératif de mobiliser les modèles de concurrence imparfaite. Pour rendre compte de l'existence d'un pouvoir de marché des entreprises, ils mobilisent deux grandes familles d'arguments :

- Le pouvoir de marché peut s'expliquer par l'existence de **barrières à l'entrée** ou à la sortie du marché ce qui affaiblit la pression concurrentielle pesant sur les entreprises en place.
- Le pouvoir de marché peut découler de la différenciation des produits vendus par les entreprises sur un marché. Si elles offrent un produit présentant des caractéristiques distinctes des autres, elles peuvent le proposer à un tarif plus avantageux dans ces conditions.

Pour apprécier le degré de concurrence existant sur un marché, le plus simple est de considérer le nombre d'entreprises y opérant. Si de nombreuses entreprises sont en compétition pour la vente du même bien, le degré de concurrence est élevé (voir le [chapitre 1](#)). Une entreprise n'ayant aucun concurrent direct sur le marché est dite en situation de **monopole**.

1.1 Les barrières à l'entrée et à la sortie des marchés

Quand un petit nombre d'entreprises opèrent sur un marché, la structure du marché est dite **oligopolistique**. Les situations de monopoles ou d'oligopoles ne signifient toutefois pas toujours que l'entreprise dispose d'un véritable pouvoir de marché. D'après les travaux de W. Baumol, J. Panzar et R. Willig

(1982), ce n'est pas le nombre d'opérateurs en place qui est l'élément déterminant pour évaluer la pression concurrentielle pesant sur les entreprises, mais la possibilité d'entrée de nouveaux concurrents. Une entreprise en situation de monopole se comporte comme si elle était sur un marché concurrentiel à condition de se sentir menacée par l'entrée de nouveaux arrivants en cas d'adoption de tarifs élevés par rapport aux prix qui prévaudraient sur un marché concurrentiel. La théorie de W. Baumol, J. Panzar et R. Willig se base ainsi sur la notion de rivalité potentielle, de **contestabilité du marché** : si d'autres entreprises peuvent librement y entrer et en sortir pour contester la ou les entreprises en place, alors la discipline du marché s'impose aux entreprises présentes même en petit nombre. Ces dernières n'adoptent donc pas un comportement concurrentiel parce qu'elles y sont forcées (discipline liée à une réglementation) mais parce que les mécanismes de marché conduisent à ce qu'une pression s'opère sur le comportement des acteurs.

Le secteur bancaire n'est pas, dans cette optique, un secteur contestable. Il existe de nombreuses barrières qui limitent l'entrée et la sortie des entreprises. Notons toutefois que l'évolution de la réglementation encadrant les pratiques des banques a conduit à une amélioration de la contestabilité du secteur.

Pour définir la « structure de marché » dans un domaine donné, il ne suffit pas de se limiter au seul nombre d'entreprises, mais on tient compte également de leurs parts de marché respectives. Par exemple, si une seule banque nationale domine largement le marché dans un pays donné, mais qu'il existe des dizaines de filiales de banques étrangères de petite taille, on ne parle pas pour autant de marché concurrentiel car en réalité, le pouvoir de marché de la banque nationale reste dominant. Par ailleurs, elle a de grandes chances d'être en quasi-monopole dans de nombreuses localités, où les filiales étrangères ne sont pas implantées. Dans ce cas, on parle de **monopole à frange** pour désigner un quasi-monopole, avec une série d'entreprises de taille marginale sur les « franges » de ce marché. Ce type de structure est très courant sur les marchés qui ont été longtemps dominés par une entreprise publique en situation de monopole, et progressivement ouverts à la concurrence, comme dans la distribution d'énergie, les transports ferroviaires, mais aussi la banque dans certains cas.

• Les barrières légales

Sur certains marchés, des **réglementations** limitent l'arrivée de nouveaux concurrents. On parle de **monopole légal** quand la loi impose l'existence d'un seul producteur sur un marché. Les **brevets** qui protègent les innovations des entreprises empêchent ainsi pendant un certain temps des concurrents de s'implanter, laissant l'entreprise innovatrice en situation de monopole temporaire. Le marché des banques est un marché réglementé et une nouvelle banque doit demander un agrément à l'autorité de contrôle prudentiel et de résolution, mise en place entre 2010 et 2013, et qui a remplacé le comité des établissements de crédit et des entreprises d'investissement (CECEI) qui gérait dans le passé cette fonction. Depuis les années 1980 et le mouvement de déréglementation, les barrières légales à l'entrée sont plus faibles mais elles restent non négligeables et limitent le degré de concurrence du secteur, car les autorités publiques considèrent que l'activité bancaire doit être réglementée pour éviter l'installation de banques potentiellement dangereuses pour leurs clients, voire pour l'économie toute entière, en cas de faillite.

• Les coûts d'entrée et de sortie

L'existence de coûts importants afin d'entrer dans le secteur est liée à la nécessité de réaliser des investissements nécessaires pour opérer sur le marché. Ces coûts constituent une autre forme de barrière à l'entrée. Une nouvelle banque doit en effet investir des montants élevés pour construire un réseau d'agences, former son personnel, etc. sans certitude sur ses recettes futures, ce qui peut la conduire à renoncer. Récemment, le développement de l'offre de banques en ligne a favorisé l'arrivée de nouveaux concurrents, les coûts d'entrée étant plus faibles pour ces établissements sans guichets, ce qui a contribué à faire baisser le prix de certains services bancaires : entre 2010 et 2016 le prix moyen des abonnements permettant de gérer ses comptes sur internet a baissé de - 77,8 %³. Les coûts d'entrée sont plus dissuasifs pour les entrants potentiels s'ils sont des **coûts irrécupérables** : il s'agit des dépenses qui ne peuvent plus être recouvertes une fois engagées, en cas de sortie du marché. Un achat d'ordinateur est un coût récupérable car on peut le revendre, mais le coût de mise au point d'un logiciel spécifique est non recouvrable. Ils désincitent fortement l'implantation de nouvelles banques, car le risque devient trop important en cas d'échec. Les banques perdent ainsi

l'investissement réalisé par exemple dans la mise au point de logiciels et ne peuvent que difficilement monnayer certains de leurs actifs comme la qualité de leurs relations avec la clientèle.

• La technologie de production

La technologie de production peut engendrer des barrières à l'entrée si elle conduit à des **économies d'échelle** ce qui est le cas pour la production des services bancaires (voir le [chapitre 3](#)). Le coût unitaire de production des services financiers diminue au fur et à mesure que la quantité augmente, par exemple les infrastructures informatiques nécessitent des coûts fixes importants, qui sont d'autant plus faciles à prendre en charge que la clientèle est nombreuse. Il existe aussi des **économies de gamme** : il est moins coûteux de produire plusieurs services bancaires conjointement que de les produire séparément, favorisant ainsi les gros opérateurs. L'ouverture du secteur bancaire à l'échelle de l'union européenne a renforcé la possibilité d'économies d'échelle car toute banque installée dans un pays de la zone euro a le droit à présent d'opérer dans l'ensemble des pays de la zone. Les concurrents potentiels ne peuvent espérer s'implanter peu à peu sur le marché car ils ne seraient jamais en mesure de concurrencer les entreprises en place qui bénéficient de coûts de production inférieurs, ils doivent être capables de produire au moins autant que les entreprises du secteur dès leur arrivée sur le marché, ce qui suppose de lourds investissements.

On parle de **monopole naturel** quand, du fait de l'existence d'économies d'échelle, un producteur unique s'impose sur le marché en produisant à lui seul les quantités demandées. Le mécanisme est aisé à comprendre : plus une entreprise a un volume de production important, plus son coût unitaire de production est faible, plus elle peut baisser ses prix et ainsi vendre davantage, etc. Les petits et moyens producteurs qui voudraient rentrer sur le marché pour concurrencer l'entreprise dominante ne seraient pas compétitives car du fait de leur plus faible volume de production, elles auraient un coût moyen de production plus élevé et ne pourraient pas pratiquer des prix aussi bas. Pour rentrer sur le marché et arriver à s'y implanter avec succès, il faut directement produire des volumes très élevés ce qui suppose une mise de fond très importante ce qui constitue une limite forte à l'entrée du marché. Ce raisonnement est valide quand les biens vendus par tous les opérateurs sont strictement identiques.

D'autres limitations de l'entrée ou de la sortie des entreprises existent sur certains marchés, quand une entreprise dispose de l'exclusivité de l'accès à certaines ressources, pour l'accès aux nappes de pétrole par exemple. Quand le marché n'est pas parfaitement contestable, les entreprises en place disposent d'un pouvoir de marché leur permettant d'agir sur le prix de vente dans un sens qui leur est profitable.

L'importance du pouvoir de marché ne dépend pas uniquement des caractéristiques de l'offre mais aussi de celles de la demande. Une demande très élastique au prix donne un plus faible pouvoir de marché aux entreprises : elles savent qu'en accroissant légèrement leur prix de vente, elles perdront de nombreux clients. Ainsi, certaines entreprises imposent à leurs clients des coûts de sortie élevés en cas de changement de fournisseurs (*switching cost*). Un client qui veut changer de banque doit supporter un certain nombre de coûts, qu'ils soient monétaires (pénalités de sortie) ou non (perte de temps pour remplir des formulaires nombreux et trouver une autre banque qui soit adaptée aux attentes du client par exemple). De telles mesures créent une **demande captive** : les demandeurs ne peuvent s'adresser à d'autres offreurs sans subir des coûts élevés, et ont intérêt à rester clients d'une firme même si son offre n'est pas la plus avantageuse. Cette situation réduit l'élasticité de la demande de services bancaires par rapport aux prix de ces services. Depuis 2005, les frais de fermeture des comptes courants ont été supprimés afin de favoriser la concurrence, mais les clients doivent toujours payer pour transférer certains produits financiers.

1.2 La différenciation des produits

Jusqu'à présent, on a considéré que les entreprises vendaient le même bien et disposaient d'un pouvoir de marché du fait d'une pression concurrentielle trop faible. Une entreprise peut également acquérir un pouvoir de marché si elle différencie ses produits de ceux vendus par les autres afin de se retrouver dans une situation de quasi-monopole. Elle devient alors la seule à produire ce bien ou service avec des caractéristiques particulières. Certaines banques proposent ainsi des cartes de crédit personnalisées, offrent des prestations spécifiques en cas d'ouverture d'un compte, etc. Il ne s'agit pas toutefois d'une situation de monopole car il existe une concurrence forte entre elles.

Chaque banque propose des cartes de crédit et même si les contrats associés à chaque carte diffèrent selon les établissements, chacune doit tenir compte de l'offre des autres banques pour fixer son tarif. On parle de **concurrence monopolistique** pour désigner une situation où les producteurs sont nombreux mais où ils disposent tout de même d'un pouvoir de marché car les biens échangés sont différenciés. Dans les modèles de concurrence monopolistique, contrairement à celui de concurrence pure et parfaite, les biens ne sont donc pas homogènes.

La différenciation des produits peut s'opérer selon diverses modalités :

- Par la différenciation des propriétés du produit : les banques ont ainsi pour politique de diversifier leur offre de produits bancaires. Cette différenciation résulte d'une innovation de produit. On en distingue deux types : d'abord la **différenciation verticale** qui porte sur la qualité des produits, telle la diversification des cartes de crédit proposée par les banques (carte standard, Gold, Black, etc.). Certains produits sont ainsi de meilleures qualités que d'autres et chaque gamme vise des consommateurs différents. La **différenciation horizontale** s'opère à qualité égale, le classement des différents produits est impossible objectivement mais dépend uniquement des goûts subjectifs des consommateurs. Certaines banques proposent ainsi à leurs clients de déposer leurs espèces dans des machines au lieu de les déposer au guichet. Ces deux modalités permettent de rendre le même service aux clients mais certains préfèrent donner leurs avoirs en main propre à une personne physique alors que d'autres apprécient davantage le dépôt en machine.
- Par la construction d'une image différente du produit à travers une politique de marketing sans que le produit soit vraiment différent des autres. La différenciation des produits dépend d'abord de la perception des consommateurs, s'ils sont persuadés qu'un produit est différent d'un autre, ils agiront en conséquence. Les banques investissent beaucoup sur leur image afin de convaincre les clients potentiels de venir chez elles en cherchant ainsi à inspirer confiance dans la qualité de leurs produits. Le sponsoring en matière sportive ou culturelle contribue par exemple à ce travail de construction d'image.
- Par la localisation : les produits vendus sont les mêmes mais les lieux de vente sont différents et les clients en tiennent compte. Ils choisissent

en effet leur banque en fonction des prestations proposées mais aussi selon l'emplacement des agences. Les agences disposent d'un **monopole local** vis-à-vis de leur clientèle proche spatialement. Cette forme de différenciation peut être assimilée une différenciation horizontale, la localisation d'un bien étant une de ses caractéristiques. La prise en compte de la notion d'espace et de l'existence de coûts de transport par la théorie économique a conduit à la publication de nombreux travaux donnant ainsi naissance à **l'économie géographique**.

Le modèle de S. Salop (1979) qui prolonge celui d'H. Hotelling (1929) éclaire les choix de localisation des entreprises. Dans ce modèle, les biens et services vendus sont identiques mais la localisation de leur point de vente compte aux yeux des consommateurs qui éprouvent du désagrément à effectuer des longs trajets pour acheter le produit désiré, d'où une perte d'utilité. Ils acceptent de s'approvisionner loin de chez eux, alors que le bien désiré est disponible dans un point de vente proche, seulement si la différence de prix compense le désagrément lié au déplacement. Une banque qui doit choisir la localisation d'une nouvelle agence intègre son éloignement des agences existantes, qui lui confère un fort pouvoir de marché. Elle sait aussi qu'en se situant loin, elle ne capte qu'une très faible partie de la demande qui s'adresse actuellement aux agences concurrentes. S. Salop part de l'hypothèse que les consommateurs sont uniformément répartis sur le territoire (représenté par un cercle) et qu'il existe un coût fixe d'implantation. Dans le cas des banques il s'apparente au coût de la construction de l'agence. Les consommateurs subissent quant à eux un coût de transport qui dépend de la distance à parcourir. S. Salop cherche à déterminer dans ce contexte quels sont les prix, quantités produites et nombre d'entreprises présentes à l'équilibre. Il montre que chaque entreprise propose un nombre d'agences supérieur au nombre efficace, si les services bancaires étaient proposés au prix concurrentiel sans prendre en compte leur localisation. Les réseaux bancaires sont surdimensionnés par rapport aux quantités produites. Les banques construisent de nouvelles agences tant qu'elles perçoivent des rentes de situation, ce qui contribue à ce que les profits diminuent peu à peu jusqu'à converger vers zéro.

• Encadré 4.1 : Concurrence imparfaite et territoire

Dans le modèle de base de la concurrence pure et parfaite, la localisation des clients comme des entreprises n'est pas prise en compte, les agents sont supposés entrer directement en contact sans agir sur un territoire donné. Or les coûts de transports et les avantages d'une localisation dans un centre urbain par exemple ont un rôle central dans un grand nombre de secteurs, comme le tourisme, le commerce ou la banque, mais aussi pour les activités industrielles, quand la compétitivité d'une firme dépend de son environnement immédiat (accès aux transports ou à un marché local d'emplois qualifiés par exemple). L'économie géographique s'est donc progressivement développée pour étudier la dimension spatiale des activités économiques, qui est un des facteurs d'imperfection de la concurrence. Parmi les travaux fondateurs, citons le modèle d'H. Hotelling (1929) : il fait l'hypothèse que les offreurs se différencient uniquement par leur localisation, et que les demandeurs subissent des coûts de transport croissants avec la distance. Deux types d'équilibres sont alors envisageables : soit les offreurs ont intérêt à optimiser les distances entre eux, pour minimiser les coûts de transport et à se répartir uniformément dans l'espace, soit ils se concentrent au contraire au centre de l'espace pour accéder au plus grand nombre de consommateurs, et se font concurrence au même lieu. La concurrence peut donc s'exercer très différemment, dès que l'on intègre la dimension spatiale. Plus récemment, les travaux de P. Krugman sur le modèle « centre-périphérie » cherchent à expliquer la force d'attraction des agglomérations, et l'inégal développement des territoires. Il reprend une intuition d'A. Marshall selon laquelle les agglomérations exercent des externalités positives (voir le [chapitre 5](#)) sur les agents, par exemple les industriels bénéficient des innovations des clients et des fournisseurs à proximité. La présence d'externalités d'agglomération et la réduction des barrières à l'échange, par exemple les droits de douane entre pays, permet de comprendre la polarisation des activités dans certaines zones.

En diversifiant leurs produits, les entreprises peuvent imposer des prix plus élevés qu'en situation de concurrence et dégager un profit supérieur. Le pouvoir de marché découle du fait que la différenciation des produits permet aux entreprises d'obtenir un monopole provisoire sur une qualité particulière de produit, tant qu'elle est seule à l'offrir, et de le vendre à un tarif plus avantageux. Ce profit est temporaire, car les entreprises s'imitent entre elles et les innovations n'ont pas d'effet durable. Si l'innovation proposée par une banque rencontre un certain succès, comme le fait de pouvoir effectuer des virements bancaires depuis son téléphone portable, elle peut tirer parti temporairement de ce succès. Mais très vite les autres banques vont intégrer cette possibilité dans leur offre et la concurrence obligera chacune à diminuer ses prix et leur profit va ainsi progressivement s'annuler. Dans cette optique, il est possible de considérer que le modèle de concurrence pure et parfaite est un cas limite où les entreprises n'innoveraient plus et proposeraient des produits similaires (voir le [chapitre 3](#)). Le profit des entreprises étant nul dans cette situation, elles essaient toujours d'éviter la concurrence, soit en construisant des barrières sur le marché soit en modifiant constamment les

produits vendus. Ce mécanisme conduit le consommateur à payer un prix plus élevé pour un service bancaire donné qu'en situation de concurrence.

Selon L. Scialom⁴, le surdimensionnement des réseaux bancaires peut se comprendre en mobilisant les modèles de concurrence monopolistique à différenciation horizontale. Un réseau bancaire désigne l'ensemble des agences dont dispose une banque sur un territoire donné. Depuis 1966, en France, le nombre de guichet par localisation n'est plus contingenté. Les banques ont eu tendance à investir massivement dans la construction d'agence afin de se trouver dans une situation de monopole au niveau local.

L'existence d'un large réseau bancaire s'explique aussi par la présence d'**externalités de réseau**. Une externalité de réseau désigne le fait que l'utilité d'un produit dépend du nombre de consommateurs qui possèdent le même bien ou utilisent le même service. Dans le cas des réseaux bancaires, les consommateurs tirent des avantages de l'existence d'un large réseau d'agences, même loin de leur domicile, ils peuvent ainsi déposer des chèques ou retirer sans frais dans toutes les agences du groupe. Les banques ont donc intérêt à construire un large réseau pour attirer plus de clients.

L'intérêt des consommateurs est d'accéder aux services bancaires pour le prix le plus bas possible, celui qui prévaut sur des marchés concurrentiels, or les prix sont plus élevés en présence d'une différenciation des produits, et une intervention publique pour diminuer les prix risque, dans ce cas de figure, d'engendrer des effets pervers. En effet, les innovations sont nécessaires à la croissance et les consommateurs apprécient l'existence d'une diversité de services, qui se réduirait si l'État imposait des tarifs plus bas. L'État peut en revanche aider les consommateurs à comparer les offres afin qu'ils aient bien conscience des différences et des similarités. Il impose ainsi aux banques depuis le 3 janvier 2008 de récapituler annuellement l'ensemble des frais de gestion liés aux services dont bénéficient leurs clients afin que les consommateurs puissent comparer les offres de façon plus précise.

2. Comportement des entreprises et structures de marché

Dans le chapitre 3, nous avons vu comment les entreprises déterminaient les quantités à produire en concurrence pure et parfaite, le prix de marché s'imposant à chacune d'elles. Si elles disposent d'un pouvoir de marché, elles peuvent décider de leur volume de production et agir sur le prix auquel elles vendent leurs biens. Le but de cette partie est d'étudier de façon plus détaillée le comportement des entreprises dans deux structures de marché particulières : l'oligopole et le monopole.

2.1 Le comportement des monopoles

Le fonctionnement du secteur bancaire n'a pas conduit à l'émergence d'un monopole naturel malgré l'importance des économies d'échelle, pour plusieurs raisons. D'abord, les autorités de contrôle surveillent les fusions entre banques afin de préserver un minimum de concurrence. Par ailleurs, les économies d'échelle disparaissent dans ce secteur au-delà d'un certain niveau de production. Les études empiriques ont montré que la courbe de coût moyen des banques avait une forme en U écrasé (voir le [chapitre 3](#) pour la définition des différents types de coûts), autrement dit le coût moyen tend à augmenter au-delà d'un certain seuil, et de ce fait une seule banque ne parvient pas à s'imposer par la réduction des coûts liées aux économies d'échelle.

Pourquoi les autorités publiques cherchent-elles généralement à éviter les situations de monopole ? Quand une entreprise est en monopole mais que le marché sur lequel elle opère est contestable, la firme adopte un comportement équivalent à celui qu'elle aurait en situation de concurrence, mais si la pression concurrentielle est moins forte, le monopoleur tire une rente sa situation au détriment des consommateurs.

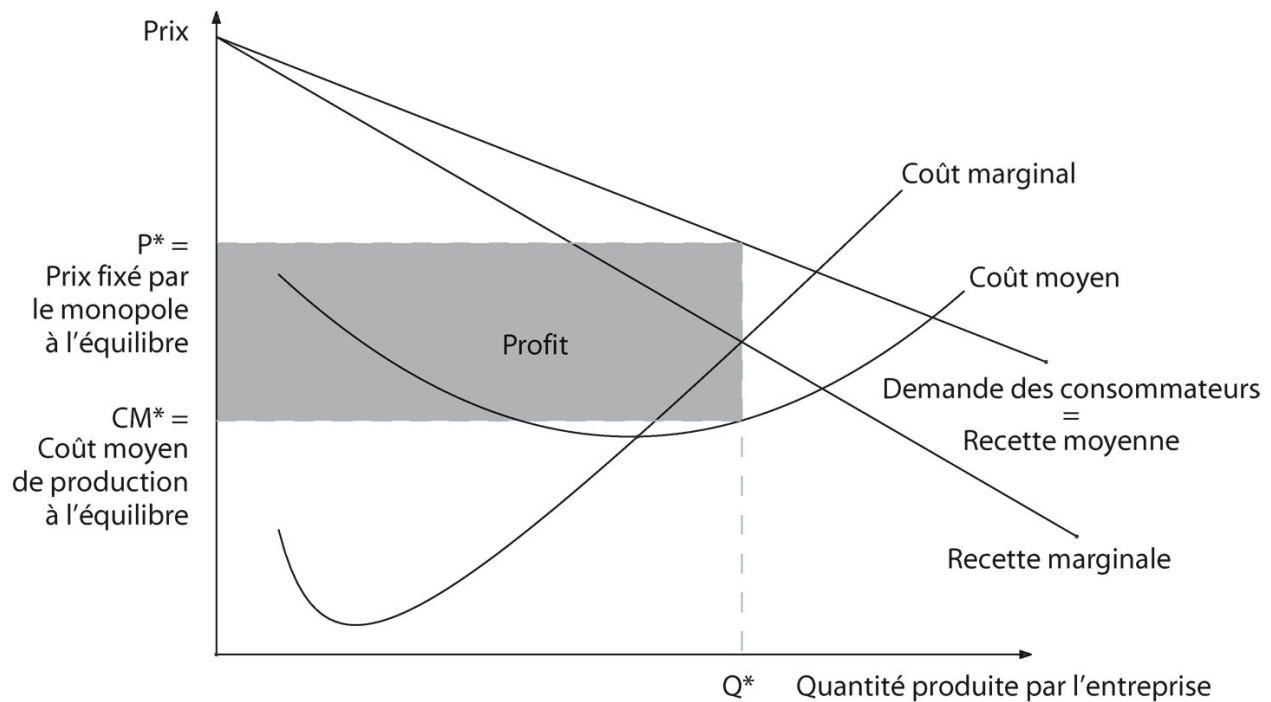
• La détermination du prix et des quantités produites

Comme en situation de concurrence, le monopoleur cherche à maximiser son

profit, soit la différence entre les recettes totales et les coûts totaux de production. Nous avons vu dans le chapitre 3 que le terme « profit » est polysémique et il est important de préciser son sens pour ne pas créer de confusion.

L'entreprise a intérêt à augmenter sa production tant que la recette marginale, c'est-à-dire la recette générée par la vente d'une unité supplémentaire, est supérieure au coût marginal, soit le coût de la dernière unité produite, car elle augmente alors son profit (voir le [chapitre 3](#)). Pour maximiser son profit, le coût marginal doit donc être égal à sa recette marginale. En concurrence pure et parfaite, la recette marginale est constante et correspond au prix fixé sur le marché. Ce n'est plus le cas en situation de monopole. L'entreprise détermine simultanément le prix de vente et la quantité produite. Le monopole doit être stratégique dans son choix et fixer ces variables en fonction de la réaction de la demande. Rien ne sert en effet de produire une grande quantité d'un bien et de le vendre très cher si personne ne l'achète. Connaissant la courbe de demande des consommateurs, le monopole sait, pour chaque niveau de production, quel prix maximal il peut demander, s'il veut écouler sa production. La courbe de demande du consommateur indique en fait la recette moyenne du monopoleur. En effet, il peut espérer vendre une quantité donnée au prix unitaire indiqué par la courbe de demande. La recette marginale n'est donc plus égale, comme en situation de concurrence parfaite, à la recette moyenne. En effet, si le producteur veut produire davantage, il doit baisser son prix de vente pour tenir compte de la demande et chaque unité supplémentaire lui rapporte moins que la précédente. Le surcroît de recette généré par la production d'une unité supplémentaire est de plus en plus faible. La droite de recette marginale est donc décroissante, et sa pente est forcément plus forte que celle de la courbe de demande comme on le voit sur le [graphique 4.1](#). Supposons par exemple que les consommateurs soient disposés à payer une certaine quantité de biens 30 €, ce prix correspond donc à la recette moyenne de la firme quand elle vend cette quantité, alors si la courbe de demande est décroissante, cela signifie que la recette moyenne diminue, donc il faut nécessairement que la recette marginale de la dernière unité vendue soit inférieure à cette recette moyenne (par exemple de 20 €), pour que cette dernière diminue avec la vente d'une unité supplémentaire.

Graphique 4.1 : Le comportement du monopole

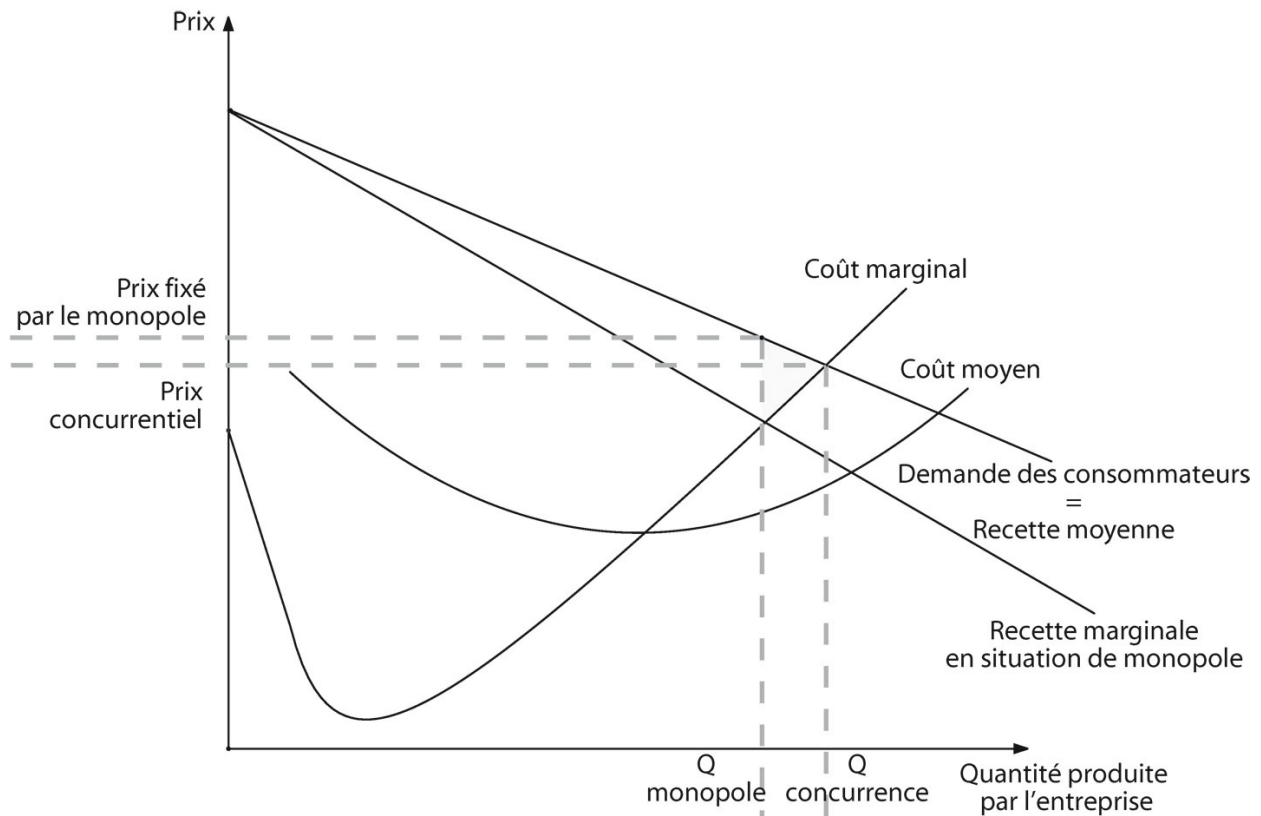


Le monopole choisit de produire la quantité Q^* pour laquelle le coût marginal est égal à la recette marginale afin de maximiser son profit (voir le [chapitre 3](#)) et il fixe le prix à P^* car cela correspond au prix maximal qu'il peut imposer aux consommateurs s'il veut écouler la quantité Q^* , puisqu'il s'agit du prix que les ménages sont disposés à payer pour cette quantité, d'après leur courbe de demande.

• Le coût social du monopole

Si le prix se fixait par le jeu de l'offre et de la demande et s'imposait à l'entreprise, comme c'est le cas sur un marché concurrentiel, l'offre du monopole serait tout à fait différente. Dans cette situation (nommée « concurrence » sur le [graphique 4.2](#)), la courbe d'offre correspond à la courbe de coût marginal selon le modèle de concurrence pure et parfaite (voir le [chapitre 3](#)). L'entreprise étant seule à produire sur le marché, l'équilibre se situe donc à l'intersection des courbes d'offre et de demande. À ce point, l'offre va se fixer sur le graphique à « Q concurrence » car il s'agit de la quantité qui maximise le profit de l'entreprise. C'est le prix que l'on observerait par exemple si une banque était seule sur un marché, mais que ce dernier était parfaitement contestable, et donc qu'elle adopterait un comportement similaire à celui sur un marché concurrentiel.

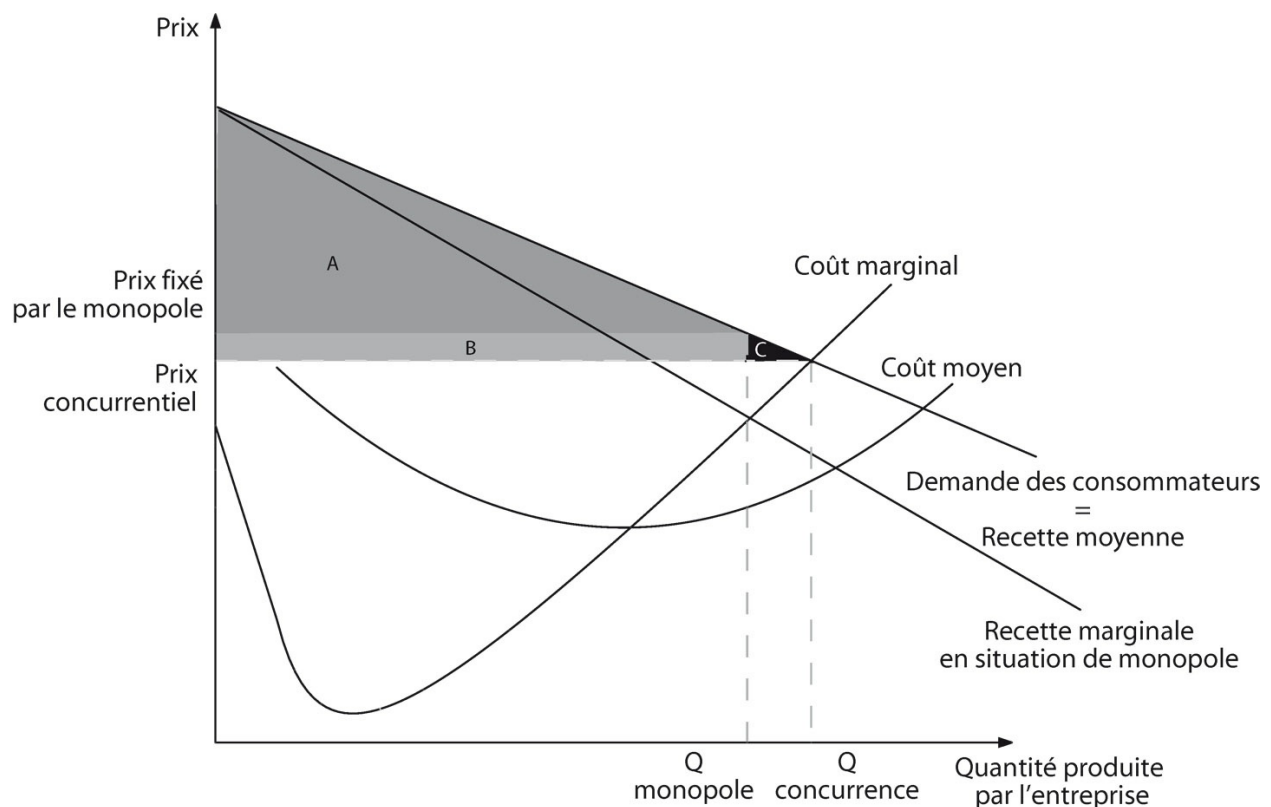
Graphique 4.2 : Monopole *versus* concurrence



En l'absence de pression concurrentielle, les quantités produites sont donc moins importantes et le prix de vente est plus élevé.

En situation de concurrence, le **surplus du consommateur** est égal à la somme des aires A, B et C (sur le [graphique 4.3](#)). En situation de monopole, le surplus du consommateur se réduit à la surface A. La zone B représente le gain pour le producteur lié à la situation de monopole. La zone C correspond à une **perte sèche** : la baisse du surplus du consommateur ne se traduit pas par une hausse égale du bien-être du producteur, et la collectivité enregistre une perte collective de surplus ce qui signifie que les gains des uns ne sont pas compensés par les pertes des autres, par rapport à la situation initiale.

Graphique 4.3 : La perte de bien-être liée au manque de concurrence



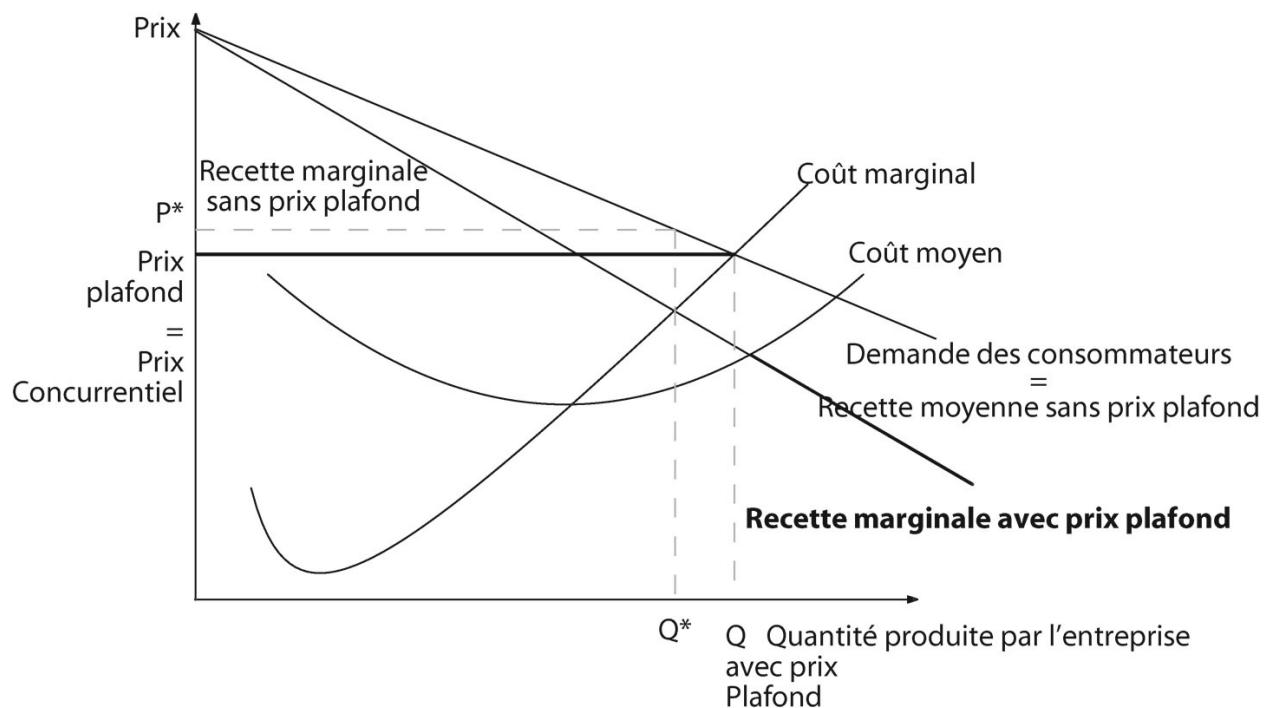
Pour défendre l'intérêt des consommateurs, les **autorités de la concurrence** essaient de maintenir une pression concurrentielle forte sur les marchés. Les banques doivent demander l'autorisation des autorités de contrôle avant de fusionner, ces dernières s'assurant que la fusion ne donnera pas naissance à un opérateur trop puissant qui pourrait abuser de sa position dominante sur le marché. Lors de la fusion des groupes Caisse d'Épargne et Banque Populaire en 2009, le conseil de la concurrence a obligé les deux parties à respecter un certain nombre d'engagements pour que cette fusion ne nuise pas à la concurrence. Les autorités de la concurrence peuvent aussi être conduites à démanteler les monopoles existants si elles n'ont pas réussi à empêcher leur émergence.

Depuis la libéralisation du secteur bancaire qui a eu lieu au début des années 1980, principalement avec la loi bancaire de 1984 en France, et l'augmentation de la concurrence entre banques qui s'en est suivie, la rentabilité des banques a diminué comme en atteste la baisse continue du taux de marge globale d'intermédiation depuis le milieu des années 1980, passant de 2,07 % en 1988 à 1,19 % en 1994⁵. Selon A. Landier et D. Thesmar (2007), l'intensification de la concurrence a réduit les rentes de situation des

banques bénéficiant de marchés protégés, et a permis le financement des firmes à un meilleur coût, tout en favorisant l'innovation financière.

Pour limiter son coût social, l'État peut aussi contrôler les prix fixés par le monopole tout en le laissant perdurer. Il existe de nombreuses règles de tarification du monopole, selon les objectifs visés et les spécificités de l'offre et de la demande. La tarification la plus connue est appelée « tarification au coût marginal » : la puissance publique impose un tarif identique à celui d'une situation de concurrence, cette situation étant la plus bénéfique pour les consommateurs.

Graphique 4.4 : L'effet d'un prix plafond



En imposant un tel prix, l'État modifie la forme des droites de recette moyenne et de recette marginale du monopole. La nouvelle droite de recette marginale est discontinue (en gras sur le [graphique 4.4](#)). La quantité qui maximise le profit du monopole (recette marginale = coût marginal) est aussi celle qui permet l'équilibre entre l'offre et la demande. Contrairement à la situation de concurrence pure et parfaite, la fixation d'un prix plafond n'implique pas forcément un rationnement de la demande en situation de monopole.

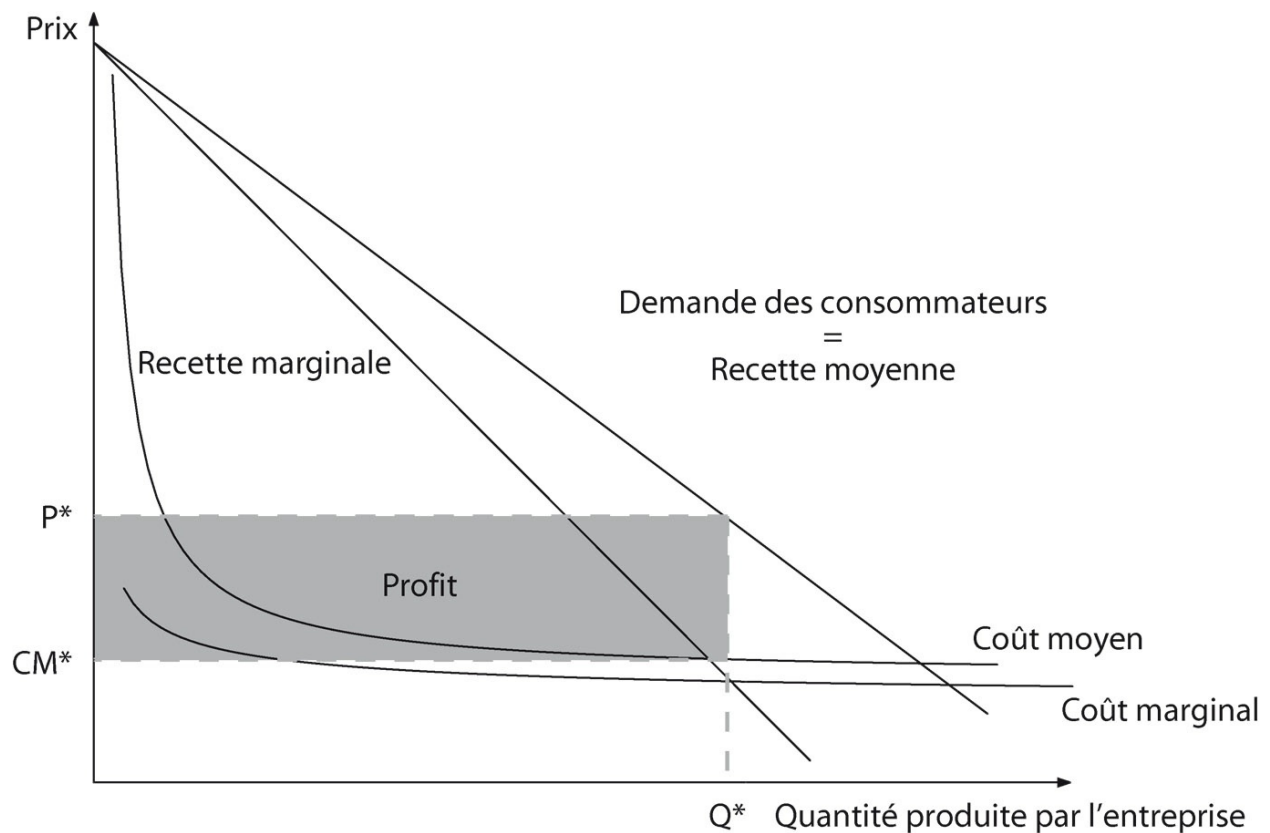
Si le prix plafond fixé est supérieur à P^* , le prix d'équilibre du monopoleur en l'absence d'intervention de l'État, les décisions du producteur ne sont pas modifiées. S'il est inférieur au prix concurrentiel, la demande sera rationnée car pour ce prix le monopole produira une quantité inférieure à la quantité demandée par les consommateurs.

• Les monopoles naturels

Un **monopole naturel** survient lorsqu'une entreprise bénéficie d'économies d'échelles suffisamment fortes sur un marché pour qu'elle puisse à elle seule répondre à la totalité de la demande ; dans ces conditions un marché sans intervention publique aboutit à la constitution d'un monopole. C'est une situation plus complexe à gérer pour les autorités car l'État n'a pas intérêt à le démanteler puisque le coût moyen augmenterait nécessairement, deux entreprises étant moins efficaces qu'une seule. L'État ne peut pas non plus imposer une tarification au coût marginal car cela entraînerait la faillite du monopole. En effet, le coût moyen est forcément supérieur au coût marginal quand il existe des économies d'échelle (voir le [chapitre 3](#)), et les recettes ne couvriraient pas les coûts de production. Mais sans aucune intervention de l'État, le monopole va fixer un tarif élevé au détriment du bien-être collectif.

Sur le [graphique 4.5](#), la courbe de coût marginal est faiblement décroissante, traduisant le fait que le coût de chaque unité supplémentaire est presque identique et relativement limité. Le coût moyen est élevé pour des faibles niveaux de production car il existe en règle générale des coûts fixes importants dans les cas de monopoles naturels. La première unité coûte très cher à produire en raison de frais d'équipement ou de recherche élevés mais ensuite les unités supplémentaires sont peu onéreuses ce qui contribue à faire diminuer le coût moyen au fur et à mesure que la quantité produite augmente. On envisage parfois de nationaliser les monopoles naturels afin que l'État contrôle leur politique tarifaire, soit en proposant des tarifs équivalents au coût marginal de production et en couvrant les pertes par des s, soit en imposant un tarif égal au coût moyen de production de telle sorte que le profit soit nul. La nationalisation des réseaux ferrés en France a été justifiée de cette manière, par exemple. Le secteur bancaire n'étant pas un monopole naturel, les nationalisations ne se sont pas posées en ces termes, mais plutôt dans le but de pouvoir contrôler les flux de financement de l'économie, notamment après la deuxième guerre mondiale.

Graphique 4.5 : Le monopole naturel



• Le monopole discriminant

Jusque-là nous avons considéré que les monopoles pratiquaient le même tarif auprès de tous leurs clients. Il est toutefois possible d'introduire une différenciation des prix en fonction des clients, on parle alors de **monopole discriminant**. La discrimination par les prix revient à proposer le même bien à des prix différents en fonction du consentement à payer des consommateurs. Ce type de stratégie n'est pas propre aux monopoles, et la **discrimination par les prix** s'observe aussi sur des marchés plus concurrentiels, si les entreprises sont en capacité de pratiquer des tarifs différenciés pour des clientèles ayant des caractéristiques distinctes. Les banques proposent parfois le même service à des prix différents selon que le client est étudiant, jeune actif, etc. Les entreprises savent en effet que tous les clients ne sont pas prêts à payer le même prix pour le même produit. En situation de concurrence, le prix s'impose à elles et elles ne sont pas en mesure d'adapter leurs prix aux clients. Quand elles ont un pouvoir de marché, elles peuvent pratiquer des taux différents afin d'accroître la clientèle et ainsi le profit de l'entreprise. Si la discrimination est parfaite, le monopole

impose à chaque consommateur le prix maximal qu'il est disposé à payer. Supposons que le coût moyen de production des services liés à une carte de crédit soit de 1 euro pour la banque par mois. Au lieu d'imposer un prix unique de 6 euros par mois à ses clients pour bénéficier d'une carte de crédit standard, elle peut ainsi demander aux 15-18 ans seulement 2 euros, 4 euros pour les étudiants, 5 euros pour les jeunes actifs et 10 euros pour les actifs plus âgés. En faisant cela, la banque vend son service aux plus jeunes qui n'auraient pas accepté de payer 6 euros par mois et accroît donc son profit tout en réalisant une marge très élevée sur les plus âgés. Une telle politique suppose que le monopole dispose d'informations très précises sur ses clients et puissent les distinguer. Ce n'est ni toujours possible, ni toujours légal : un supermarché n'a par exemple pas le droit de vendre ses produits à des prix différents selon les caractéristiques de ses clients. Précisons que la discrimination par les prix peut être aussi pratiquée par des entreprises sur un marché oligopolistique et pas seulement en situation de monopole, dès lors qu'elles ont suffisamment de pouvoir de marché pour l'imposer, comme c'est le cas dans le domaine bancaire.

2.2 Le comportement des oligopoles

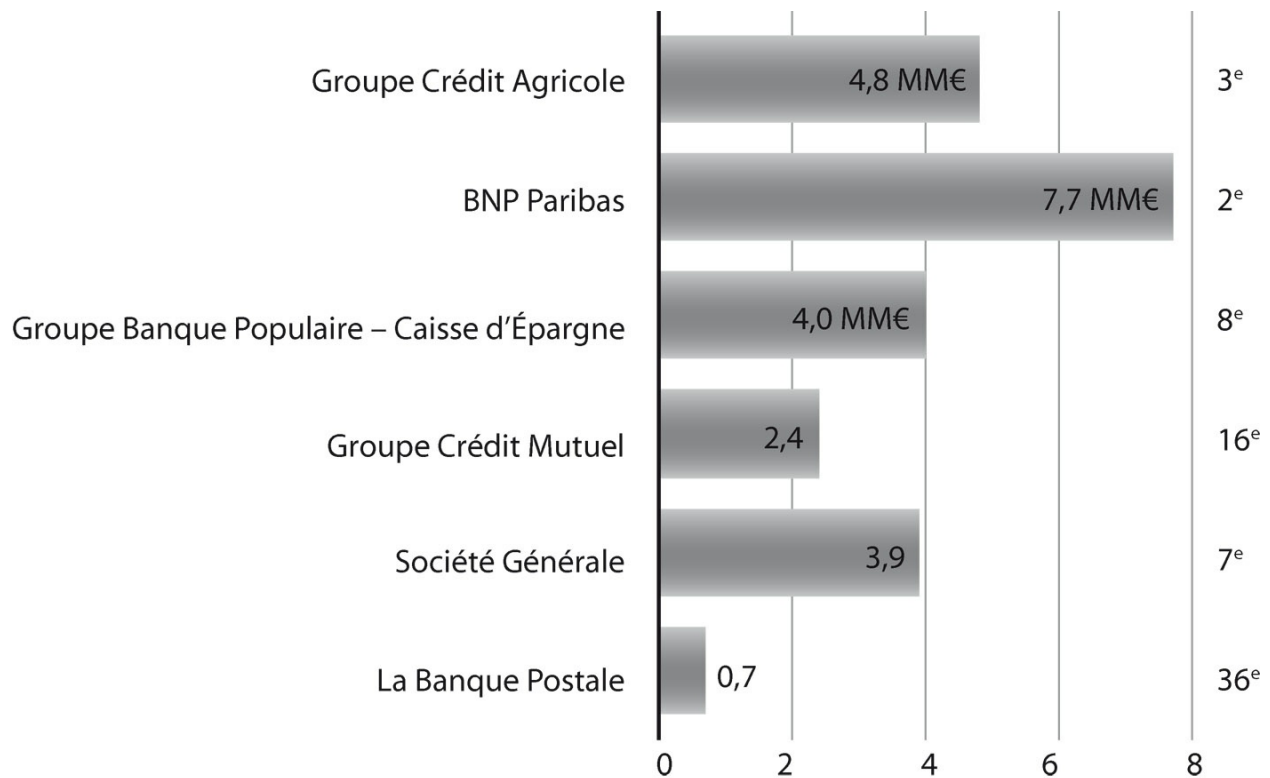
Il est rare qu'il existe un seul producteur sur un marché. Le secteur bancaire n'est pas caractérisé par une situation de monopole, mais par une structure oligopolistique : s'il existe en apparence un grand nombre de firmes, puisqu'on compte 149 établissements de crédit en France⁶, la plupart sont de petite taille ou des filiales d'un des six groupes bancaires qui dominent le marché français. Un **groupe d'entreprises** est un ensemble de sociétés liées entre elles par des participations au capital, qui en permettent le contrôle. Comme on le voit sur le **document 4.1**, quatre de ces six groupes se situent dans les « top 10 » des groupes bancaires européens, et le montant total de leur résultat net, c'est-à-dire leur bénéfice après impôt, avoisine les 23,5 milliards d'euros en 2016, soit près de 1 % du PIB pour fixer les idées.

Un marché sur lequel il y a peu de producteurs et beaucoup d'acheteurs est en situation d'**oligopole**. Une certaine concurrence existe entre les firmes mais ces dernières possèdent tout de même un pouvoir de marché, plus faible toutefois qu'en situation de monopole.

En situation d'oligopole, les entreprises peuvent adopter des stratégies très

variées selon les caractéristiques du marché et la stratégie adoptée, le profit de chacune dépendant dans tous les cas du comportement des autres.

**Document 4.1 : Le résultat net des six principaux
groupes bancaires français en 2016 et leur classement
parmi les banques européennes**



Lecture : en 2016, le Groupe BNP Paribas a réalisé un résultat net de 7,7 milliards d'euros, et s'est classé 2^e groupe bancaire européen en termes de volume d'actifs.

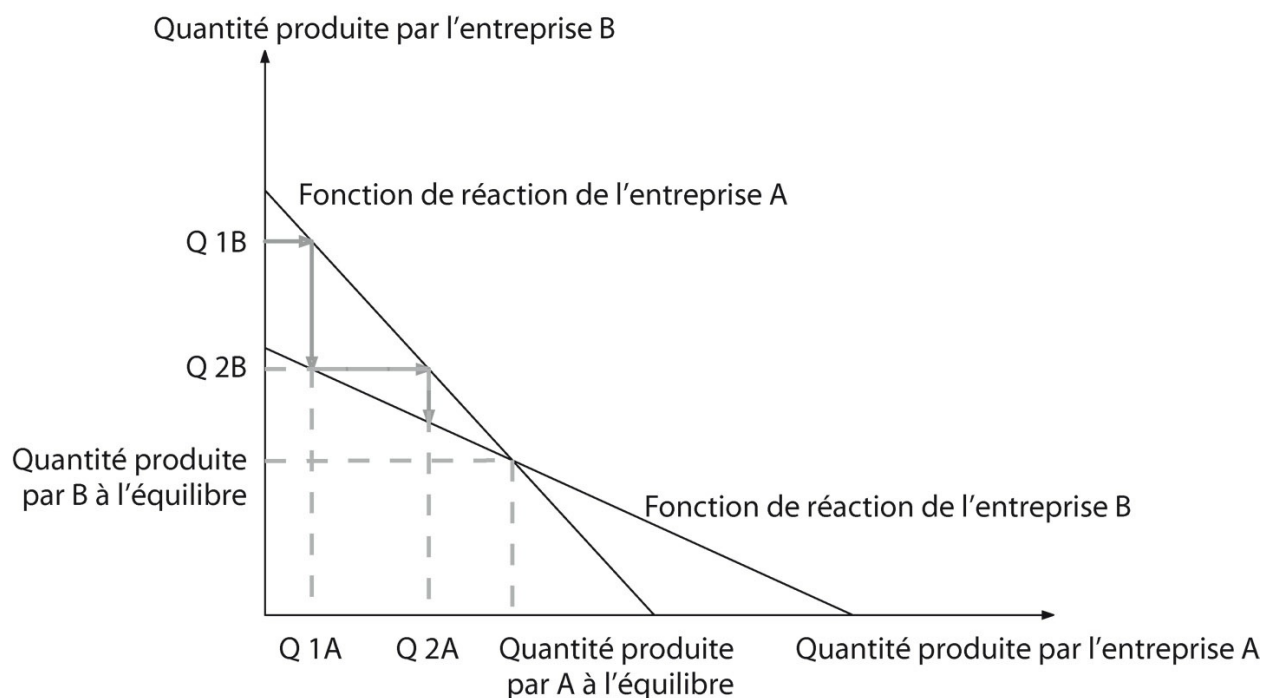
Sources : Classement 2016 des 50 premiers groupes européens selon Standard & Poors.

• Le duopole

Le **duopole**, marché sur lequel s'affrontent deux entreprises, est le cas le plus simple d'oligopole et a donné lieu à une littérature abondante, pour rendre compte des stratégies des firmes, chacune devant intégrer l'action de l'autre dans ses choix.

Le modèle de A. Cournot (1838) analyse un duopole dans lequel les deux entreprises se font concurrence sur les quantités, le prix de vente étant fixé sur le marché en fonction de l'offre globale. Chaque producteur considère le volume de production de l'autre entreprise comme donné et détermine la quantité qu'il va produire pour maximiser son profit. On appelle fonction de réaction d'une entreprise A la relation qui permet de déterminer la quantité produite de l'entreprise A en fonction de celle de l'entreprise B. Sur le [graphique 4.6](#), les deux fonctions de réaction sont présentées en indiquant sur les axes les quantités produites par chacune d'entre elles.

Graphique 4.6 : Le duopole de Cournot



Supposons que l'entreprise A anticipe la production de la quantité Q_{1B} par sa concurrente B. De ce fait, elle décide de produire la quantité Q_{1A} d'après sa fonction de réaction, car cela permet de maximiser son profit. Voyant que l'entreprise A produit la quantité Q_{1A} , l'entreprise B décide de produire la quantité Q_{2B} dans la même logique. L'entreprise A réagit alors en produisant la quantité Q_{2A} . Ce mécanisme continue jusqu'à une situation d'équilibre où les quantités produites par chaque entreprise maximisent le profit de chacune d'entre elles. La quantité globale produite sur le marché est égale à la somme de la quantité produite par A et B et le prix de vente découle de la rencontre de cette offre avec la demande. On peut démontrer que la quantité globale produite est inférieure au volume de production en situation de concurrence car ce dernier ne correspond pas à une situation qui maximise le profit des entreprises en duopole. Le prix de vente est ainsi plus élevé qu'en situation de concurrence : les producteurs limitent leur offre afin de pouvoir maintenir un profit élevé. Il n'y a pas cependant d'entente directe entre eux mais ils tiennent simplement compte de la réaction de leur concurrent. Le modèle de Cournot permet donc de comprendre qu'un duopole est moins favorable aux consommateurs qu'une situation plus concurrentielle, même si les entreprises se livrent une réelle compétition.

Toutefois, l'hypothèse d'un prix unique pratiqué par les deux firmes ne paraît pas la plus réaliste : en situation de duopole, deux banques peuvent très bien décider de pratiquer un tarif plus bas que l'autre pour conquérir des parts de marchés. Le modèle de J. Bertrand (1883) est basé sur ce principe. Il fait l'hypothèse que la stratégie des entreprises ne porte pas sur les quantités mais le prix de vente. Les duopoleurs mènent ici une guerre des prix dans le but d'accaparer toute la demande en faisant baisser leur prix de vente. Dans ce cas, la concurrence entre les entreprises conduit à ce que le prix de vente soit égal au coût marginal de production : chacune des deux banques peut baisser son propre tarif tant qu'il est supérieur à son coût marginal et qu'elle réalise encore un profit. On retrouve le résultat observé sur les marchés concurrentiels. C'est le **paradoxe de Bertrand** : même sur un marché sur lequel n'opèrent que deux producteurs, sans entente entre eux, l'équilibre du marché est équivalent à celui d'une situation de concurrence pure et parfaite. Le modèle de Bertrand nous permet ainsi de comprendre pourquoi les entreprises essayent de passer des accords pour éviter de se faire directement concurrence, car une guerre des prix ne leur est pas bénéfique. Les modèles de Cournot ou de Bertrand restent cependant très simplifiés, car ils n'envisagent que des situations où les deux firmes ont exactement la même stratégie, et ne s'adaptent pas à la stratégie suivie par l'autre. Or la confrontation de deux entreprises donne souvent lieu à de multiples rebondissements, et c'est pourquoi ces modèles, issus des premiers travaux néo-classiques au XIX^e siècle, ne sont pas suffisants pour comprendre les duopoles.

• L'apport de la théorie des jeux

La **théorie des jeux** étudie d'une façon générale les comportements stratégiques des agents en fonction des objectifs qu'ils cherchent à atteindre et des stratégies des autres joueurs. Le développement de cette branche de l'économie a révolutionné l'étude du comportement des acteurs, notamment en présence d'asymétrie d'information. On peut définir un **jeu** par la confrontation des choix stratégiques d'au moins deux agents, qui peuvent porter sur tous types de décisions économiques : produire ou ne pas produire, ouvrir ou non une filiale à l'étranger, pratiquer un prix élevé ou bas, taxer ou non un produit, etc. Chaque agent choisit au minimum entre deux stratégies, et un jeu associe à chaque combinaison de stratégies des différents agents les

gains et les pertes que chacun subit. On le voit, la théorie des jeux a une portée bien plus générale que les modèles précédents, parce qu'elle peut s'appliquer à de multiples interactions entre agents, et que les décisions ne portent pas uniquement sur le prix et la quantité produite. Appliquée à l'étude du comportement des entreprises, la théorie des jeux permet de comprendre la diversité des stratégies que ces dernières mettent en œuvre pour maximiser leur profit. La notion de stratégie d'entreprise a un sens quand il existe un pouvoir de marché des entreprises et qu'elles peuvent donc influencer l'équilibre du marché par leur comportement.

Les entreprises peuvent alors s'entendre entre elles afin de préserver leur profit : on dit qu'il y a **entente illicite** ou **collusion** dans ce cas de figure, lorsque l'entente entre les entreprises enfreint la loi, et pénalise les consommateurs en leur imposant des conditions plus défavorables qu'en situation de concurrence. Un **cartel** est une entente illicite au sein de laquelle les entreprises fixent entre elles la quantité produite et/ou le prix pratiqué par chacune : en se coordonnant elles parviennent à éliminer la concurrence. Les collusions sont interdites en Europe par le droit de la concurrence, mais force est de constater qu'elles restent fréquentes, car il est très avantageux pour les firmes de s'entendre et d'éviter une guerre des prix. En 2010, l'autorité de la concurrence a ainsi condamné en France 11 banques pour entente illégale. Ces dernières ont dû acquitter une amende globale de 384,9 millions d'euros pour s'être concertée afin d'imposer pendant 5 ans une commission interbancaire sur le chèque. Les montants des amendes paraissent souvent très élevés, mais ils sont calculés en fonction du préjudice subi par les consommateurs, et les études montrent qu'ils restent souvent sous-évalués. Selon E. Combe⁷ les amendes sont très souvent inférieures au profit réalisé par les cartels. Au-delà des accords explicites entre entreprises, ces dernières peuvent se comporter comme s'il y avait entente. Une banque annonce par exemple une augmentation de tarif et choisit de la réaliser ou non, en fonction de la réaction des autres entreprises : si elles annoncent à leur tour des augmentations, tous les offreurs vont augmenter leur prix accroissant ainsi leur profit sans perte de compétitivité.

Les collusions sont difficiles à maintenir dans le temps car elles supposent des **coûts de surveillance** très importants pour s'assurer que chaque entreprise respecte sa part de l'accord (problème du principal-agent, voir le chapitre 3). En effet, selon les situations, les entreprises ont parfois intérêt à tricher pour accroître leur profit. La théorie des jeux a permis d'éclairer de

façon très stimulante dans quelles configurations les accords étaient stables, aucune partie contractante n'ayant intérêt à en dévier.

En situation d'oligopole, le comportement de chaque entreprise est à même de modifier l'équilibre du marché et les entreprises intègrent les stratégies mises en œuvre par leurs concurrentes pour décider de leur propre stratégie. On distingue deux grands types de jeux : les **jeux coopératifs**, dans lesquels les participants décident collectivement de la stratégie que chacun doit mettre en œuvre, et les **jeux non coopératifs**, où chaque joueur détermine sa stratégie isolément. Les premiers travaux en théorie des jeux sont anciens (publication en 1944 du livre fondateur de cette théorie par O. Von Neumann et J. Morgenstern, *Theory of games and economic behaviour*) mais l'intérêt pour cette approche s'est vraiment développé depuis la fin des années 1980.

Considérons un cas simple où seules deux banques (A et B) opèrent sur le marché et décident isolément de la stratégie à mener : c'est donc un jeu non coopératif. Le jeu présenté ici est appelé un « **dilemme du prisonnier** », modélisé par A.W. Tucker en 1950, à partir des travaux de J. Nash. Un exemple est présenté dans le **document 4.2**.

Partant d'une situation donnée où les prix de vente et les profits des deux entreprises sont faibles, les entreprises ont intérêt à relever leurs prix simultanément afin d'accroître leur profit sans perdre en compétitivité. Le tableau récapitule les gains réalisés par chaque banque en fonction de la stratégie de l'autre entreprise, chaque case donnant les gains respectifs de B (chiffre de gauche) et de A (chiffre de droite).

Document 4.2 : Le dilemme du prisonnier

		Banque A	
		Augmentation des prix	Prix inchangé
Banque B	Augmentation des prix	(15 ; 15)	(0 ; 20)
	Prix inchangé	(20 ; 0)	(3 ; 3)

Si la banque A augmente seule ses prix, tous les consommateurs vont

s'adresser à la banque B qui propose des tarifs plus attractifs. La banque A verra ses gains s'annuler alors que ceux de la banque B augmenteront fortement (gains de 20 dans le tableau). Si les deux entreprises augmentent simultanément leur prix, elles en bénéficient toutes les deux et réalisent chacune un profit de 15. La théorie des jeux montre pourquoi cette solution ne va pas émerger, en l'absence de coopération. En effet, quand la banque A décide de sa stratégie, elle tient compte de la position de B. Elle considère tout d'abord le cas où l'entreprise B laisse ses prix inchangés : pratiquer des prix inchangés est alors une meilleure stratégie pour A (gains de 3 plutôt que 0). Elle considère ensuite le cas où l'entreprise B augmente ses prix : maintenir ses prix inchangés est également la meilleure stratégie (gains de 20 plutôt que 15). Quelle que soit l'action de la banque B, la banque A a intérêt à laisser ses prix inchangés, on parle alors de stratégie dominante, c'est une stratégie toujours plus avantageuse pour le joueur considéré, quel que soit le choix de l'autre joueur. La stratégie « augmentation des prix » est dite à l'inverse dominée : elle est toujours moins avantageuse que la stratégie alternative, quelle que soit la stratégie adoptée par l'autre. La banque B adopte le même raisonnement que la banque A, et dans ces conditions les interactions sur le marché imposent la stratégie des prix bas pour les deux acteurs, même si elle n'est pas la meilleure solution pour chacune d'entre elles, si elles avaient la possibilité de coopérer. Il s'agit d'un **équilibre de Nash**, c'est-à-dire une situation stable où aucun acteur n'a intérêt à changer seul de stratégie compte tenu de celles des autres acteurs. Cet équilibre est sous optimal car il serait possible d'améliorer la situation de la banque A tout en augmentant celle de la banque B si elles choisissaient une autre combinaison de stratégies. Face à une situation de type « dilemme du prisonnier », les banques pourraient s'entendre et déterminer ensemble leur stratégie. Pourtant dans ce cas, aucune entreprise n'aurait intérêt à respecter son engagement puisque son profit serait très élevé si l'autre banque est la seule à augmenter ses prix. En l'état, l'entente est donc très improbable. Toutefois, nous avons considéré jusqu'à présent que les entreprises prenaient une décision une fois pour toutes, or un changement de cadre temporel modifie la donne stratégique. Si le jeu se déroule sur plusieurs périodes, il est moins évident que les entreprises n'aient pas intérêt à respecter leur accord car elles savent que si elles ne le font pas, à la période suivante l'autre entreprise baissera ses prix en réaction contribuant ainsi à la baisse du profit de l'entreprise qui a dévié en premier. On parle de **jeux dynamiques** quand

on intègre l'existence de plusieurs périodes.

Face aux risques de collusion, les autorités de la concurrence ont cherché à mettre en place une législation afin d'inciter les entreprises à ne pas respecter les accords qu'elles auraient pu éventuellement passer. En effet, bien que les cartels soient interdits, de nombreuses collusions existent et l'Union européenne a mis en place un programme de clémence qui donne l'immunité (ou au moins une réduction forte de la peine) à la première entreprise qui dénonce l'entente créant ainsi une incitation à la dénonciation, chaque entreprise risquant d'être dénoncée avant d'avoir dénoncé elle-même.

Il existe de nombreux autres modèles dans la théorie des jeux, qui éclairent la complexité des interactions stratégiques, et les processus qui conduisent les agents à prendre des décisions qui ne sont pas toujours conformes à ce qu'un individu parfaitement rationnel devrait faire. La théorie des jeux présente l'avantage de proposer un cadre de réflexion qui facilite l'évaluation expérimentale, en modélisant des processus de décision et en confrontant ensuite la solution « théorique » du jeu, celle que choisirait un agent parfaitement rationnel, et les stratégies adoptées par les agents participant à une expérience (voir le [chapitre 2](#) pour une présentation du raisonnement expérimental). Prenons le cas du « jeu de l'ultimatum »⁸ : on suppose que le joueur A reçoit une somme de 100 € à répartir entre lui et le joueur B. Il décide unilatéralement de la répartition, et le joueur B peut ensuite accepter ou refuser cette répartition : s'il accepte « l'ultimatum » le partage a lieu selon la proposition du joueur A, s'il le refuse le gain de chaque joueur est nul. La solution théorique du jeu de l'ultimatum, c'est qu'un joueur A parfaitement rationnel devrait proposer 1 € à l'autre joueur, et garder les 99 € restant, comptant sur le fait que le joueur B a intérêt à accepter un gain minime de 1 € plutôt que d'empocher 0 € en cas de refus. Mais les nombreuses expériences menées sur la base de ce jeu montrent que les deux joueurs ne se comportent pas ainsi : en moyenne, le joueur A propose à l'autre 40 % du total, le joueur B refuse une fois sur deux une part des gains inférieure à 20 %, et il refuse quasi systématiquement la solution théorique (le plus petit partage possible). Ce type d'expérience met en évidence l'importance du sentiment d'équité dans les négociations, et permet de comprendre que les agents en tiennent compte et attendent des autres un traitement plus équitable que ce que le strict rapport de force permet d'imposer. Les difficiles négociations entre l'État grec et ses créanciers, principalement des banques européennes, notamment allemandes, suite à la

crise bancaire de 2008 en offrent une illustration. Une faillite de l'État grec n'est favorable ni pour lui, ni pour ses créanciers, mais ces derniers ne peuvent imposer pour autant que la totalité de la dette soit remboursée par les Grecs, sans créer un profond sentiment d'injustice dans l'opinion. La seule solution, pour éviter que les deux « joueurs » supportent le coût d'une faillite, c'est de négocier une restructuration de la dette grecque, comme ce fut le cas en 2012, avec un abandon de 53,5 % de la dette détenue par les créanciers privés.

La prise en compte, grâce à la théorie des jeux notamment, de la psychologie des agents financiers et de la manière dont leur comportement effectif s'éloignait de celui d'agents parfaitement rationnels est à l'origine de la **finance comportementale**, qui s'est développée depuis les années 1970 et permet de mieux comprendre les conséquences des imperfections des marchés financiers sur les comportements des agents, à l'image des travaux de Robert Shiller, Prix Nobel en 2013. Avec George Akerlof, il met en évidence dans *Les esprits animaux* (2009) l'importance de la psychologie des acteurs, et notamment de leurs émotions et de leur sentiment d'équité, pour comprendre l'instabilité intrinsèque des marchés financiers, et la nécessité d'une régulation des activités bancaires et financières.

• Encadré 4.2 : La politique de la concurrence

La politique de la concurrence est un type de politique structurelle dont l'objectif est de maintenir ou instaurer la concurrence sur un marché. Elle vise à contrôler les ententes entre entreprises (cartels, etc.) et les concentrations ; elle lutte contre les abus de position dominante et contribue à la diffusion de l'information à destination des consommateurs afin qu'ils puissent comparer de façon précise la qualité des différents produits qui leur sont proposés à la vente. La politique de la concurrence émerge tardivement en Europe par rapport aux États-Unis, suite à l'adoption du traité de Rome en 1957, qui pose les bases d'un marché commun et donne à la Commission Européenne un pouvoir de régulation de la concurrence. Il faut attendre les années 1980 pour que la France développe une véritable politique de la concurrence avec, entre autres, la mise en place du Conseil de la concurrence en 1986.

Le pouvoir de marché des entreprises peut être limité en veillant à l'atomicité du marché (nombreux vendeurs et demandeurs) ou en s'assurant que la pression concurrentielle est forte sur les marchés oligopolistiques ou monopolistiques (suppression des barrières à l'entrée ou à la sortie des marchés afin qu'il n'y ait pas abus de position dominante). De ce fait, la politique de la concurrence permet une baisse des prix et contribue à l'augmentation du pouvoir d'achat des consommateurs. Soumises à une concurrence accrue, les entreprises sont poussées à innover et à être efficaces pour rester compétitives. La politique de la concurrence génère donc à long terme des gains de productivité et est favorable à la croissance et à l'emploi. Il y a cependant un débat relatif à l'articulation des politiques de la concurrence et des politiques de croissance. Dans la mesure où la

croissance repose sur les rendements croissants et sur l'innovation (protection de la propriété industrielle), elle suppose donc dans une certaine mesure une limitation de la concurrence (voir le [chapitre 3](#)).

La politique de la concurrence lutte contre les **accords horizontaux**, notamment la constitution de cartels, et les **accords verticaux**, qui concernent les relations entre une entreprise et ses fournisseurs ou ses distributeurs. Ainsi, une entreprise ne peut imposer à ses distributeurs un prix de revente aux consommateurs. À l'inverse, la loi protège les fournisseurs qui se verraient imposer par une entreprise très puissante des prix d'achat trop bas, comme des centrales d'achat des supermarchés vis-à-vis de leurs fournisseurs.

Les autorités de la concurrence contrôlent aussi les concentrations. Les fusions d'entreprises qui pourraient conduire à l'émergence d'une firme disposant d'un trop fort pouvoir de marché sont interdites. Elles peuvent imposer le démantèlement d'une entreprise dont la taille est trop importante.

3. Les asymétries d'information

Le modèle de concurrence pure et parfaite est basé sur l'hypothèse selon laquelle tous les agents disposent de l'information nécessaire à leurs décisions sans aucun coût, ils connaissent parfaitement les caractéristiques des transactions sur le marché. Que se passe-t-il quand cette condition n'est pas vérifiée ? Une **asymétrie d'information** désigne une situation où l'information n'est pas distribuée de façon uniforme entre les parties prenantes. Par exemple, un banquier cherche à vendre un produit financier à un client qui en connaît mal les contours. Le banquier en sait plus sur le produit que l'acheteur potentiel. Les marchés fonctionnent-ils toujours efficacement quand il existe des asymétries d'information entre vendeurs et acheteurs ?

3.1 Le risque d'antisélection

Lors de la réalisation d'une transaction, il est important que chaque partie évalue la fiabilité de l'autre partie prenante et la qualité du produit échangé. Le risque de sélectionner les mauvais produits ou partenaires du fait d'une asymétrie d'information est appelé risque **d'antisélection** (ou **sélection adverse**). Les banques font face à ce problème de façon récurrente quand elles proposent des offres de crédit à leurs clients : il leur faut impérativement évaluer de façon pertinente la solvabilité de ces derniers. En gérant les comptes de leurs clients sur la durée, les banques disposent de ce point de vue d'une situation privilégiée pour évaluer le risque de défaut. L'historique des opérations de leurs clients leur donne en effet de précieuses informations pour évaluer la situation financière des demandeurs de crédit. Le rôle d'intermédiation de bilan des banques, c'est-à-dire la collecte de fonds auprès d'agents à capacité de financement pour octroyer des crédits aux agents à besoin de financement, est en fait lié à l'existence d'asymétries d'information. En effet, les agents à capacité de financement peuvent être réticents à prêter directement aux agents à besoin de financement *via* les marchés financiers car ils disposent de très peu d'informations sur ces derniers. Quand ils décident toutefois de le faire, ils font ainsi souvent appel à

des conseillers pour les guider dans leurs choix et suivent les évaluations fournies par les agences de notation sur la qualité des différents emprunteurs. L'intermédiation de bilan pratiquée par les banques leur permet d'utiliser au mieux les informations dont elles disposent sur les emprunteurs.

Aucune évaluation n'est toutefois parfaite et celles pratiquées par les banques conduisent parfois à des dysfonctionnements. J. Stiglitz et A. Weiss (1981) ont montré qu'en fixant des taux d'intérêts élevés pour décourager les emprunteurs peu fiables, les banques allaient en fait attirer davantage les clients dont le projet est très risqué. Les autres, qui savent que leur projet est fiable refuseront de payer des sommes importantes compte tenu de la qualité de leur projet. Ainsi quand le taux d'intérêt augmente, le risque de crédit moyen augmente aussi et il se produit un phénomène d'antisélection : les agents présentant de « bon risques » sont évincés par des taux trop élevés. Les banques n'ont donc pas intérêt de proposer le prix le plus haut possible, cela peut nuire à leur rentabilité s'il en découle une plus grande fréquence de défauts de remboursement. Avec un taux d'intérêt plus modéré les banques maximisent leur profit, mais ce taux ne permet pas forcément d'égaliser l'offre et la demande : à ce niveau la demande sera plus forte que l'offre et aucun mécanisme ne va conduire à l'égalisation de ces deux valeurs car les banques n'ont pas intérêt à augmenter leur prix pour faire diminuer la demande. C'est une situation de **rationnement du crédit**, c'est-à-dire d'ajustement par les quantités et non par les prix.

L'existence de relations de clientèle durables est donc nécessaire pour réduire l'asymétrie d'information envers leurs clients, ce qui limite les inefficiences découlant des imperfections de l'information entre prêteurs et emprunteurs.

Les asymétries d'information peuvent même conduire à la disparition du marché. Un exemple célèbre a été donné par G. Akerlof (1970), qui a étudié le marché de la voiture d'occasion appelé « *lemons* » en argot. Les acheteurs ne pouvant évaluer parfaitement les voitures proposées à la vente, ils refusent de payer un prix trop élevé pour une voiture d'occasion. Les détenteurs de voiture en bon état vont de ce fait se retirer du marché, préférant vendre leur voiture par d'autres réseaux, par exemple grâce à des concessionnaires qui garantissent les voitures d'occasion, ou même la garder. Seules les voitures les moins fiables restent sur le marché, puisque les propriétaires des voitures de bonne qualité se retirent progressivement, et la qualité moyenne des

véhicules se dégrade dans ces conditions. Constatant les déconvenues de certains acheteurs, les autres risquent de perdre confiance dans le marché et renoncer à acheter une voiture d'occasion contribuant peu à peu à la disparition du marché, à mesure que les clients se tournent vers des réseaux de distribution plus fiables.

Un mécanisme analogue a joué dans la crise de liquidité en 2008-2009, et a conduit à la paralysie du marché interbancaire suite à la crise des *subprimes*. Chaque banque a besoin de liquidités en monnaie centrale, c'est-à-dire en monnaie émise par la banque centrale, pour faire face à ses engagements vis-à-vis de ses clients et de leurs banques. Le **marché interbancaire** est le marché réservé aux prêts de liquidité entre les banques. Quand une banque veut emprunter des liquidités à une autre, elle doit offrir en garantie des titres d'un montant équivalent à la somme empruntée. La banque qui octroie le prêt garde ses titres en pension pendant la durée du prêt et les rend à l'autre banque quand cette dernière rembourse son crédit. L'évaluation de la qualité d'un titre est difficile du fait de l'existence d'asymétries d'information. La crise des *subprimes* a renforcé la méfiance des banques envers les titres des autres banques, aucune institution n'arrivant à identifier clairement les risques associés à chaque banque. En effet, les produits structurés issus de la titrisation des crédits *subprimes* sont très opaques et mélangent en leur sein des crédits de bonne qualité à des crédits de faible qualité. Les agences de notation ont d'ailleurs été largement critiquées pour leur évaluation défaillante de ces titres. De ce fait les banques qui disposaient de fonds prêtables ne voulaient plus en faire bénéficier les banques qui en avaient besoin. Il a fallu que la banque centrale intervienne pour restaurer la confiance entre les banques. Sans cela, le marché risquait de s'effondrer (MARTEAU, 2016).

Les entreprises ne sont pas en reste pour tenter d'accentuer les asymétries d'information et l'avantage qu'elles leur donnent, en matière financière tout particulièrement. Comme le montrent G. Akerlof et R. Shiller dans *Marchés de dupe* (2016), les asymétries d'information sont en partie produites par des stratégies délibérées des grandes firmes afin de rendre le marché plus opaque et de limiter la capacité des consommateurs à procéder à des évaluations cohérentes des produits proposées, en augmentant par exemple artificiellement leur complexité. L'existence d'asymétrie d'information nuit au final à l'efficacité du marché et la puissance publique doit intervenir pour en améliorer le fonctionnement. L'État peut contribuer à réduire l'asymétrie

d'information en imposant plus de transparence. La création de labels ou de dispositifs de certification est un exemple de ce type d'intervention. L'État oblige par exemple les banques à détailler auprès de leurs clients tous les risques associés aux produits vendus et prévoit un délai légal de rétractation le temps que le client réfléchisse à son engagement et lise tous les documents fournis par la banque afin que son choix s'opère en toute lucidité.

3.2 Le risque d'aléa moral

Un agent adopte un comportement **opportuniste** s'il ne tient pas ses engagements ou dissimule volontairement des informations qui pourraient jouer en sa défaveur. Ce type de comportement apparaît quand il existe une asymétrie d'information. Le risque d'**aléa moral** survient quand une des parties prenantes adopte un comportement opportuniste à la suite d'une transaction entre elles, du fait qu'il ne peut être contrôlé parfaitement. Il s'agit d'un risque *ex post* alors que le risque de sélection adverse a lieu *ex ante*. Une banque peut craindre que le client auquel elle a octroyé un crédit n'utilise pas la somme pour l'usage prévu et accroisse ainsi son risque de défaut. Prenons l'exemple d'un particulier qui demande à sa banque un crédit de 20 000 euros pour acheter un parking qu'il projette de louer. La banque peut intégrer le revenu supplémentaire lié à la location du parking, et en cas de problème elle pourra toujours faire saisir le garage. Si la banque accorde le prêt et que le client s'en sert pour financer un voyage autour du monde avec sa famille, il a fait preuve d'opportunisme. Il profite du fait que le banquier ne peut contrôler l'utilisation de la somme prêtée. Les banques ont mis en place des techniques pour surveiller plus efficacement leurs débiteurs (techniques de monitoring). Les agents chargés de compte reçoivent ainsi des alertes quand un débiteur se comporte d'une façon préoccupante (par exemple en cas de dépassement d'autorisation de découvert) et ces derniers vont entrer immédiatement en contact avec le client en question pour l'inciter à éviter un défaut de paiement. Le chargé de compte peut par exemple menacer une entreprise de refuser le paiement d'un effet si elle ne lui délivre pas des informations plus précises, et ne fait pas d'effort pour améliorer sa situation.

Le risque d'aléa moral existe dans de nombreuses situations et son importance dépend pour beaucoup de l'environnement institutionnel.

A. Landier et D. Thesmar⁹ ont ainsi montré que la libéralisation du secteur bancaire dans les années 1980 a privé les banques du filet de sécurité étatique les conduisant à être beaucoup plus soucieuse de la qualité des crédits octroyés, les banques n'étant plus systématiquement renflouées en cas de difficultés financières, liées notamment à des défauts de paiement. Le changement de la législation a rendu moins intéressant pour les banques l'adoption d'un comportement risqué en matière de crédit. L'aléa moral reste fort car les banques centrales viennent souvent en aide aux banques de grande taille lorsqu'elles sont en difficultés pour éviter une crise systémique. Les banques importantes sont couvertes par le principe du *Too big to fail* : elles sont tentées d'adopter des comportements très risqués car en cas de réussite, elles bénéficient de gains substantiels alors qu'en cas de perte, elles sont aidées par la puissance publique, qui estime que leur faillite ferait courir un risque trop grand à l'ensemble des agents. Un tel comportement opportuniste est possible car il est très difficile d'évaluer de l'extérieur le risque pris par les banques et la banque centrale ne peut ainsi qu'imparfaitement limiter les prises de risque.

Selon D. Marteau¹⁰, le concept d'aléa moral est indispensable pour comprendre la crise économique qui a débuté en 2007 et ce pour deux raisons majeures.

Tout d'abord, les banques ont été incitées à être peu regardantes sur la qualité des crédits qu'elles octroyaient car elles pouvaient les céder à des investisseurs à travers des opérations de titrisation. Les banques ne supportaient ainsi pas le risque découlant des prêts octroyés. Le sommet du G20 à Pittsburgh en 2009 en a conclu qu'il fallait réduire l'aléa moral sur les marchés financiers. Une des propositions du sommet est que les banques à l'origine d'une opération de titrisation conservent une fraction des prêts accordés les plus risqués. L'idée de cette mesure est que les banques seraient incitées à être plus attentives à la qualité des créances titrisées car elles seraient concernées par un éventuel défaut de paiement. D'autre part, le système de rémunération des opérateurs de marché au sein des banques leur attribue souvent des primes en cas d'opérations réussies mais aucune sanction en cas de perte. D'où le risque élevé d'aléa moral, car ce système incite les opérateurs à prendre des risques excessifs pour augmenter leurs primes, quitte à user de dissimulation ou de mensonges.

La crise de 2008 a mis en évidence les limites du contrôle exercé sur les

banques par les autorités bancaires européennes, jusqu'à présent différentes pour chaque pays, ce qui imposait un comportement asymétrique, les grandes banques européennes jouant de leur implantation dans plusieurs pays, alors que chaque régulateur n'agissait que sur son marché national. D'où le projet d'Union bancaire européenne adopté en 2014, qui consiste à créer un « mécanisme de supervision unique » (MSU) des banques implantées au sein de la zone euro, chargé de les contrôler et de les sanctionner. Progressivement, ce mécanisme est complété par un « mécanisme de résolution unique », un fonds de garantie en cas de faillite bancaire, financé par les banques elles-mêmes, afin de minimiser le recours à l'argent public pour sauver une banque « *too big to fail* » (BEITONE et RODRIGUES, 2017).

Pour limiter le risque d'aléa moral, il faut réduire l'asymétrie d'information, mais tout contrôle a un coût et il est impossible de les supprimer complètement. Plutôt que de chercher à les contrôler, certains économistes proposent de concevoir des contrats incitant les individus à en respecter les termes et à renoncer à adopter un comportement opportuniste. On parle de **contrats incitatifs**. La théorie des incitations, développée par J.-J. Laffont et J. Tirole (Prix Nobel en 2014), vise à élaborer des modèles pour déterminer quelles sont les structures des contrats les plus efficaces pour inciter les agents à suivre l'objectif fixé. La plupart des banques proposent par exemple aujourd'hui à leurs clients des contrats d'assurance automobile auxquels sont associés des systèmes de bonus et de malus. Un tarif de base de l'assurance est défini entre la banque et son client et en cas de mauvais comportements, le tarif à payer pour s'assurer va augmenter. À l'inverse, si la personne n'a pas d'accidents pendant un certain temps, elle verra son bonus augmenter et paiera des cotisations moins élevées. Un tel mécanisme incite les individus à ne pas être trop imprudents même s'ils sont assurés.

Notes

- [1.](#) Les comptes des actifs bancaires sont consolidés : ils intègrent les actifs détenus en France aussi bien qu'à l'étranger. Source : Les chiffres du marché français de la banque et de l'assurance, rapport 2016 de l'ACPR-Banque de France.
- [2.](#) Rappelons que l'indice est obtenu en faisant le carré de la somme des parts de marchés des firmes du secteur, plus il est élevé, plus la concentration est grande (source : rapport 2017 de la BCE sur les structures bancaires).
- [3.](#) Source : rapport de l'observatoire des tarifs bancaires, CCSF, 2016.
- [4.](#) L. Scialom, *Économie bancaire*, La Découverte, coll. « Repères », 2013.
- [5.](#) Source : rapport de la commission des finances, « Banques : votre santé nous intéresse », n° 52, Alain Lambert, 1996/1997.
- [6.](#) Données 2016, Les chiffres du marché français de la banque et de l'assurance, rapport de l'Autorité de Contrôle Prudentiel et de Régulation.
- [7.](#) E. Combe, « Quelles sanctions contre les cartels ? », *Revue internationale de droit économique*, 2006, p. 11-46.
- [8.](#) W. Güth, R. Schmittberger et B. Schwarze, « An experimental analysis of ultimatum bargaining », *Journal of Economic Behavior and Organization*, 1982.
- [9.](#) A. Landier et D. Thesmar, *Le grand méchant marché*, Flammarion, coll. « Champs », 2007.
- [10.](#) D. Marteau, *Le marché des capitaux*, Armand Colin, coll. « Cursus », 2016.

Chapitre 5

Biens collectifs et externalités

L'exemple de la gestion de l'eau

1. CARACTÉRISTIQUES DES BIENS ET FONCTIONNEMENT DU MARCHÉ

2. EXTERNALITÉS ET DÉCISIONS ÉCONOMIQUES DES AGENTS

3. QUELLE RÉGULATION PUBLIQUE EN CAS DE DÉFAILLANCE DU MARCHÉ ?

L'eau est un bien de consommation courante, qui nécessite une organisation économique assez sophistiquée pour être mis à disposition de tous. La complexité du processus, pas toujours perçue par le consommateur final, vient des propriétés économiques de l'eau, qui en font un bien particulier, et aussi de son impact sur de très nombreuses activités. La gestion de l'eau nécessite notamment de fixer précisément quelles sont les tâches prises en charge par des acteurs privés, et celles revenant à des acteurs publics. Pour mieux le comprendre, prenons l'exemple de la gestion de l'eau dans une commune française, inspiré d'un exemple réel. Elle repose sur un processus en cinq étapes, que l'on retrouve dans toutes les communes, mais qui peut donner lieu à des choix collectifs très différents qui ne dépendent pas seulement de contraintes techniques. Dans cette commune, la *collecte de l'eau* s'effectue dans deux barrages appartenant à la municipalité, entretenus par son personnel. Deuxième étape, la *production d'eau potable* en stations de traitement. Elles appartiennent à la ville, mais leur fonctionnement est assuré par une compagnie privée, dans le cadre d'une « délégation de service public » (l'entreprise est désignée ci-dessous comme le « délégataire »). La ville prend en charge les investissements nécessaires au bon traitement de l'eau, le délégataire paie les dépenses de fonctionnement et les salaires de ses salariés. L'État intervient également, pour contrôler la qualité de l'eau et la conformité du processus aux textes réglementaires.

Troisième étape, la *distribution de l'eau* potable aux consommateurs finaux : ce bien étant indispensable, la quasi-totalité des foyers sont raccordés, à l'exception d'une dizaine de fermes isolées. L'entretien courant et le bon fonctionnement du réseau d'alimentation est à la charge du délégataire, tandis que les travaux importants sont assurés par la ville.

Le réseau d'égouts municipaux assure la *collecte des eaux usées*. Il existe un deuxième réseau de collecte dédié aux eaux d'écoulement (pluie, etc.),

mais l'un comme l'autre sont loin de couvrir l'ensemble des usagers, qui ne sont donc pas systématiquement facturés pour cette prestation. Un délégataire assure le bon fonctionnement du réseau, et la ville s'occupe des travaux d'investissement et de remise en état.

Dernière étape, *l'assainissement des eaux usées* est assuré par une usine très récente appartenant à la ville, dont le bon fonctionnement est pris également en charge par un délégataire, une société différente de celle à l'œuvre dans les étapes précédentes. L'installation permet de brûler les boues récupérées lors du traitement. Cette opération est source de pollution mais le coût pour l'environnement est moindre que celui d'autres méthodes de traitement, comme l'épandage des boues.

L'utilisateur n'a pas besoin de connaître la complexité du processus car il n'achète pas séparément chaque étape de la gestion de l'eau : il reçoit une seule facture, qui couvre sa consommation individuelle, et sa contribution au financement collectif. La loi prévoit en effet que le service de l'eau doit être financé par les usagers et non *via* les impôts locaux. En revanche, les administrations en charge du contrôle de la conformité aux textes en vigueur du service de l'eau sont financées par l'État. On voit dans cet exemple une solution parmi d'autres permettant de fournir à tous un accès à l'eau. D'autres communes proposent un service entièrement municipal, ou au contraire ont poussé encore plus loin l'intervention d'entreprises privées. Si l'eau est au final « vendue » aux consommateurs, puisqu'ils paient proportionnellement à ce qu'ils consomment, pourquoi est-ce que le marché ne fournit pas spontanément une solution efficace pour répondre à la demande en eau ?

Deux propriétés microéconomiques de l'eau permettent de comprendre pourquoi elle est aussi complexe à gérer, et nécessite une prise en charge au moins partielle par les autorités publiques.

D'abord, certaines étapes de la gestion de l'eau ne peuvent pas être fournies efficacement par le marché, car elles n'ont pas les propriétés économiques des biens qu'on y échange d'ordinaire. Il faudra définir précisément ces caractéristiques, afin de mieux comprendre pourquoi certains biens sont produits ou gérés par les autorités publiques. C'est l'objet de la première partie.

L'eau a une seconde propriété qui explique les spécificités de sa prise en charge : l'assainissement des eaux usées affecte tous les acteurs, pas uniquement celui qui en a acheté puis rejeté une certaine quantité. On parle

d'**externalité** pour désigner les conséquences économiques du comportement d'un agent sur la situation d'un autre agent, qui ne sont pas mis en relation par une transaction marchande. Les rejets d'eaux polluées par une entreprise affectent ainsi les agriculteurs qui l'entourent ou la santé des ménages voisins, sans que cette action ne fasse l'objet d'une transaction marchande entre l'entreprise et les autres agents. Les modalités de gestion de l'eau permettent la prise en charge de certaines externalités, grâce à une régulation publique, comme on le voit dans la deuxième partie.

Quand les biens ont certaines caractéristiques ou quand il existe des externalités, le fonctionnement du marché ne conduit pas à un équilibre efficace d'un point de vue économique. On parle de **défaillances du marché**, qui nécessitent une régulation publique. Ce terme est aussi employé en présence d'un monopole naturel et d'asymétries d'information (chapitre 4) puisque le libre jeu des mécanismes de marché conduit à un équilibre inefficace. En cas de défaillance du marché, les autorités doivent notamment fixer la frontière entre ce qui doit être produit sur un marché, et ce qui relève d'une production publique. Elles peuvent aussi mettre en place une réglementation et une fiscalité adaptées pour améliorer le fonctionnement du marché. Comme le suggère l'exemple de l'eau, les enjeux sont considérables, sur le plan environnemental par exemple. La troisième partie porte donc sur les modalités de l'intervention publique en cas de défaillance du marché.

1. Caractéristiques des biens et fonctionnement du marché

1.1 Les différents types de biens économiques

• Biens excluables et biens rivaux

Les biens économiques que l'on consomme le plus fréquemment, appelés **biens privatifs**, ont deux propriétés particulières : ils sont rivaux et excluables.

Un bien est dit rival si sa consommation prive les autres de son usage. Ainsi, quand un agent consomme un litre d'eau, il ne peut plus être consommé par un autre, il y a donc **rivalité** pour accéder à ce bien. Les biens privatifs sont par définition rivaux, qu'il s'agisse de biens matériels (une automobile, un lieu d'habitation ou un kilo de pommes), ou de biens immatériels : une séance d'épilation est rivale, car si l'esthéticienne accorde une prestation à une personne elle ne peut plus l'accorder à une autre personne au même moment.

Un bien est **excluable** si le vendeur peut en réserver l'accès à ceux qui ont payé le prix fixé en échange de son usage. Une automobile est excluable car le vendeur peut réserver son accès à celui qui l'achète ou la loue. La distribution d'eau potable est également excluable, car son accès peut être fermé à celui qui ne paie plus sa facture. En revanche, l'eau d'un fleuve est plus difficilement excluable, de même qu'une nappe phréatique, car certains ont le droit d'y accéder par un puits, et on ne peut pas toujours contrôler la quantité consommée.

Précisons que certains ouvrages parlent de « biens privés » plutôt que de biens privatifs, mais cette expression s'avère trompeuse, car elle entraîne la confusion entre les caractéristiques économiques des biens et le fait qu'ils soient offerts par un agent privé ou par un agent public (État, collectivités locales, etc.). Or un bien privatif peut parfaitement être offert par un agent public. Certains biens privatifs sont en fait produits sous la tutelle de l'État, il

s'agit des **biens tutélaires** qui sont réalisés directement par une structure publique, ou dans des conditions strictement contrôlées par l'État. Un bien tutélaire pouvant être à la fois rival et excluable, il pourrait être fourni par le marché, mais l'État considère qu'un contrôle public est nécessaire en raison des externalités générées (voir la [deuxième partie](#)), ou bien pour faciliter son accès à toute la population, pour des questions de justice sociale.

On peut douter à première vue de l'intérêt des concepts de rivalité et d'excluabilité, tant les qualités des biens courants semblent aller de soi, mais en fait ces notions sont essentielles pour comprendre les problèmes spécifiques que posent la production et l'accès à certains biens économiques, en raison de leurs caractéristiques propres. En croisant les deux propriétés identifiées, on peut définir quatre catégories de biens, comme le montre le **document 5.1**.

Document 5.1 : Typologie des biens économiques

	Excluabilité : l'accès au bien peut être rendu payant	Non excluabilité : l'accès au bien ne peut pas être contrôlé par un prix
Rivalité : l'usage d'une unité de bien en prive les autres	Biens privatifs : – vêtements, coupe de cheveux, restaurant	Biens communs : – banc de poissons – les ressources forestières
Non rivalité : l'usage d'une unité de bien n'en prive pas les autres	Biens de club : programme de télévision crypté salles de spectacle autoroute à péage connaissances protégées par la propriété intellectuelle	Biens collectifs : la défense nationale les phares maritimes les radios libres les connaissances non brevetées

Chaque type de bien a des propriétés différentes :

- Comme on l'a vu, les biens privatifs sont à la fois rivaux et excluables. En raison de ces propriétés, ils peuvent faire l'objet d'une production destinée à la vente sur un marché. En effet, le vendeur peut exiger un

prix en échange d'un bien excluable, ce qui lui permet de rémunérer sa production. En situation de concurrence pure et parfaite, le marché conduit à une gestion efficace de ce type de bien (chapitres 1 et 7).

- Les **biens collectifs** sont définis par l'absence de rivalité et d'excluabilité. La défense nationale est par exemple un service rendu à la collectivité qui n'est pas rival, au sens où la protection assurée à un citoyen par l'armée n'en prive pas les autres. De même, la défense nationale est un service non excluable, car on ne peut discriminer entre ceux qui ne souhaitent pas être protégés et ceux qui le souhaitent par un système de prix : soit le service est fourni à tous, soit à aucun. Pour désigner les biens non rivaux et non excluables, il nous semble préférable d'utiliser l'expression bien collectif et non la traduction littérale du « *public good* » anglo-saxon, soit « bien public » car ce dernier qualificatif fait davantage référence en français à celui qui fournit le service (une administration publique) qu'à la nature du service considéré, les biens collectifs pouvant être produits par des agents privés, comme le sont les programmes des radios libres par exemple..
- Les **biens communs** forment une catégorie intermédiaire, ils sont rivaux comme les biens privatifs, mais non excluables comme les biens collectifs. Il s'agit souvent de ressources disponibles en quantité limitée, dont l'accès est difficilement contrôlable par un mécanisme marchand. Par exemple, un ban de poissons est rival puisque la ressource peut se tarir, mais il s'avère difficilement excluable car l'accès à la ressource est libre en haute mer.
- Les **biens de clubs** présentent des propriétés inverses aux biens communs : ils sont excluables mais non-rivaux. Une chaîne de télévision cryptée par exemple est non rivale, au sens où un consommateur supplémentaire ne réduit pas le service disponible pour les autres. En revanche, on ne peut accéder à ses programmes qu'à condition de payer une redevance.

Certains économistes, à la suite de P. Samuelson, utilisent le terme bien collectif dans un sens plus extensif, en parlant de biens collectifs « mixtes » pour désigner les biens communs et les biens de clubs, et de « biens collectifs purs » pour désigner les biens qui sont à la fois non rivaux et non excluables.

Nous préférons réserver le terme bien collectif à cette dernière catégorie, pour éviter toute confusion.

- **Des frontières mouvantes selon l'évolution technologique, institutionnelle et politique**

En première analyse, les caractéristiques des biens semblent liées à leur nature même : une nappe phréatique n'a manifestement pas les mêmes propriétés qu'un morceau de musique ou un service de voirie. Pourtant, les frontières entre les différentes catégories de biens n'ont rien de naturel. Elles résultent en fait largement de l'évolution technologique et institutionnelle.

La technologie disponible joue un rôle essentiel : le choix entre produire un bien en le réservant à certains consommateurs, ou au contraire en l'offrant simultanément à tous, dépend des possibilités techniques et de l'inventivité des acteurs. Par exemple, la mise en bouteille d'eau de source, avec une marque déposée, a permis de transformer des ressources communes en biens privatifs. De même, la gestion collective des eaux usées est une technologie qui a été développée très précocement, sous la Rome antique au ^{vi}^e siècle avant notre ère, et qui a connu une longue période de quasi-disparition en Europe, avant de se développer à nouveau au ^{xix}^e siècle. Entre les deux périodes, la possibilité technique de fournir un service d'assainissement évitant à chacun de trouver une solution privée, avait tout simplement disparu. L'amélioration des connaissances technologiques permet donc de faire émerger certains types de biens, ou de déplacer les frontières entre les biens privatifs et les autres, selon la façon dont la rivalité ou l'excluabilité est mise en œuvre. La facturation d'eau en fonction de la quantité consommée suppose qu'il est techniquement possible de mettre en place des compteurs individuels. La technologie s'est développée progressivement au ^{xix}^e siècle, accompagnant un changement majeur dans la gestion de l'eau (KONSTANTINOS, 2006) : on passe d'une distribution principalement gratuite par des fontaines collectives à une consommation individualisable pour chaque immeuble, puis plus tard chaque consommateur, grâce aux compteurs appropriés. La technologie a donc rendu la distribution d'eau plus facilement excluable.

Les frontières entre les biens privatifs et les autres catégories de biens reposent également sur des facteurs institutionnels. Les règles de propriété ont ainsi une incidence sur les caractéristiques des biens. Les ressources en

eau douce par exemple sont excluables si les droits de propriété permettent à un individu de contrôler l'accès à la ressource. L'utilisation d'un puits ou l'accès à l'eau des rivières sont en France très réglementés : on ne peut décider de détourner une rivière qui passe sur sa propriété personnelle. Les prélèvements effectués ne doivent pas affecter le débit du cours d'eau. L'excluabilité dépend donc des règles institutionnelles adoptées sur un territoire donné. Les débats sur le logiciel libre ou la brevetabilité du vivant montrent également que la frontière entre un bien non rival et librement accessible et un bien non-rival, faisant l'objet d'une appropriation, sont également soumises aux décisions du législateur. Au final, les autorités politiques ont un rôle important dans la construction des biens non-privatifs, en agissant sur les droits de propriété, et les modalités institutionnelles d'accès à ces biens.

• L'eau, quels biens économiques ?

L'étude de l'eau, dans ses différentes dimensions, nécessite une typologie rigoureuse des biens économiques. Différents usages lui sont associés en effet par le « cycle de l'eau », c'est-à-dire la succession d'étapes qui conduisent les eaux d'écoulement à nourrir les réserves d'eau douce, puis à être consommées, et ensuite rejetées après traitement, avant de rejoindre le réseau hydrique naturel à nouveau. L'eau renvoie donc à des biens ou services multiples aux caractéristiques économiques différentes.

Les *ressources en eau douce* constituent manifestement un bien commun. L'eau est une ressource naturelle renouvelable, qui se reconstitue au cours du temps mais peut se dégrader ou s'épuiser en raison des activités humaines constatées localement (ROTILLON, 2010). Les ressources en eau à l'échelle d'un territoire sont donc bien rivales, dans la mesure où l'eau consommée par l'un n'est plus consommable par d'autres. L'excluabilité des ressources en eau est réduite, car les prélèvements dans les cours d'eau ou les nappes phréatiques sont autorisés, même si la loi les réglemente.

En revanche, l'*eau potable* distribuée par un réseau à des consommateurs individuels a les caractéristiques d'un bien privatif : une quantité donnée d'eau potable est rivale, et il est possible d'exclure un consommateur qui ne paie pas sa facture. Mais l'eau étant un bien dont l'accès est jugé essentiel à tous, sa distribution est généralement placée sous la tutelle de l'État, pour

éviter de priver une partie de la population d'un bien indispensable : c'est donc un bien tutélaire. La loi fait obligation aux mairies de permettre à tous l'accès à l'eau potable.

L'*assainissement des eaux usées* est clairement une activité non excluable, dans la mesure où il n'est pas possible de réserver ce service aux consommateurs qui le paient : le réseau est conçu de manière à récupérer l'ensemble des eaux usées d'un territoire. Certains ménages ne peuvent être raccordés au réseau pour des raisons techniques, et la loi les oblige à mettre en place une installation individuelle d'assainissement, qui est alors un bien privatif contrairement à une installation collective. Le service d'assainissement est faiblement rival, dans la mesure où le raccordement d'un consommateur supplémentaire a un coût marginal très faible. Il nécessite cependant des investissements importants pour être entretenu et étendu au besoin.

La *qualité de l'eau* assainie et rejetée dans le milieu naturel est en fait un bien collectif, dans la mesure où elle est ni rivale ni excluable : la qualité des eaux disponibles dans un lac par exemple est non rivale, car disposer d'une eau de qualité n'en prive pas les autres, et elle est également non excluable car on ne peut réserver une qualité d'eau donnée à certains. Elle est définie par des normes sanitaires qui s'imposent aux communes ayant la responsabilité de ce service, et sont par exemple dans l'obligation d'intervenir quand la qualité d'une eau de rivière se dégrade.

La gestion d'un bien se conçoit sur un territoire donné, et à ce titre les services rendus par l'eau peuvent être abordés à des niveaux très différents. On considère parfois que l'eau est un **bien public mondial**², terme popularisé par les organisations internationales, mais cette expression est moins précise que le vocabulaire des économistes, car elle confond selon les cas des biens collectifs ou des biens communs, définis à l'échelle mondiale. Ainsi, les ressources en eau sont un bien commun mondial dans la mesure où les ressources en eau douce de la planète sont en quantité limitée, et elles nécessitent une coopération internationale pour être partagées correctement. De même, la qualité de l'air est un « bien collectif mondial » dans la mesure où elle est non rivale et non excluable, et se définit au niveau mondial car le climat ne connaît pas les frontières.

Dans de nombreux pays, la gestion de l'eau est assurée localement par la municipalité, comme c'est le cas en France. L'expression de **bien public**

local³ est parfois employée pour désigner un bien disponible sur un territoire donné et financé par les autorités publiques à l'échelle de ce territoire. Comme pour les biens publics mondiaux, l'expression peut s'avérer trompeuse car elle recouvre selon les cas des biens aux caractéristiques économiques très différentes. L'assainissement de l'eau est par exemple un bien public local : les réseaux de récupération et de traitement sont bâtis à l'échelle d'une commune ou d'un regroupement intercommunal, et cette dimension locale joue un rôle déterminant dans le comportement des agents, car ils peuvent décider de se déplacer d'un territoire à l'autre en fonction des caractéristiques des biens publics locaux proposés, et de leur tarification. Une entreprise industrielle peut décider par exemple de changer de commune parce qu'on lui impose une redevance trop élevée pour ses rejets d'eaux usées.

1.2 La production de biens collectifs, de biens communs ou de biens de club peut-elle être assurée par le marché ?

Dans les conditions d'équilibre d'un marché concurrentiel, et en l'absence d'imperfections, les producteurs fournissent le niveau optimal de biens (voir les chapitres [1](#) et [7](#)), sans qu'une intervention publique ne soit nécessaire. Mais pour les autres catégories de biens, les marchés sont défaillants, car les agents ont tendance à produire une quantité insuffisante ou au contraire excessive de ces biens.

• Les biens communs et le risque de surexploitation

Si une ressource est non excluable, les agents peuvent l'exploiter sans avoir à en payer le prix. Les ressources naturelles en eau sont bien dans ce cas de figure, elles peuvent être utilisées par les agents sans avoir systématiquement à l'acheter, en puisant directement dans une nappe phréatique, une rivière ou un lac. Ce n'est pas le cas de la distribution d'eau potable, qui est excluable.

Les ressources naturelles en eau ont une valeur économique qui dépend de leur rareté. Si elles s'épuisent, il devient d'autant plus précieux d'y accéder : un agriculteur a un besoin urgent de prélever de l'eau souterraine quand son approvisionnement habituel en surface s'épuise. Si la ressource est librement accessible, des agents rationnels ont intérêt à prélever ce dont ils ont besoin,

sans tenir compte du **coût social** de leur décision c'est-à-dire l'ensemble des coûts estimés, pour eux-mêmes et pour les autres. Le coût social intègre donc le coût privé, mais va au-delà en comptabilisant les conséquences indirectes sur les coûts subis par tous les autres agents. La prise en compte des coûts sociaux est centrale dans le domaine environnemental, dès lors qu'un producteur peut affecter le milieu naturel, et donc toute la collectivité, par ses décisions individuelles (pollution, réduction de la biodiversité, épuisement d'une ressource, etc.). Ainsi, les agriculteurs irriguent leurs champs parce que le **coût privé** est faible puisqu'ils ne paient pas l'eau, et ils épuisent la ressource ce qui entraîne une perte collective élevée si la production agricole chute faute d'irrigation. C'est ce qu'on nomme la « **tragédie des communs** », selon le titre d'un célèbre article de G. Hardin⁴, soit la tendance à la surexploitation d'un bien commun par des agents rationnels.

Quand un bien privatif est davantage demandé et que son prix augmente, les individus sont incités à le produire en plus grande quantité. On pourrait imaginer par exemple qu'un investisseur privé détourne une rivière pour alimenter une réserve d'eau, afin de répondre aux besoins d'irrigation quand sa raréfaction lui donne une valeur supplémentaire. Mais si l'accès à l'eau est non excluable, l'investisseur n'a aucune raison de produire ce bien car il ne sera pas rémunéré pour son effort. Si la ressource est disponible gratuitement dans la nature, les agents n'ont pas plus besoin de payer pour en disposer, il leur suffit de continuer à la prélever gratuitement. C'est un comportement de **passager clandestin**, soit un agent qui bénéficie de l'accès à une ressource non excluable sans en payer le prix. Les incitations à produire sont donc insuffisantes dans ce cas de figure.

De façon générale, le marché est défaillant à un double titre, en l'absence d'excluabilité : les agents risquent de surexploiter les ressources disponibles, et la production de ces biens peut s'avérer insuffisante. La régulation publique de la production est donc généralement nécessaire, comme nous le verrons dans la troisième partie. De nombreux biens communs posent les mêmes problèmes d'efficacité, comme la gestion de l'assainissement des eaux, mais aussi celle des forêts, des réserves de pêche, ou des infrastructures publiques dont l'accès est libre, comme l'aménagement des bords de mer.

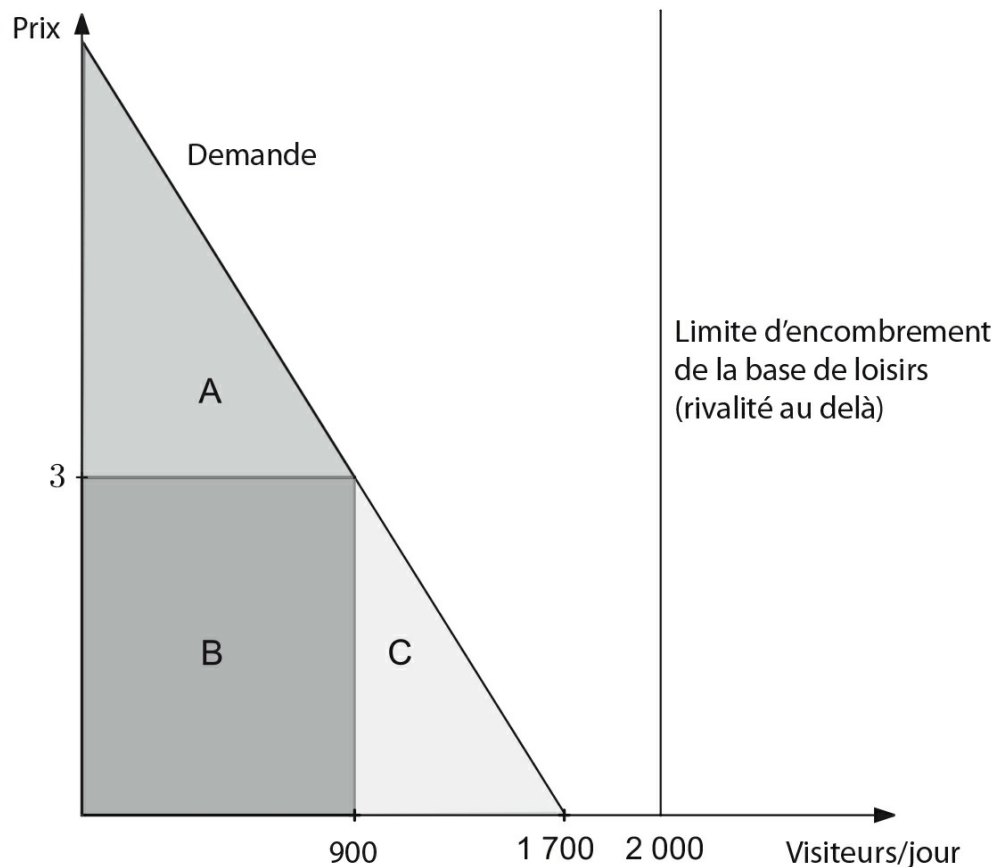
Toutefois, les biens communs ne font pas l'objet d'une surexploitation systématique par les agents, comme le suggère le travail d'E. Ostrom (2010) : de nombreuses communautés humaines ont su s'organiser pour éviter la

surexploitation d'un bien commun renouvelable, sans que l'intervention coercitive de l'État soit nécessaire pour protéger la ressource. C'est le cas en particulier de l'eau destinée à l'irrigation, utilisée collectivement dans certaines régions pour en contrôler la surexploitation, comme on le verra dans la troisième partie.

- **Les biens de club et les restrictions d'accès aux biens**

Les biens de club sont non rivaux, donc ils peuvent être consommés par plusieurs personnes simultanément, sans effet sur la consommation des autres, et de surcroît l'accès à ces biens peut être réservé à ceux qui sont prêts à les payer. Une base de loisir au bord d'un lac répond à ces conditions, tant que l'encombrement généré par les visiteurs est négligeable, autrement dit tant qu'ils ne se gênent pas mutuellement. Les biens de club sont très souvent soumis à des limites d'encombrement, c'est-à-dire qu'ils deviennent rivaux au-delà d'un certain nombre d'utilisateurs (tunnel à péage, réseau ADSL, etc.). En dessous de ce seuil, le coût marginal d'un utilisateur supplémentaire est nul, car il n'a pas d'effet sur le coût du bien mis à disposition : une fois que la base de loisir est construite, un visiteur supplémentaire n'entraîne pas d'augmentation des coûts, tant qu'il n'y a pas d'encombrement.

**Graphique 5.1 : Le cas d'une base de loisir, un bien de club
à l'origine d'une perte sèche**



Dans ces conditions, la production privée d'un bien de club entraîne un accès restreint au bien, qui se traduit par une perte d'efficacité. Pour le comprendre, le [graphique 5.1](#) présente le cas d'une base de loisirs aux bords d'un lac, par hypothèse non rivale en dessous de 2 000 visiteurs par jours, car elle ne présente pas d'encombrement dans ces conditions.

Si un producteur privé aménage ce site, il a intérêt à faire payer l'accès à la base de loisir. Supposons que le prix est fixé à 3 € par visiteur¹, prix pour lequel, d'après la courbe de demande, 900 visiteurs sont disposés à payer chaque jour. La zone B correspond à la recette du producteur, qui encaisse 2 700 € chaque jour (3×900). La zone A correspond au surplus des consommateurs qui sont disposés à payer pour l'accès à ce bien, et la zone C correspond enfin à la perte sèche subie par les consommateurs (voir le [chapitre 1](#)). Faire payer 3 euros par visiteur prive 800 personnes chaque jour (1 700-900) de l'accès à un bien qui est trop onéreux pour eux, alors que leur présence sur la base de loisir n'entraînerait aucune dépense supplémentaire pour le producteur, leur coût marginal étant nul.

Quand un bien est non rival, le prix imposé par le producteur restreint donc son accès alors qu'il ne subit pas de coût supplémentaire si la consommation augmente, et entraîne une perte de surplus pour la collectivité : la quantité produite est trop faible, et donc inefficace. Ce problème existe pour de nombreux bien de clubs, quand un producteur privé a intérêt à imposer un péage. Ainsi, les tunnels à péage que l'on observe dans certaines villes sont souvent sous-utilisés en raison de leur prix élevé, ce qui entraîne un coût collectif, accentué par les effets en chaîne, par exemple l'encombrement sur d'autres routes.

• La sous-production des biens collectifs

Un bassin hydrographique est une région dont l'ensemble des cours d'eau se jettent dans le même fleuve, dans ces conditions une dégradation de l'eau en amont peut affecter toutes les ressources en eau situées en aval. La qualité de l'eau dans un bassin hydrographique est donc un bien collectif, au sens où il n'est pas possible de faire accéder seulement certains consommateurs à ce bien et d'exclure les autres, la qualité d'un cours d'eau étant la même pour tous. Par ailleurs, elle est non rivale, car la consommation d'une eau de bonne qualité par les uns n'en prive pas les autres.

Les biens collectifs, à l'image de la qualité de l'eau, cumulent les deux problèmes d'efficacité présentés ci-dessous. Étant non excluables, les individus sont incités à en bénéficier sans en payer le prix. Étant non rivaux, on ne peut priver certains individus de leur consommation sans en priver les autres. Dans ces conditions, le marché ne peut généralement pas mettre à disposition les biens collectifs en quantité suffisante. Aucun acteur ne va en effet de lui-même produire un service permettant d'améliorer la qualité de l'eau s'il ne peut ensuite faire payer les utilisateurs, et aucun consommateur n'a intérêt à payer pour que le service soit produit car la somme nécessaire pour couvrir les coûts de production est élevée et qu'ensuite, tous les acteurs pourront en profiter gratuitement sans que cela en dégrade la qualité. Détaillons le raisonnement à travers l'exemple d'un service qui améliorerait la qualité de l'eau d'un lac. Il s'agit d'un bien collectif car une fois le service produit, tous les acteurs peuvent profiter sans payer de l'amélioration de la qualité de l'eau et le fait qu'un individu profite de la qualité de l'eau du lac n'empêche pas les autres acteurs d'en profiter. Dans le cas des biens privés, les mécanismes du marché conduisent à ce que les producteurs augmentent

leur volume de production tant que le coût marginal est inférieur au prix (voir le [chapitre 3](#)) et le prix d'équilibre se fixe à un niveau tel que le cout marginal de production est égal à la disposition marginale à payer des consommateurs (voir le [chapitre 1](#)). La situation est différente pour les biens collectifs. Supposons que le coût marginal pour obtenir une eau de meilleure qualité dans un village de 50 habitants soit de 1 000 €. Si on laisse faire le marché, il faudra espérer qu'au moins un agent soit disposé à payer 1 000 € pour bénéficier d'une meilleure eau, ce qui a peu de chances d'être le cas : personne ne réalise donc cette dépense, et l'eau reste de qualité médiocre. Supposons que chaque agent soit disposé à payer 20 € pour bénéficier d'une eau de bonne qualité : individuellement, aucun n'est prêt à payer seul 1 000 euros, la disposition à payer étant inférieure au coût marginal, mais puisqu'il s'agit d'un bien non rival, payer pour une eau de meilleure qualité bénéficiera à tous, par conséquent, les 50 habitants ont intérêt à produire collectivement cette eau de bonne qualité, puisque la somme de leurs dispositions à payer couvre le coût marginal. On voit à travers cet exemple que les biens collectifs sont fournis généralement dans une quantité trop faible. Le marché est donc défaillant, et il est logique que les biens collectifs soient financés par la puissance publique dans ces conditions. C'est le cas de nombreux biens collectifs comme la sécurité nationale, la justice, le nettoyage des rues, la prévention des épidémies ou la signalisation routière.

Il faut noter cependant que la mise à disposition de certains bien collectifs peut être le fait de l'initiative privée. R. Coase montre dans un article célèbre⁵ que les phares, prototype du bien collectif, se sont en fait développés à partir du ^{xvii}e siècle en Angleterre grâce à l'initiative privée. Les modalités de gestion des biens collectifs restent donc assez plastiques.

Si les biens collectifs peuvent être pris en charge par une autorité publique, cela ne règle pas le problème de savoir dans quelle quantité il faut les produire.

2. Externalités et décisions économiques des agents

Les ressources en eau au sein d'une région, et la qualité de cette eau, sont à l'origine de nombreuses externalités sur les activités humaines les plus diverses, comme le tourisme, qui pâtit fortement d'une dégradation de l'une ou de l'autre. À l'inverse, de nombreuses activités sont à l'origine d'externalités subies par l'eau, à commencer par la pollution résultant des activités industrielles ou agricoles. Dans ces conditions, quel est l'impact des externalités sur la production de biens et sur l'efficacité des marchés ?

2.1 Les caractéristiques des externalités

- **Les externalités générées par la production d'un bien**

Une **externalité** intervient lorsque le comportement d'un agent économique a un effet sur un autre agent économique, sans que cette action ne fasse l'objet d'une transaction marchande. On peut distinguer les **externalités positives**, qui sont dues à une amélioration de la situation d'un agent consécutivement aux décisions d'un autre, sans qu'ils soient reliés par un marché, et les **externalités négatives**, dans le cas inverse. Les externalités peuvent concerner les relations entre les agents les plus divers (producteur, consommateur, administrations publiques, etc.). La gestion de l'eau permet de bien saisir la fécondité du concept. En s'inspirant d'un exemple d'[J.E. Meade⁶](#), un des fondateurs de la théorie des externalités, supposons qu'une société agricole située dans une région assez aride décide de passer à une production agricole moins intensive en eau, économisant ainsi celle de la rivière sur laquelle elle est installée. Elle peut agir ainsi en raison du coût de l'eau, afin d'économiser cette ressource à titre individuel, ou pour d'autres raisons, mais rien ne l'incite à tenir compte de l'impact de son action sur les autres entreprises du secteur. Or, si elle se montre plus économe en eau, elle réduit les risques que l'eau manque en aval, pour d'autres activités sur la rivière comme le tourisme, la pêche ou l'irrigation sans que les entreprises liées à ces secteurs ne subventionnent l'entreprise agricole pour l'aider à économiser l'eau. Elle est donc à l'origine d'une externalité positive sur les

acteurs exerçant d'autres activités situées en aval.

Comme on le voit sur le **document 5.2**, les acteurs utilisant l'eau peuvent subir des externalités positives ou négatives, quand les ressources en eau sont affectées par les comportements d'agents économiques exerçant des activités n'ayant pas de rapport direct avec l'eau (en rejetant des déchets découlant de la production d'un bien dans une rivière par exemple) ; par ailleurs, une modification de la gestion de l'eau peut aussi générer des externalités pour les autres agents. Par exemple, l'amélioration du traitement des eaux usées près d'un lac favorise le tourisme, sans que les hôteliers ou les restaurateurs concernés aient payé spécifiquement pour cela. Mais les ressources en eau subissent aussi de nombreuses externalités qui affectent tous les agents économiques ayant besoin de cette ressource, la pollution de l'eau en étant le prototype. Un seul industriel ou un seul agriculteur peut dégrader la qualité de l'eau de toute une région en polluant une rivière, tous les autres utilisateurs étant touchés.

Document 5.2 : Les externalités liées à l'eau

	Externalités positives avec l'amélioration de la qualité de l'eau	Externalités négatives avec la dégradation de la qualité de l'eau
Externalités dues à la gestion de l'eau	Hausse du tourisme Amélioration de la valeur esthétique du patrimoine naturel Préservation de la biodiversité	Nuisances au voisinage d'une station d'eaux usées Hausse des risques sanitaires pour une eau de mauvaise qualité
Externalités subies par les ressources en eau	Réduction des rejets des ménages Passage à une agriculture biologique	Pollution industrielle des cours d'eau Rejet de nitrates par les éleveurs

Il existe aussi d'autres formes d'externalités, qui ne concernent pas directement le cas de l'eau, notamment les **externalités de réseau**. Elles apparaissent lorsque la valeur d'un bien est supérieure pour un individu si les autres adoptent également ce bien : par exemple avec l'adoption d'un

standard commun facilitant la communication (le standard GSM des téléphones), ou avec l'accès à un vaste réseau bancaire (voir le [chapitre 4](#)).

Les ressources en eau, leur préservation et leur qualité, sont donc au centre de multiples externalités, positives ou négatives, ce qui accroît l'importance économique de leur gestion. D'autres biens économiques ont une valeur particulière en raison de l'importance des externalités qui sont associés à leur production, comme les services de santé ou d'éducation par exemple : favoriser l'apprentissage de la lecture ne bénéficie pas uniquement aux nouveaux instruits, mais aussi à tous ceux qui voient leur situation s'améliorer indirectement, parce que la diffusion d'informations écrites est désormais facilitée pour tous. La présence d'externalités est une des raisons de l'attention du législateur à l'attribution des droits sur l'eau, afin d'établir les responsabilités individuelles ou collectives, quand le comportement d'un agent peut avoir des conséquences sur la situation de beaucoup d'autres. La présence d'une eau de bonne qualité dans une région ayant des effets manifestes sur son environnement, au sens le plus large du terme, elle est protégée par la loi, qui fait de l'eau un élément du patrimoine national depuis la loi sur l'eau de 1992.

• Comment donner une valeur aux externalités ?

Pour tenir compte des externalités, il faut s'accorder sur leur valeur économique, or elle ne peut être déterminée par leur valeur sur le marché, puisque précisément personne n'achète ou ne vend cette externalité. L'analyse économique de la pollution s'est confrontée très tôt à cette question, car il est essentiel d'attribuer un coût économique aux externalités négatives comme la pollution, mais aussi aux externalités positives. Cette notion perd en effet une grande partie de son intérêt, si en pratique on ne peut pas proposer une évaluation précise des **coûts externes** ainsi que des **bénéfices externes** entraînés par l'action d'un agent, c'est-à-dire les coûts ou les bénéfices que son action génère pour d'autres, sans verser ou recevoir de compensation. Il est plus difficile de demander aux agents de tenir compte des externalités économiques qu'ils génèrent s'il est impossible de leur donner une valeur.

Pour évaluer monétairement les externalités, les économistes adoptent une définition plus large de la valeur que pour les biens échangés sur un marché,

appelée **valeur économique totale**. Elle intègre non seulement la valeur d'usage direct d'un bien, si on le vend sur le marché, mais aussi sa valeur d'option (les usages qu'il permettra dans le futur si on le conserve) ainsi que sa valeur de non usage (par exemple la valeur accordée à son existence sans avoir à l'utiliser, comme c'est le cas d'un paysage). L'écart entre valeur économique totale et valeur d'un bien sur le marché est lié aux externalités qu'il génère : de ce fait, si l'on est capable de déterminer quelle valeur économique totale est attribuée à un bien, on peut en déduire par différence le coût des externalités qui lui sont associées.

Il existe de nombreuses méthodes d'évaluation de la valeur des externalités (BONTEMS, ROTILLON, 2013), et pour donner un simple aperçu, deux grandes familles de méthode se distinguent :

- Les méthodes indirectes évaluent les conséquences économiques d'une action, pour en déduire la valeur de l'externalité subie par les agents. On peut quantifier par exemple les coûts de dommage subis par la collectivité à cause de la pollution d'une nappe phréatique par des nitrates sur la santé publique, mais cela suppose de connaître précisément les données scientifiques du problème. D'autres méthodes indirectes s'appuient sur l'évaluation des coûts de remplacement, soit les dépenses nécessaires pour rétablir la situation initiale, par exemple en traitant la pollution de l'eau.
- Les méthodes directes d'évaluation des externalités consistent à interroger les agents sur leurs préférences pour en déduire la valeur économique qu'ils accordent à une externalité. Par exemple, pour estimer les coûts entraînés par la pollution d'une nappe phréatique, on interroge les agents sur leur disposition à payer pour une eau de meilleure qualité, en leur précisant dans l'enquête sous quelles modalités (mode de financement par exemple). C'est ce qu'on appelle une **évaluation contingente**. Une fois connue la disposition à payer moyenne, il est possible d'en déduire la valeur économique accordée au phénomène étudié. Par exemple, une enquête en Alsace montre que les ménages seraient prêts à dépenser entre 113 et 132 € par an (en euros 2005) pour préserver la qualité de la nappe phréatique⁷.

Les méthodes en question permettent de mesurer précisément la valeur d'une externalité, dans un contexte très précis mais les résultats sont fort peu

généralisables à d'autres situations. Pour pouvoir comparer de multiples externalités, il est nécessaire de bâtir des indicateurs ayant une portée plus générale. C'est tout l'intérêt par exemple de l'estimation de la **valeur tutélaire du carbone**, qui désigne le prix, fixé par la puissance publique, d'une tonne de gaz carbonique émis en fonction de son effet potentiel à long terme sur l'environnement, quelle que soit sa source. Sa détermination fait l'objet de débats au sein du monde scientifique, car il est très complexe d'estimer l'effet moyen de l'émission d'une tonne de carbone sur le réchauffement climatique et par là, sur l'environnement. Une fois déterminée la « valeur tutélaire du carbone », les externalités liées à la pollution ou à l'utilisation d'énergie peuvent être ramenées à leur équivalent en émission de gaz carbonique. Si par exemple une usine rejette un million de tonnes de gaz carbonique par an, et que sa valeur tutélaire est de 32 €, comme le préconisait le rapport de la Commission Quinet en 2009⁸, alors on peut estimer le coût de l'externalité liée à la pollution à 32 millions d'euros. La valeur tutélaire du carbone est vouée à augmenter progressivement, de façon à inciter à la diminution des émissions de carbone, de façon progressive afin de rendre l'évolution acceptable, avec un horizon d'une taxe fixée autour de 100 € la tonne en 2030, référence retenue jusqu'à présent. Ce mécanisme permet de mener une politique environnementale cohérente quelle que soit la source de pollution ou de consommation d'énergie, en attribuant toujours la même valeur à une quantité donnée de pollution. Cependant, il n'y a pas de consensus aujourd'hui pour fixer la valeur tutélaire du carbone, dont la progressivité fait débat, et les autorités politiques rencontrent généralement des difficultés pour imposer à leur opinion publique une valeur tutélaire conforme aux préconisations des experts.

2.2 La défaillance des marchés en présence d'externalités

La présence d'externalités conduit généralement les individus à prendre des décisions non optimales, car ils ne tiennent pas compte de tous les effets, positifs ou négatifs, de leurs actions. Dans certains cas, les agents peuvent intégrer les externalités qu'ils subissent ou dont ils sont la cause, sans intervention publique, mais sous des conditions qu'il convient de préciser.

- Une production non optimale en présence d'externalités

Pour prendre une décision, les agents économiques comparent par hypothèse les coûts privés qu'ils subissent et les bénéfices privés qu'ils retirent de chaque solution envisagée. Ils ont intérêt à produire tant que le bénéfice privé d'une unité supplémentaire est supérieur au supplément de coût privé pour produire cette unité, et à l'équilibre le coût marginal de la dernière unité produite est égal à son bénéfice marginal. Par hypothèse, si le prix d'équilibre d'une bouteille d'eau minérale sur un marché concurrentiel est de 1 € pour une quantité de 100 000 bouteilles, un consommateur est donc disposé à payer 1 € pour bénéficier de cette 100 000^e bouteille et son bénéfice marginal privé est égal à 1 €, alors que le coût marginal du producteur de cette bouteille est aussi égal à 1 €. En l'absence d'externalités, les décisions privées conduisent sur un marché concurrentiel à utiliser les ressources efficacement, du point de vue collectif (voir le [chapitre 1](#)). Ce n'est plus le cas en présence d'externalités où le fonctionnement du marché ne conduit plus à un **optimum social**. Pour le montrer, on peut s'appuyer sur un modèle qui confronte les coûts marginaux sociaux et les bénéfices marginaux sociaux. On suppose ici que le bien en question génère une externalité négative, une pollution par exemple, et qu'elle augmente avec la quantité produite.

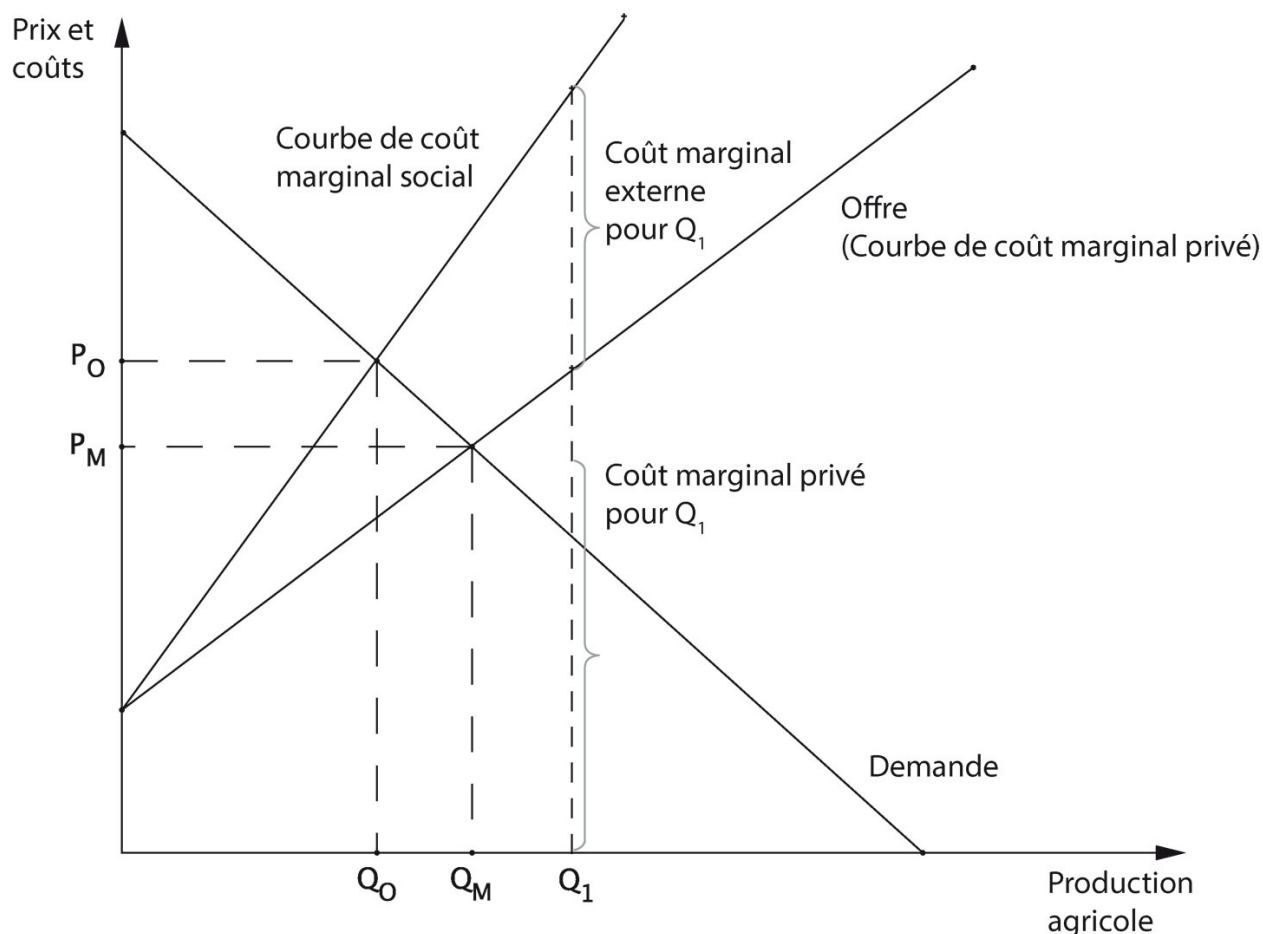
Le **coût social** d'un bien ou d'une activité représente l'ensemble des coûts supportés par les agents du fait du déroulement de cette activité ou de la production du bien. Il englobe les coûts privés et les coûts externes subis par les autres agents. En présence d'externalités négatives, le coût social est supérieur au coût privé des agents producteurs. Ils sont égaux quand il n'y a pas d'externalités ou en présence d'externalités positives. Le **coût marginal social** du bien est égal au coût marginal de production auquel s'ajoute le coût marginal externe. Cette notion peut paraître assez abstraite, mais on peut l'estimer en utilisant les méthodes d'évaluation des externalités vues précédemment.

Symétriquement, le **bénéfice social** d'un bien ou d'une activité représente l'ensemble des avantages⁹ dont profitent les agents du fait du déroulement de cette activité ou de la consommation du bien. Il englobe ainsi les **bénéfices privés** que retirent les personnes qui participent à cette activité ou consomment ce bien mais aussi les bénéfices externes dont peuvent profiter d'autres agents. Ainsi, quand une personne se fait vacciner, le vaccin ne lui profite pas uniquement mais est aussi bénéfique aux autres puisqu'ils voient

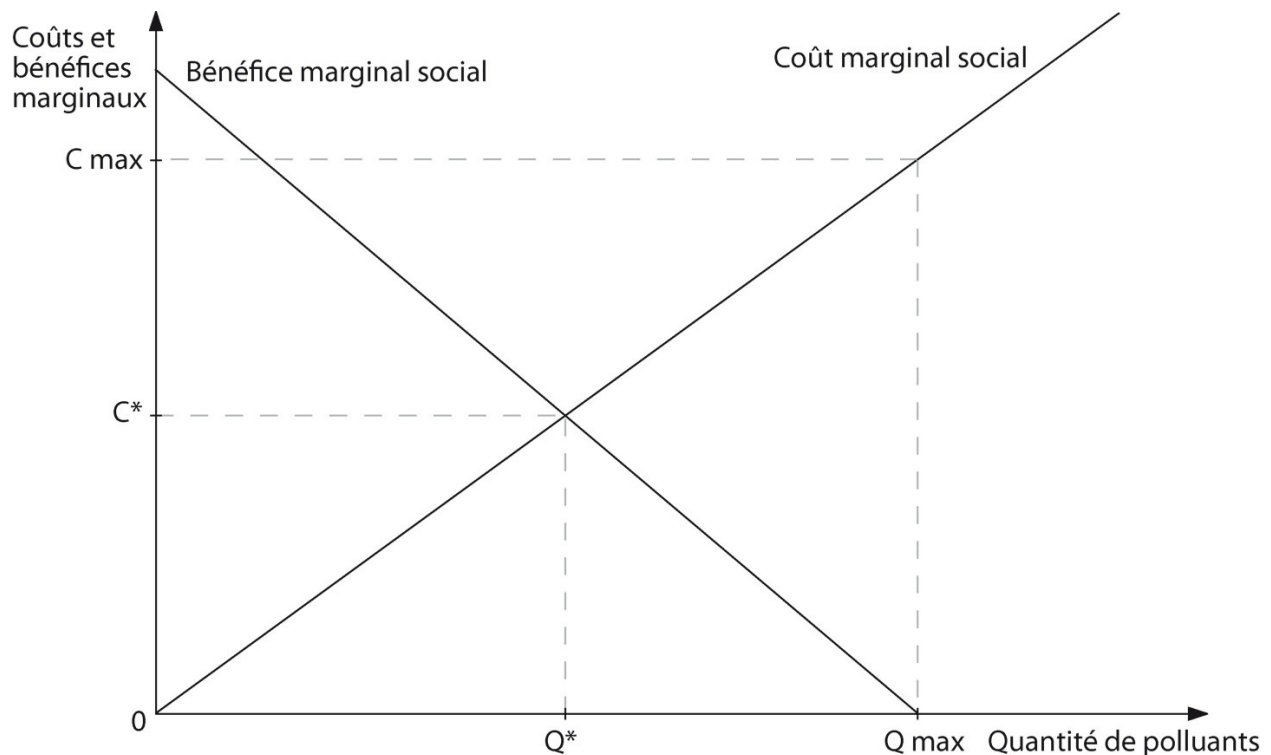
réduire la probabilité de se faire contaminer. Selon les activités concernées, le bénéfice marginal social prend des formes très diverses, il désigne tout type d'avantage et il ne faut donc surtout pas le confondre avec le « bénéfice » d'une entreprise, au sens comptable du terme. En présence d'externalités positives, le **bénéfice social** est supérieur au bénéfice privé à cause des avantages externes dont profitent les agents. Le **bénéfice marginal social** est égal au bénéfice marginal que retirent individuellement les consommateurs d'un bien ou d'une activité auquel s'ajoute le bénéfice marginal externe.

Prenons l'exemple de la pollution aux nitrates des ressources en eau par les activités agricoles. L'utilisation massive d'engrais et les lisiers (mélange de déjections animales) résultant de l'élevage intensif entraînent un excès de nitrates qui s'infiltre dans les sols et se concentre dans les eaux souterraines et les cours d'eau. Au-delà d'un certain seuil, les nitrates sont nocifs pour la santé humaine et contribuent au déséquilibre des milieux aquatiques en favorisant la multiplication de certaines algues, au détriment des autres espèces. Dans ce cas de figure, le coût marginal social de la production d'une quantité Q_1 de biens agricoles est supérieur à son coût marginal privé, comme on le voit sur le [graphique 5.2](#). Si on laisse faire le marché, les agriculteurs produiront Q_M , à l'intersection des courbes d'offre et de demande, mais le choix optimal, du point de vue collectif, serait de produire Q_O , de façon à éviter que le coût marginal social de la production agricole soit supérieur aux bénéfices marginaux qui en sont tirés.

Graphique 5.2 : L'effet d'une externalité négative sur l'équilibre d'un marché agricole



Dans certains cas, on ne peut pas facilement mesurer le coût externe associé à chaque quantité produite, mais on peut en revanche plus aisément évaluer les externalités associées à chaque unité de pollution. Il est alors préférable de raisonner non par quantité de produit, mais par quantité de polluant émis. Le **coût social de la pollution** aux nitrates englobe l'ensemble des coûts qui découlent de l'existence d'une telle pollution. Il inclut l'évaluation monétaire des dommages engendrés par la pollution et la perte en bien-être liée par exemple à la dégradation des plages par la prolifération des algues. Le **coût marginal social de la pollution** renvoie ainsi au coût supplémentaire imposé à la société dans son ensemble du fait de l'augmentation d'une unité de pollution. Le coût marginal est généralement croissant car la nature est de plus en plus affectée par la pollution quand celle-ci s'accroît. Les producteurs responsables de la pollution aux nitrates ne supportent pas le coût social de leur activité : il s'agit en effet d'un pur coût externe car il n'existe pas de coût privé de production de la pollution, celle-ci étant simplement la conséquence annexe de la production d'un bien.



Le **bénéfice social de la pollution** aux nitrates renvoie aux avantages que retirent les agents du fait de l'existence d'un certain niveau de pollution. Il recouvre ici les économies de dépenses réalisées par les producteurs car ils évitent des investissements coûteux dans des technologies plus propres. Les avantages liés à la pollution ne profitent qu'aux producteurs car ce sont les seuls à bénéficier de la réduction des coûts liée à une pollution plus importante : il n'y a pas de bénéfices externes. Le **bénéfice marginal social de la pollution** est le gain supplémentaire que la société retire d'une unité supplémentaire de pollution. Ce dernier est généralement décroissant car réduire la pollution est moins coûteux pour des niveaux élevés que des niveaux faibles de pollution et ainsi les économies de dépenses sont de plus en plus faibles à mesure que le niveau de pollution augmente.

Sur le [graphique 5.3](#), les courbes de coût et de bénéfice marginal social de la pollution sont représentées. Pour la société, le **niveau optimal de pollution** est atteint en « P optimal » (voir le [chapitre 7](#)), car le bénéfice marginal social et le coût marginal social sont égaux. Tant que l'on se situe en dessous de ce seuil de pollution, il est socialement préférable d'augmenter la pollution, dans la mesure où le bénéfice marginal social qui en résulte est

supérieur au coût marginal social subi. La courbe de bénéfice marginal social se situe en effet au-dessus de celle de coût marginal social. C'est donc le niveau optimal, en intégrant l'ensemble des externalités dues à la pollution. Cette idée peut paraître « choquante » car elle implique qu'un certain niveau de pollution est acceptable : du point de vue économique, réduire la pollution à zéro serait très coûteux, car on se passerait des avantages tirés d'une production modérément polluante (suppression par exemple de toute la production agricole dans la branche considérée).

Le fonctionnement « spontané » du marché entraîne une pollution supérieure à ce seuil, car les producteurs ne tiennent pas compte du coût externe de leur action et polluent de plus en plus tant que le bénéfice marginal qu'ils retirent de ce surcroît de pollution est positif. En effet, les producteurs n'ont aucun intérêt à engager des dépenses pour réduire la pollution en dessous de $Q_{\max}$. Le marché conduit donc à ce que ce soit la quantité $Q_{\max}$ de pollution qui soit émise et non Q^* . Le marché est ainsi défaillant en présence d'externalité, car il entraîne un niveau excessif de pollution en $Q_{\max}$, alors que l'**optimum social** se situe en « Q^* ».

Un bien comme l'eau étant au cœur de multiples externalités, il devient essentiel pour la puissance publique d'en contrôler la production. La directive cadre européenne de 2000, qui fixe les principes généraux de la gestion de l'eau en Europe, fait obligation aux États membres de respecter des objectifs environnementaux, et donc de veiller aux externalités liées à l'usage de l'eau. L'État dispose pour cela de multiples instruments, comme on le verra dans la troisième partie.

• Encadré 5.1 : Soutenabilité forte, soutenabilité faible, et externalités

La portée du raisonnement néo-classique, qui sert de fondement à l'économie de l'environnement fait l'objet de débats. La fixation du niveau optimal de pollution repose sur un raisonnement en équilibre partiel (voir le [chapitre 1](#)), valable pour un polluant spécifique, à une période donnée. Si le niveau de pollution optimal, établi selon cette méthode, excède celui qui peut être absorbé à long terme par la nature, il va conduire à une dégradation de la capacité d'absorption à la période suivante, ce qui risque d'entraîner une hausse des externalités négatives dues à la pollution. Par exemple, une pollution « optimale » d'une rivière peut conduire à modifier l'équilibre d'un milieu naturel, ou bien à la raréfaction d'une espèce de poisson et entraîner de proche en proche une surpêche d'une autre espèce en compensation, etc. Il serait donc préférable, pour certains environnementalistes comme L. Brown¹⁰, de fixer des normes strictes à l'usage des ressources polluantes, de façon à se situer en dessous du seuil correspondant aux capacités d'assimilation et de régénération de la nature, quitte à limiter fortement les activités économiques. On parle de soutenabilité forte pour définir cette approche des problèmes environnementaux fondée sur l'adoption de

normes strictes de limitation des activités économiques dégradant l'environnement, en contraignant les marchés à se soumettre aux règles assurant la conservation du patrimoine naturel. Cela risque de conduire à limiter fortement la croissance, pour protéger les ressources naturelles. Par opposition, l'optique de la soutenabilité faible, qui sous-tend le raisonnement néo-classique, considère que l'on peut s'appuyer sur les mécanismes de marché et les progrès technologiques, les agents faisant progressivement des choix plus économes en ressources par la substitution de facteurs (voir le [chapitre 3](#)), à mesure que les coûts de certains biens s'élèvent en raison de la dégradation de certaines ressources. Les agents doivent alors investir et développer des technologies plus respectueuses à cause des changements de prix relatifs pénalisant ceux qui utilisent les ressources naturelles devenues plus rares et plus chères. L'évolution des prix relatifs et des technologies n'est cependant pas assez rapide pour protéger l'environnement puisque les agents ne tiennent pas compte des externalités, et dans l'optique de la soutenabilité faible, on considère que des politiques environnementales sont indispensables pour accélérer le changement, en créant des incitations fiscales par exemple.

• Le marché peut-il intégrer les externalités sans intervention publique ?

Quand une externalité concerne un petit nombre d'agents, une négociation entre ces derniers est envisageable. Elle donne la possibilité aux agents d'intégrer les externalités en trouvant un accord privé, fondé sur une compensation monétaire des coûts ou des bénéfices externes. Prenons l'exemple d'un complexe touristique situé sur un lac et d'une industrie polluante en amont de la rivière qui alimente ce lac. Considérons que d'un point de vue juridique, l'industrie a le droit de polluer la rivière. Le complexe touristique subit une externalité due à la pollution, dont il peut identifier la cause. S'il est capable d'estimer les coûts du préjudice subi, il a intérêt à engager une négociation avec la firme afin de trouver un accord dans lequel l'industrie s'engage à réduire la pollution en l'échange d'une aide financière du complexe touristique. Ce dernier est prêt à payer pour la réduction d'une unité de pollution tant que le bénéfice qu'il retire de la dépollution est supérieur à l'indemnité qu'il doit verser à l'entreprise polluante pour qu'elle émette une unité de pollution en moins. Cette dernière va accepter de réduire sa pollution d'une unité tant que la somme versée pour réduire d'une unité la pollution émise est supérieure au coût marginal de dépollution. Cet arrangement les amène à intégrer dans leur prise de décision les bénéfices et les coûts externes de leur action. Ils vont ainsi négocier un accord tel que le niveau optimal de pollution est atteint. Le marché n'est donc pas défaillant dans ce cas de figure. Le **théorème de Coase**¹¹ stipule ainsi que si les coûts de transaction (voir le [chapitre 3](#)) sont nuls, les agents sont capables de

trouver une solution efficace pour internaliser les externalités, sans que l'intervention de l'État soit nécessaire. L'internalisation d'une externalité signifie que chacun intègre les coûts et les bénéfices externes de son action dans sa prise de décision.

Ce résultat est valable quelle que soit la façon dont sont définis les droits de propriété. Reprenons l'exemple précédent. Le fait que l'industrie polluante rejette ce que bon lui semble dans la rivière sans que le complexe touristique puisse protester revient à considérer que l'industrie a un droit de propriété sur la rivière puisqu'elle peut en faire libre usage (le droit d'usage est un des droits liés à la propriété). Supposons à présent que la rivière et le lac soient la propriété du complexe touristique. Ce dernier peut interdire à l'industrie le rejet de toute émission polluante. Le complexe subissant un préjudice à la pollution, sans négociation, il ne va pas autoriser l'industrie à émettre la moindre pollution. Mais cette dernière est disposée à verser une somme au complexe pour la dédommager de la pollution subie. L'industrie sera prête à payer tant que le bénéfice marginal qu'elle retire de l'émission d'une unité supplémentaire de pollution (économie de frais de traitement de ses eaux usées, etc.) est supérieur à la dépense supplémentaire qu'elle doit engager pour dédommager le complexe. Ce dernier va autoriser l'industrie à émettre une unité supplémentaire de pollution tant que la somme versée pour autoriser ce surcroît de pollution compense les coûts qu'elle génère. Au final, la négociation va aboutir à une situation équivalente au premier cas de figure du point de vue de la quantité d'émission polluante. En revanche, du point de vue de la répartition des revenus, les deux situations ne sont pas équivalentes. Dans le premier cas de figure, le complexe touristique verse une somme à l'industrie pour qu'elle réduise sa pollution alors que dans le second, c'est l'industrie polluante qui indemnise le complexe pour avoir l'autorisation de polluer.

Le raisonnement présenté ci-dessus pour des externalités négatives vaut aussi en présence d'externalités positives.

Dans cette optique, quand les coûts de transaction sont nuls, les problèmes d'externalités peuvent être gérés par une simple définition claire des droits de propriété par la puissance publique. La régulation marchande, qui repose sur la prise de décision privée décentralisée, n'est ainsi pas toujours défailante en présence d'externalités.

Toutefois, l'hypothèse d'une absence de coûts de transaction, nécessaire

pour valider le théorème de Coase, est en fait très restrictive, car les situations d'externalités sont souvent très complexes, et nécessitent des coûts importants pour négocier un accord, vérifier sa mise en œuvre, etc. R. Coase considérait lui-même que ce théorème était peu réaliste, dans la mesure où la présence de coûts de transaction est plutôt la règle. Quand la pollution résulte des interactions entre des millions de producteurs et d'usagers, comme c'est le cas pour la pollution de l'eau par les nitrates, les coûts de transaction liés aux négociations sont très élevés : un ménage insatisfait de la pollution en nitrates ne peut entamer une négociation avec tous les agriculteurs à l'origine de son préjudice. Dans ce cas, le marché est mis en échec et l'intervention de l'État est nécessaire pour internaliser les externalités afin que l'**optimum social** soit atteint.

3. Quelle régulation publique en cas de défaillance du marché ?

En présence de défaillances du marché dues à des externalités ou à des caractéristiques spécifiques des biens, une intervention publique est généralement nécessaire pour améliorer l'efficacité des décisions économiques. Une gestion privée de l'eau, sans que l'État impose des règles pour défendre l'intérêt collectif, accroît considérablement les risques de subir des dommages pour la santé publique, l'environnement ou la cohésion sociale. Mais les modalités d'intervention publique sont très variées, et parfois opposées entre elles, alors comment choisir les plus efficaces, selon les problèmes posés par la gestion de l'eau ?

La puissance publique peut agir de deux façons : soit l'État contrôle la production des biens concernés en les produisant directement ou en encadrant strictement cette production, soit il cherche à modifier le comportement des agents sur les marchés à travers des systèmes d'incitations et de contraintes.

3.1 Quelles alternatives au recours au marché, en présence de défaillances ?

La solution la plus efficace pour produire certains biens, en raison de leur caractère non rival ou non excluable, ou encore des externalités qui les accompagnent, consiste à les placer sous le contrôle de l'autorité publique. C'est le cas lorsqu'une initiative privée conduirait à produire en quantité insuffisante ces biens, du point de vue du bien-être collectif, comme on l'a vu précédemment. La distribution d'eau potable et l'assainissement des eaux usées en offrent de bons exemples, qui ne sont d'ailleurs jamais laissés au seul marché dans les pays développés. Toutefois, le contrôle exercé sur ces activités peut prendre des formes très diverses. Selon les communes, les différentes étapes de la gestion de l'eau sont confiées soit à des institutions publiques, une **régie** municipale par exemple qui est un service de la ville assurant directement la production d'un bien, soit à des entreprises privées, dans des conditions strictement fixées par un cahier des charges. C'est ce qu'on appelle en France la **délégation de service public**, modèle original qui

s'est développé précisément sur le marché de l'eau, au ^{XIX}^e siècle (BAUBY, SIMILIE, 2007). Il ne faut en effet pas confondre la notion juridique de **service public**, qui place un ensemble d'activités de nature très variable sous le contrôle de l'État au nom de l'intérêt général et de l'égalité d'accès de tous au service considéré, et la nature publique ou privée de l'opérateur en charge de ces activités. Ainsi, la loi fait obligation aux municipalités d'organiser la distribution de l'eau potable, mais elles sont libres de choisir leur mode de gestion.

Nous allons donc examiner le choix entre opérateur privé et public, en fonction notamment de la nature économique des biens concernés, puis nous étudierons le cas particulier de certains biens communs qui échappent au dilemme entre l'État et le marché, par la mise en place d'une gouvernance collective de la communauté concernée, dans la lignée des travaux d'E. Ostrom.

• La production par une institution publique

La production par une institution publique concerne des biens de nature économique très variée, certains sont des biens marchands, vendus à leurs utilisateurs et pris en charge par des entreprises publiques comme les biens tutélaires, d'autres sont non marchands et confiés à des administrations publiques. En effet, une entreprise, publique ou privée, est une institution qui réalise une production destinée à être vendue sur un marché, tandis qu'une administration produit des services non-marchands.

Les agents privés n'ont généralement pas intérêt à produire des biens collectifs, comme nous l'avons vu et ils font donc souvent l'objet d'une **production publique** s'ils sont jugés utiles par la collectivité. C'est le cas des grands services régaliens, comme la justice ou la défense nationale, mais aussi des services administratifs chargés de représenter et de défendre l'intérêt général dans chaque domaine d'activité. Différentes administrations publiques interviennent ainsi dans le processus de gestion de l'eau, comme les directions départementales de la cohésion sociale et de la protection des populations (anciennement DDASS), chargées notamment de la surveillance sanitaire, elles contrôlent la qualité de l'eau fournie dans chaque commune. De même, les agences de l'eau sont des établissements publics autonomes créées par la loi sur l'eau de 1964, qui ont la responsabilité de la gestion des

ressources en eau sur les six grands bassins versants correspondant aux grands fleuves desservis par les rivières françaises. Les biens collectifs fournis par ces administrations (contrôle et application des textes réglementaires, protection des ressources et pilotage de l'aménagement des ressources en eau) ne peuvent être délégués à des agents privés, et ne peuvent être vendus, ces biens étant non excluables. Les ressources disponibles pour les financer proviennent donc de l'impôt.

Les biens de club ou les biens tutélaires peuvent également faire l'objet d'une production publique, même s'ils sont par nature excluables, et pourraient donc donner lieu à une production privée. La production et la distribution d'eau potable relèvent de cette catégorie, elles sont directement prises en charge par une régie municipale dans certaines communes. La production publique permet également de mettre sur le marché la quantité souhaitée d'un bien pour éviter le risque de sous-consommation propre aux biens non rivaux, et éventuellement de le proposer gratuitement ou à un tarif situé en dessous de son coût unitaire, grâce à une subvention de sa production, car il n'est pas rentable de le produire dans ces conditions.

Certains biens sous le contrôle de la puissance publique relèvent d'un **monopole naturel** (voir le [chapitre 4](#)). La distribution de l'eau en présente certaines caractéristiques, les coûts fixes forment une part essentielle des coûts supportés, et le coût marginal est décroissant, à l'échelle d'une agglomération en tout cas. L'eau potable est vendue aux ménages à un prix réglementé, fixé par la puissance publique, ce qui pose le problème du choix d'une tarification appropriée pour couvrir les coûts d'une part et assurer un accès au plus grand nombre d'autre part. En France, la règle générale est de répercuter les coûts de l'eau sur les utilisateurs, de façon à assurer l'équilibre des recettes et des coûts. Ils paient les dépenses de fonctionnement et les investissements courants ainsi que différentes redevances aux organismes publics de gestion de l'eau, en revanche certains investissements lourds dans les infrastructures peuvent être subventionnés par la collectivité (construction d'un barrage par exemple).

Le type de tarification pratiqué est très variable d'une commune à l'autre, le modèle le plus courant est celui de la **tarification binôme** : le consommateur paie un abonnement fixe, qui couvre une part des coûts de fonctionnement du réseau et des dépenses d'investissement, et une part variable en fonction de la quantité consommée. Ce mode de facturation incite

à modérer la consommation d'eau, ce qui est une source d'externalités positives sur l'environnement, mais il crée des inégalités, dans la mesure où le prix moyen de l'eau est de ce fait plus élevé pour les petits consommateurs, qui paient surtout la part fixe. Par ailleurs, les inégalités sont grandes d'une région à l'autre, le prix de l'eau variant en France du simple au double d'un département à l'autre (de 3,50 à 4,57 € le m³ en 2014¹²), ce qui reflète les différences de coûts d'usage de la ressource, mais génère de fortes inégalités territoriales. Il y a donc un débat sur le « prix de l'eau », certains proposent par exemple une tarification progressive pour éviter de pénaliser les foyers modestes, qui consomment moins.

Contrairement à l'eau, certains biens produits par un agent public sont mis à disposition gratuitement, comme les services d'éducation ou de santé. Ces services pourraient être tarifés (et certains le sont), puisqu'ils sont excluables, mais l'exclusion de certains agents en raison de leur prix implique de les priver d'un bien jugé essentiel à tous, principe à la base du droit à l'éducation ou à la santé. Par ailleurs la production publique de biens ou services générant des externalités positives comme l'éducation est une façon d'internaliser ces externalités, l'État décidant de la quantité produite.

Cependant la production publique peut rencontrer des problèmes importants de mise en œuvre, ce qui rend crédible la solution alternative, la confier à un agent privé. Le producteur public n'a pas toujours accès aux solutions techniques les plus performantes, pour diverses raisons : économies d'échelle limitées par rapport à des firmes de taille mondiale, investissement dans la recherche et développement réduits par ses capacités financières, etc. Certaines régies municipales de l'eau sont de trop petite taille par exemple pour offrir des solutions techniques efficaces, car elles ne peuvent employer un personnel compétent à temps plein, tandis qu'une entreprise privée peut partager son personnel entre plusieurs sites municipaux. De ce fait, la régie ne produit pas toujours au meilleur coût. Par ailleurs, une situation de monopole local n'incite pas toujours à l'innovation et à la recherche de gains de productivités. À la suite des travaux de **l'école des choix publics**, l'hypothèse selon laquelle l'État producteur n'est pas forcément « bienveillant » mérite d'être étudiée. Cette école s'intéresse au fonctionnement des institutions publiques dans une économie de marché, et postule que les agents publics poursuivent eux aussi leur intérêt, qui n'est pas forcément en ligne avec l'intérêt collectif, ce qui est une source importante d'inefficacité, l'État ne se comporte donc pas toujours comme un agent bienveillant, agissant

conformément à l'intérêt général (LANDIER, THESMAR, 2010). Le personnel politique peut favoriser par exemple le sur-emploi dans certains services, de façon à se donner une manne de postes à distribuer, afin de peser sur sa réélection, il peut protéger des intérêts corporatistes au détriment de solutions organisationnelles plus efficaces, afin de préserver le climat social dans son propre intérêt, etc.

- **Confier la production au secteur privé, sous la tutelle publique**

Pour gérer les biens placés sous leur tutelle, les autorités publiques ont la possibilité d'en confier la production à un agent privé, afin de pallier les problèmes d'efficacité parfois rencontrés par une production publique. Le droit français a accordé progressivement une place importante à un modèle original, la **délégation de service public (DSP)** : elle consiste à attribuer à une entreprise privée appelée délégataire, pour une période donnée, le monopole de la fourniture d'un bien sur un territoire, dans des conditions fixées par un cahier des charges, déterminant le prix et les caractéristiques des prestations fournies. La délégation de service public est apparue dès l'Ancien Régime mais elle se développe à partir du milieu du XIX^e siècle, avec l'expansion rapide de la fourniture d'eau par des compagnies privées, maîtrisant de mieux en mieux le processus technique. De grands groupes se sont développés dans ce secteur, comme la Générale des eaux devenue Véolia, et la Lyonnaise des Eaux-Suez, ils proposent une large gamme de services aux communes, et se sont fortement internationalisés. Aujourd'hui, la majorité du secteur reste dominé par les DSP, qui couvrent 61 % de la population pour l'eau potable, en 2014 mais seulement 31 % des services communaux ou intercommunaux, car les petites communes ont tendance à privilégier la gestion en régie directe¹³. On observe au cours des années 2000 une inflexion de tendance assez sensible et un retour à la gestion en régie, après plusieurs décennies de développement des DSP. La délégation de service public est très répandue, car elle présente plusieurs avantages pour les communes, liés à la nature de la distribution et de l'assainissement de l'eau. Au niveau local, ces activités ont les caractéristiques d'**industries de réseau** : ce sont des monopoles naturels basés sur la mise en œuvre d'une infrastructure d'interconnexion entre utilisateurs, comme dans l'électricité, les transports ou les télécommunications. Ces industries nécessitent la

présence active d'un **régulateur public**, institution chargée d'organiser la concurrence sur le marché, afin d'éviter une tarification trop élevée par le monopole, et permettre l'accès du plus grand nombre au bien considéré (voir le [chapitre 4](#)). Le régulateur fixe ainsi les conditions d'accès au marché, comme le nombre d'opérateurs, peut encadrer les tarifs pratiqués et peut fixer la durée pendant laquelle une activité est confiée à ces opérateurs (par exemple le contrat de délégation d'un service d'assainissement de l'eau), de façon à rendre le monopole « contestable » périodiquement.

Pourtant, le retour spectaculaire à la régie dans certaines communes, à l'image de Paris et Rouen, découle de la mise en évidence des effets pervers liés à la délégation de service public, qui ont notamment été identifiés par l'économie publique de la réglementation (LÉVÊQUE, 2004). Le délégataire bénéficie d'asymétries d'information importantes, face à une municipalité de taille réduite parfois, à l'expertise limitée, il connaît bien mieux les coûts réels et donc les rentes de situation qu'il pourrait capter. Il risque d'adopter un comportement opportuniste en tirant parti de sa situation pour mentir ou tromper délibérément le donneur d'ordre (voir le [chapitre 3](#)). Le contrôle effectif sur son activité est donc difficile à mettre en œuvre pour les communes. Enfin, la concurrence oligopolistique entre les grands groupes des secteurs peut pousser à des ententes (voir le [chapitre 4](#)) ou à des phénomènes de corruption des élus, pour capter les rentes de situation une fois installés. Les grands groupes du secteur ont participé au financement illicite des partis politiques dans les années 1980-1990, et leur comportement est remis en cause dans certains pays émergents, avec de nombreux mouvements de contestation de la privatisation de l'eau¹⁴ et de ses implications.

Face aux asymétries d'information et au risque de comportement opportuniste d'un agent privé, une solution consiste à mettre en place des mécanismes d'incitations, pour pousser le délégataire à révéler ses informations et à adopter un comportement plus efficace socialement. La législation a progressivement obligé les délégataires à publier un rapport annuel, et à proposer une batterie d'indicateurs économiques pour réduire les asymétries d'information. Par ailleurs, les contrats de délégation prévoient souvent aujourd'hui un tarif réglementé qui augmente chaque année de l'inflation et d'un indice tenant compte des gains de productivité du délégataire : la seule solution pour accroître ses profits, c'est alors de réaliser des gains de productivité supérieurs à ceux intégrés à l'indice, ce qui le pousse donc à rechercher l'efficacité. Les travaux de la nouvelle économie

industrielle développés par J.-J. Laffont et J. Tirole, Prix Nobel en 2014 (Tirole 2016), montrent qu'une solution possible aux asymétries entre réglementeur et entreprises consiste à proposer plusieurs contrats au choix de ces dernières, de façon à inciter les meilleures firmes à choisir le contrat le plus rémunérateur mais le plus exigeant en termes de performances. Par leurs choix contractuels, les entreprises révèlent donc des informations sur leur efficacité productive qu'elles ne souhaitent pas communiquer directement.

• Une production gouvernée en commun ?

Une solution alternative à la production par l'État ou un délégataire désigné par la puissance publique est envisageable dans certains cas, en s'appuyant sur l'existence des communautés humaines liées par la gestion commune d'un bien, comme par exemple les agriculteurs qui dépendent d'une même source d'approvisionnement en eau. Les formes de régulation privée ne sont pas toujours marchandes et peuvent être de type communautaire. Les travaux d'E. Ostrom ont ainsi montré que dans de nombreuses communautés humaines, les biens communs constitués de ressources renouvelables, qui ont tendance à s'épuiser en cas de surexploitation, échappent à la « tragédie des communs » présentée en première partie. Les agents mettent en place des institutions appropriées à la gestion commune de la ressource, qui permettent de la préserver, plutôt que de laisser faire le marché ou de passer à une production publique.

Un des exemples les plus célèbres sur lequel s'est appuyée E. Ostrom (2010) concerne la gestion de l'eau, dans un pays où elle peut facilement manquer à certaines périodes, compte tenu du climat : il s'agit du système des huertas espagnol, mis en place à l'époque musulmane et conservé depuis un millénaire. Cette institution s'appuie sur des règles très précises d'accès aux canaux d'irrigation, de façon à attribuer l'eau à chaque membre à tour de rôle. Un système de gérants élus surveille de façon très précise l'usage de la ressource, et des sanctions sont prévues pour les contrevenants. Elles sont cependant très légères pour éviter que les cultivateurs se retournent contre l'institution. L'adhésion volontaire et la co-organisation des usagers garantissent la pérennité du système.

L'analyse d'E. Ostrom, récompensée du Prix Nobel en 2009, a le mérite de mettre en lumière le rôle des institutions dans la gestion de ressources

partagées, et d'inciter les économistes à l'étude d'organisations plus complexes que le face à face classique entre l'État et le marché, en particulier dans le domaine des nouvelles technologies de l'information, où la question de la « gestion des communs » (logiciels libres, plate-forme collaboratives) se pose avec acuité.

3.2 Quels outils pour réguler les comportements des agents sur les marchés, en présence de défaillances ?

Pour pallier les défaillances du marché, les autorités publiques peuvent agir sur les comportements des agents présents, plutôt que d'intervenir directement sur la production. Il n'est en effet pas envisageable que l'État assure la production de tous les biens générant des externalités ou présentant des caractéristiques spécifiques. Le nombre de biens concernés est trop important et la régulation marchande présente un certain nombre d'avantages. En présence d'asymétries d'information, l'État n'est pas en mesure d'évaluer parfaitement les préférences des consommateurs et peut ainsi produire des biens et proposer des prix inadaptés. Dans les économies planifiées, les exemples de sous-production de biens désirés par les consommateurs et de sur-production de biens jugés moins intéressants sont légion.

Comment alors encadrer le fonctionnement du marché afin de pallier ces défaillances ? Les autorités publiques disposent de différents instruments pour agir sur un marché : elles peuvent imposer des règles pour contrôler la production, telles des quotas, mettre en place un système de taxes ou subventions, etc.

On distingue plus précisément trois instruments pour agir sur un marché, à savoir la réglementation, l'action sur les prix par la taxation et l'action sur les quantités par des quotas ou des permis d'émission (voir le [chapitre 1](#)). Quels sont les effets de ces différents instruments, pour limiter les défaillances du marché ? Quels sont les avantages et les inconvénients de chacun d'entre eux ?

• La réglementation est-elle efficace ?

Le recours à la **réglementation** consiste à définir des normes qui s'imposent à tous les agents, sous le contrôle des autorités publiques. C'est le principal

instrument utilisé actuellement pour réguler les prélèvements d'eaux souterraines en France (MONTGINOUL, RINAUDO, 2009)¹⁵. Les prélèvements individuels dans les nappes phréatiques offrent un bon exemple des problèmes collectifs posés par la gestion d'une ressource commune qui risque de s'épuiser localement ou de se dégrader, et qui est à l'origine de nombreuses externalités. Une surexploitation d'une nappe phréatique entraîne par exemple des externalités de coûts de pompage (il faut pomper plus profondément), diminue la possibilité de s'appuyer sur la nappe phréatique pour faire face aux fluctuations des eaux de surface, et accroît le risque de dégradation de la ressource (par exemple, une nappe qui se vide au bord des côtes est rapidement polluée par l'eau de mer qui l'envahit). Les normes en vigueur prévoient par exemple la déclaration obligatoire des forages, et la fixation d'un **quota de prélèvements**. Dans les périodes de raréfaction de l'eau, les prélèvements sont interdits pour une durée progressive, de façon à assurer la reconstitution des réserves.

Ce type d'instruments est en apparence peu coûteux pour l'État contrairement à des subventions par exemple, et de nature à éviter l'épuisement des nappes phréatiques et les externalités négatives que cela génère. Mais en réalité, il se révèle peu efficace, conformément aux limites repérées depuis longtemps par l'analyse économique. Les normes ne valent d'abord que si elles sont appliquées, or plus elles se révèlent contraignantes et difficiles à vérifier, et plus elles incitent les agents à y échapper : dans le Roussillon par exemple, région particulièrement touchée par les risques d'épuisement des eaux souterraines, seuls 5 à 10 % des forages seraient déclarés. Si on accroît les contrôles, les normes seront davantage appliquées, mais les coûts de surveillance peuvent se révéler très élevés. Les quotas entraînent souvent des effets distorsifs importants, car une norme qui s'applique uniformément à tous les agents ne tient pas compte du fait que certains peuvent s'y soumettre plus facilement que d'autres. Ainsi, pour un objectif donné de réduction des émissions polluantes, il est plus efficace d'un point de vue économique de demander un effort relativement plus important aux entreprises pour lesquelles le coût de dépollution est moindre. Imposer à toutes les entreprises le même quota d'émissions polluantes ne permet pas d'atteindre l'objectif à moindre coût, car elles sont toutes obligées de se soumettre au même objectif alors qu'elles ont des contraintes de coûts très différentes. Il en est de même si l'on considère une norme de prélèvement maximal en eau souterraine par exploitation agricole et par hectare cultivé.

Elle aura des effets très variés selon que la culture réclame peu ou beaucoup d'eau. La réduction des prélèvements est ainsi beaucoup plus coûteuse pour certaines exploitations que pour d'autres.

Il est par ailleurs très difficile d'évaluer le niveau auquel l'État doit fixer les quotas. En présence d'asymétries d'information, l'État ne connaît pas le niveau optimal de production car il ne peut s'appuyer sur les transactions effectuées (voir le [chapitre 4](#)).

Ce type d'instrument est toutefois adapté pour les émissions polluantes dangereuses où une réglementation stricte est nécessaire car il importe que toutes les entreprises, quelle qu'elles soient et où qu'elles se situent respectent ces normes, pour des raisons de santé publique.

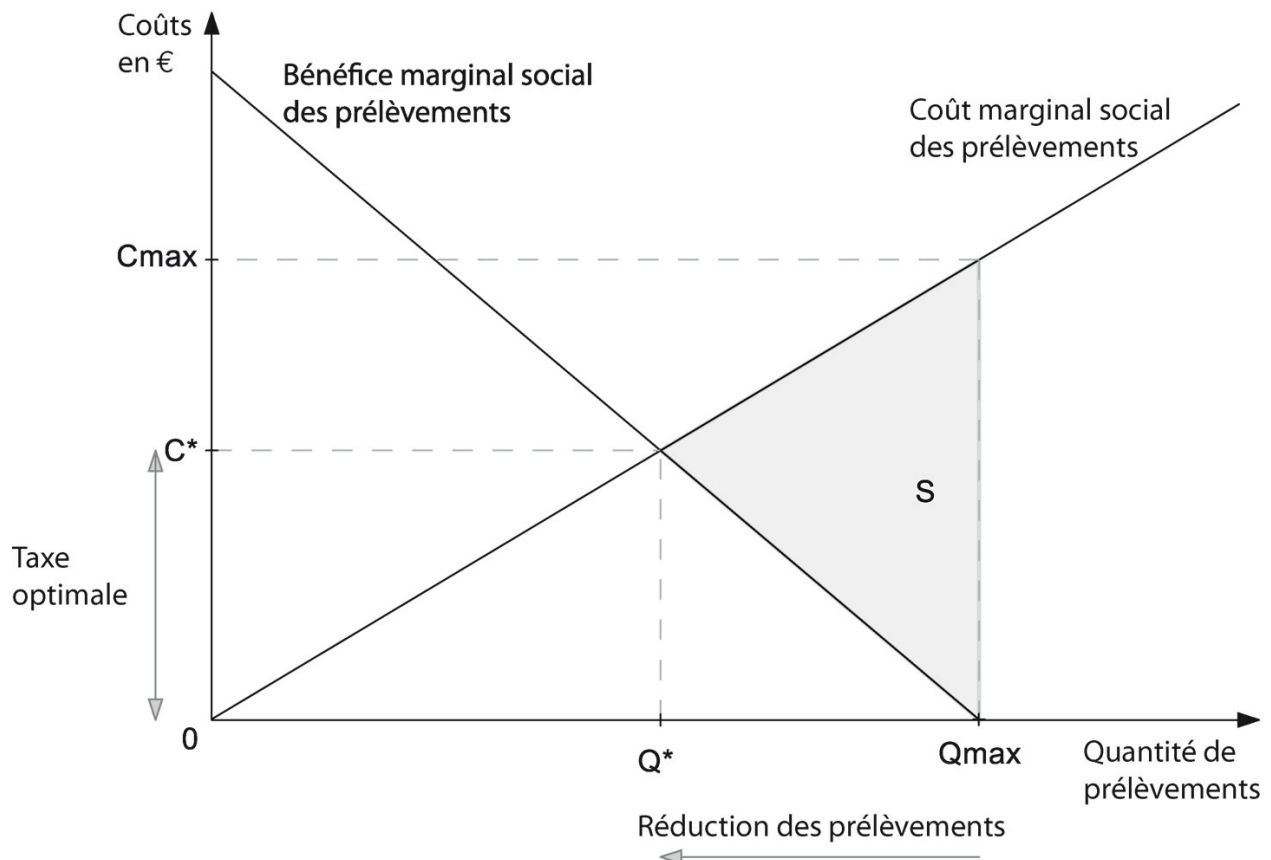
• Les taxes et subventions

Une alternative à la régulation par des normes consiste à agir sur les prix. En taxant les producteurs responsables d'externalités négatives ou en subventionnant les producteurs à l'origine d'externalités positives, les agents sont conduits à internaliser les effets externes de leur action.

Dans cette optique, il faut d'abord évaluer le coût ou les gains générés par les externalités pour déterminer le niveau de la taxe ou de la subvention. Considérons le cas d'industries qui rejettent dans les rivières des produits toxiques utilisés au cours du processus de production. L'État peut instaurer une taxe sur chaque tonne de produits rejetés par ces entreprises afin qu'elles intègrent dans le calcul de leur coût de production l'externalité négative que leur production génère. Les entreprises cherchent à améliorer leur technique de production afin de polluer moins et d'avoir à payer moins de taxe pour un volume de production donné mais cette opération a un coût et la taxe conduit d'une façon générale à une augmentation des coûts de production. La taxe va ainsi conduire à diminuer l'offre des biens que ces entreprises proposent à la vente pour un prix de vente donné et ainsi moins de biens vont se vendre et ce à un prix plus élevé. La taxe a conduit à diminuer les émissions polluantes mais est en partie reportée sur le consommateur, selon les principes de l'incidence fiscale (voir le [chapitre 1](#)). Une **taxe pigouvienne**¹⁶ consiste à taxer l'agent à l'origine d'une externalité négative de façon que son montant soit égal à la différence entre les coûts privés et le coût social de l'activité qui génère des externalités. Supposons que l'on souhaite taxer les prélèvements

en eau souterraine réalisés par les agriculteurs, de façon à atteindre le niveau optimal de prélèvement, compte tenu des externalités négatives qu'ils génèrent. Si on reprend la logique de l'optimum de pollution présenté en deuxième partie du chapitre, on voit sur le [graphique 5.4](#) que les agents ont intérêt à prélever $Q_{\max}$ en l'absence de taxation, ce qui correspond en fait à la quantité prélevée d'eau s'ils ne tiennent compte que de leurs avantages, car ils ont intérêt à augmenter leurs prélèvements tant qu'ils réalisent un bénéfice marginal supérieur à zéro. Du point de vue collectif, la quantité de prélèvements en eaux souterraines optimale est Q^* , pour laquelle le bénéfice marginal social d'une unité supplémentaire compense exactement son coût marginal social.

Graphique 5.4 : La taxe pigouvienne



Si la puissance publique impose une taxe d'un montant C^* , c'est-à-dire une taxe égale au coût marginal social correspondant à la quantité optimale de prélèvements, elle incite les producteurs à réduire leur prélèvement de Q_{max} à Q^* . En effet, le bénéfice marginal réalisé par les agriculteurs est désormais diminué du montant de la taxe, il devient donc nul pour Q^* et il est négatif au-delà de cette quantité, donc les producteurs n'ont pas intérêt à prélever au-delà de Q^* , qui correspond à une production optimale. La taxe les oblige à tenir compte du coût marginal social de la pollution.

En internalisant les coûts sociaux dus aux effets externes, les agents se comportent désormais comme s'ils subissaient des coûts privés équivalents aux coûts sociaux : ils ont intérêt à choisir le niveau de prélèvement optimal. La surface S représente le gain social réalisé grâce à la taxe, soit la différence entre le supplément de coûts sociaux subis et les bénéfices sociaux résultant des prélèvements en eaux souterraines.

La taxation pigouvienne présente de nombreux avantages théoriques. Elle conduit d'une part les agents à faire le choix de pollution optimale et conduit à ce que les pollueurs supportent les coûts sociaux de la pollution : c'est le

principe du pollueur-payeur. Du fait, des mécanismes de marché, notons néanmoins comme nous l'avons vu que les pollueurs ne supportent pas en fait seuls le poids de la taxe.

On considère aussi qu'elle peut générer un **double bénéfice** (on parle aussi de double dividende), si elle entraîne à la fois une réduction des externalités, et un accroissement des recettes publiques.

Cependant, la taxation pigouvienne rencontre aussi des problèmes de mise en œuvre qui conduisent à s'interroger sur son efficacité réelle. Il faut supposer que la puissance publique dispose d'un excellent niveau d'information, de manière à déterminer précisément les coûts sociaux des externalités, faute de quoi elle pourrait introduire une taxation ayant plus d'effets perturbateurs que correctifs. D'où la question cruciale de l'évaluation des externalités, abordée en deuxième partie. Par ailleurs, une taxation élevée, efficace du point de vue collectif mais dont l'effet est perceptible par chaque agent, pose aussi des problèmes d'acceptabilité sociale.

Du point de vue d'une taxation pigouvienne, le niveau actuel de la fiscalité sur les prélèvements des eaux souterraines est clairement insuffisant (MONTGINOUL, RINAUDO, 2009) : la redevance versée à l'Agence de l'eau, que doivent payer ceux qui prélèvent des eaux souterraines (quand ils sont déclarés et disposent d'un compteur d'eau), est de l'ordre en 2016 de 0,03 à 0,06 € le m³, et les agriculteurs bénéficient en plus d'une décote importante par rapport à ce prix déjà très faible. La taxation n'est pas pensée pour réduire sensiblement les prélèvements, elle sert surtout à financer certains investissements sur les cours d'eau. Au-delà de cet exemple, l'expérience des dernières années montre que l'acceptabilité sociale d'une taxation écologique ne va pas de soi, comme le suggère l'abandon de la Contribution Climat Énergie, qui devait taxer les émissions carbone associées à un certain nombre de biens. Elle s'est heurtée en 2009, après son adoption au Parlement, à l'avis négatif du Conseil constitutionnel, qui s'opposait à la rupture d'égalité devant l'impôt de la taxe en raison des nombreuses exemptions envisagées, or ces exemptions avaient précisément été négociées au départ... pour favoriser son acceptabilité.

• Les marchés de quotas

Une autre solution envisageable pour réduire les externalités est la mise en

place d'un **marché de quotas**. Au lieu d'imposer à chaque agent un quota individuel d'émission polluante, la puissance publique définit au niveau global un volume maximal de pollution, par exemple 10 000 tonnes d'émissions de carbone par an. En fonction de cet objectif, elle va créer des droits d'émission, par exemple 10 000 titres donnant chacun l'autorisation d'émettre 1 tonne de carbone par an. Une entreprise qui désire émettre 30 tonnes de carbone par an doit impérativement posséder 30 titres. L'État peut initialement donner ou vendre ces titres aux entreprises, dans les deux cas elles sont en mesure de les vendre entre elles sur un marché, dit marchés de quotas. Supposons que le prix d'un titre d'une tonne d'émission de carbone se fixe à 100 euros. Les entreprises achètent un titre si le coût de réduction de ses émissions d'une tonne de carbone est supérieur à 100 euros. Sinon, l'entreprise a intérêt à modifier ces techniques de production afin de réduire sa pollution et d'économiser le coût d'achat d'un titre. Les entreprises qui possèdent des titres vont les revendre si leur prix est supérieur au coût de réduction d'une tonne de pollution. En fonction des coûts de dépollution de chaque entreprise, le prix du titre va se fixer plus ou moins haut. L'objectif de pollution va être atteint à moindre coût car au final les entreprises pour lesquelles les coûts de dépollution sont faibles vont plutôt choisir de dépolluer plutôt que d'acheter des titres alors que celles qui subissent des coûts élevés de dépollution vont chercher à acquérir des quotas d'émission. Cette solution semble plus efficace qu'un système de quotas individuels. Si le marché des quotas est ouvert à tous les acteurs (et pas seulement réservé aux entreprises polluantes), les individus qui désirent voir le niveau de pollution diminuer peuvent acheter des titres afin de les stocker et réduire le niveau global de pollution. Il est ainsi possible d'imaginer qu'une association écologique retire 1 000 titres du marché afin de faire porter à 9 000 le nombre de tonnes de carbone émises par an. En faisant cela, elle révèle la valeur de sa préférence pour l'environnement. Dans un tel cas de figure, le mécanisme du marché conduit le prix à se fixer de façon telle que la pollution se situe à son niveau optimal.

Nous avons vu qu'en présence d'asymétries d'information entre la puissance publique et les acteurs privés sur les coûts et bénéfices des différentes actions, il est impossible pour l'État de définir de façon optimale le niveau d'une taxe. Dans ce cas de figure, le marché de quotas joue un rôle efficace. Cependant, ce résultat repose sur l'hypothèse que le marché de quotas ne présente pas d'imperfections. Cette solution s'inspire en fait des

travaux de Coase sur l'internalisation des externalités par les transactions entre les agents, évoquée précédemment. De tels systèmes sont déjà à l'œuvre, comme le marché européen du carbone, créé en 2005 mais sont généralement réservés aux entreprises polluantes. Le marché européen du carbone a connu des vicissitudes importantes (LAURENT, CACHEUX, 2015). Une partie des problèmes vient du fait que les titres ont été initialement émis en trop grand nombre, ce qui a eu pour conséquence un prix trop faible de la tonne de carbone n'incitant pas les entreprises à réduire leurs émissions. Par ailleurs, ces marchés sont souvent caractérisés par une instabilité endogène, les comportements spéculatifs et les anticipations des agents conduisant à de fortes variations des cours qui peuvent perturber le bon fonctionnement du marché, et sa capacité à fixer les prix permettant d'atteindre le niveau optimal de pollution. Notons toutefois qu'il est possible de limiter ces fluctuations en régulant le fonctionnement du marché, comme en imposant par exemple un prix plancher et un prix plafond au prix du carbone.

Dans le cas des eaux souterraines, la mise en place d'un marché des droits de prélèvements sur cette ressource est envisageable, sur le modèle de ce que l'on observe déjà dans certains pays comme l'Australie. Il s'agit de fixer une quantité maximale de prélèvement des eaux souterraines à une échelle locale pertinente du point de vue de la gestion durable de la ressource, puis de distribuer aux usagers, et principalement aux agriculteurs, des droits de prélèvement en fonction de cette quantité globale et de l'historique de leurs prélèvements antérieurs, pour ne pas pénaliser les agriculteurs les plus productifs initialement. Aux périodes suivantes, si un agriculteur a besoin de prélever davantage qu'un autre, il peut acquérir des permis de prélèvement, ou au contraire vendre ses permis en surplus quand il n'en utilise pas la totalité (MONTGINOUL, RINAUDO, 2009).

Le choix entre les différents instruments ne peut pas être fixé en soi et dépend du problème précis qu'il s'agit de résoudre.

Cependant, quelques règles peuvent être énoncées : il faut d'abord un système efficace de contrôles, organisé par la puissance publique, pour que ces diverses solutions entraînent les effets attendus, en prévoyant des pénalités incitatives en cas de non-respect des normes, de non-paiement d'une taxe ou de production polluante sans disposer d'un permis d'émission. Une évaluation rigoureuse des coûts externes est nécessaire pour fixer le niveau de pollution optimal. Le processus de négociation politique des

contrats doit aussi être examiné avec soin, pour éviter que certains agents bénéficient de la négociation, en imposant le choix d'un instrument à leur avantage (Tirole, 2016). En fait, les autorités publiques ont souvent intérêt à mettre en place des solutions mixtes, combinant les différents instruments, car il est rarement possible de mettre en œuvre un seul mécanisme et de s'assurer pleinement de son efficacité.

Notes

- [1.](#) La détermination du prix fixé par le producteur dépend de son pouvoir de marché dans ces conditions, voir le chapitre 4.
- [2.](#) Il serait plus rigoureux de parler de « bien collectif mondial » ou de « bien commun mondial » mais la traduction littérale du terme anglais de « *global public good* » s'est imposée, notamment par sa diffusion *via* les organisations internationales comme l'ONU.
- [3.](#) La notion est introduite par C. Tiebout, « A Pure Theory of Local Expenditures », *Journal of Political Economy*, 1956.
- [4.](#) Garrett Hardin, « The tragedy of commons », *Science*, vol. 162, n° 3859, 1968.
- [5.](#) R. Coase, « Le phare en économie », *The Journal of Law and Economics*, 17, n° 2, 1974, repris in *L'entreprise, le marché et le droit*, Éditions d'Organisation, 2005.
- [6.](#) James E. Meade, « External economies and diseconomies in a competitive situation », *Economic Journal*, vol. 62, n° 245, 1952.
- [7.](#) A. Rozan, A. Stenger et M. Willinger, *Valeur de préservation et transférabilité des bénéfices : application à la nappe phréatique d'Alsace*, Rapport de contrat pour le GIP Hydrosystèmes, 1996.
- [8.](#) Rapport de la commission présidée par Alain Quinet (2009), « La valeur tutélaire du carbone », Centre d'analyse stratégique – La Documentation française.
- [9.](#) Le terme « avantage » traduit plus fidèlement l'expression anglaise *benefit* même si l'usage du terme « bénéfice » s'est largement répandu. Le mot « bénéfice » n'a donc pas ici le sens courant de la différence entre la recette liée à la vente d'un bien ou service et son coût de production.
- [10.](#) L. R. Brown, *Éco-économie : une autre croissance est possible, écologique et durable*, Le Seuil, 2003.
- [11.](#) Le « théorème de Coase » est en réalité formulé par G. Stigler, à partir de son commentaire d'un article de Coase consacré au « problème du coût social ». Voir G. Stigler, *The Theory of price*, New York, MacMillan, 1966 (trad. fr. aux éditions Bordas, 1980).
- [12.](#) Prix moyens des départements métropolitains, selon le rapport 2017 de l'Observatoire des services publics d'eau et d'assainissement.
- [13.](#) Source : rapport 2017 de l'Observatoire des services publics d'eau et

d'assainissement.

[14.](#) O. Hoedeman et S. Kishimoto, *L'eau, un bien public – Alternatives démocratiques à la privatisation de l'eau dans le monde entier*, Éditions C.-L. Mayer, 2010.

[15.](#) Les développements présentés dans cette partie à propos des prélèvements en eau souterraine s'inspirent largement de cet article.

[16.](#) Le nom fait référence à A. C. Pigou, précurseur de l'économie publique et de la théorie des externalités, dans son livre *The Economics of Welfare* (1920).

Chapitre 6

Microéconomie du marché du travail

L'exemple des jeunes

1. LES DÉTERMINANTS DE L'OFFRE DE TRAVAIL

2. LES DÉTERMINANTS DU SALAIRE

3. L'EXPLICATION MICROÉCONOMIQUE DU CHÔMAGE

En 2016 en France, le **taux d'activité** des 15-24 ans est de 36,9 %, ce qui signifie que 36,9 % des individus âgés entre 15 et 24 ans sont des actifs, c'est-à-dire qu'ils occupent ou cherchent à occuper un emploi. Parmi ces actifs, 24,6 % sont au chômage en 2016, le **taux de chômage** étant le rapport entre le nombre de chômeurs, c'est-à-dire le nombre de personnes cherchant un emploi et disponibles pour travailler, et celui des actifs, pour une tranche d'âge donnée. Les jeunes sont les premiers touchés dans les périodes de crise économique, et la progression du chômage des jeunes observée après la crise de 2008-2009 le confirme. Au total, le **taux d'emploi** des 15-24 ans est de 27,8 % en 2013, ce qui signifie que 27,8 % des individus ayant entre 15 et 24 ans occupent un emploi. Les jeunes participent moins que leurs aînés au marché du travail (taux d'activité de 87,9 % chez les 25-49 ans en 2016) mais sont bien plus touchés qu'eux par le chômage lorsqu'ils le font : par comparaison, le taux de chômage des 25-49 ans est de 9,3 %. Par ailleurs, le type d'emploi occupé n'est pas le même : en 2016, 32,1 % des salariés âgés entre 15 et 24 ans ont un contrat à durée déterminée (CDD) contre 9,5 % des 25-49 ans¹. D'une façon générale, les jeunes définis conventionnellement par la tranche d'âge des 15-24 ans², occupent des emplois plus précaires que leurs aînés. Les jeunes travailleurs ont aussi en moyenne des salaires plus faibles que ces derniers. Le salaire net annuel moyen en équivalent temps plein en 2014 était de 17 010 euros pour les 15-24 ans contre 24 110 euros pour les 25-39 ans et 28 790 euros pour les 40-49 ans³.

Pourquoi les jeunes éprouvent-ils plus de difficultés que leurs aînés à trouver un emploi et sont-ils le plus souvent inactifs ? Pourquoi occupent-ils souvent des emplois précaires et faiblement rémunérés ?

Ces questions, forts complexes, ont fait l'objet de nombreuses études aussi bien en science économique qu'en sociologie. L'objet de ce chapitre est de montrer comment l'analyse microéconomique peut être mobilisée pour

comprendre le fonctionnement du marché du travail, et plus spécifiquement les comportements des jeunes sur ce marché. Les microéconomistes ont depuis longtemps développé des modèles afin d'aider à comprendre les décisions de participation au marché du travail, les déterminants du salaire et de l'emploi. Dans quelle mesure ces modèles peuvent-ils être mobilisés pour comprendre la situation des jeunes sur le marché du travail ?

1. Les déterminants de l'offre de travail

Sur le marché du travail, les offreurs sont les travailleurs qui cherchent à vendre une certaine quantité de travail aux entreprises, mesurée en heures par exemple, moyennant le paiement d'un prix unitaire, le **taux de salaire**. Une **offre de travail** équivaut ainsi à une demande d'emploi. L'objet de cette partie est de comprendre les déterminants de l'offre de travail. Pourquoi un individu va-t-il décider de participer au marché du travail ? En fonction de quels critères va-t-il chercher plutôt à occuper un travail à temps partiel ou à temps plein ?

1.1 L'arbitrage consommation/loisir

• Rémunération du travail et renonciation au loisir

En France, les jeunes de plus de 16 ans, l'âge requis légalement, peuvent chercher un travail pendant l'été ou en parallèle de leurs études pour accroître leur revenu et ainsi leur consommation. Consommer plus leur procure de la satisfaction, de l'utilité dans le langage microéconomique : ils souhaitent acquérir des vêtements, sortir dans des bars ou restaurants, etc. Laissons de côté pour l'instant les jeunes étudiants : la plupart des autres n'ont pas une aide suffisante de leurs parents pour atteindre le niveau de consommation désiré, ils cherchent à travailler pour accroître leur capacité de consommation et ainsi leur satisfaction. Symétriquement, les emplois occupés n'étant pas toujours intéressants et le temps passé à travailler n'étant plus disponible pour sortir avec ses amis, se reposer, etc., le travail réduit donc la satisfaction. L'analyse microéconomique part du principe que les individus doivent réaliser un arbitrage entre consommation et loisir : s'ils veulent consommer beaucoup, il leur faut un revenu élevé et pour cela travailler beaucoup et ainsi renoncer à un niveau élevé de loisir. Les économistes considèrent que l'arbitrage entre consommation et loisir ne porte pas sur la totalité du temps disponible. On exclut en effet ce qu'on appelle le « temps physiologique » qui désigne le temps nécessaire à un individu pour dormir, se nourrir, etc. La différence entre le temps total et le temps physiologique permet de déduire le

temps arbitrage entre travail et loisir. Le terme de loisir n'est pas employé en économie au sens courant du terme. Il ne renvoie pas nécessairement à l'oisiveté : un individu peut en effet utiliser son temps de loisir pour étudier par exemple, et chaque étudiant sait qu'une heure de « job » réduit le temps disponible pour ses études. Le **loisir** désigne ainsi simplement la part du temps arbitrage non travaillée : une jeune mère qui décide de s'occuper de ses enfants plutôt que de travailler fait elle aussi un arbitrage entre travail et loisir. Le **salaire** est généralement la principale source de revenus, si on ne peut compter sur d'autres ressources : pour consommer davantage, ils doivent travailler. En considérant que les individus ne peuvent travailler plus de 48 h par semaine, limite légale en Europe en intégrant les heures supplémentaires, ces derniers doivent décider s'ils désirent travailler 25 heures, 35 heures, 40 heures, etc. sachant que chaque heure travaillée leur rapporte un certain taux de salaire. Le taux de salaire réel représente ainsi le **coût d'opportunité** d'une heure de loisir : chaque heure de loisir supplémentaire implique de renoncer à un certain niveau de consommation. Désirer travailler 35 heures par semaine ne garantit en rien de trouver une entreprise qui propose un emploi pour ce temps-là. Nous nous intéressons ici simplement aux souhaits des individus, qu'ils soient réalisés ou non par la suite, car ils déterminent le volume de travail qu'ils sont prêts à offrir. Tous les individus n'ont pas le même rapport à la consommation et au travail : certains valorisent très fortement le fait de pouvoir consommer alors que d'autres ont un rapport plus distancié à la consommation. Certains sont aussi très rétifs au travail alors que d'autres ne voient pas trop d'inconvénient à renoncer à du temps de loisir. Les individus ont donc des préférences différentes pour la consommation et le loisir. À l'instar des préférences pour les différents paniers de consommation (voir [chapitre 2](#)), un graphique donne une représentation des goûts des individus à travers des courbes d'indifférence, comme on le voit sur le [graphique 6.1](#).

L'analyse microéconomique part du principe que les individus cherchent à maximiser leur satisfaction. En fonction de leur préférence pour la consommation et le loisir, du taux de salaire réel auquel ils peuvent prétendre sur le marché et des revenus dont ils bénéficient en ne travaillant pas, les individus arbitrent entre consommation et loisir, selon la même logique que leur arbitrage entre différents paniers de consommation étudiés dans le chapitre 2. Supposons qu'un jeune puisse prétendre à un salaire réel de 10 euros net de l'heure s'il travaille, qu'il dispose d'un revenu non salarial

sous forme d'une aide de ses parents de 150 euros par semaine et qu'il puisse au maximum travailler 48 heures par semaine. Compte tenu de ses préférences pour la consommation et le loisir, il choisit de travailler le nombre d'heures par jour qui lui permet de maximiser sa satisfaction.

Il gagne au plus 630 euros par semaine s'il travaille pendant 48 heures ($48 \times 10 + 150$) et son revenu est égal à 150 euros s'il ne travaille pas. La droite de budget représente l'ensemble des combinaisons entre nombre d'heures de loisir (et donc de travail par différence avec l'horaire maximal) et niveau de consommation, et sa pente dépend du taux de salaire réel. Nous voyons sur le [graphique 6.1](#) qu'en travaillant 13 heures par semaine, cet individu dispose d'un revenu de 280 euros ($13 \times 10 + 150$) et maximise sa satisfaction, puisqu'au point A il atteint la courbe d'indifférence la plus élevée possible. À l'optimum, le rapport entre l'utilité marginale du loisir et de la consommation est égal au taux de salaire réel.

Les économistes considèrent que la décision d'**offre de travail** se scinde en fait en deux temps. L'individu se demande tout d'abord s'il participe au marché du travail : on parle d'une décision le long de la **marge extensive** car c'est le nombre de travailleurs qui augmente alors. S'ils ont décidé de travailler, ils déterminent ensuite le nombre d'heures qui leur convient le mieux, il s'agit alors d'une décision le long de la **marge intensive**. On distingue ces deux temps au lieu de considérer une décision unique car il existe un coût fixe de participation au marché du travail, quel que soit ensuite le nombre d'heures travaillées. Un étudiant désirant travailler devra ainsi par exemple peut-être avoir à s'acheter un costume pour passer des entretiens d'embauche qu'il travaille ensuite une heure par semaine ou 30 heures dans l'entreprise. Ce coût n'est pas seulement monétaire, il renvoie aussi à une perte de satisfaction liée par exemple à l'impossibilité de décider pleinement de la date de ses jours de congés.

• Les caractéristiques de l'offre de travail

Lorsque le taux de salaire réel auquel peut prétendre un individu change, le nombre d'heures de travail qui maximise sa satisfaction évolue. L'effet d'une augmentation du taux de salaire sur l'offre de travail est ambigu : pour un meilleur salaire, un étudiant peut d'une part profiter de ce supplément de pouvoir d'achat pour travailler moins d'heures et maintenir le même niveau

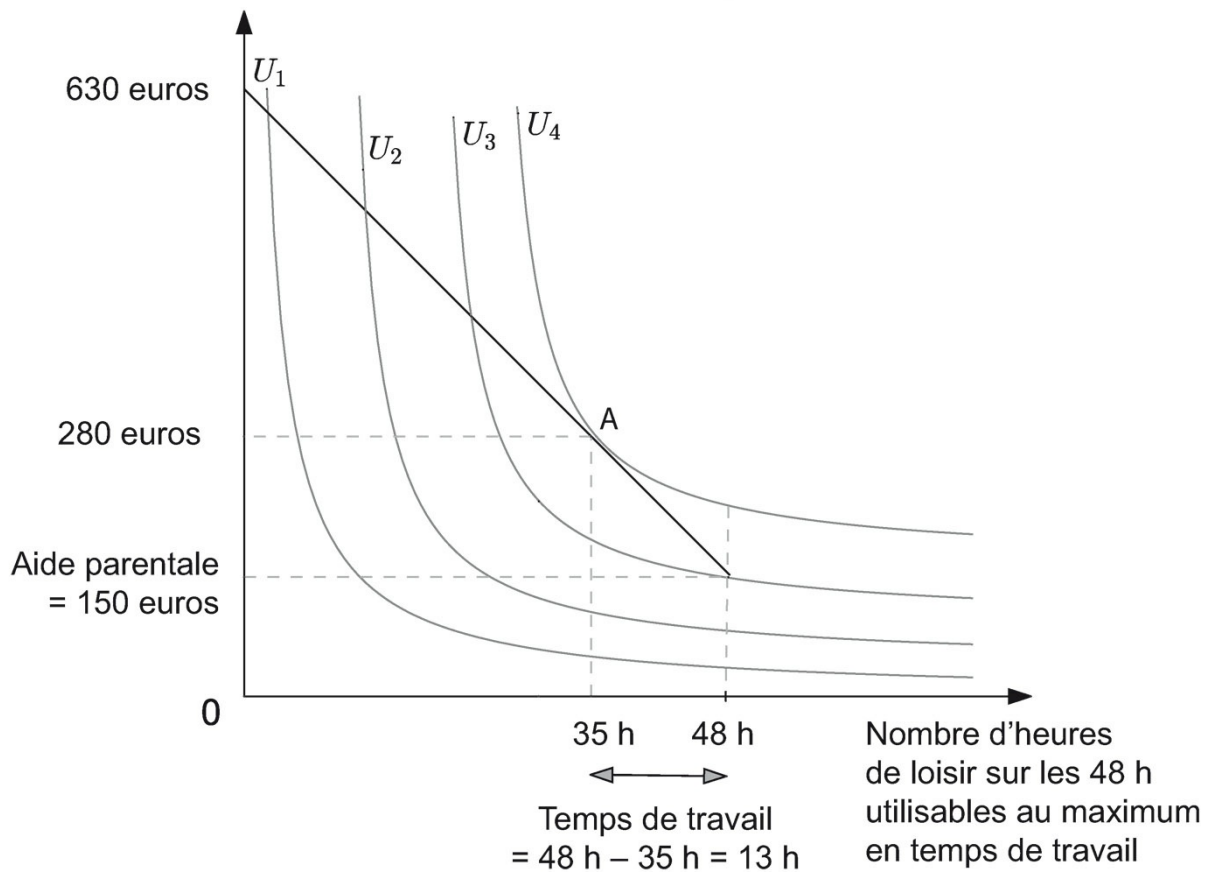
de consommation. L'**effet revenu** désigne ici l'impact de la variation du taux de salaire sur le pouvoir d'achat de l'agent, qui peut l'utiliser pour modifier son arbitrage entre travail et loisir. La hausse du taux de salaire rend plus attractive chaque heure travaillée car elle est mieux rémunérée, ce qui accroît le coût d'opportunité du loisir : c'est l'**effet de substitution**. Comme pour les choix de consommation, la somme de l'effet revenu et de l'effet de substitution donne l'effet global d'une variation du taux de salaire réel sur l'offre de travail (voir le [chapitre 2](#)).

Comme on l'a vu précédemment, avec un taux de salaire réel de 10 €, l'étudiant ayant les préférences représentées par les courbes d'indifférence tracées sur le [graphique 6.2](#) travaille 13 heures pour un revenu hebdomadaire de 280 €. Si le taux de salaire augmente jusqu'à 16 €, la nouvelle droite de budget a une pente plus élevée, et l'étudiant décide alors de travailler 10 heures par semaine pour atteindre une meilleure satisfaction au point B. L'augmentation du taux de salaire a eu ici pour effet de diminuer son offre de travail, l'effet revenu étant supérieur à l'effet de substitution. Bien que le coût d'opportunité du loisir ait augmenté, l'individu travaille moins. Il dispose pourtant d'un niveau de consommation plus élevé que dans le cas précédent car chaque heure travaillée lui rapporte plus (310 euros contre 280 euros dans la situation précédente).

L'**élasticité de l'offre de travail au taux de salaire** permet de mesurer les effets d'une variation relative du taux de salaire sur l'offre de travail relative des individus. On parle aussi de l'**élasticité des heures travaillées au taux de salaire**. Elle se calcule en rapportant la variation relative de la quantité d'heures travaillées désirée à la variation relative du taux de salaire. Cette valeur n'est pas forcément constante pour tous les niveaux de salaire. Un individu peut être très sensible à la variation du taux de salaire quand ce dernier est faible pour ensuite être moins sensible quand son taux de salaire est élevé. C'est pour cela, par exemple, que l'on distingue dans les évaluations empiriques l'élasticité de l'offre de travail des hauts revenus de celle des bas revenus.

Graphique 6.1 : L'arbitrage consommation-loisir

Revenu disponible
à la consommation



La **courbe d'offre de travail individuelle** associe chaque taux de salaire envisageable au nombre d'heures de travail souhaitées par l'agent. La forme de la courbe d'offre de travail dépend de ses préférences. Elle peut par exemple être croissante pour de faibles taux de salaire puis décroissante pour des taux plus élevés : c'est alors une courbe d'offre de travail coudée. Ce cas de figure n'est pas improbable, par exemple pour de jeunes mères célibataires qui doivent, sur leur temps de loisir (au sens économique du terme) s'occuper de leur enfant, et considèrent au-delà d'un certain taux de salaire que le supplément de revenu leur permet de réduire leur offre de travail et de consacrer un peu plus de temps à leur enfant. Les études économétriques montrent que les mères isolées se trouvent parfois dans ce cas de figure (CAHUC, ZYLBERBERG, 2015).

Sur le [graphique 6.3](#), la quantité de travail offerte augmente avec le taux de salaire : l'effet de substitution domine l'effet revenu pour tous les niveaux de salaire. Ainsi, quand le taux de salaire augmente de 10 euros net de l'heure

à 11 euros, l'offre de travail de l'individu augmente : ce dernier décide de travailler 8 heures de plus. Le taux de salaire en dessous duquel un individu ne désire pas participer au marché du travail est appelé **salaire de réserve**. Dans cette optique, la faible participation des jeunes au marché du travail pourrait s'interpréter comme étant le signe qu'ils ont un salaire de réserve élevé. Cette explication semble ici peu pertinente pour expliquer ce phénomène et nous montrerons plus loin dans le chapitre que d'autres types d'explications peuvent être mobilisés pour comprendre un tel phénomène.

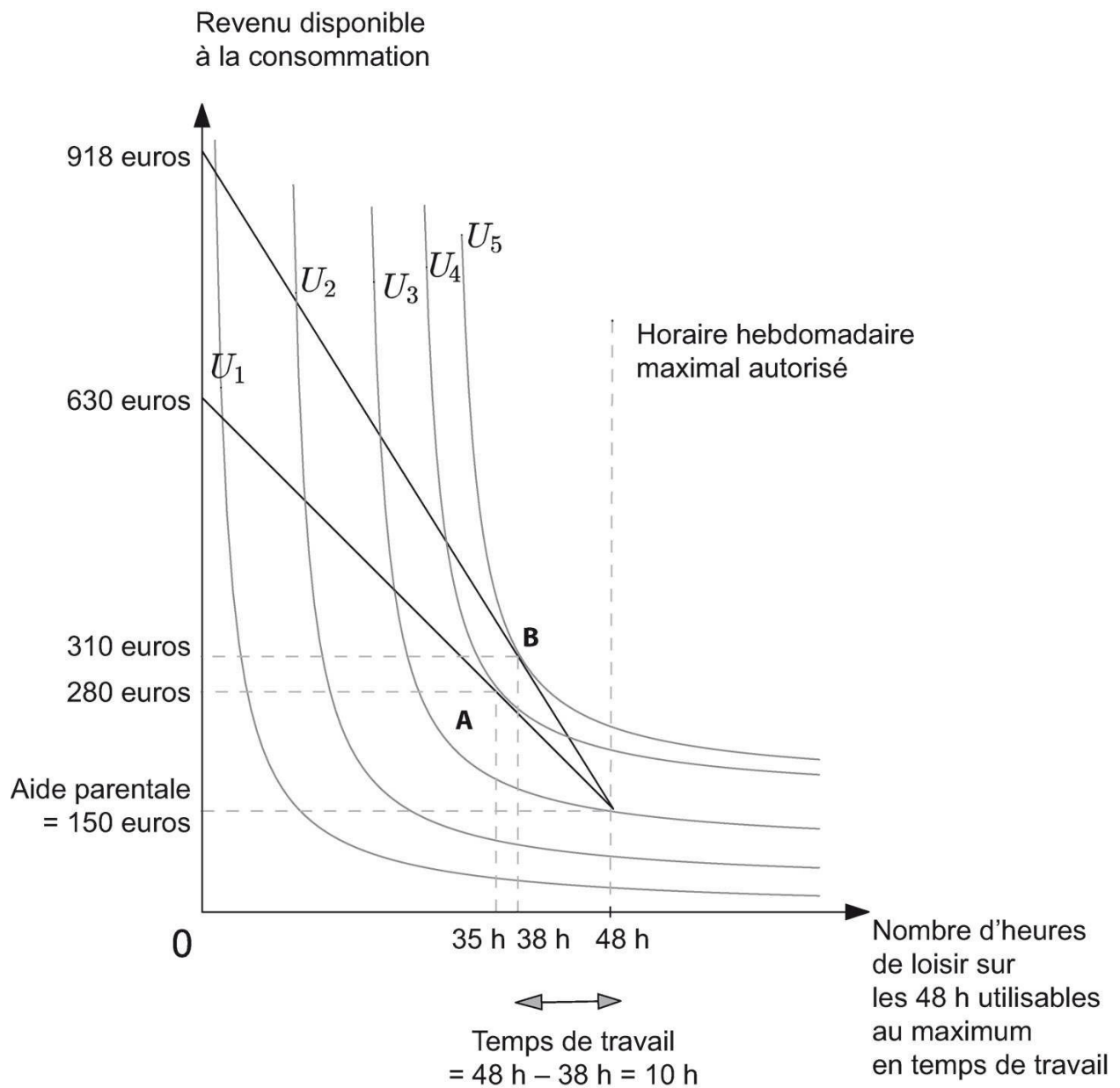
En sommant le nombre d'heures que chaque personne désire travailler pour un taux de salaire donné, on construit la **courbe d'offre de travail agrégée** qui permet de modéliser l'offre totale de travail sur le marché en fonction du taux de salaire. Mais de fait, il n'y a pas vraiment de sens d'additionner l'offre de travail d'un plombier à celle d'un médecin et il est préférable de distinguer une pluralité de marchés du travail en fonction du type de qualification requise pour occuper les différents types d'emplois. Un jeune qui sort d'une école d'ingénieur n'offre pas le même type de travail qu'un autre qui sort d'une école d'esthétique. On peut tenir compte de ces différences de qualification en distinguant simplement le travail qualifié du travail non qualifié. Ce type d'approche très générale est utile pour comprendre certains phénomènes touchant l'ensemble des personnes non qualifiées, quel que soit ensuite leur domaine d'activité.

L'offre de travail globale dépend des évolutions démographiques : si le nombre de personnes ayant entre 15 et 24 ans augmente, il y a fort à parier que le nombre de jeunes désirant travailler va augmenter par un simple effet mécanique. Cela va conduire à un déplacement vers la droite de la courbe d'offre de travail agrégée. L'évolution des normes sociales et des politiques publiques a aussi un impact sur l'offre de travail. En ouvrant l'accès aux études supérieures à un public plus large, les politiques éducatives de massification scolaire ont conduit à réduire l'offre de travail des jeunes. En effet, la plupart se sont consacrés entièrement à leurs études ou n'ont accepté que des emplois à temps partiel pour avoir le temps d'étudier en parallèle.

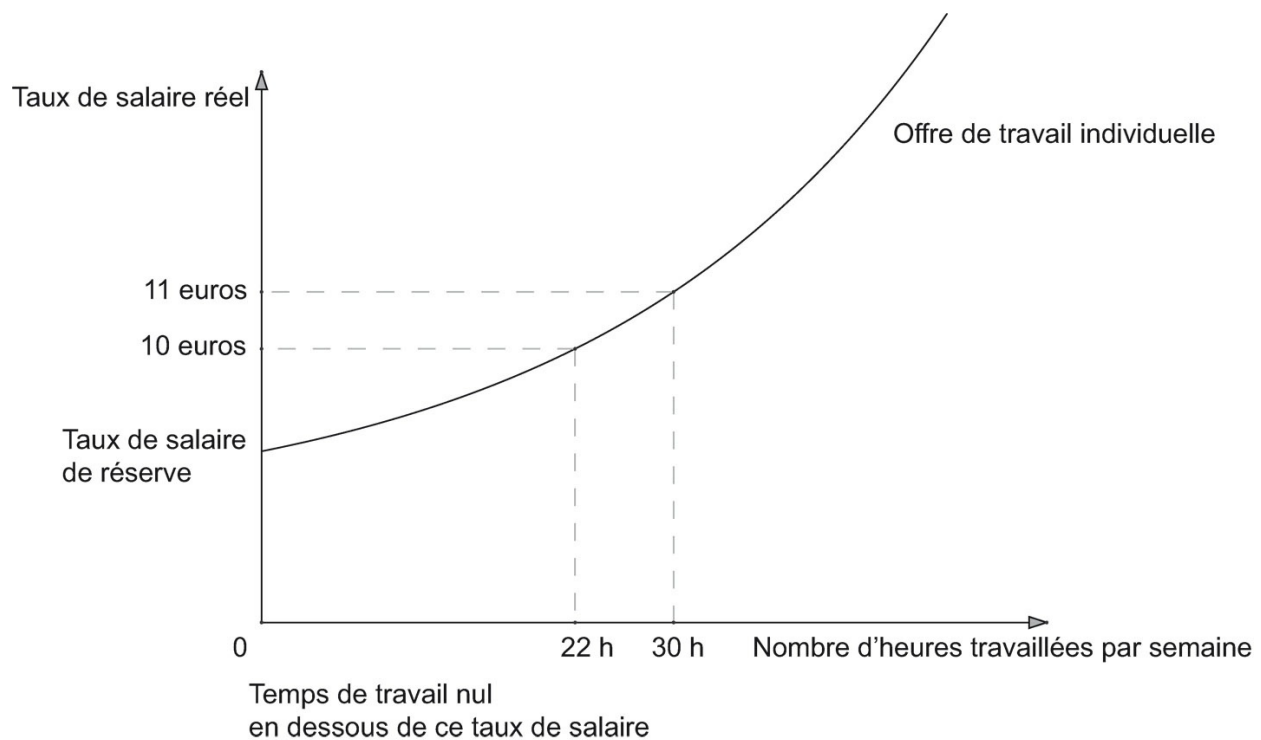
En observant le comportement des individus, les économètres peuvent estimer empiriquement la façon dont les individus modifient leur comportement en fonction des variations de leur taux de salaire. On distingue **l'élasticité de la participation au marché du travail** au taux de salaire qui mesure l'impact d'une variation relative du taux de salaire sur la décision de

participation et renvoie à la décision le long de la marge extensive, de **l'élasticité des heures travaillées en fonction du taux de salaire** qui renvoie à la décision le long de la marge intensive.

Graphique 6.2 : Les effets d'une augmentation du taux de salaire



Graphique 6.3 : Un exemple de courbe d'offre de travail individuelle



Il est ainsi possible de calculer la façon dont varie relativement le taux d'activité des 15-24 ans suite à une variation du taux de salaire (par exemple une augmentation de 20 %). Évidemment, tous les jeunes ayant entre 15 et 24 ans ne peuvent pas prétendre aux mêmes taux de salaire et n'ont pas la même façon de réagir aux variations de taux de salaire : une augmentation de 20 % de ce dernier n'a pas le même impact sur l'offre de travail des jeunes ayant comme seul diplôme le BEPC et sur l'offre de travail de ceux qui sont titulaires d'une licence. Les économètres en tiennent compte et proposent une estimation d'un effet moyen d'une variation relative des salaires sur la décision de participer au marché du travail.

Très peu de travaux empiriques ont été dédiés à l'estimation des comportements d'offre de travail propres aux jeunes. Ils semblent indiquer que l'offre de travail de ces derniers est peu sensible aux variations du taux de salaire réel⁴.

Les valeurs de l'élasticité de participation et des heures travaillées sont des paramètres cruciaux pour comprendre les effets de la taxation des revenus du travail et des prestations sociales sur l'offre de travail. En effet, une élasticité des heures travaillées positive et élevée signifie que l'offre de travail est très sensible à la rémunération et qu'une baisse de cette dernière va conduire à

une diminution de l'offre de travail. Dans ce contexte, une taxation importante des revenus diminue la somme dont disposent les travailleurs après chaque heure travaillée et les incite à moins travailler. En revanche, si l'élasticité des heures travaillées est négative, la taxation des revenus incite les individus à travailler plus. Nous détaillons dans le chapitre 7 les effets de la redistribution des revenus sur l'offre de travail et les débats autour de la détermination d'un taux optimal de taxation des revenus (ainsi que des transferts sociaux).

1.2 La théorie de la prospection

Nous avons jusque-là fait l'hypothèse qu'un individu désirant travailler au taux de salaire en vigueur trouvait un emploi instantanément. Or la recherche d'un emploi peut être longue et coûteuse et l'individu ne trouve pas toujours d'emploi, quand persiste un certain taux de chômage⁵.

Dans quelle mesure la décision de participation au marché du travail des individus est-elle modifiée par l'existence de coûts de recherche d'emploi ? Lorsque l'information est imparfaite, la recherche d'un emploi est plus complexe. Un étudiant qui désire travailler doit recenser les offres disponibles, obtenir des informations sur les caractéristiques des postes (salaire, durée et conditions de travail, etc.), ce qui demande du temps et même parfois une dépense pour des achats de journaux regroupant des petites annonces, des frais de communication, etc. À quelle condition est-il alors rentable de chercher un travail et de sortir ainsi de l'inactivité ? À quelle condition les chômeurs vont-ils accepter une offre de travail et passer ainsi du statut d'actif inoccupé ou statut d'actif occupé ? La théorie de la prospection d'emploi fondée par G. Stigler permet d'éclairer ce débat.

Pour expliquer la décision de participation au marché du travail, faisons l'hypothèse que les individus comparent leur perspective de satisfaction entre l'inactivité et la recherche d'un emploi. Pour cela, ils tiennent compte des ressources monétaires dont ils disposent en étant inactifs (minima sociaux, aides familiales, etc.), mais aussi des ressources dont ils peuvent bénéficier en cherchant un emploi (allocation chômage), des coûts de recherche d'un emploi, des perspectives d'emploi, etc. À partir de ces informations, ils calculent l'espérance d'utilité associée à chaque situation⁶. Si la satisfaction espérée de la situation d'activité est supérieure à celle de la situation

d'inactivité, les individus vont décider de participer au marché du travail.

Les chômeurs découragés, appelés aussi parfois travailleurs découragés, sont des inactifs qui accepteraient de travailler pour les salaires existants mais ne cherchent plus de travail du fait de coûts de recherche trop élevés et de perspectives d'emploi trop faibles. Une dégradation de la conjoncture associée à une hausse du chômage rend moins probable et plus coûteux le fait de trouver un emploi et contribue ainsi à la baisse de l'utilité espérée associée à la recherche d'un emploi. Davantage d'individus choisissent ainsi l'inactivité quand les perspectives d'emplois se dégradent. À l'inverse, quand la conjoncture s'améliore, certains chômeurs découragés cherchent à nouveau un travail et passent du statut d'inactif à celui d'actif. Les observations empiriques confirment ces conjectures. On parle de l'existence d'une flexion conjoncturelle des taux d'activité : le **taux de flexion** désigne le rapport entre la baisse du nombre de chômeurs et l'augmentation du nombre d'emplois. Quand des emplois se créent, le nombre de chômeurs ne diminue pas d'autant car des chômeurs découragés qui étaient devenus inactifs se mettent à chercher un travail et deviennent ainsi des chômeurs, et donc des actifs. Quand le taux de flexion est faible, cela signifie qu'il existe de nombreux chômeurs découragés susceptibles de retourner vers l'activité en cas d'amélioration de la conjoncture et donc d'augmentation du nombre d'emplois.

Des estimations de la part des travailleurs découragés ont été réalisées dans la plupart des pays de l'OCDE. On voit dans le **document 6.1** que la part des travailleurs découragés reste modérée en France, mais qu'elle augmente en période de crise quand la recherche d'emploi se révèle plus infructueuse.

Document 6.1 : Part des travailleurs découragés en France

En %		2001	2004	2007	2009	2010	2013
Ensemble	Part de la population active	0,07	0,13	0,16	0,18	0,13	0,3
	Part de la population totale	0,05	0,09	0,11	0,13	0,09	0,2
	Part de la population active	0,05	0,32	0,42	0,52	0,34	0,4

15-24 ans							
	Part de la population totale	0,02	0,12	0,16	0,21	0,13	0,2

Source : OCDE.

On évalue ainsi en France en 2013 que le nombre de travailleurs découragés ayant entre 15 et 24 ans représente 0,4 % de la population active.

Pour expliciter les conditions dans lesquelles un chômeur va accepter une offre d'emploi, on fait l'hypothèse que les chômeurs évaluent les gains associés à chaque offre d'emploi qu'ils reçoivent (salaire mais aussi conditions de travail, perspectives de carrières, etc.). Ces informations leur permettent de déduire l'espérance d'utilité intertemporelle associée à un poste, soit la satisfaction anticipée dans le futur s'ils acceptent ce poste, qui est corrélée positivement au salaire.

Les chômeurs comparent cette utilité espérée avec la perspective de satisfaction liée au fait de rester sans emploi. Compte tenu de leurs ressources financières, du coût de la poursuite de recherche d'un emploi et de la probabilité estimée de trouver un autre emploi, un individu va refuser ou accepter une offre de travail. Le **salaire de réservation** correspond au salaire en dessous duquel un chômeur n'accepte pas une offre de travail, à ne pas confondre avec le salaire de réserve qui concerne la décision de participer ou non au marché du travail. Une augmentation des aides auxquelles a droit un individu sans emploi va avoir pour conséquence d'accroître son salaire de réservation et va ainsi contribuer à allonger sa période de recherche d'emploi. Si les perspectives économiques s'améliorent, le salaire de réservation augmente car la probabilité de trouver un autre emploi mieux rémunéré est plus forte que par le passé.

L'augmentation de la durée moyenne de la recherche d'emploi n'est pas forcément négative. Il existe certes un risque que les travailleurs perdent en productivité du fait d'un éloignement durable de l'emploi, et deviennent progressivement inemployables. Mais il faut aussi tenir compte du fait qu'ils trouvent des emplois plus adaptés à leurs attentes car ils sont plus sélectifs dans leur choix. Des allocations trop faibles et trop fortement dégressives dans le temps pousseraient les travailleurs à accepter des offres d'emplois inadaptées compte tenu de leur qualification et de leurs attentes ce qui n'est pas efficace d'un point de vue économique. Les allocations chômage sont

dans cette perspective une forme de subvention à la recherche d'emploi. D. Acemoglu et R. Shimer (2000) mettent ainsi en avant le fait qu'en améliorant la qualité des appariements en offre et demande de travail, les allocations chômage peuvent conduire à un accroissement du produit global de l'économie. Notons que la théorie de la prospection ignore le fait que les individus peuvent aussi chercher du travail en étant déjà en poste (*on-the-job search*) et qu'ils peuvent fournir des efforts plus ou moins intenses dans leur recherche d'emploi, la probabilité de trouver un emploi étant de ce fait plus ou moins grande. De nombreux modèles ont ainsi été développés dans la lignée du modèle de Stigler pour tenir compte de ces éléments. Un problème d'**aléa moral** (voir le [chapitre 4](#)) se pose entre le régulateur public qui désire fournir aux travailleurs une assurance contre le chômage pour compenser la perte de revenus, et les chômeurs qui risquent de diminuer l'intensité de leur recherche d'emploi, en adoptant un comportement opportuniste.

Les travaux d'économie comportementale montrent cependant que les comportements sur le marché du travail sont caractérisés par de nombreux biais de rationalité (voir le [chapitre 2](#)). Par exemple, les agents ne réalisent pas une recherche d'emploi d'une durée suffisante parce qu'ils sont plus sensibles aux pertes qu'aux gains (Villeval, 2016) : ils ont tendance à arrêter trop rapidement leur recherche, pour éviter de prolonger une démarche qu'ils jugent trop coûteuse, même si les gains obtenus à l'issue de cette recherche sont objectivement supérieurs aux pertes, car un euro de perdu dans la recherche d'emploi leur semble plus pénalisant psychologiquement qu'un euro de gagné en salaire supplémentaire. Cette plus grande aversion pour la perte que pour les gains est un biais de rationalité qui s'observe dans un grand nombre de domaines de la vie économique.

2. Les déterminants du salaire

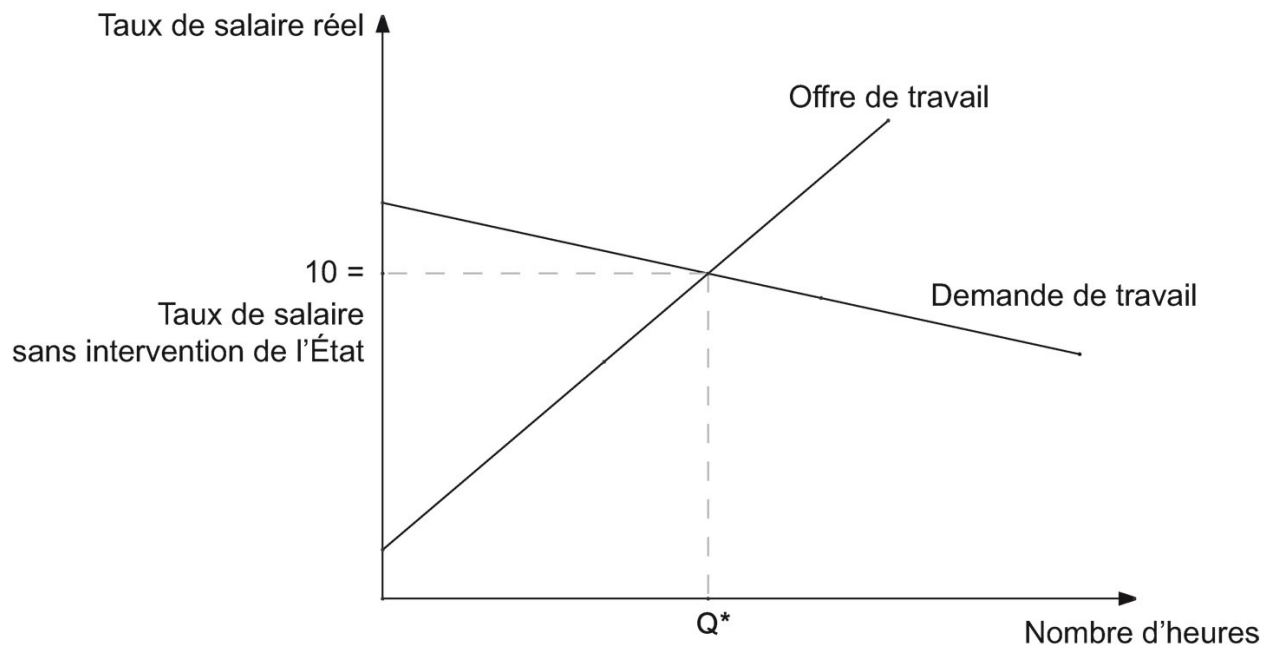
Comment comprendre que certains jeunes puissent prétendre à des taux de salaire élevés pour leur premier emploi alors que d'autres ne peuvent espérer mieux qu'un emploi payé au salaire minimum ? L'analyse des déterminants des salaires a fait l'objet de nombreux travaux en microéconomie et nous allons présenter ci-dessous les principaux modèles explicatifs qui ont été proposés jusqu'à présent.

2.1 Rémunération et qualité de la main-d'œuvre

- **La rémunération à la productivité marginale**

La **demande de travail**, désigne la quantité de travail demandée par les entreprises pour un taux de salaire donné. Elle dépend du coût relatif du travail par rapport aux autres facteurs, ainsi que de la productivité marginale relative du travail. Le choix de la combinaison productive dépend en effet du coût relatif des facteurs et du rapport des productivités marginales de ces derniers (voir le [chapitre 3](#)). Le **coût du travail** englobe l'ensemble des coûts supportés par une entreprise pour disposer de main-d'œuvre : salaires, primes, cotisations sociales, etc. En faisant l'hypothèse d'une absence d'intervention publique et en simplifiant la présentation du coût du travail, il peut être approché par le taux de salaire réel. Compte tenu de sa valeur, un producteur va augmenter la quantité de travail utilisée tant que la **productivité marginale du travail** est supérieure au taux de salaire réel. La productivité marginale étant par hypothèse décroissante (loi des rendements marginaux décroissants), la demande de travail est une fonction décroissante du taux de salaire réel. Si nous considérons le cas classique d'une courbe d'offre de travail croissante en fonction du taux de salaire réel et si le marché du travail se trouve en concurrence pure et parfaite, le taux de salaire réel du marché se fixe alors à l'intersection de la courbe d'offre et de demande.

Graphique 6.4 : La détermination du taux de salaire d'équilibre

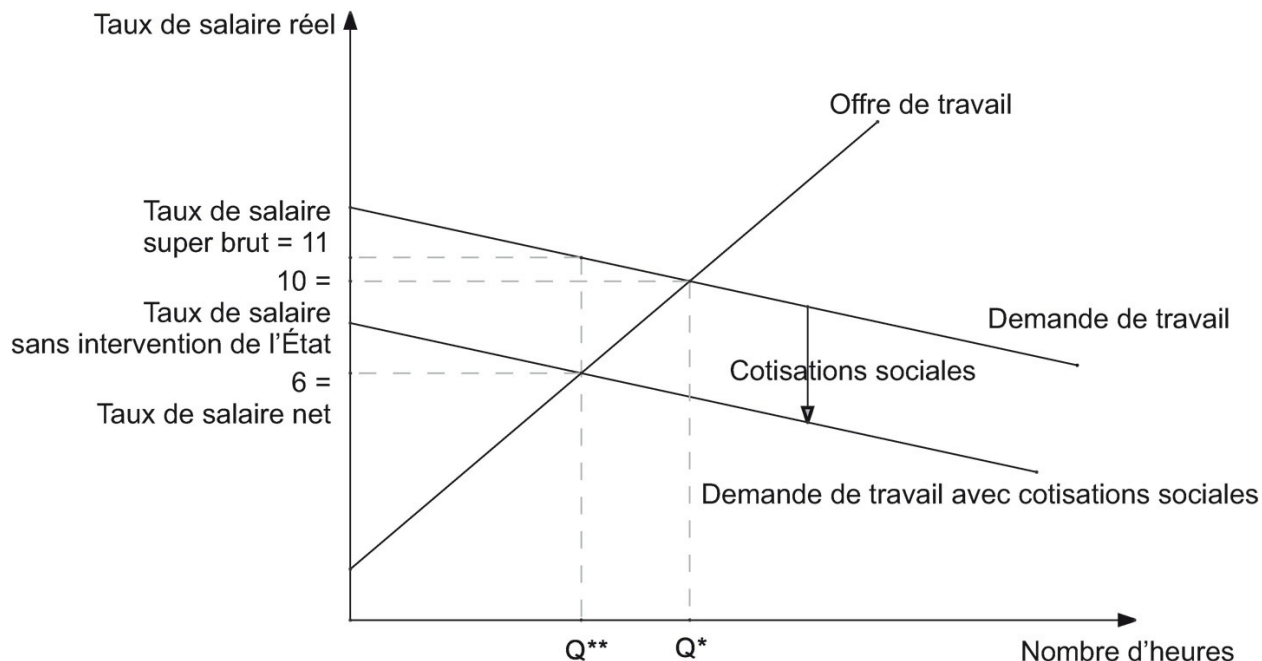


Lorsque le marché du travail est parfaitement concurrentiel, le taux de salaire réel se fixe à un niveau tel que la quantité de travail offerte est égale à la quantité de travail demandée. Le travail est rémunéré à sa productivité marginale. Dans l'exemple proposé ci-dessus, un taux de salaire de 10 euros de l'heure permet d'équilibrer l'offre et la demande de travail.

Des cotisations sociales sont généralement prélevées sur les salaires afin de financer la protection sociale. On appelle **coût socio-fiscal** la différence entre le coût du travail pour l'employeur qui correspond pour simplifier au **salaire super brut** (salaire brut augmenté des cotisations patronales) et la somme reçue par le salarié après prélèvement des cotisations salariales, le salaire net. Il renvoie à la somme des cotisations sociales salariales et patronales. L'existence de prélèvements sur les salaires va modifier le taux de salaire d'équilibre. La façon dont sont réparties formellement, sur le papier, les cotisations entre salariés et employeurs est sans importance car la répartition qui aura lieu *in fine* ne résulte pas de ce partage formel mais de l'élasticité de l'offre et de la demande de travail au taux de salaire. Ce résultat est à nuancer s'il existe un salaire minimum qui porte sur le salaire brut, car pour les salaires situés à ce niveau, cela fige la possibilité d'ajustement à la baisse du salaire net en cas de hausse des cotisations patronales. Considérons que l'État fixe les cotisations sociales patronales à 5 euros par heure travaillée et à 0 euro pour les cotisations sociales salariales. Sur le papier, les employeurs

supportent l'intégralité des cotisations sociales. En réalité, il n'en est rien.

Graphique 6.5 : l'effet des cotisations sociales sur la détermination du taux de salaire d'équilibre



Dans le cas présenté sur le graphique ci-dessus, l'offre de travail est moyennement élastique alors que la demande de travail est plus sensible au taux de salaire. Les cotisations sociales diminuent la demande de travail des entreprises ce qui conduit à une baisse du taux de salaire perçu par les salariés, de 10 euros de l'heure à 6 euros de l'heure. Le coût du travail n'augmente lui que de 1 euro, malgré un taux de salaire super brut élevé, car égal à 11 euros, à comparer au taux d'équilibre initial à 10 euros, sans intervention publique. En règle générale, les salariés supportent l'essentiel du poids des cotisations sociales patronales car les entreprises en reportent la charge sur les salaires nets.

Seule importe la valeur totale des cotisations sociales et non la façon dont elles sont formellement réparties entre salariés et employeurs (entre cotisations salariales et cotisations patronales donc). L'analyse de l'incidence fiscale (voir le [chapitre 1](#)) des cotisations sociales est ainsi essentielle pour sortir des faux débats sur la répartition de ces dernières entre les travailleurs et le patronat et nous invite à nous concentrer sur leur montant global et leur impact sur le niveau du taux de salaire, qui dépend de l'élasticité de l'offre et

de la demande de travail.

• Le rôle du capital humain

Les jeunes de 16 à 24 ans ne doivent pas uniquement choisir à quelles conditions ils participent au marché du travail, ils décident également de poursuivre ou non leur formation initiale, de façon à améliorer les compétences négociables sur le marché du travail.

Le travail n'est pas un bien homogène et les emplois sont plus ou moins qualifiés. En général, plus un travailleur dispose de qualifications importantes, mieux il sera rémunéré car ses compétences ont plus de valeur pour l'entreprise.

La théorie du capital humain développée par G. Becker explique d'un point de vue microéconomique les différences de productivité entre les actifs. Le **capital humain** désigne l'ensemble des compétences qu'un individu peut valoriser sur le marché du travail. Une partie du capital humain renvoie à des compétences générales qui peuvent être valorisées dans différents types d'emploi (par exemple maîtriser les logiciels de base de traitement de texte) d'autres sont plus spécifiques et sont plus difficilement monnayables sur le marché du travail en dehors d'emplois très particuliers (par exemple la maîtrise d'un logiciel qui est utilisé dans très peu d'entreprises).

Le capital humain dont dispose un individu dépend à la fois de l'expérience cumulée dans l'emploi et de sa formation initiale. La poursuite d'étude représente ainsi un investissement en capital humain. L'accumulation de capital humain dépend également des formations suivies par les individus pendant leur vie active.

Empiriquement, on constate l'existence d'une forte corrélation entre le niveau de diplôme et le salaire.

Les données statistiques montrent ainsi qu'en début de carrière, les jeunes actifs sortis en 2013 de leur formation initiale connaissent en 2016, trois ans après leur entrée sur le marché du travail, des disparités de revenus déjà fortes : le salaire mensuel net moyen des garçons non-diplômés est de 1270 €, ceux qui ont atteint le niveau bac touchent un salaire de 1400 € et ceux qui ont atteint un niveau bac+5 accèdent à un salaire net moyen de 2 200 €⁷.

À quelle condition un étudiant va-t-il accepter de poursuivre ses études au

lieu de chercher du travail immédiatement ? Pour G. Becker, l'arbitrage entre formation et emploi dépend du rendement espéré de l'investissement en capital humain et du coût d'investissement dans la formation. Un étudiant titulaire d'une licence qui envisage de poursuivre en master compare le coût d'opportunité de la poursuite en master, qui dépend des frais d'inscription et du salaire auquel il renonce en ne travaillant pas immédiatement, au supplément de salaire qu'il pourra obtenir en continuant ses études. Toutes choses étant égales par ailleurs, plus le rendement de l'éducation est important, plus les individus ont intérêt à se former. À l'inverse, plus le coût des études est élevé, à rentabilité du diplôme donné, moins les individus ont intérêt à se former. Le niveau d'investissement optimal en éducation est celui pour lequel le coût marginal d'une année supplémentaire est égal à son rendement marginal.

Il est rationnel que l'investissement dans la formation se réalise principalement en début de carrière afin de profiter pendant la période la plus longue possible des augmentations de salaire qui découlent de l'accroissement du capital humain : la rentabilité de l'investissement est plus forte.

Les individus font donc le choix de se former en début de vie, et la théorie du capital humain permet ainsi d'expliquer la forte proportion de jeunes qui font le choix de l'inactivité : ce n'est pas parce qu'ils ont une préférence forte pour le loisir mais parce qu'ils investissent dans la formation de capital humain. L'investissement en capital humain correspond ainsi à une façon particulière d'occuper son temps de loisir, qui renvoie au temps non travaillé. L'individu doit ainsi arbitrer dans la gestion de son temps entre trois possibilités : travailler, étudier et l'oisiveté (loisir au sens courant du terme).

Il est rationnel que les individus financent l'accumulation de **capital humain général** puisqu'ils pourront faire valoir ce capital dans d'autres entreprises. Une entreprise qui financerait une telle formation s'exposerait au risque que son salarié la quitte, et de ce fait son investissement bénéficierait à d'autres. Il est en revanche rationnel que les entreprises financent l'investissement en **capital humain spécifique**. En effet un individu qui investirait dans une telle formation ne serait pas en mesure de négocier une augmentation salariale auprès de son entreprise car cette dernière sait que cette compétence n'a pas de valeur ailleurs. Cet investissement est donc trop risqué pour un individu.

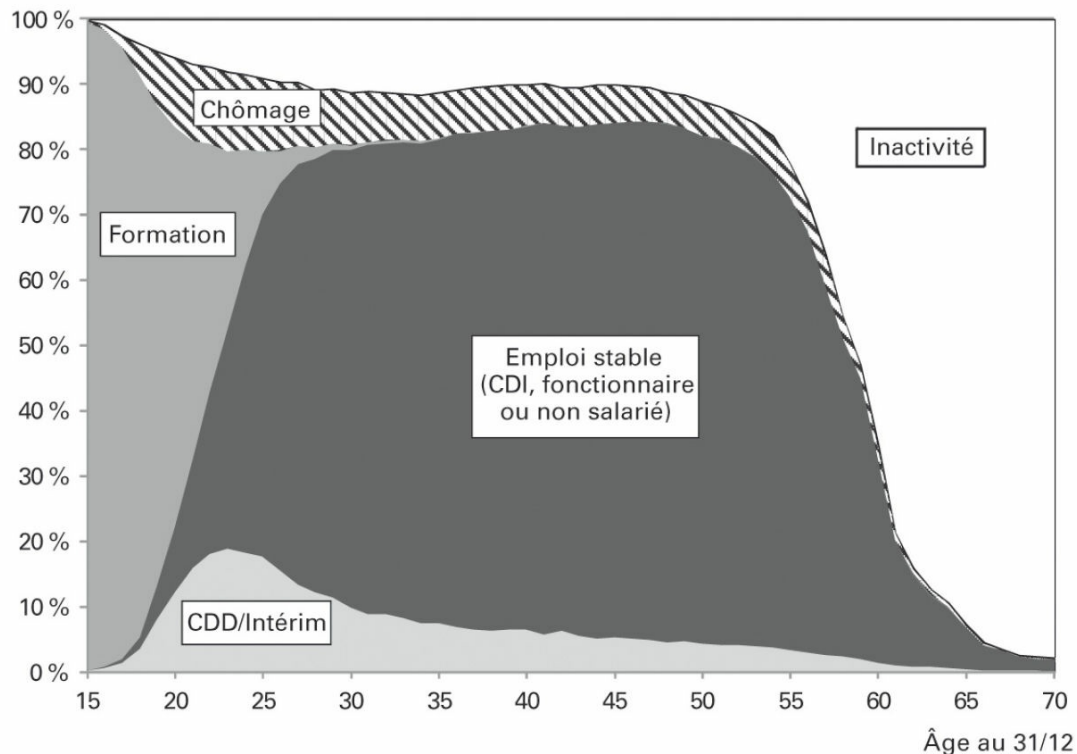
Les gains et les coûts de la formation n'étant pas les mêmes pour tous, on comprend mieux que les individus ne font pas les mêmes efforts pour améliorer leur capital humain. Une personne avec des difficultés scolaires a moins de chances de valider ses examens en fin d'année et peut considérer que le risque de poursuivre une année d'étude supplémentaire est trop élevé. Un diplôme n'a pas toujours la même rentabilité selon le réseau social de l'individu. La théorie du capital humain permet donc d'expliquer en partie le moindre investissement dans l'éducation des enfants d'origine populaire : pour ces individus, le coût de l'investissement dans l'éducation est plus important et le rendement des diplômes plus faible. D'autres explications ont été avancées par des économistes (imperfection du marché du crédit, voir ci-dessous) ou par des sociologues de l'éducation, qui insistent sur l'effet des inégalités de ressources fournies par la socialisation familiale (DURU-BELLAT et VAN ZANTEN, 2009), mais l'approche microéconomique permet de saisir l'impact propre à l'accumulation de capital humain.

L'accumulation de capital humain peut aussi provenir d'un effet d'expérience. En effet, les individus apprennent aussi par la pratique (*learning by doing*) et accroissent ainsi leurs compétences et leur productivité en travaillant. Les stages sont un moyen pour les étudiants d'acquérir de l'expérience et d'accroître leur capital humain. Même s'ils sont en général très mal rémunérés, ils constituent un bon investissement afin de pouvoir postuler dans un emploi plus valorisé ensuite.

Les études empiriques disponibles montrent que le niveau de diplôme des individus et l'expérience accumulée mesurée à travers le nombre d'années travaillées sont deux des principaux facteurs explicatifs du niveau de salaire d'un individu. Une explication alternative à celle de la théorie du capital humain postule que le diplôme est en fait un « signal » pour l'employeur d'une capacité à s'investir et s'imposer dans son travail, plus que d'une compétence particulière. Selon M. Spence (1973), en situation d'asymétrie d'information entre employeur et demandeur d'emploi, le fait de détenir un diplôme reconnu permet de signaler à l'employeur son sérieux et sa capacité de travail. Le diplôme fournit ainsi à l'employeur des informations sur les candidats. Les employeurs préfèrent embaucher les personnes diplômées, même si la formation reçue n'est pas vraiment en adéquation avec le poste proposé et sont prêts à payer plus pour disposer d'un salarié considéré comme plus fiable. Quand l'information est imparfaite, le diplôme constitue donc un signal qui permet à l'employeur de diminuer le risque de **sélection**

adverse, c'est-à-dire de voir les meilleurs candidats retirer leur candidature, faute de pouvoir valoriser leur compétences (voir le [chapitre 4](#)).

Document 6.2 : Situation au regard de l'activité par âge



Source : Enquêtes Emploi en Continu (EEC), moyenne 2003-2010, INSEE, calculs DG-Trésor in BATARD P.-E. et SAILLARD E., 2011.

Tous les individus ne peuvent pas investir de façon optimale dans l'éducation du fait notamment de l'imperfection du marché du crédit. En effet, les étudiants qui ne peuvent bénéficier de l'aide de leur famille doivent emprunter pour financer leurs études. Or il existe des asymétries d'information entre les banques et les emprunteurs, et la demande de crédit est donc rationnée (voir le [chapitre 4](#)) : certains individus ne parviennent pas à obtenir le prêt nécessaire au financement de leurs études, et doivent entrer de façon précoce sur le marché du travail. L'État aide donc au financement des formations, mais son intervention est liée également au fait que l'investissement en capital humain génère des **externalités positives** : les avantages collectifs, et donc le rendement social de l'investissement éducatif, sont supérieurs à son rendement privé (voir le [chapitre 5](#)). Les individus sous-investissent ainsi de façon structurelle dans leur formation car ils tiennent

compte seulement des rendements privés de ces investissements. L'État subventionne donc l'éducation pour accroître ses avantages collectifs au-delà de l'effort que feraient spontanément les individus.

• Les théories du salaire d'efficience

Nous avons jusqu'ici considéré que la productivité des individus était uniquement modifiée par les investissements en capital humain, et qu'ils bénéficiaient de salaires plus ou moins élevés en fonction de leur productivité. Selon les théories du **salaire d'efficience**, le taux de salaire peut modifier le niveau de productivité d'un individu, il existe donc un taux de salaire optimal qui maximise les efforts des agents, et donc leur productivité, tout en limitant le coût pour la firme. Prenons le cas d'un étudiant à qui un centre d'appel propose un travail plus rémunérateur que chez la concurrence, à condition de faire un effort d'implication dans son travail. Dans ces conditions, son salaire est supposé agir sur sa productivité, hypothèse au cœur des théories du salaire d'efficience. Elles proposent en fait différentes pistes pour étudier les facteurs explicatifs de la relation entre productivité et rémunération des individus.

A. Weiss (1980) part du principe qu'il existe une asymétrie d'information entre l'employeur et les individus qui postulent pour un emploi. Il est en effet très difficile d'évaluer de façon précise la qualité d'un postulant. En proposant des salaires faibles, l'employeur a toutes les chances de voir la qualité moyenne des candidats se dégrader car les individus qui se savent très productifs vont refuser de postuler à ce type d'emploi. Proposer des salaires élevés est ainsi une façon d'attirer les meilleurs candidats et de diminuer le risque d'embaucher un salarié peu productif. Des salaires élevés permettent ainsi de limiter le risque de **sélection adverse** (voir le [chapitre 4](#)).

C. Shapiro et J. Stiglitz (1984) proposent une autre explication du lien entre salaire et productivité. Ils partent eux aussi du principe qu'il existe une asymétrie d'information entre l'employeur et le salarié mais s'intéressent aux effets de cette asymétrie une fois que le contrat de travail a été signé. Si l'employeur ne peut qu'imparfaitement surveiller ses salariés, ces derniers risquent de « tirer au flanc » : risque d'aléa moral lié à un opportunisme *ex post* (voir le [chapitre 4](#)). L'employeur a intérêt à payer ses salariés plus que ses concurrents pour que ces derniers s'investissent pleinement dans leur

travail du fait qu'ils craignent d'être licenciés s'ils sont pris en train d'adopter un comportement opportuniste et de perdre ainsi ce surcroît de salaire. Ces considérations rejoignent ainsi les arguments mobilisés par la théorie des incitations présentée aux chapitres 3 et 4. Le salaire est un instrument d'incitation pour les salariés et en fonction du mode de rémunération choisi par l'employeur l'incitation à l'effort est plus ou moins efficace. Selon la théorie principal-agent, une prime d'incitation à l'effort en fonction d'un indicateur de résultat incite les travailleurs à maximiser leur effort (Laffont, 1991).

S. Salop (1979) met lui en avant le fait qu'une politique de bas salaire favorise le *turn over* de la main-d'œuvre et ainsi une faible productivité des travailleurs qui n'ont pas pu accumuler de l'expérience. Pour éviter cela les employeurs (ou au moins certains d'entre eux) proposent donc des salaires élevés. Un *fast-food* qui emploie des étudiants a intérêt à payer davantage ses salariés que ses concurrents afin de bénéficier d'une main-d'œuvre stable qui maîtrisera parfaitement les différentes tâches à effectuer, et pourra ainsi servir la clientèle de façon plus efficace.

G. Akerlof et J. Yellen (1988) proposent enfin un autre type d'explication basée sur le sentiment de justice. Des travailleurs mieux payés se sentent plus valorisés dans l'entreprise ce qui favorise leur implication et leur productivité. Un jeune constatant que son salaire augmente peut y voir une reconnaissance de sa qualité par son employeur, et en retour il peut lui témoigner sa reconnaissance en s'impliquant fortement dans l'entreprise.

2.2 Rémunération, type d'emploi occupé et pouvoir de négociation

Les considérations liées à la productivité des travailleurs ne sont pas les seules pertinentes pour expliquer les inégalités de salaire. L'économie du travail est un champ de recherche très dynamique, et de multiples travaux, théoriques et empiriques, portent sur la diversité des déterminants du salaire.

• La prise en compte de la pénibilité selon la théorie des différences compensatrices

La pénibilité de l'emploi permet de comprendre certaines différences de salaire. Si les employeurs proposaient le même salaire pour des emplois à

pénibilité différente, tous les salariés voudraient occuper les emplois peu pénibles et personne ne postulerait aux autres emplois. Ce mouvement de la demande devrait avoir pour effet de faire croître les salaires des emplois pénibles afin d'attirer la main-d'œuvre.

La **théorie des différences compensatrices** (ROSEN, 1986) repose sur l'hypothèse d'un fonctionnement concurrentiel du marché du travail. Les différences de salaire découleraient de différences de compétences et/ou de pénibilité des tâches. O. Godechot et M. Gurgand (2000) montrent ainsi, en exploitant l'enquête *Travail et mode de vie*, que les inhalations dangereuses et la tension nerveuse font l'objet d'une compensation financière positive significative pour les hommes. Selon eux, ce n'est cependant pas toujours le cas et la compensation est même une exception plutôt que la règle. Ainsi, les femmes ne bénéficient d'aucune compensation pour l'exposition à des nuisances et les fortes amplitudes thermiques entraînent un salaire significativement plus faible que pour les hommes et non plus élevé (– 6 %). Les travaux sociologiques de Ch. Baudelot et M. Gollac (1993) montrent que la pénibilité engendre un surcroît de salaire lorsqu'elle est explicitement reconnue par la direction de l'entreprise. Pour comprendre pourquoi des bas salaires et des mauvaises conditions de travail sont souvent associés, la théorie de la segmentation du marché du travail offre une grille de lecture féconde.

• La théorie de la segmentation du marché du travail

Jusqu'à présent, nous avons considéré que pour un niveau donné de productivité, il existait un marché du travail unifié. M. Piore et P. Doeringer (1971) font l'hypothèse que le marché du travail est dualiste, et a tendance à se polariser autour de deux grands segments : le **marché primaire** où les conditions de travail sont bonnes, la stabilité de l'emploi, des hauts salaires et des perspectives de promotion ; le **marché secondaire** où les salaires sont faibles, la précarité de l'emploi forte et les conditions de travail dégradées.

Lorsque les étudiants entrent sur le marché du travail, ils évoluent bien souvent sur le marché secondaire en cumulant des emplois précaires (stages, CDD, etc.) et des salaires faibles. Moins d'un jeune salarié sur deux a accès à un contrat à durée indéterminée ou CDI (44 % des 15-24 ans) alors que c'est le cas de 87,5 % des 25-49 ans⁸. Signalons que la théorie de la segmentation se

situe clairement dans une perspective hétérodoxe et critique vis-à-vis de la théorie néo-classique, en insistant notamment sur les fondements institutionnels du marché du travail. P.-E. Batard et E. Saillard, (2011) expliquent ainsi que les difficultés d'insertion des jeunes ne sont pas liées à un facteur d'âge mais sont simplement le reflet du dualisme du marché du travail. Quel que soit l'âge, l'accès à l'emploi stable est difficile pour les entrants.

• Les modèles de négociation

Le niveau de salaire d'un individu dépend aussi de sa place dans la hiérarchie et de sa capacité de négociation. Il ne découle pas d'une simple rencontre entre l'offre et la demande de travail mais est aussi le résultat d'un rapport de force. Ainsi deux individus à qualification égale peuvent disposer d'un salaire différent, s'ils ont des capacités de négociation différentes. Selon les estimations économétriques du sociologue O. Godechot (2011), le capital humain et les primes d'incitation à l'effort mises en œuvre afin d'accroître la productivité du travail ne rendent pas pleinement compte des inégalités : les primes versées à certains salariés sont largement supérieures au montant nécessaire pour les stimuler. Le gain de productivité ne compense donc pas le surcoût que génèrent de telles primes pour l'entreprise. Pour expliquer certains salaires particulièrement élevés, il met en avant l'existence de phénomènes de rente dont bénéficient certains métiers. Selon lui, « les métiers diffèrent par le volume d'actifs que les salariés peuvent s'approprier (les savoirs, les savoir-faire, les logiciels, les clients, les équipes) et par leur transférabilité d'une entreprise à une autre ». Dans le secteur de la finance, les salariés des métiers *front office* (*traders* et vendeurs) cumulent à la fois une forte appropriation de l'activité et une forte transférabilité ce qui leur permet de négocier des salaires bien plus élevés car leur départ serait préjudiciable à l'entreprise. Toutefois, les négociations salariales sont rarement faites au niveau individuel. Dans la plupart des situations, elles ont lieu au niveau collectif et sont menées par les syndicats. En France, la plupart des salaires dépendent ainsi de conventions collectives.

Différents modèles ont été développés pour représenter les négociations salariales. Ils partent de l'hypothèse qu'il existe un surplus à partager entre employeurs et salariés : le profit est positif en lien notamment avec l'existence d'imperfections de marché. L'issue de la négociation salariale va

ainsi dépendre du pouvoir de négociation des différentes parties prenantes. La théorie des jeux a permis d'éclairer de façon pertinente la façon dont le marchandage entre syndicat et employeur avait lieu (voir le [chapitre 4](#)). Deux cas ont été mis en avant : celui où les syndicats et les employeurs décident conjointement du niveau des salaires et de l'emploi en fonction de leur pouvoir de négociation (modèle du contrat efficace, MACDONALD et SOLOW, 1981) ; celui où syndicats et employeurs fixent le niveau des salaires en fonction de leur pouvoir de négociation et l'employeur décide ensuite seul de la quantité de travail qu'il demande (modèle dit de droit à gérer, NICKELL et ANDREWS, 1983).

• La discrimination

Les écarts de salaire peuvent aussi s'expliquer par l'existence de discriminations sur le marché du travail. Les estimations empiriques montrent que toutes choses étant égales par ailleurs (niveau de diplôme, âge, catégorie socioprofessionnelle, type de contrat de travail, temps de travail, secteur d'activité et taille de l'entreprise, etc.) les femmes gagnaient en moyenne, en 2014, 8 % de moins que les hommes⁹.

Si les pratiques discriminatoires sont souvent évoquées pour mettre en avant les difficultés rencontrées par les femmes ou les minorités ethniques, l'existence d'une discrimination en fonction de l'âge a fait l'objet d'une reconnaissance croissante (voir POLI, LEFRANÇOIS, CARADEC, 2009). Si la discrimination subie par les seniors est celle qui est la plus souvent mise en avant, l'idée que les jeunes sont discriminés sur le marché du travail est de plus en plus débattue. Au-delà de leur manque d'expérience, qui peut justifier des salaires plus faibles, les jeunes seraient, à compétence égale, moins payés que leurs aînés ce qui permet de conclure à une discrimination en fonction de l'âge.

Différents modèles ont été développés pour comprendre les pratiques discriminatoires. Deux grandes catégories peuvent être distinguées : la première comporte, à la suite des travaux de G. Becker (1957), les théories fondées sur des préférences discriminatoires (un « goût » pour la discrimination). Le deuxième type de modèle se situe dans la lignée de l'analyse de E. Phelps (1972) et fait reposer les différences de traitement sur des écarts de performances moyennes dans un contexte d'asymétrie

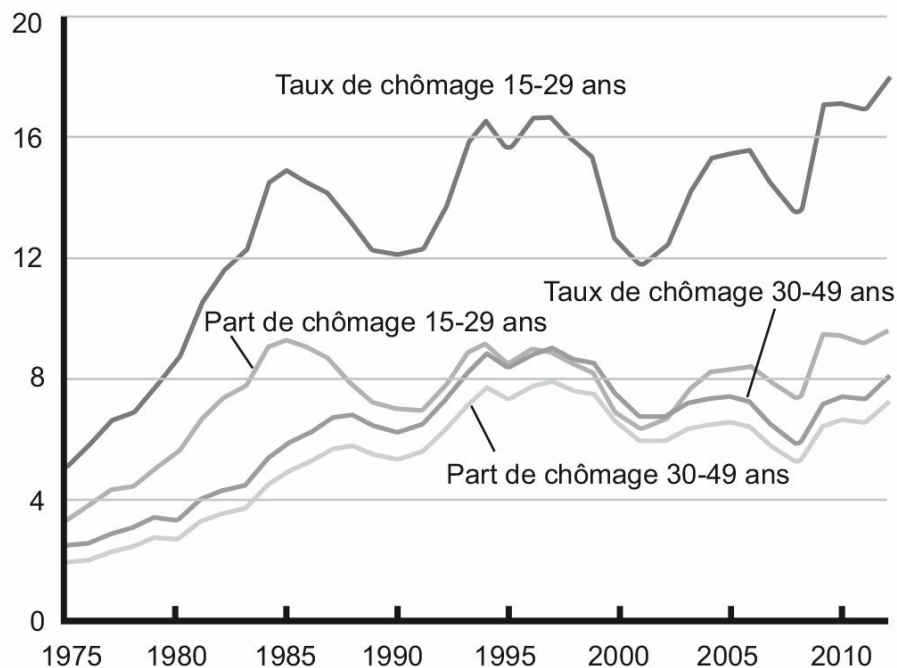
d'information entre employeurs et candidats à l'embauche. Les entreprises ne peuvent évaluer parfaitement la qualité des personnes qu'elles recrutent. Un employeur devant embaucher une personne, ou fixer son taux de salaire, fonde sa décision sur la base de la productivité moyenne de la catégorie à laquelle l'individu appartient (par exemple le groupe des femmes dont l'âge est compris entre 15 et 24 ans). Si les jeunes femmes sont en moyenne moins productives que les jeunes hommes, les employeurs vont moins les embaucher et moins les payer à durée de travail égale. Une jeune femme qui serait très productive aurait ainsi un taux de salaire inférieur à sa productivité du fait qu'elle appartienne à un groupe qui en moyenne est moins productif qu'un autre. On parle de **discrimination statistique** car c'est à travers les jugements statistiques que les employeurs projettent sur les individus certaines caractéristiques de leur groupe.

3. L'explication microéconomique du chômage

Le chômage est sans doute l'une des questions les plus débattues en science économique. Les jeunes sont particulièrement touchés par ce phénomène, comme on le voit sur le **document 6.3**.

Document 6.3 : L'évolution du taux de chômage et de la part du chômage par groupe d'âge

Taux et part de chômage des 15-29 ans et des 30-49 ans, en %



Lecture : le taux de chômage d'un groupe d'âge rapporte les chômeurs au nombre d'actifs de ce groupe d'âge, tandis que la part du chômage rapporte les chômeurs à la population totale du groupe d'âge, inactifs inclus.

Champ : France métropolitaine.

Source : INSEE, « l'insertion des jeunes », in *Formations et emploi*, édition 2013.

L'explication du chômage des jeunes semble cruciale pour pouvoir concevoir des politiques efficaces. Les études macroéconomiques ont permis de mettre en avant différents facteurs jouant sur le niveau de l'emploi : les chocs conjoncturels, la croissance, le dynamisme de la consommation,

l'investissement, l'inflation, etc. L'analyse microéconomique se situe à un niveau différent et cherche à expliquer le chômage en partant des décisions individuelles des agents. Il s'agit de comprendre pourquoi certaines catégories d'individus sont davantage au chômage que d'autres. La distinction entre explication microéconomique et macroéconomique du chômage est cependant en partie artificielle, dans la mesure où les modèles macroéconomiques contemporains ont très souvent des fondements microéconomiques.

Le but de cette partie est de présenter différents modèles microéconomiques qui permettent de saisir les enjeux liés à la question du chômage.

3.1 La question du coût du travail

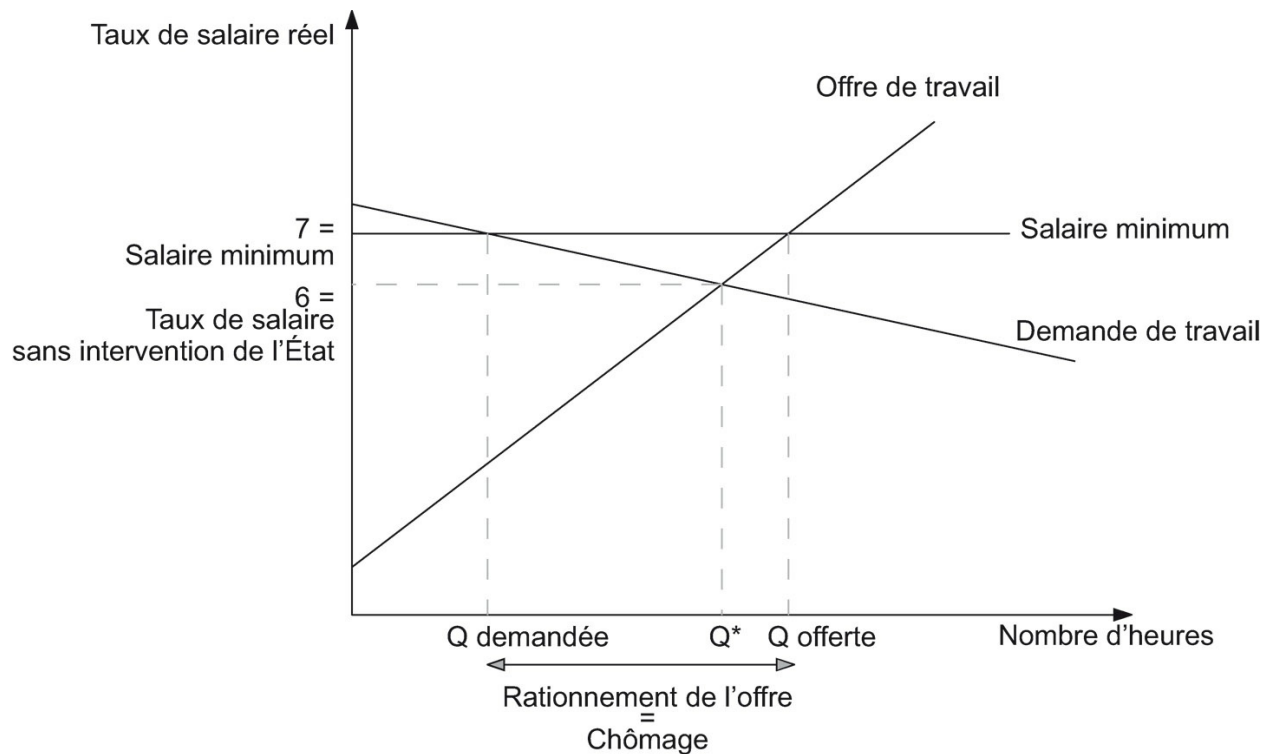
Nous avons vu que la demande de travail dépendait du coût relatif de ce dernier ainsi que de la productivité marginale relative du travail. Un coût du travail trop élevé est souvent présenté comme un facteur de chômage. Le coût du travail englobe l'ensemble des coûts que doit supporter une entreprise pour disposer de main-d'œuvre : salaires, primes, cotisations sociales, etc.

• Un coût élevé du travail : un facteur de chômage ?

Le modèle microéconomique de base part de l'hypothèse que le marché du travail est un marché comme les autres : en situation de concurrence et sans intervention de l'État, il n'y a pas de chômage à l'équilibre. Cette situation n'est pas la meilleure pour tous, car certains inactifs sont exclus du marché du travail, le taux de salaire d'équilibre étant trop faible pour qu'ils offrent leur travail. Seuls les individus qui désirent travailler au taux de salaire du marché le peuvent. L'expression de « chômeurs volontaires » est parfois employée pour désigner les personnes qui décident de ne pas travailler au taux de salaire d'équilibre mais cette expression est erronée. En effet, ces individus ne cherchent pas de travail, ils ne sont donc pas des chômeurs mais bien des inactifs.

Dans ce contexte, la négociation collective d'un taux de salaire minimum supérieur au taux de salaire d'équilibre peut conduire à l'apparition de chômage, comme le suggère le [**graphique 6.6**](#).

Graphique 6.6 : L'effet de l'instauration d'un salaire minimum



Au taux de salaire minimum la quantité de travail demandée est inférieure à la quantité de travail offerte, ce qui signifie qu'un certain nombre d'individus voudraient travailler mais ne le peuvent pas à ce taux. Le salaire minimum a ainsi créé du chômage.

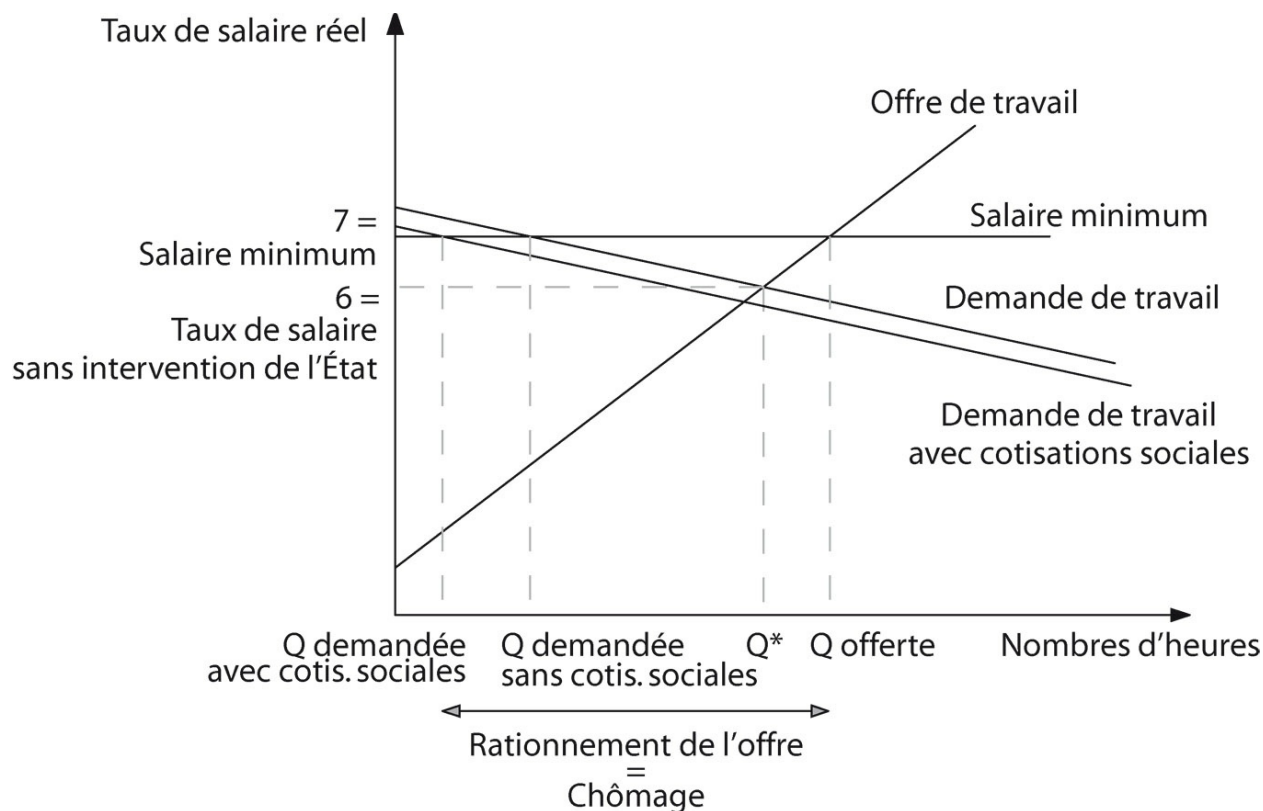
L'existence de salaires trop élevés par rapport au niveau concurrentiel peut aussi s'expliquer par le rôle des syndicats. Nous avons vu que les syndicats participent aux négociations salariales collectives. En fonction de leur pouvoir de négociation et du taux de couverture des négociations collectives, leur impact est plus ou moins fort sur la fixation des salaires.

L'existence de cotisations sociales élevées accroît également le coût du travail. En présence d'un salaire minimum, le taux de salaire net ne peut diminuer en dessous du minimum légal et les cotisations sociales renforcent ainsi le chômage en diminuant la quantité de travail demandée par les employeurs, comme on le voit sur le [graphique 6.7](#).

Dans cette optique, abaisser le niveau du SMIC ou alléger les cotisations sociales peut être un moyen de lutter contre le chômage. Depuis le gouvernement Balladur en 1993, les gouvernements successifs ont concentré leur action sur l'allégement des charges pesant sur les bas salaires car le coût

du travail pénalise d'abord l'emploi de ces catégories. La mise en place du « crédit impôt compétitivité emploi » en 2012 puis du « pacte de responsabilité » par le gouvernement Valls en 2014, s'inscrit dans cette même stratégie d'action sur l'emploi par l'allègement des cotisations sociales, de façon à rendre les entreprises plus compétitives. Il s'agit de diminuer les cotisations patronales progressivement, ce qui représente un allègement de l'ordre de 30 à 35 milliards d'euros, avec un barème dégressif d'allègement jusqu'à 3,5 SMIC. Certains économistes ont critiqué le fait que cette mesure n'était pas assez centrée sur le bas de l'échelle salariale, ce qui réduirait son impact sur l'emploi¹⁰.

Graphique 6.7 : L'effet des cotisations sociales avec salaire minimum



Si le SMIC empêche l'ajustement à la baisse seulement pour les bas salaires, c'est précisément à ce niveau que l'effet négatif des cotisations sociales sur l'emploi joue le plus. L'existence d'un salaire minimum pèse surtout sur l'emploi des travailleurs peu qualifiés dont la productivité marginale peut être inférieure à son montant alors que les travailleurs plus qualifiés ont en général une productivité supérieure au taux de salaire minimum et auraient de toute façon été payés à un niveau supérieur au minimum légal. Enfin, les travailleurs non qualifiés subissent plus fortement que les autres la concurrence des travailleurs à bas salaires exerçant dans les pays en voie de développement et un coût du travail trop élevé peut entraîner des délocalisations d'entreprises vers ces pays et diminuer ainsi la demande de travail, même si le phénomène reste limité d'après les évaluations des emplois effectivement délocalisés.

Les arguments présentés ci-dessus éclairent le taux supérieur de chômage des travailleurs peu qualifiés par rapport aux autres. Il existe ainsi une relation très nette entre taux de chômage des jeunes et niveau de diplôme. D'après l'enquête Génération 2013 du Céreq, qui interroge les jeunes trois ans après leur sortie de formation initiale en 2013, le taux de chômage des

hommes non-diplômés atteint 51 %, contre 27 % pour les titulaires d'un CAP ou BEP et seulement 8 % pour les Bac+5¹¹.

Le taux particulièrement élevé de chômage des jeunes de moins de 24 ans peut s'expliquer par le fait qu'ils sont en général peu qualifiés. En effet, les individus qui optent pour des études longues (master, grandes écoles, etc.) ne sont pas encore sortis du système scolaire à cet âge et seuls sont présents sur le marché du travail ceux qui ont un diplôme moins élevé.

En mobilisant la théorie du salaire d'efficience présentée précédemment, la relation entre chômage et niveau des salaires s'éclaire sous un jour différent. Nous avons vu en effet que les employeurs ont intérêt à augmenter le salaire de leurs salariés afin d'améliorer leur productivité. Les entreprises offrent ainsi des salaires supérieurs au taux de salaire qui équilibre l'offre et la demande. Dans ces conditions, le chômage se développe du fait d'un taux de salaire plus élevé que le taux d'équilibre. Le chômage ne provient pas ici d'une intervention de l'État ou du rôle des syndicats mais du fonctionnement du marché du travail du fait notamment de l'existence d'asymétries d'information. Dans cette optique, il est inutile de baisser le SMIC ou les cotisations sociales pour résoudre la question du chômage, celui-ci résulte du fonctionnement même du marché du travail.

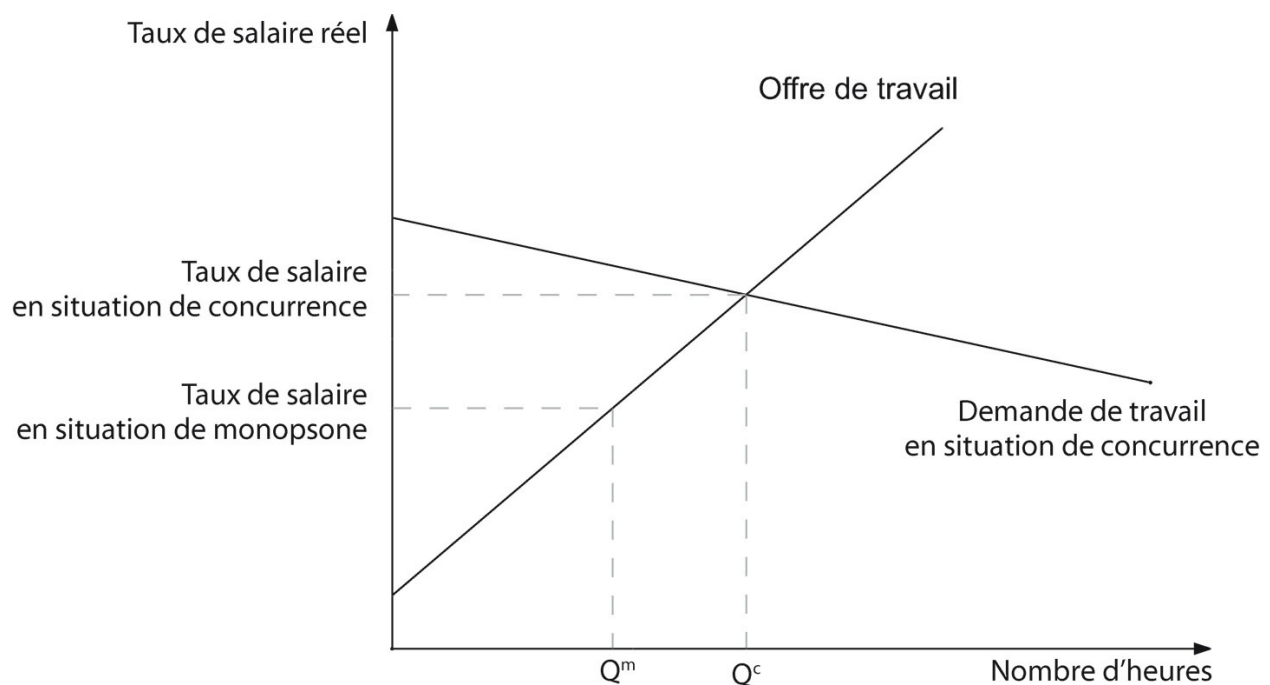
• Coût du travail et emploi : une relation non mécanique

L'effet d'une hausse du SMIC sur l'emploi est l'objet de nombreux débats. Tout d'abord, d'un point de vue macroéconomique, le SMIC peut contribuer au soutien de la consommation des ménages et ainsi au dynamisme économique. Le raisonnement présenté ci-dessus se situe ainsi en équilibre partiel (voir le [chapitre 1](#)) et ne tient pas compte des effets positifs sur la demande globale générés par le SMIC. Notons toutefois, que la demande globale peut être soutenue de façon plus efficace par des politiques de redistribution par exemple. Les distorsions générées par un système de taxes et de transferts sont en règle générale moins importantes que celles générées par la mise en place d'un salaire minimum.

Nous avons fait l'hypothèse jusqu'ici que le marché du travail était concurrentiel. En tenant compte de l'existence d'un pouvoir de marché des entreprises, l'effet du SMIC peut être bénéfique sur l'emploi. Un **monopsonne** désigne une situation où il existe un seul demandeur et de nombreux offreurs.

En l'absence de syndicats et de conventions collectives qui étendent à tous les travailleurs le résultat des négociations, les travailleurs sont en situation de faiblesse pour négocier leur salaire et le modèle du monopsonne est plus adapté pour rendre compte de la fixation des salaires. Le monopsonne (l'employeur) fixe le salaire le plus faible possible lui permettant d'embaucher le nombre de travailleurs nécessaire pour réaliser le volume de production désiré. Au final, la productivité marginale du travail est supérieure au taux de salaire fixé.

Graphique 6.8 : L'effet du SMIC en situation de monopsonne



Si l'État fixe le SMIC entre le taux de salaire en situation de concurrence et le taux de salaire en situation de monopsonne, comme on le voit sur le [graphique 6.8](#), le salaire minimum augmente alors l'emploi car les entreprises ont intérêt à embaucher à ce niveau de salaire, qui reste inférieur à la productivité marginale du travail, égale au taux de salaire de concurrence. En revanche au-delà du taux de salaire concurrentiel, le salaire minimum aura un effet négatif sur l'emploi. D. Card et A. Krueger (1994) ont réalisé une étude expérimentale très connue sur l'effet d'une hausse du salaire minimum de 4,25 \$ de l'heure à 5,05 \$ en avril 1992 dans l'État du New Jersey. Elle a entraîné une augmentation de la demande de travail dans les *fast-foods* du New Jersey, secteur où l'emploi est majoritairement non qualifié et où les

salariés sont payés au salaire minimum. Dans ce secteur, les salariés ne disposent d'aucun pouvoir de négociation et on considère généralement que les fast-foods disposent d'un pouvoir de monopsonne. Face à la hausse du salaire minimum, les employeurs ont préféré augmenter la quantité de travail demandée pour compenser le coût supplémentaire par une hausse d'activité, par exemple en étendant les horaires d'ouverture.

La hausse du SMIC a donc des effets différenciés sur l'emploi en fonction du contexte économique et du degré de concurrence sur le marché du travail, ainsi que d'autres paramètres comme la valeur des cotisations sociales, la productivité du travail, l'indice des prix, etc.

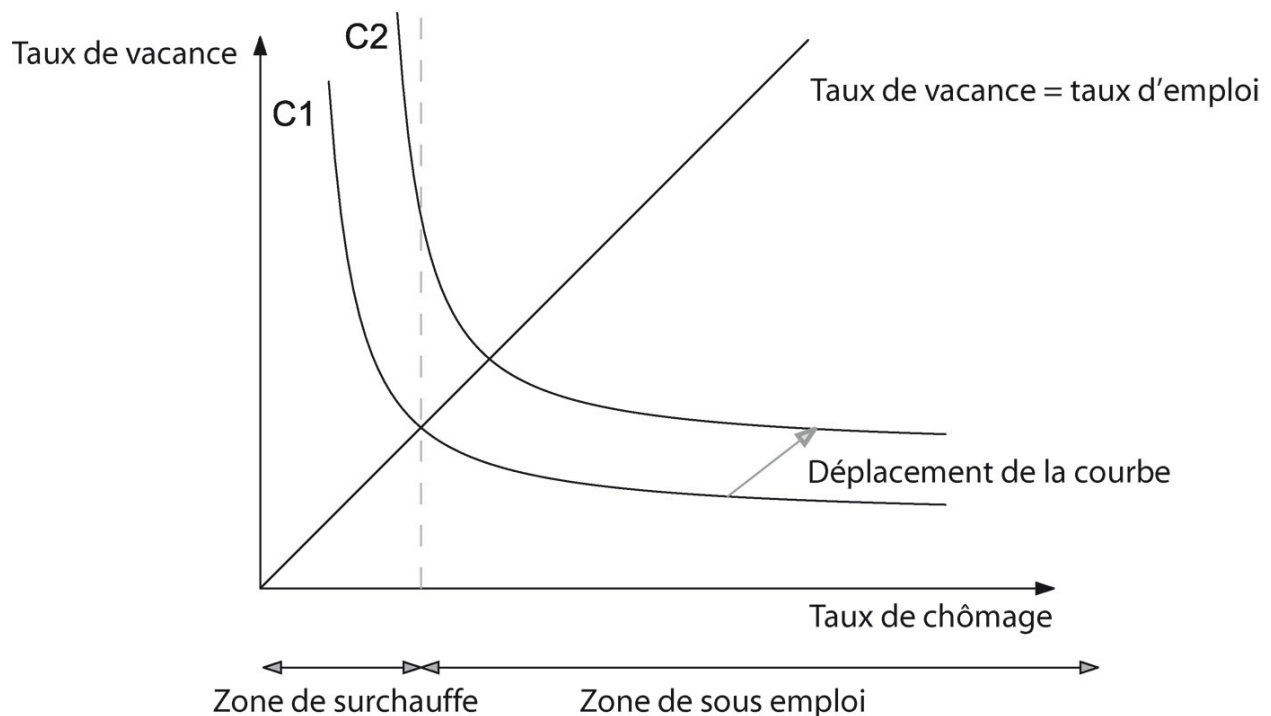
3.2 Chômage frictionnel et rôle des institutions

Le fonctionnement du marché du travail est imparfait, l'offre et la demande de travail ne se rencontrent d'ailleurs pas instantanément. Comme le montre la théorie de la prospection d'emploi, il est coûteux, en temps et financièrement, de chercher un emploi. De même, il est coûteux pour une entreprise de chercher un salarié du fait de l'existence d'asymétries d'informations. De plus, le travail étant hétérogène, il peut exister une inadéquation entre le type de travailleurs recherchés par les entreprises et le type de main-d'œuvre disponible. Dans certains cas, la présence de chômeurs et d'emplois vacants cohabite donc sur le marché du travail. On parle alors de **chômage frictionnel** car il résulte de l'existence de frictions empêchant une rencontre spontanée entre l'offre et la demande. La **courbe de Beveridge** permet de représenter le taux de vacance des emplois qui désigne le rapport entre le nombre d'emplois vacants et le nombre total d'emplois en fonction du taux de chômage, comme on le voit sur le [graphique 6.9](#).

Considérons que dans la situation initiale la courbe de Beveridge est C1. Lorsque le taux de chômage est faible et le taux de vacance des emplois élevé, l'économie est en situation de surchauffe ce qui se traduit généralement par des tensions inflationnistes. À l'inverse, quand le taux de chômage est élevé et le taux d'emploi faible, on est en situation de sous-emploi, ce qui se traduit généralement par des pressions à la baisse sur les taux de salaires. Une courbe de Beveridge représente à un moment donné un certain fonctionnement du marché du travail et la politique gouvernementale consiste à stimuler l'économie en situation de sous-emploi et à lutter contre

les tensions inflationnistes en situation de surchauffe. La politique de l'emploi doit se fixer un objectif, par exemple que le taux de chômage soit égal au taux d'emplois vacants. Lorsque la courbe se déplace, par exemple de C1 à C2, cela signifie que le fonctionnement structurel du marché du travail se modifie. Plus la courbe est éloignée de l'origine, plus il est difficile de faire correspondre les chômeurs et les emplois vacants.

Graphique 6.9 : La courbe théorique de Beveridge



Les données existantes indiquent une tendance structurelle à une importante réallocation des emplois et de la main-d'œuvre sur le marché du travail. Chaque jour de nombreux emplois sont détruits et d'autres se créent : ce sont les flux d'emploi. Les flux de main-d'œuvre désignent l'ensemble des départs et des arrivées de salariés. Si les deux phénomènes sont liés, il ne faut pas les confondre : les créations et destructions d'emplois ne concernent pas forcément les mêmes personnes et le titulaire d'un poste peut changer. Les flux de main-d'œuvre sont en général plus importants que les flux d'emploi, et on peut les évaluer par le **taux de rotation** de la main d'œuvre, qui est la demi-somme des flux annuels d'entrée et de sortie rapporté au nombre de salariés. Les enquêtes de l'INSEE montrent que ce taux a été multiplié par cinq durant les trente dernières années, ce qui est un bon indicateur des

transformations structurelles du marché du travail, avec davantage de contrats courts et de segmentation du marché du travail¹². P. Cahuc et F. Kramarz (2005) montrent qu'en moyenne, une entreprise qui crée un emploi au cours d'une année embauche cinq personnes et en licencie quatre autres durant la même période. En revanche, la suppression d'un poste de travail se traduit en moyenne par trois embauches et quatre départs. L'importance des flux d'emploi et de main-d'œuvre dépend du contexte institutionnel : une protection élevée de l'emploi réduit la possibilité pour les entreprises d'adapter la quantité de main-d'œuvre utilisée, notamment du fait de coûts d'ajustement élevés (primes de licenciement importantes, démarches administratives complexes, etc.). Tous les travailleurs n'ont pas par ailleurs la même mobilité, notamment en fonction de leur position dans le cycle de vie. Les jeunes actifs font ainsi preuve d'une mobilité plus forte que la moyenne des travailleurs.

D. Mortensen et Ch. Pissarides (1994), Prix Nobel en 2010, ont développé un modèle d'**appariement** afin d'analyser les problèmes liés aux asymétries d'information et à la mobilité de la main-d'œuvre. Dans leur modèle, flux d'emploi et de main-d'œuvre ne sont pas distingués et le stock de chômage dépend de la différence entre les flux d'embauche et de licenciements. Comme pour la théorie de la prospection d'emploi seuls les chômeurs cherchent un travail, les travailleurs cherchant à changer d'emploi ne sont ainsi pas pris en compte. L'appariement entre l'offre et la demande va notamment dépendre des efforts de recherche des salariés. Le nombre d'emplois proposés par les entreprises dépend des coûts de création d'un poste et de recherche d'un salarié ainsi que de la productivité du travail. La probabilité de sortie du chômage dépend du nombre d'emplois vacants et du nombre de chômeurs. En accroissant les indemnités de chômage, l'État peut contribuer à allonger la période de recherche d'emploi et accroître ainsi la valeur du chômage frictionnel, ce qui n'est pas nécessairement une mauvaise chose si in fine les appariements sont de meilleures qualités.

Afin de réduire le chômage frictionnel, la diffusion de l'information sur les postes existants par une agence de placement, comme Pôle emploi en France, facilite la rencontre entre l'offre et la demande de travail. L'État peut aussi financer des programmes de formation de la main-d'œuvre afin que l'offre de travail soit plus en adéquation avec la demande de travail. B. Crépon, E. Duflo., M. Gurgand, R. Rathelot et P. Zamora (2011) ont ainsi montré que l'accompagnement vers l'emploi de jeunes de niveau bac+2 minimum par

des opérateurs privés de placement (OPP) de 2007 à 2009 sur certains territoires a permis d'accélérer l'accès à l'emploi des bénéficiaires en les accompagnant dans leurs démarches, notamment pour rechercher des offres disponibles sur le marché.

Nous avons considéré jusqu'à présent l'existence de coûts liés à l'embauche et non de coûts liés au licenciement. En contribuant à l'augmentation du coût du travail, ces derniers conduisent à modifier la demande de travail des entreprises. Des modèles microéconomiques ont ainsi été développés pour intégrer ce type de coût dans les modèles d'appariement.

La protection de l'emploi risque donc de diminuer la demande de travail et de freiner l'adaptation de la structure des emplois aux mutations de l'économie. La réallocation de la main-d'œuvre est un facteur important de croissance de la productivité puisque les entreprises moins productives tendent à détruire plus d'emplois alors que les entreprises plus productives en créent davantage (OCDE, 2009). Augmenter la flexibilité de l'emploi permettrait dans cette optique de réduire le chômage. C'est le sens du débat sur la réforme du code du travail, qui a donné lieu à la loi El Khomri en 2016, puis aux réformes par ordonnances adoptées en septembre 2017. Ces dernières prévoient par exemple un plafonnement des indemnités de licenciement en cas de recours aux Prud'hommes pour licenciement abusif, de façon à donner aux entreprises plus de prévisibilité sur le coût maximal d'un licenciement. L'efficacité de ces réformes fait toutefois débat, car leur impact sur le niveau de chômage est difficile à évaluer (Askenazy et Erhel, 2017).

Nombre de propositions ont été faites pour concilier la flexibilisation des parcours professionnels avec la volonté de sécurité et de stabilité des salariés. Le modèle de la « **flexisécurité** » consiste à compenser l'instabilité de l'emploi par des mesures qui sécurisent les travailleurs entre deux emplois, comme l'ont fait des pays d'Europe du Nord. Ce système repose sur l'idée qu'il faut avant tout protéger l'employabilité de l'individu plus que son emploi. La possibilité accrue donnée aux entreprises de licencier leurs employés doit ainsi s'accompagner d'indemnités chômage généreuse et d'une véritable politique de formation pour aider les chômeurs à se reconverter.

Au-delà de la question de la flexibilité de l'emploi, les institutions existantes dans les différents pays impliquent des fonctionnements du marché du travail très différents. De nombreux débats existent pour savoir s'il est

possible de déterminer un arrangement institutionnel qui serait plus efficace qu'un autre, notamment pour obtenir un bas taux de chômage. Il semble en effet crucial de ne pas se contenter d'évaluer les différents dispositifs existants de façon isolée mais de les considérer dans leur ensemble. Il existe en effet des complémentarités institutionnelles entre les institutions d'un pays, qui se renforcent et se combinent mutuellement. On ne peut donc juger de l'efficacité d'une institution, par exemple de la nature d'un dispositif d'allocation pour les chômeurs, qu'à condition de tenir compte des autres institutions existantes, afin de voir si elle s'insère dans un ensemble cohérent ou non. B. Amable (2005) distingue ainsi cinq formes de capitalisme fonctionnant sur des modèles institutionnels différents ayant chacun leur logique propre. Ainsi, un même dispositif peut être efficace dans un arrangement institutionnel particulier et inefficace dans un autre.

Notes

1. Les données sont issues des enquêtes emploi de l'INSEE.
2. La délimitation de la catégorie des jeunes entre 15 ans révolus et 24 ans ne doit rien au hasard : il faut avoir plus de 15 ans pour pouvoir entrer sur le marché du travail, à l'exception de cas particuliers en apprentissage notamment. Par ailleurs, l'accès à certaines prestations sociales dépend également de l'âge, et certaines ne s'ouvrent pleinement aux jeunes qu'après 25 ans.
3. Source : Emploi, chômage, revenus du travail, INSEE Références, édition 2017.
4. D.S. Hamermesh, *La substitution entre les différentes catégories du travail, salaires relatifs et chômage des jeunes*, OCDE, disponible à l'adresse <http://www.oecd.org/dataoecd/9/13/35483268.pdf>
5. Pour une présentation détaillée de ces questions, on peut consulter l'ouvrage de D. Anne et P. L'Horty, *Économie de l'emploi et du chômage*, Armand Colin, coll. « Cursus », 2013.
6. L'espérance renvoie ici au concept mathématique. Quand les gains associés à une action ne sont pas certains (comme les gains liés au fait de jouer au quinté), on suppose que les individus déterminent le comportement à adopter en maximisant l'espérance d'utilité, ce qui revient à maximiser la somme des satisfactions ressenties dans les différentes configurations envisageables en pondérant chaque utilité par la probabilité d'occurrence de l'événement associé.
7. Source : Centre d'Etude et de Recherche sur les Qualification (Cereq) (2017), *Quand l'école est finie. Premiers pas dans la vie active de la Génération 2013*, Cereq Enquêtes.
8. Source : INSEE, enquête emploi, données 2016.
9. Source : INSEE, *Emploi, chômage, revenus du travail*, Insee Références, édition 2017.
10. Bruno Ducoudré *et al.*, « CICE et Pacte de responsabilité : une évaluation selon la position dans le cycle. Simulations *ex-ante* à partir d'un modèle macroéconomique », *Revue de l'OFCE* 2016/2 (n° 146), et Haithem Ben Hassine et Claude Mathieu (2017), « Préfinancement du CICE et ses effets sur l'emploi : une évaluation *ex-post* », *France Stratégie*, Document de travail n°2017-10.

[11.](#) Centre d'Étude et de Recherche sur les Qualification (Cereq) (2017), *Quand l'école est finie. Premiers pas dans la vie active de la Génération 2013*, Cereq Enquêtes.

[12.](#) C. Picart, « Une rotation de la main-d'œuvre presque quintuplée en 30 ans : plus qu'un essor des formes particulières d'emploi, un profond changement de leur usage », INSEE, *Emploi et salaires*, édition 2014.

Chapitre 7

Modifier l'équilibre du marché

L'exemple de la redistribution des revenus

1. L'ÉQUILIBRE GÉNÉRAL CONCURRENTIEL

2. QUELS CRITÈRES DE JUSTICE ?

3. INTRODUCTION À LA TAXATION OPTIMALE DES REVENUS

L'analyse économique ne se limite pas à la question de l'efficacité des systèmes économiques mais pose aussi celle de leur caractère plus ou moins juste. L'économie normative est une branche de l'économie qui introduit des considérations relatives à la justice sociale, elle propose des critères normatifs pour évaluer l'équilibre atteint par le marché et s'interroger sur l'existence de situations préférables à d'autres, du point de vue collectif. Les critères de justice mis en avant par les économistes sont très largement inspirés des travaux menés en philosophie, notamment ceux de J. Rawls¹ qui ont profondément modifié l'approche de cette question. À la différence de la philosophie cependant, l'économie normative associe réflexion sur les critères de justice et analyse positive des avantages et inconvénients économiques associés aux différentes politiques envisageables. Les économistes n'ont pas plus de légitimité que d'autres citoyens pour se prononcer sur ce qui est juste et ils n'ont pas vocation à fonder scientifiquement la légitimité d'un critère de justice plutôt qu'un autre. Le véritable apport des économistes se situe dans l'analyse des effets des différents outils économiques que les responsables politiques peuvent utiliser.

Dans la plupart des démocraties modernes, l'État a pour objectif de réduire les inégalités de revenus. Il dispose pour ce faire de plusieurs outils :

- agir sur la répartition des revenus primaires qui découle de la participation à l'activité productive. La mise en place du SMIC est un exemple de ce type de politique (voir le [chapitre 6](#)).
- agir sur la répartition des revenus disponibles en jouant sur les taxes et les transferts (impôt progressif sur le revenu, minima sociaux, etc.), chaque impôt ayant des effets distincts selon la situation des agents.

La réduction des inégalités de revenus passe par des politiques très variées : le soutien aux bas revenus peut s'appuyer sur des instruments

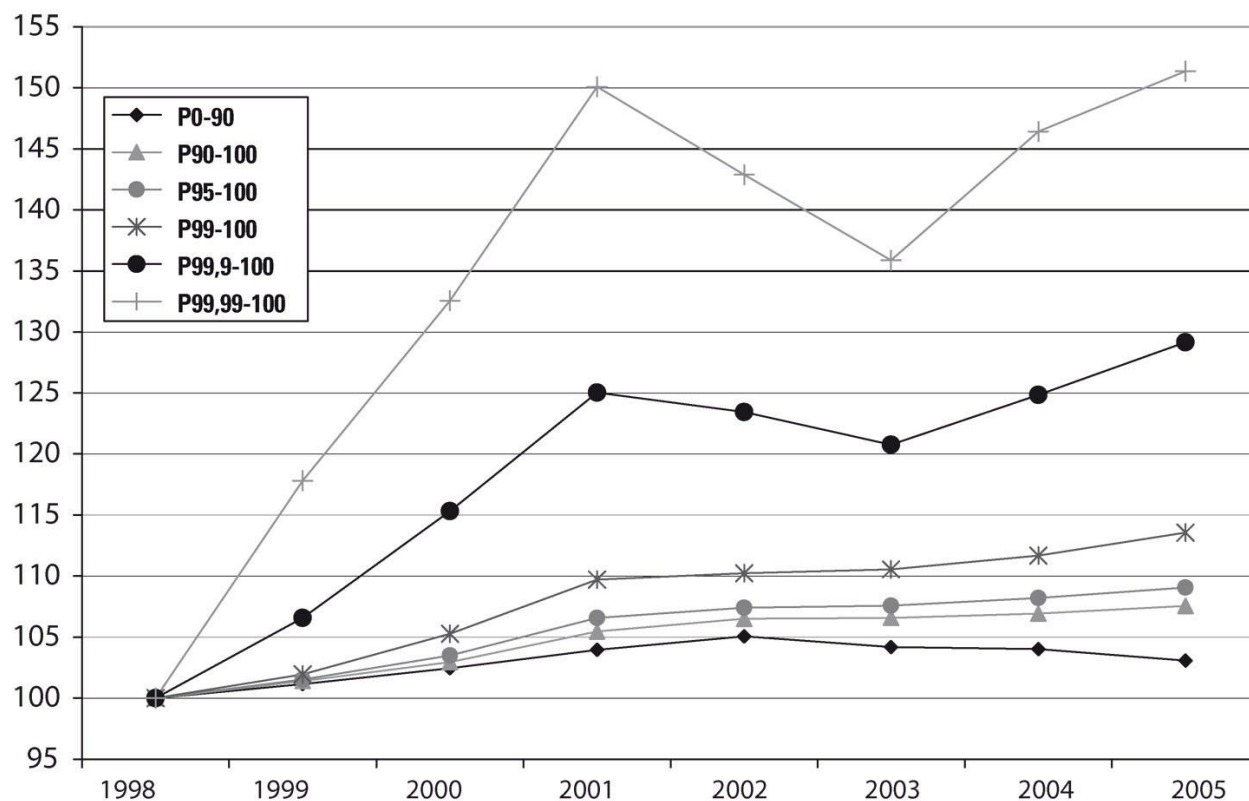
comme le RSA, la baisse des cotisations salariales, les allocations logement, etc. Si l'impôt progressif sur le revenu est un instrument privilégié de ces politiques, il reste encore à choisir entre une imposition commune ou différenciée des revenus du capital et du travail, une taxe linéaire ou avec taux marginal d'imposition croissant, etc. Le travail de l'économiste consiste à dire quelle est la méthode la plus efficace, d'un point de vue économique, pour réduire les inégalités de revenus compte tenu de l'idée de « la justice fiscale » que l'on se donne au départ. De ce fait, il se place dans une démarche scientifique.

Nous allons nous concentrer dans ce chapitre sur la question des revenus du travail, et plus particulièrement des salaires. Les travaux empiriques récents, notamment de C. Landais², montrent que les hauts salaires ont fortement augmenté ces dix dernières années alors que les salaires faibles et moyens ont plutôt stagné.

Nous pouvons ainsi lire dans le **document 7.1** que les 0,1 % salariés les mieux rémunérés ont vu leur salaire croître de 29,2 % entre 1998 et 2005 contre 3,1 % pour les 90 % des salariés les moins bien payés.

Les inégalités salariales sont aujourd'hui fortes comme en témoigne le **document 7.2, à la fois sous l'angle des différences de qualification et entre les sexes.**

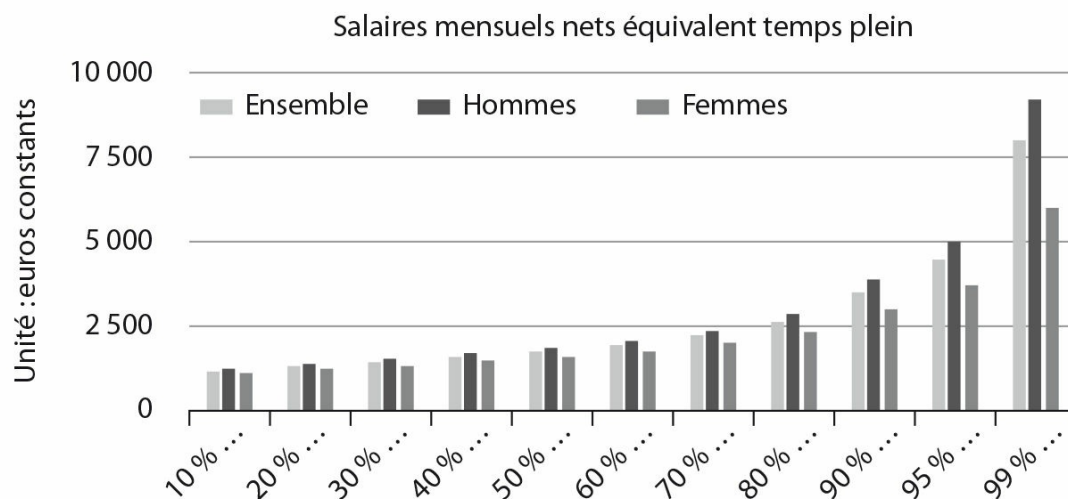
Document 7.1 : Évolution du salaire moyen de différents fractiles de salaires



Lecture : en 2001, l'indice du salaire moyen des 1 % les mieux rémunérés (le fractile P99-P100) atteignait 110, base 100 en 1998, il a donc augmenté de $(110-100)/100$ soit 10 % entre 1998 et 2001.

Source : Camille Landais, 2007.

Document 7.2 : Niveaux de salaires par sexe et par décile



Champ : France. Salariés du secteur privé et semi-public y compris les bénéficiaires de contrats aidés. Sont exclus les apprentis, les stagiaires, les salariés agricoles et les salariés des particuliers employeurs.

Lecture : le salaire mensuel net en équivalent temps plein des 10 % les plus pauvres est de 1 200 euros.

Source : Insee – Données 2013.

De nombreux citoyens et responsables politiques sont favorables à une réduction des inégalités notamment à travers une politique de redistribution des revenus en direction des plus démunis. Certains craignent néanmoins qu'une forte taxation des hauts revenus désincite leur détenteur à travailler et que des minima sociaux trop élevés encouragent l'inactivité. Quel regard l'économiste porte-il sur ce débat ? Comment concevoir un système de transferts et de taxes qui soit à la fois juste et efficace ?

1. L'équilibre général concurrentiel

Pour étudier les effets économiques de la redistribution des revenus, les économistes ont besoin de construire des modèles afin de rendre compte de leur répartition initiale résultant des mécanismes de marché. Nous nous focaliserons ici sur la présentation de l'équilibre atteint en situation de concurrence pure et parfaite. Cette démarche peut surprendre car ce cadre n'est pas le plus réaliste pour étudier les inégalités dans les économies contemporaines, caractérisées par un grand nombre d'imperfections. Mais il offre l'avantage de clairement séparer les problèmes de « **redistribution pure** », c'est-à-dire de redistribution pour de pures considérations de justice sociale au sein d'une économie où l'allocation des ressources est assurée efficacement par le marché, comme c'est le cas en concurrence pure et parfaite, et les problèmes de « **redistribution efficace** », soit de redistribution motivée par un souci d'amélioration du fonctionnement du système économique, dans une économie où les marchés sont imparfaits.

1.1 Une allocation optimale des ressources

L'allocation des ressources renvoie à la répartition des facteurs de production entre leurs différents usages possibles, par exemple utiliser plus ou moins de capital et de travail dans l'industrie agro-alimentaire ou dans la fabrication des ordinateurs. Une situation est dite **optimale au sens de Pareto** s'il n'est pas possible d'améliorer la satisfaction d'un individu sans diminuer celle d'un autre (voir le [chapitre 1](#)). Le critère de Pareto est un critère d'efficacité, il implique qu'aucune ressource n'a été gaspillée ou est inutilisée et que les facteurs sont mobilisés pour produire des biens et services considérés comme utiles par les individus. En effet, si ce n'était pas le cas, il serait possible de mobiliser plus de facteurs ou de revoir leur allocation afin de produire plus et de réorienter la production dans un sens plus conforme aux attentes des individus ce qui permettrait nécessairement d'augmenter la satisfaction de certains sans que d'autres y perdent. Le **premier théorème de l'économie du bien-être** établit que tout équilibre général en concurrence pure et parfaite est nécessairement un optimum de Pareto. Le comportement maximisateur des

agents, quand les marchés sont parfaitement concurrentiels, conduit spontanément à un équilibre optimal au sens de Pareto par le biais du mécanisme des prix. À travers les quantités demandées pour différents niveaux de prix, les consommateurs révèlent en effet leurs préférences et ces dernières orientent ainsi la façon dont les ressources vont être allouées. Bien entendu, dans une économie où les marchés sont imparfaits, l'allocation des ressources n'est pas optimale et peut être améliorée par une intervention publique (voir le [chapitre 5](#)).

1.2 Un équilibre qui peut être très inégalitaire

La nature et la quantité de biens et services accessibles pour les différents individus à l'équilibre et le niveau de leur bien-être dépend fortement de la dotation initiale de chacun en facteurs. Pour simplifier considérons deux facteurs de production, le travail et le capital.

Un individu qui dispose d'un patrimoine élevé du fait d'un héritage important a un revenu plus élevé à l'équilibre qu'un individu tirant ses revenus de son seul travail, et peut ainsi accéder à un nombre plus important de biens et services. De plus, si chaque individu dispose d'une force de travail, la qualité de son travail est différente. Les économistes considèrent ainsi qu'il existe des individus de type différents, du point de vue du marché du travail, qualifiés de plus ou moins « talentueux ». Certains sont plus aptes que d'autres à être productifs et cette capacité dépend de multiples facteurs comme l'environnement familial et social dans lequel ils ont évolué, leur personnalité, aptitude physique, parcours scolaire, etc. Précisons qu'il n'y a aucun jugement normatif sur les individus ayant moins de « talent » que d'autres, mais uniquement la construction d'un modèle pour identifier les effets de cette variable sur la répartition des revenus, quelle que soit l'origine des inégalités de « talent ». Cette variable renvoie à la capacité d'un individu à être productif mais ne peut être directement assimilée à la productivité. En effet, à « talent » donné, chaque individu peut ensuite s'investir plus ou moins pour valoriser ce talent en s'impliquant dans son travail. La **productivité** d'un individu résulte ainsi de son « talent » et de l'effort fourni. Certains individus « talentueux » ne s'investissent pas dans leur travail et au final ont la même productivité horaire que d'autres, moins talentueux mais pleinement investis. Prenons l'exemple d'un individu mal voyant : du fait de sa condition physique, il sera plus difficile pour lui d'effectuer certaines

tâches. À effort et temps de travail égal, il est vraisemblable que pour un certain nombre d'activités, il produira moins qu'un individu n'ayant pas cette déficience et aura donc une productivité plus faible. Les enquêtes montrent d'ailleurs que les travailleurs handicapés ont un taux de chômage plus de deux fois supérieur au reste de la population, et la puissance publique peut s'interroger par exemple sur la manière de compenser ces inégalités³. On fait l'hypothèse que la dotation initiale de chaque individu en « talent » détermine le taux de salaire auquel les différents individus peuvent prétendre : sur des marchés concurrentiels, les facteurs de production sont rémunérés à leur productivité marginale (voir les chapitres 3 et 6). Les différences de salaire entre individus reflètent donc les différences de productivité et de volume d'heures travaillées. Les individus à faible « talent », même s'ils fournissent un effort maximal, ne peuvent pas obtenir des taux de salaire élevés. Si la dotation initiale en facteurs est très inégalitaire, l'équilibre atteint par le marché le sera tout autant, et les individus auront des revenus et des niveaux de satisfaction très hétérogènes. Le fait que l'équilibre soit optimal n'implique donc pas qu'il soit juste. Si tout le monde est plus ou également satisfait dans une situation par rapport à une autre, elle semble collectivement préférable mais cela ne signifie pas pour autant qu'elle est juste, car tout dépend en réalité du critère de justice adopté. Par exemple, le fait que les personnes handicapées, pour des raisons qui ne dépendent pas de leur volonté, disposent au final d'un revenu plus faible que les autres à cause d'une productivité inférieure peut être considéré comme injuste, même si cet état de fait résulte d'une allocation optimale des ressources. En plus des considérations sur la répartition des revenus, on peut vouloir modifier l'allocation des ressources opérée par le marché afin d'orienter la production vers des activités jugées socialement plus utiles, par exemple privilégier la production des soins médicaux plutôt que la production d'armes, quand bien même l'allocation opérée par le marché est efficace.

Par ailleurs, plusieurs allocations optimales des ressources sont envisageables, et le critère de Pareto ne permet pas de les classer. Il s'agit d'un critère d'ordre incomplet, car il ne propose pas de hiérarchiser toutes les situations possibles mais seulement de départager celles qui sont efficaces des autres. Il est donc crucial de se doter de critères de justice permettant de classer différents états possibles de la société afin de déterminer celui qui semble le plus désirable, et décider des politiques à mettre en œuvre pour l'atteindre.

2. Quels critères de justice ?

Qu'est-ce qu'une allocation juste des ressources ? Comment devraient être répartis les revenus ? Il ne s'agit pas de débattre des différentes options philosophiques évoquées dans cette partie, mais de présenter simplement les critères mobilisés par les économistes. L'accent sera mis sur l'approche proposée par l'économie du bien-être, la plus usuelle en microéconomie. D'autres ont été développées et d'importantes critiques ont pu être formulées vis-à-vis de la conception bien-être de la justice mais elle a l'avantage de faire l'objet de nombreuses applications en économie publique, notamment sur la taxation optimale des revenus. Un lecteur curieux pourra se reporter à l'ouvrage de M. Fleurbaey (1996) ou à l'article de D. Maguain (2002) pour plus d'informations sur les autres approches.

2.1 L'approche « bien-être »

• Les fonctions de bien-être social

Le courant de l'**économie du bien-être** (*welfare economics*) se propose d'évaluer les différentes allocations des ressources en comparant la satisfaction ressentie par chacun dans chaque situation. L'objectif est de s'appuyer sur les satisfactions individuelles pour construire une mesure de la satisfaction globale. Par exemple, on mesure le bien-être de chacun pour différentes répartitions des revenus et en agrégeant les satisfactions individuelles, on en déduit la satisfaction globale associée à chaque type de répartition. L'allocation qui permet de maximiser le bien-être collectif est alors considérée comme la meilleure solution du point de vue « social », c'est-à-dire pour la collectivité considérée. Cette approche peut ainsi être qualifiée d'individualiste car elle est basée sur les satisfactions individuelles. Tout l'enjeu réside dans la façon dont on agrège ces utilités individuelles. Les économistes utilisent des **fonctions de bien-être social**, qui sont des fonctions dont les arguments (variables) sont les utilités individuelles et dont la valeur représente le bien-être social. Les fonctions de bien-être social servent à classer l'ensemble des allocations possibles et à déterminer laquelle

est la plus juste. La forme fonctionnelle choisie n'est pas neutre : en fonction des critères choisis, l'allocation qui maximise la valeur de la fonction et ainsi le bien-être social n'est pas la même et la situation considérée comme la plus juste sera différente.

Les fonctions de bien-être social de base reposent sur une **conception cardinale de l'utilité** : les individus sont capables de mesurer quantitativement leur degré de satisfaction, et par hypothèse les échelles d'évaluation des différents individus sont comparables entre elles (voir le [chapitre 2](#)). Les développements récents en économie normative ont permis de proposer des fonctions de bien-être social plus réalistes qui ne reposent pas sur une conception cardinale de l'utilité et ne nécessitent pas que les utilités soient comparables entre elles mais nous ne détaillerons pas ces travaux ici et renvoyons le lecteur intéressé à l'article de M. Fleurbaey et F. Maniquet (1999).

• L'approche utilitariste classique

L'utilitarisme classique mesure le bien-être collectif à travers la somme des niveaux de bien-être individuels. Ce courant de pensée est issu de la philosophie morale du ^{xviii}e siècle et a été nourri par les travaux de J. Bentham, J.-S. Mill ou encore H. Sidgwick. Si la somme des utilités individuelles est maximale, l'allocation est considérée comme juste. Notons que dans ce cas de figure la situation est optimale au sens de Pareto. En effet, toute modification de l'allocation impliquerait une perte d'utilité au niveau global et donc forcément une perte d'utilité pour au moins une personne.

Considérons un exemple simple où deux individus sont amenés à comparer deux répartitions possibles des revenus. Dans la situation 1, Paul a un revenu disponible de 1 000 euros par mois et José 5 000 euros. Dans la situation 2, Paul dispose de 2 000 euros et José 4 000 euros. Chaque individu est capable par hypothèse d'évaluer de façon chiffrée la satisfaction ressentie dans chaque situation (conception cardinale de l'utilité, par souci de simplification). Paul chiffre la valeur de son bien-être à 500 dans la situation 1 et à 900 dans la situation 2. Le doublement de son revenu a presque doublé sa satisfaction. José estime que son bien-être est de 2 000 dans la situation 1 et de 1 600 dans la situation 2. En perdant 1 000 euros de l'une à l'autre, il voit sa satisfaction diminuer de 400. Le bien-être global est

de 2 500 dans la situation 1 et de 2 600 dans la situation 2. D'un point de vue utilitariste, la répartition n° 2 doit être préférée, non parce qu'elle est moins inégalitaire mais en raison de la maximisation du bien-être collectif. La raison pour laquelle la variation de la satisfaction n'est pas proportionnelle à la variation du revenu vient du fait qu'on considère généralement que l'utilité marginale du revenu est décroissante (voir le [chapitre 2](#)). Un euro supplémentaire apporte plus de satisfaction à un individu qui dispose d'un faible revenu qu'à un individu qui bénéficie d'un revenu conséquent. Les répartitions très inégalitaires maximisent ainsi rarement le bien-être collectif. Cette approche a le mérite de considérer avec la même importance la satisfaction de chacun et ne juge pas les préférences individuelles : il n'y a pas de critère moral qui s'impose aux individus.

Le critère utilitariste classique a été très largement critiqué dans les débats théoriques sur la justice sociale. Il est en effet indifférent à l'inégalité des satisfactions individuelles. Peu importe que certains individus tirent au final une satisfaction beaucoup plus forte d'une allocation que d'autres du moment que la somme des utilités individuelles est maximale. Que se passe-t-il, par exemple, si la répartition maximisant la satisfaction collective prévoit en fait d'exclure une minorité d'individus de l'accès à toute richesse : est-ce une situation acceptable ? Ces critiques ont conduit les économistes à revoir les fonctions de bien-être social utilisées.

• Introduire de l'aversion pour les inégalités

La solution adoptée consiste à définir la fonction de bien-être social comme la somme pondérée des utilités, les faibles niveaux d'utilité étant plus valorisés que les forts niveaux d'utilité. Le cas utilitariste classique correspond au cas où il n'y a aucune aversion pour l'inégalité et où tous les individus sont pondérés de la même façon. On peut aller jusqu'à envisager d'attribuer un poids nul à toutes les utilités à l'exception de la plus faible, le but de la société devenant alors de maximiser l'utilité des moins bien lotis. Cette fonction d'utilité est parfois qualifiée de « fonction de bien-être rawlsienne », car J. Rawls préconisait de maximiser la situation des plus faibles, selon le « principe du maximin ». Cette appellation est cependant abusive, Rawls ayant construit sa théorie en opposition à la théorie utilitariste. Ce critère de justice n'est pas égalitariste dans la mesure où l'égalité parfaite des situations n'est pas visée. Des inégalités seront

acceptées si elles sont favorables aux plus pauvres. Ainsi une augmentation trop forte de la taxation des hauts revenus du travail qui découragerait le travail des individus talentueux et conduirait à une baisse des ressources fiscales et de la production, et *in fine* à offrir aux plus démunis des prestations sociales faibles, n'est pas valorisée dans une optique rawlsienne. Une taxation des hauts revenus un peu moins élevée en étant plus favorable au travail des plus talentueux permet de financer des aides sociales plus généreuses et profite aux plus pauvres.

2.2 Les limites de l'approche bien-être

La conception bien-être de la justice a été fortement critiquée. Un certain nombre d'informations qui peuvent sembler pertinentes pour évaluer la justice d'une situation sont en effet passées sous silence. Par exemple, la question du respect des droits de l'homme n'est pas prise en compte à travers un critère basé sur l'utilité. Si une majorité d'individus voit sa satisfaction augmenter en raison de la ségrégation d'un groupe social minoritaire, alors une situation où un petit nombre d'individus est discriminé peut être jugé meilleure d'un point de vue utilitariste. Inversement, certains éléments qui semblent peu pertinents sont valorisés : un individu n'éprouvant de plaisir qu'en consommant des biens très onéreux aurait à revenu égal, un plus faible niveau d'utilité qu'un individu ayant des goûts plus simples, et il devrait être subventionné pour que sa faible utilité soit compensée. Ne vaut-il pas mieux directement considérer les inégalités de revenus plutôt que les inégalités d'utilités ?

Se limiter à la question de la répartition des revenus pose néanmoins aussi problème : certains auteurs, dont A. Sen, ont fait valoir le fait que tous les individus n'ont pas la même capacité à transformer une ressource en bien-être. Pour lui, l'essentiel n'est pas qu'un individu dispose de telle ou telle ressource mais ce qu'il est en capacité de faire avec elle. L'approche de Sen est ainsi qualifiée d'approche par les « **capacités**⁴ ». L'État doit donc compenser, selon Sen, les différences de certaines capacités fondamentales (principe de compensation) sans tenir compte ensuite du niveau d'effort choisi par les individus et donc du niveau d'utilité qu'ils auront (principe de responsabilité). Ainsi, un individu qui serait pauvre malgré ses capacités à obtenir un revenu correct n'aurait pas à être aidé alors qu'un individu pauvre

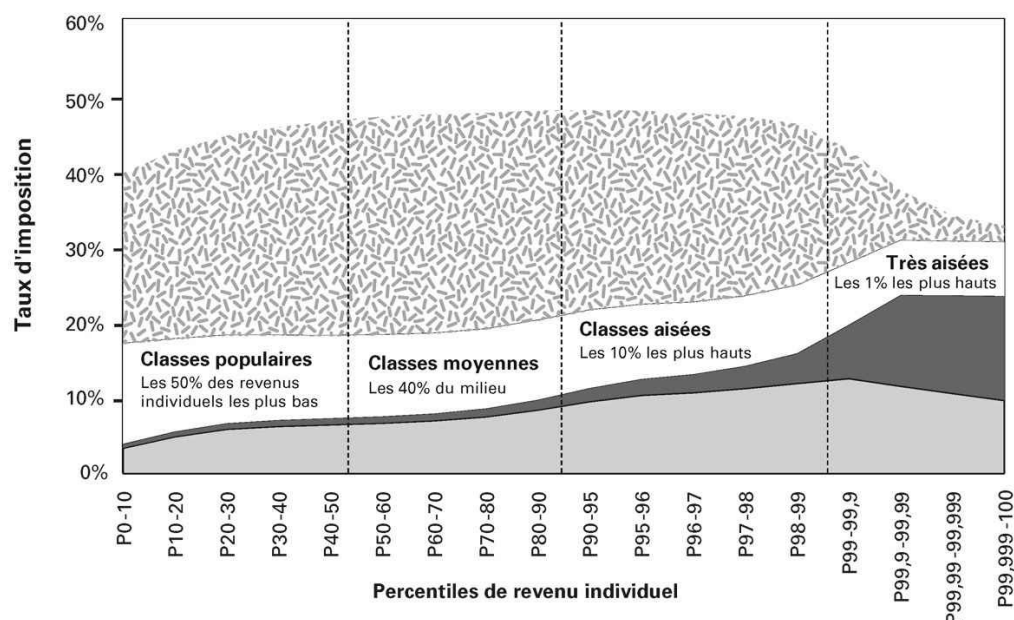
du fait d'un handicap par exemple doit recevoir une subvention.

3. Introduction à la taxation optimale des revenus

Supposons que l'équilibre atteint par les mécanismes de marché soit considéré comme injuste et que la puissance publique cherche à le modifier. Comment doit-elle s'y prendre ? Nous allons nous focaliser ici sur la répartition des revenus du travail. À l'équilibre les salaires peuvent être très inégalitaires en raison des différences de « talents ». L'État peut alors modifier la répartition des revenus par un système de taxes et de transferts. Il existe bien d'autres outils que la fiscalité, par exemple la mise en place d'une politique éducative de démocratisation scolaire agit directement sur les inégalités de talents, pour que les revenus du travail qui en découlent soient moins inégaux. Dans cette partie seuls les outils de redistribution dont dispose l'État sont examinés, la répartition des **revenus primaires**, soit les revenus liés à la participation au processus productif, est donc donnée. Alors comment l'État peut-il faire en sorte que la répartition des **revenus disponibles** soit moins inégalitaire ? Les revenus disponibles résultent de l'action redistributive sur les revenus primaires, auxquels on retire les impôts directs et cotisations sociales, et on ajoute les transferts sociaux (allocations, aides sociales, etc.). Le système actuel de redistribution est considéré par beaucoup comme injuste et inefficace. Injuste car il ne serait pas assez redistributif et inefficace car il engendrerait des effets pervers.

Document 7.3 : Les prélèvements obligatoires en France en fonction du revenu

- ▣ Cotisations sociales et taxes sur les salaires
- Impôts sur la consommation : taxe sur la valeur ajoutée (TVA)+autres
- Impôts sur le capital : impôt sur les bénéfices des sociétés (IS)+taxe foncière (TF)+impôt sur la fortune (ISF)+droits de succession (DMTG)
- ▨ Impôts sur le revenu : contribution sociale généralisée (CSG)+impôt sur le revenu des personnes physiques (IRPP)



Source : LANDAIS, PIKETTY, SAEZ, 2011.

C. Landais, T. Piketty et E. Saez (2011) ont calculé pour chaque niveau de revenu le montant des impôts ou cotisations sociales acquittés par les individus afin d'en déduire un **taux global effectif** de prélèvement (appelé improprement taux d'imposition dans le **document 7.3** par niveau de revenu, c'est-à-dire la part de leur revenu que les individus acquittent effectivement en prélèvements obligatoires. Il ne faut pas confondre le taux effectif et le taux légal d'imposition pour un revenu donné : supposons qu'une personne soit soumise à une taxe de 10 %, mais en raison de l'utilisation de son mode de calcul ou de différents dispositifs de réduction de l'impôt (par exemple des placements déductibles dans des produits d'épargne) elle paie un impôt plus faible, alors le taux effectif de cette taxe, soit le rapport entre l'impôt effectivement versé et son revenu, sera bien inférieur à 10 %. Les auteurs ont ensuite calculé le taux de prélèvement acquitté par chaque centile de revenu (percentile en anglais), c'est-à-dire le taux moyen d'imposition effectivement acquitté par les 1 % les plus pauvres, les 2 % les plus pauvres, etc. Nous pouvons lire ici les résultats de ces calculs.

D'après le **document 7.3**, les prélèvements obligatoires sont dégressifs pour les revenus les plus élevés. Les individus appartenant au 0,001 % les

plus fortunés acquittent en proportion moins que les 10 % les moins fortunés. Les impôts directs (impôts sur le capital et le revenu) sont globalement progressifs contrairement aux impôts sur la consommation dont la part dans le revenu diminue à mesure que celui-ci augmente. C. Landais, E. Saez et T. Piketty en appellent pour cette raison à une « révolution fiscale »⁵ afin de rétablir une réelle progressivité du système de prélèvement. En tenant compte de l'effet des transferts sociaux, il est toutefois important de noter la progressivité globale du système français de redistribution, bien qu'elle reste insuffisante aux yeux de certains. Il ne faut en effet pas oublier que ces prélèvements servent notamment à financer des prestations sociales qui, si elles ne sont pas toutes réservées aux ménages à bas revenus, profitent en moyenne plus à ces derniers et contribuent ainsi fortement à la réduction des inégalités de revenus disponibles.

3.1 La démarche de la taxation optimale

• Optimalité, justice et information parfaite

Si la puissance publique a identifié à partir de critères de justice qui lui sont propres, la situation à atteindre, elle doit chercher à le faire de façon la plus efficace possible. Si l'État connaissait le « talent » de chaque individu, il lui suffirait de mettre en œuvre des transferts forfaitaires entre les individus et de laisser jouer ensuite les mécanismes de marché. Un **impôt forfaitaire** (ou un transfert forfaitaire) est un impôt (transfert) dont le montant est défini en fonction de la valeur d'une variable que les individus ne peuvent modifier. Pour chaque valeur de la variable, la somme à payer (ou à recevoir) est fixe et ne dépend pas du comportement de l'individu. Un système dans lequel tous les individus âgés entre 30 et 60 ans devraient payer une taxe de 100 euros par an, les autres n'ayant rien à payer est un exemple d'impôt forfaitaire. Les individus ne peuvent modifier leur âge et pour chaque âge une somme fixe doit être acquittée quel que soit leur revenu, leur situation professionnelle, etc. Dans le cas présent, le « talent » de chacun est une donnée que les agents ne peuvent modifier en fonction du montant de la taxe à payer ou du transfert à recevoir. L'État pourrait ainsi par exemple imposer une taxe de 1 500 euros par mois aux individus les plus talentueux, quel que soit le nombre d'heures travaillées, et verser une subvention fixe aux

individus à faible « talent ». Ce type d'impôt et de transfert est considéré comme **non distorsif** car il n'entraîne pas de perte sèche, (voir le [chapitre 1](#)) et ne modifie pas la structure des prix relatifs. Les individus décident en effet de leur offre de travail en fonction du gain marginal associé à chaque heure supplémentaire travaillée. Ce dernier n'ayant pas changé (travailler une heure de plus rapporte toujours la même somme qu'avant), la décision de travailler ne dépend pas du système de taxes et de transferts (voir le [chapitre 6](#)). Un individu qui travaillerait beaucoup n'aurait pas à payer plus de taxes qu'un individu travaillant peu, la taxe ne dépendant que du « talent » de chacun. De ce fait, si l'État connaît le talent de chacun, il peut définir un système de taxe et de transferts qui soit non distorsif afin de lutter contre les inégalités. Quand l'information est parfaite, il n'y a pas d'arbitrage entre efficacité et justice. Partant d'une situation efficace découlant du fonctionnement des marchés concurrentiels, l'action publique permet d'atteindre une autre allocation plus juste sans qu'il y ait de pertes d'efficacité. Le nouvel équilibre est ainsi optimal au sens de Pareto.

Au-delà de la question de la répartition des revenus, en modifiant les dotations initiales en facteurs par un système de transferts forfaitaires et en laissant ensuite jouer les mécanismes de marché, l'État peut faire en sorte que le fonctionnement du marché aboutisse à un équilibre choisi à l'avance en fonction de critères de justice. Les transferts forfaitaires permettent de « décentraliser » l'optimum atteint par le marché vers une allocation juste des ressources, c'est-à-dire d'atteindre un autre optimum du fait d'une modification de la dotation initiale. Cette situation optimale est qualifiée d'optimum de premier rang. Le **second théorème de l'économie du bien-être** stipule ainsi qu'à tout équilibre de Pareto peut être associé un équilibre concurrentiel. L'État peut donc théoriquement par son action faire en sorte que l'économie converge vers une situation à la fois juste et efficace.

• L'arbitrage efficacité / justice

L'information étant distribuée de façon imparfaite, une telle solution est impossible en pratique. L'État ne connaît pas le talent des individus et ne peut mettre en œuvre des transferts forfaitaires en fonction de ces derniers. S'il observe le revenu de chacun et le nombre d'heures travaillées et peut donc en déduire la productivité horaire moyenne des individus, l'État ne peut savoir quelle est la part de la productivité qui découle de l'effort fourni et

celle qui dépend de leur « talent ». La taxation du travail risque alors de générer des **distorsions**, c'est-à-dire de conduire à une allocation non optimale des ressources. Deux possibilités s'offrent alors à l'État pour modifier la répartition des revenus.

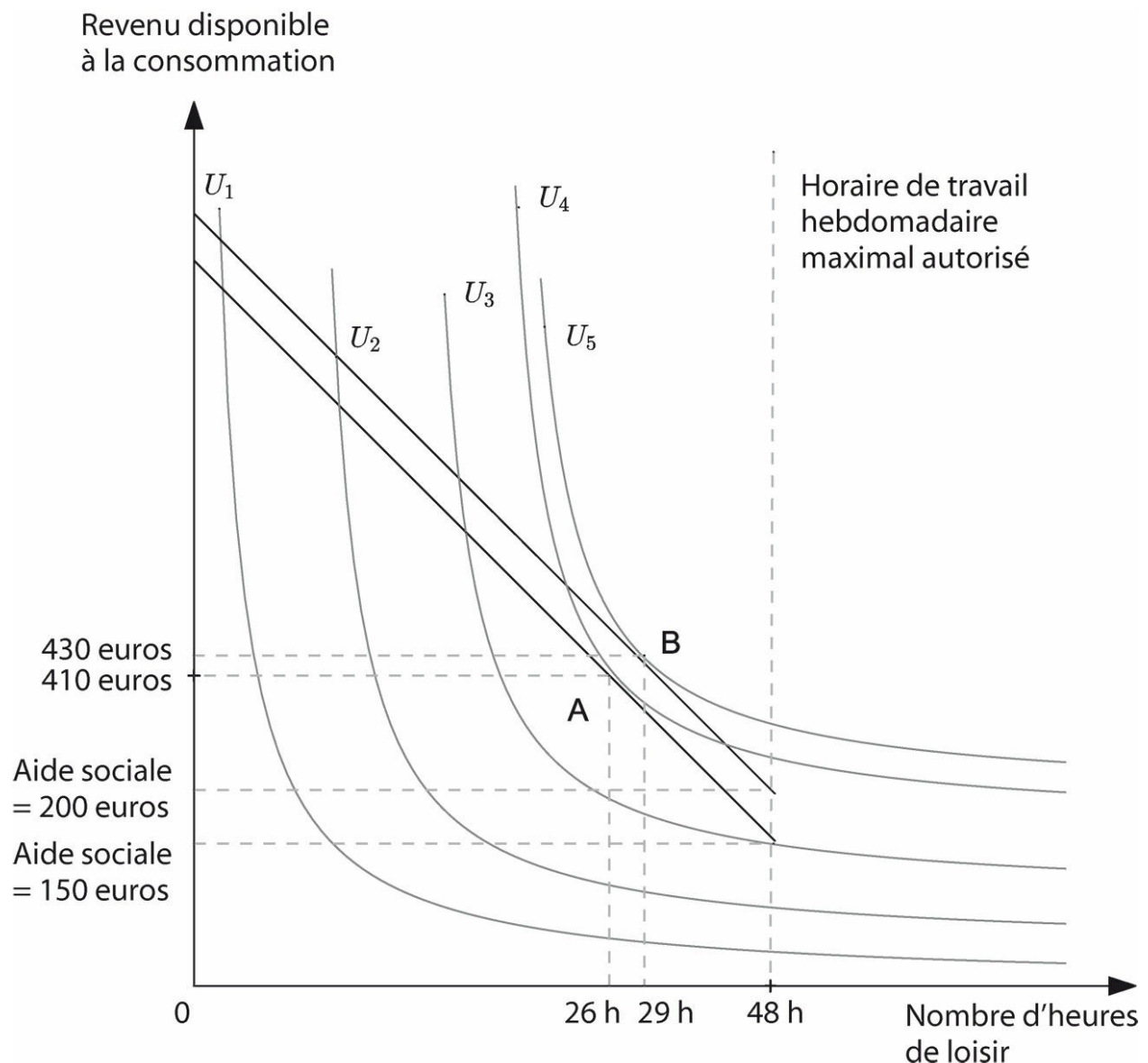
La première a été proposée par G. Akerlof (1978). L'État ne sachant identifier directement le « talent » de chaque individu, il fixe ses taxes et transferts à partir de paramètres observables qui ne dépendent pas de l'offre de travail et qui sont corrélés aux compétences et à la capacité de travailler de chacun tels l'âge, l'existence d'un handicap, le statut familial, etc. Cette solution a l'avantage de ne pas introduire de distorsions mais permet une redistribution beaucoup moins précise des revenus. Les variables corrélées au revenu mais non manipulables par les individus sont en effet peu nombreuses et ne permettent pas de cibler parfaitement la redistribution opérée par le système de taxes et de transferts : il est impossible d'atteindre parfaitement l'objectif de baser la fiscalité sur le « talent » des individus par cette méthode.

Deuxième possibilité, accepter que le système de taxes et transferts génère des distorsions pour atteindre de façon précise l'objectif fixé, tout en minimisant ces distorsions. L'État base son système fiscal sur le revenu déclaré par les personnes sans savoir si un individu qui déclare un revenu faible est un individu peu compétent mais fortement impliqué dans son travail ou un individu « talentueux » mais peu motivé. L'intervention de l'État va de fait influencer l'offre de travail des individus, et la puissance publique va faire en sorte que son objectif de justice soit atteint de la façon la moins inefficace possible en tenant compte dans la détermination de son barème fiscal des effets de son action sur les comportements.

Afin d'illustrer l'impact que peut avoir la fiscalité ou les transferts sociaux sur l'**offre de travail**, considérons les effets d'une augmentation des aides sociales sur cette dernière (voir le [chapitre 6](#)). Supposons que l'État verse de façon hebdomadaire 150 euros aux individus quel que soit le nombre d'heures travaillées et que ces derniers puissent prétendre à un taux de salaire net de 10 euros de l'heure en travaillant. Que se passe-t-il si l'État décide d'augmenter cette aide à 200 euros ? Nous mobilisons le cadre théorique présenté au chapitre 6 pour répondre à cette question. L'augmentation de l'aide décale vers le haut la contrainte budgétaire subie par le travailleur.

Graphique 7.1 : Les effets d'une augmentation des aides sociales

sur l'offre de travail



Nous pouvons voir sur le [graphique 7.1](#) que l'augmentation de l'aide sociale entraîne la baisse de l'offre de travail de ces personnes, l'arbitrage travail-loisir passant du point A au point B. Sans aide, elles travaillaient 22 heures par semaine (48-26) avec l'aide, elles ne travaillent plus que 19 heures (48-29). Leur revenu est plus important avec l'aide que sans l'aide. L'aide a contribué à augmenter le revenu de ces personnes au prix d'une diminution de l'offre de travail. Dans cette logique, certains individus risquent de sortir complètement du marché du travail si l'aide est conséquente, et si son montant diminue s'ils travaillent : c'est le problème des **trappes à inactivité**. L'impact des aides sur l'offre de travail dépend de

la sensibilité de cette dernière au revenu.

Les individus fortement taxés sont également incités à travailler moins car plus ils travailleront plus ils seront taxés.

Si l'État cherche à atteindre son objectif de justice de la façon la moins distorsive possible, et optimise les choix fiscaux en fonction des contraintes économiques existantes, c'est-à-dire en tenant compte de la réaction des agents, on dit qu'il poursuit un **optimum de second rang**, c'est-à-dire une situation qui n'est pas optimale au sens de Pareto mais qui est la moins inefficace possible. Le système fiscal à privilégier pour atteindre l'objectif fixé doit maximiser le bien-être social, sous contrainte d'équilibre budgétaire, les prélèvements devant compenser les transferts. Puisque l'offre de travail dépend du système fiscal, il faut intégrer ces effets. Sur le marché du travail, les agents ont deux possibilités : soit ils changent leur décision de participation au marché du travail, ce qu'on appelle le choix au niveau de la marge extensive entre activité et inactivité (voir le [chapitre 6](#)) ; soit ils modifient le nombre d'heures travaillées ou l'effort fourni pendant leur travail ce qui, à « talent » donné, va modifier leur revenu, c'est le choix au niveau de la marge intensive. En fonction de l'**élasticité de participation de l'offre de travail** – plus ou moins grande sensibilité de la décision de participation au marché du travail par rapport au revenu disponible – et de l'**élasticité-revenu de l'offre de travail** – plus ou moins grande sensibilité du choix du nombre d'heures travaillées par rapport au revenu disponible –, la réaction des agents au système fiscal est plus ou moins forte. Le barème fiscal optimal dépend ainsi de la valeur de ces élasticités.

La branche de l'économie publique qui s'intéresse à ce type de problématique est la théorie de la « **taxation optimale** ». Une taxe est dite optimale si elle permet d'atteindre l'objectif visé de la façon la moins distorsive possible. La théorie de la taxation optimale est issue des travaux fondateurs de J. Mirrlees (1971) qui a reçu le prix Nobel d'économie en 1996. Les économistes font généralement l'hypothèse que l'État doit maximiser une fonction de bien-être social utilitariste classique ou avec aversion pour les inégalités et c'est pourquoi nous allons présenter brièvement les résultats principaux issus de ces travaux⁶.

3.2 Les principaux résultats en taxation optimale

De nombreux travaux ont été menés en taxation optimale depuis ceux de J. Mirlees. Des modèles de plus en plus sophistiqués ont été développés pour tenir compte de différents paramètres (existence de frictions sur le marché du travail, existence d'incertitude sur le revenu, de possibilité de migration pour motif fiscal, etc.). Nous présentons ici une simple sélection de modèles consacrés aux revenus du travail, pour montrer la fécondité de cette approche théorique.

• La taxation des hauts revenus

La réduction des inégalités de revenu suppose de définir les principes de taxation des revenus les plus élevés, pour déterminer comment redistribuer la somme prélevée aux moins bien lotis. Des études empiriques diverses ont montré que l'élasticité de participation (décision au niveau de la marge extensive) des plus aisés est quasi nulle et que l'élasticité de l'offre de travail au salaire net d'impôt (décision au niveau de la marge intensive) varie en fonction des professions tout en restant relativement faible. Saez (2001) a calculé le montant du taux marginal optimal de taxation des hauts revenus pour différentes valeurs de l'élasticité de l'offre de travail et pour différents objectifs. Il montre que si le gouvernement cherche à maximiser la situation des plus pauvres, les taux marginaux optimaux sont plus élevés que s'il adopte un critère utilitariste classique de maximisation de la somme des utilités individuelles. Par ailleurs, moins les individus sont sensibles à la taxation, plus les taux marginaux optimaux sont élevés. Ainsi, quand le gouvernement valorise fortement l'égalité et que l'élasticité de l'offre du travail est faible, le taux marginal optimal dépasse les 70 %. En revanche, si l'élasticité est plus importante et que l'État adopte un critère utilitariste classique, le taux marginal optimal de taxation des hauts revenus se situe aux alentours de 50 %. Dans tous les cas, ces chiffres sont plus élevés que le taux marginal supérieur de taxation des hauts revenus pratiqué sur les revenus actuellement en France, soit un taux de 45 % sur la part du revenu annuel imposable par part supérieure à 152 000 euros en 2017.

Le modèle de E. Saez ne tient pas compte du risque de voir les hauts revenus quitter le territoire français et s'installer sur des terres fiscalement plus clémentes. L'émigration d'origine fiscale doit être prise en compte car elle représente une perte en termes de production et de recettes fiscales pour le pays. Il n'existe pas aujourd'hui d'évaluation empirique de l'importance du

phénomène mais l'accroissement de la mobilité du travail est un fait et il est raisonnable de penser que les motifs fiscaux puissent jouer un rôle, même faible, dans les décisions de mobilité. Les travaux d'A. Trannoy et L. Simula (2010) permettent d'intégrer l'effet de « l'évasion fiscale » sur la détermination des taux marginaux optimaux de taxation des hauts revenus (TRANNOY, 2010). Selon eux, si le gouvernement veut maximiser le bien-être des plus défavorisés, alors que les coûts de migration sont élevés et que l'individu a la possibilité de se localiser dans un paradis fiscal (pas de taxation des revenus), alors le taux marginal optimal de taxation des revenus se situe plutôt autour de 40 %.

La possibilité pour les États de maintenir un impôt progressif est donc fortement liée à la lutte contre les paradis fiscaux et à l'existence d'une certaine harmonisation fiscale entre les pays⁷.

• Quelle politique de soutien aux bas revenus ?

Une fois que des recettes fiscales ont été collectées, comment peuvent-elles être redistribuées aux personnes dont les revenus sont plus faibles, afin d'accroître leur revenu disponible ? Les politiques de soutien aux bas revenus sont très variées et ne passent pas toujours par un système de **transferts monétaires**. Nous nous focalisons uniquement sur cet aspect et examinons les différentes possibilités qui s'offrent à l'État.

En France, de nombreux **minima sociaux** (*welfare benefits*) ont été mis en place pour soutenir les bas revenus (minimum vieillesse, allocation parent isolé, RMI, etc.). Ces minima sociaux, s'ils ont permis de réduire de façon significative le taux de pauvreté en France, ont fait l'objet de vifs débats. Dans la plupart des cas il s'agit d'allocations différentielles, ce qui signifie que leur montant dépend de la différence entre le revenu de la personne et le revenu minimal désiré. Par exemple, un individu n'ayant aucun revenu et touchant une allocation de 400 euros par mois qui se mettrait à travailler quelques heures et gagnerait 100 euros par mois verrait le montant de son aide diminuer de 100 euros par mois de telle sorte que son revenu disponible resterait à 400 euros. C'est comme si son revenu était taxé au taux marginal de 100 %, ce qui décourage la reprise d'activité.

Le retour à l'emploi des bénéficiaires de minima sociaux est lié à de multiples facteurs qui ne dépendent pas des choix des individus. Cependant

un dispositif d'aide efficace doit faire en sorte de ne pas décourager le retour à l'emploi. Dès la création du RMI, ses concepteurs avaient anticipé ce problème et mis en place des mécanismes d'intéressement. Ces mécanismes, qui visent à autoriser un cumul partiel et temporaire des allocations avec des revenus du travail en cas de prise d'emploi, ont été par la suite redéfinis et développés. Restant néanmoins temporaires, beaucoup ont estimé que leurs effets incitatifs étaient trop faibles. C'est ainsi que depuis la fin des années 1990, de multiples réformes ont été mises en œuvre en France comme dans de nombreux pays développés pour valoriser l'accès à l'emploi, le principe étant que le travail doit être réellement rémunérateur.

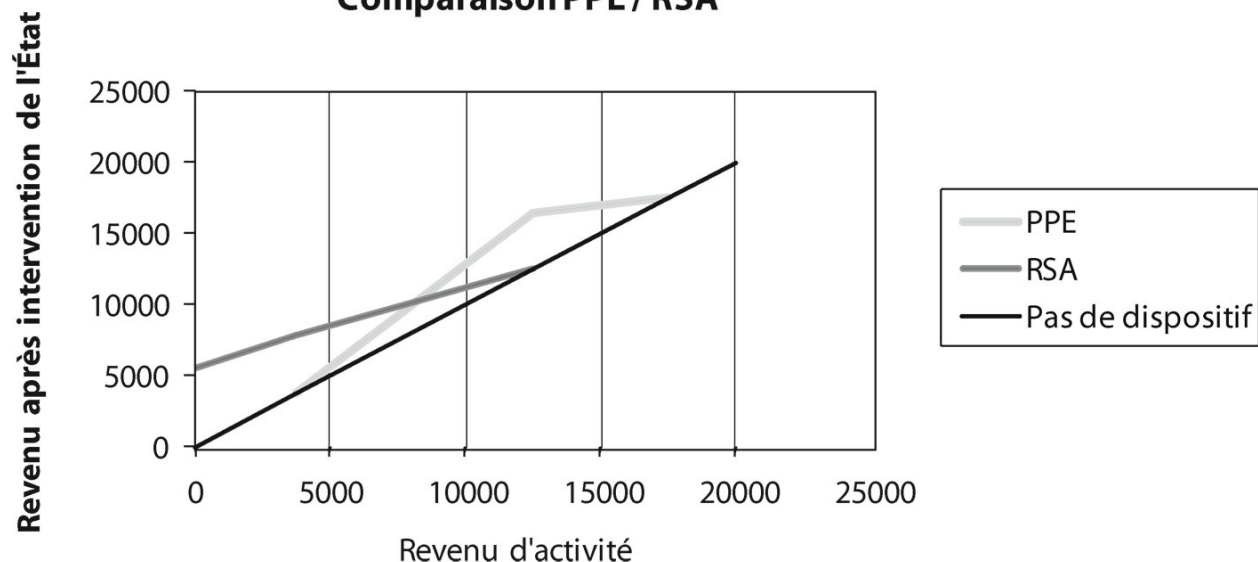
Plusieurs solutions sont envisageables. D'une part, des transferts monétaires réservés aux personnes à faibles revenus mais en emploi (*in-work benefits*), tels que la Prime pour l'emploi (PPE) instaurée entre 2001 et 2016. La PPE était une aide au retour à l'emploi et au maintien de l'activité professionnelle qui prenait la forme d'un crédit d'impôt. Elle n'était versée qu'aux travailleurs appartenant à un foyer fiscal disposant de revenus ne dépassant pas certaines limites. Aux États-Unis, un système de crédit d'impôt pour les travailleurs à bas revenus a été instauré dès 1975 le *Earned Income Tax Credit* (EITC), tout comme en Angleterre en 1999 avec le *Working Family Tax Credit* (WFTC). Un tel dispositif, s'il n'est couplé à aucun autre, revient à adopter des taux de taxation effective négatifs pour les bas revenus. Par exemple, une personne seule au chômage (revenu disponible de 0 €) qui trouve un travail à temps plein (salaire net de 1 000 €) recevrait une aide de 200 € du fait que le revenu de son travail est faible. Ce système revient à taxer son gain marginal (1 000 euros) au taux marginal de - 20 %.

Autre solution, revoir la façon dont sont calculés le montant des *welfare benefits* auquel un individu a droit. Ce sont des aides proposées à tout individu disposant d'un revenu inférieur à un certain montant et qui ne sont pas réservées aux personnes en emploi. L'allocation compensatrice de revenu, le Revenu Minimum d'Insertion (RMI) remplacé par le Revenu de solidarité active (RSA) sont de ce type. Le RSA, instauré le 1^{er} juin 2009, est une allocation qui se substitue au RMI, à l'Allocation parent isolé (API) et aux dispositifs d'intéressements associés. Lorsqu'un individu reprend un travail le montant des minima qu'il touche est diminué mais seulement d'une fraction des revenus du travail (38 %). Ce système revient à imposer un taux marginal de taxation des bas revenus de 38 %. Reprenons le cas de notre individu qui était au chômage (revenu de 0 €) et qui trouve un travail à temps

plein (salaire net mensuel de 1 000 €). Avec un système de type RSA, dans la situation initiale son revenu disponible est de 400 € alors qu'il est de $1\,000 + 400 - 0,38 \times 1\,000 = 1\,220$ quand il est en emploi. Le RSA est ainsi dégressif : il diminue jusqu'à s'éteindre à un certain niveau de revenu salarial, à savoir le SMIC à temps plein pour une personne seule. Le RSA bénéficie ainsi à la fois aux personnes en emploi mais aussi aux personnes au chômage. Le RSA socle désigne la somme dont un individu bénéficie pour compléter la différence entre son revenu et le revenu minimal défini par l'État et le RSA activité désigne la somme versée aux travailleurs en supplément du montant nécessaire pour combler l'écart entre leur revenu et le revenu minimum. Ce système peut être assimilé à un système « d'impôt négatif » : le montant des transferts diminue à chaque fois que le revenu de l'individu augmente mais de façon moins rapide de sorte que le revenu disponible des individus est toujours strictement croissant à mesure qu'ils travaillent plus.

Document 7.4 : Prime pour l'emploi et Revenu de solidarité active

Comparaison PPE / RSA



Source : graphique des auteurs, données : www.service-public.fr, 2011.

Nous pouvons modéliser de façon simple les effets de la prime pour l'emploi et du RSA sur le revenu des individus. Les revenus indiqués dans le [document 7.4](#) sont des revenus annuels et nous considérons le cas d'un individu vivant seul.

Sur le graphique, on lit en abscisse les revenus d'activité et en ordonnée le

revenu après intervention de l'État. En l'absence d'intervention, le revenu d'activité serait égal au revenu après intervention et on se situerait sur la bissectrice (revenu d'activité en abscisse = revenu après impôt en ordonnée). Les courbes correspondant au RSA et à la PPE indiquent l'ampleur correctrice de la redistribution. On observe que le montant du RSA diminue au fur et à mesure que le revenu de l'individu augmente pour devenir nul quand l'individu dispose d'un revenu équivalent à un SMIC à temps complet. Le montant de la Prime pour l'emploi est nul pour des revenus d'activité très faibles et augmente à mesure que le revenu croît jusqu'à un certain seuil. La PPE est maximale pour un individu payé au SMIC et diminue ensuite.

La théorie de la taxation optimale fournit un cadre précieux de réflexion pour nous aider à déterminer le montant des *welfare* et des *in-work benefits* à accorder. Le budget de l'État est en effet souvent limité et il est rarement possible de distribuer à la fois des *welfare benefits* importants (par exemple une aide inconditionnelle de 1 000 euros) tout en maintenant des incitations à la reprise d'activité, ce qui impliquerait de maintenir un niveau d'aide important pour les personnes en emploi. Quel est le meilleur choix entre des *welfare benefits* relativement importants couplés à des taux marginaux plus ou moins élevés au bas de la distribution (système de type impôt négatif) et des *in-work benefits* plus importants que les *welfare benefits*, et donc des taux marginaux négatifs au bas de la distribution (transferts de type EITC) ? La réponse à cette question dépend de la sensibilité du comportement des individus et de l'objectif du gouvernement. Plus les individus ont une élasticité de participation au marché du travail forte, plus les taux marginaux de taxation et les minima sociaux devront être faibles voire négatifs. Au contraire si les décisions de participation au marché du travail sont peu distordues par les transferts sociaux, des taux marginaux très élevés au bas de la distribution semblent adaptés (minima sociaux relativement importants et *in-work benefits* très faibles). Les études empiriques indiquent que l'élasticité de l'offre de travail au salaire net des bas revenus est faible (marge intensive) alors que l'élasticité de participation n'est pas négligeable (marge extensive). Saez (2002) montre que pour une élasticité de participation de 0,5 et une élasticité revenu de l'offre de travail de 0,25, ce qui correspond à des valeurs crédibles des élasticités, les taux marginaux optimaux pour les États-Unis ne sont pas négatifs pour les bas revenus : ils sont situés entre 12 % et 69 % selon la préférence pour la redistribution du gouvernement. Si l'on considère une préférence moyenne pour la redistribution, le taux marginal optimal est

de 37 % et le revenu annuel minimum garanti égal à 7 300 \$. Le point de sortie du dispositif d'aide est 12 900 \$, ce qui signifie qu'au-dessus de ce revenu annuel, plus aucune aide n'est versée. Le revenu disponible des personnes dont le revenu est inférieur à 12 900 \$ est alors déterminé de la manière suivante : $\text{Revenu disponible} = (1 - 0,37) \times \text{Revenu du travail} + 7\,300$. Cela suggère que le mécanisme du RSA tel qu'il est en vigueur en France aujourd'hui est assez proche des valeurs optimales, en faisant l'hypothèse que les paramètres français sont d'une valeur proche des paramètres américains.

Notons toutefois que depuis 2016, le RSA activité et la PPE ont fusionné pour former la prime d'activité. Elle concerne tous les actifs salariés (à temps plein ou à temps partiel) mais aussi les travailleurs indépendants ayant plus de 18 ans qui perçoivent une rémunération inférieure à un certain plafond. Pour un célibataire sans enfant, ce plafond est fixé à 1,3 Smic. La prime est versée tous les mois. Le montant de la prime d'activité est calculé en prenant en compte :

- un montant forfaitaire qui varie selon la composition du foyer et le nombre d'enfants à charge auquel il faut ajouter 62 % du montant des revenus professionnels du foyer. Cette somme pourra être bonifiée compte tenu des revenus professionnels de chaque travailleur ;
- les ressources prises en compte du foyer.

La formule de calcul est la suivante : $\text{montant forfaitaire (= montant variable + 62 \% des revenus professionnels + bonifications individuelles)} - \text{les ressources prises en compte du foyer}$. Le montant de base de la prime d'activité pour une personne seule équivaut à 526,25.

Notes

1. J. Rawls, *A Theory of Justice*, Harvard University Press, 1971, traduction au Seuil, coll. « Points ».
2. Voir notamment « Les hauts revenus en France (1998-2006) : une explosion des inégalités ? », Document de travail PSE, 2007.
3. Source : « Emploi et chômage des personnes handicapées », *Synthèse Stat.* n° 1, Dares, novembre 2012.
4. On parle aussi parfois d'approche par les capacités, le terme capacités étant un néologisme issu du terme anglo-saxon *capabilities*. Les termes de capacités et de capabilités sont ainsi synonymes dans l'approche de Sen.
5. C. Landais, T. Piketty et E. Saez, *Pour une révolution fiscale*, Le Seuil, coll. « La République des idées », 2011.
6. Force est de constater que les développements récents en économie normative n'ont pas encore eu beaucoup d'applications en économie publique. L'approche « bien-être » reste dominante. Certains auteurs s'y attèlent néanmoins et le lecteur curieux pourra découvrir ces travaux précurseurs en lisant les articles de L. Jacquet et D. Van de Gaer (2009) ou de M. Fleurbaey et F. Maniquet (1999).
7. Sur la question des paradis fiscaux et des conséquences sur la capacité à lever l'impôt, voir G. Zucman, *La richesse cachée des nations*, Le Seuil, coll. « La République des idées », 2013.

Bibliographie

Chapitre 1

- BÉNASSY-QUÉRÉ A., COEURÉ B., JACQUET P., PISANI-FERRY J. (2017), *Politique économique*, chapitre 7 : « La politique fiscale », De Boeck.
- FACK G. (2005), « Pourquoi les ménages à bas revenus paient-ils des loyers de plus en plus élevés ? L'incidence des aides au logement en France (1973-2002) », *Économie et statistique*, n° 381-382.
- INSEE, Les conditions de logement en France, coll. « INSEE Références », édition 2017.
- LAFERRÈRE A., LE BLANC D. (2002), « Comment les aides au logement affectent-elles les loyers ? », *Économie et statistique*, n° 351.
- MISTRAL J., PLAGNOL V. (2008), *Loger les classes moyennes – La demande, l'offre et l'équilibre du marché du logement*, Rapport du Conseil d'Analyse Économique, La Documentation française.
- Regards Croisés sur l'Économie (2011), *Pour sortir de la crise du logement*, n° 9, La Découverte.
- TRANNOY A., WASMER E. (2013), *La politique du logement locatif*, Les notes du conseil d'analyse économique, n° 10.
- WASMER E. (2007), « Analyse économique du marché du logement locatif », *Revue économique*, vol. 58, 2007/6.

Chapitre 2

- BARBERIS I. et POIRSON M. (2013), *L'économie du spectacle vivant*, PUF, coll. « Que-sais-je ? ».
- BENHAMOU F. (2017), *L'économie de la culture*, La Découverte, coll. « Repères ».
- BENHAMOU F., THESMAR D. (2011), *Valoriser le patrimoine culturel de la France*, Rapport du Conseil d'Analyse Économique, La Documentation française.
- Cahiers Français (2014), *Économie et politique de la culture*, La Documentation française, n° 382, septembre-octobre 2014.
- DAN A. (2008), *C'est (vraiment ?) moi qui décide*, Flammarion.
- GREFFE X. (2010), « L'économie de la culture est-elle particulière ? », *Revue d'économie politique*, vol. 120, 2010/1.
- GUILLON O. (2009), « La fidélité des consommateurs de biens culturels : une analyse empirique des comportements d'assiduité et d'adhésion au Parc de la Villette », *Revue d'économie politique*, vol. 119, 2009/2.
- KARPIK L. (2007), *L'économie des singularités*, Gallimard.
- LÉVY-GARBOUA L., MONTMARQUETTE C. (2010), « The Demand for the Arts », *Handbook of Cultural Economics*, UK, Edward Elgar Publishing.
- MAIRESSE F. et ROCHELANDET (2015), *Économie des arts et de la culture*, Armand Colin, coll. « U ».
- MOUREAU N., VIDAL M. (2009), « Quand préférences révélées et préférences déclarées s'opposent : le téléspectateur, un cas paradoxal pour l'économiste », *Revue française de socio-économie*, La Découverte, n° 4.
- MOUREAU N., SAGOT-DUVAUROY D. (2010), *Le marché de l'art contemporain*, La Découverte, coll. « Repères ».
- PERONA M., POUYET J. (2010), *Le prix unique du livre à l'heure du numérique*, Cepremap, Éditions Rue d'Ulm.
- RAGOT S. (2010), « Le cinéma est-il devenu un loisir de luxe ? Une étude économétrique de la demande française », *Revue d'économie politique*, vol. 120, 2010/1.
- STEINER P. (coord.) (2008), « Regards croisés sur *L'économie des singularités* de Lucien Karpik », *Revue française de sociologie*, 2008/2, vol. 49.

THALER R., SUNSTEIN C. (2010), *Nudge – La méthode douce pour inspirer la bonne décision*, Vuibert.

URRUTIAGUER D., HENRY P., DUCHÊNE C. (2012), « Territoires et ressources des compagnies en France », *Culture-Études*, n° 2012-1, DEPS-Ministère de la Culture et de la Communication.

Chapitre 3

- AGHION P., ROULET A. (2011), *Repenser l'État – Pour une social-démocratie de l'innovation*, Le Seuil, coll. « La République des idées ».
- AGRESTE (2012), « Un million d'actifs permanents dans l'agriculture », *Agreste primeur*, n° 276, janvier 2012.
- BAUDRY B., CHASSAGNON V. (2014), *Les théories économiques de l'entreprise*, La Découverte, coll. « Repères ».
- BESSIÈRE C. *et alii* (2011), « Les agriculteurs et leur patrimoine : des indépendants comme les autres ? », *Économie & Statistique*, n° 444-445.
- BUREAU J.-C. (2007), *La politique agricole commune*, La Découverte, coll. « Repères ».
- CAVAILHÈS J., HILAL M., WAVRESKY P. (2011), « L'influence urbaine sur le prix des terres agricoles et ses conséquences pour l'agriculture », *Économie & Statistique*, n° 444-445.
- CHAUDEY M. (2014) *Analyse économique de la firme*, Armand Colin, coll. « Cursus ».
- CHEVALIER B. (2007), « Les agriculteurs recourent de plus en plus à des prestataires de services », *INSEE Première*, n° 1160, octobre 2007.
- COURLEUX F., (2011), « Augmentation de la part des terres agricoles en location : échec ou réussite de la politique foncière ? », *Économie & Statistiques*, n° 444-445.
- CŒURDEROY R. (2007), *Économie des organisations*, Vuibert.
- DEDIEU M.-F. (2011), « Les coopératives agricoles : un modèle d'organisation économique des producteurs », *Analyse – Centre d'études et de prospective*, n° 36, novembre 2011.
- DESRIERS M. *et alii* (2009), « Le poids des aides directes dans le revenu des exploitations agricoles », in *Les revenus d'activité des indépendants, édition 2009*, INSEE.
- EYMARD-DUVERNAY F. (2004), *Économie politique de l'entreprise*, La Découverte, coll. « Repères ».
- KRUGMAN P., WELLS R. (2013), *Microéconomie*, De Boeck, parties 4 et 6.
- INSEE (2007), *L'agriculture, nouveaux défis*, Éditions de l'INSEE.
- OCDE (2011), *Renforcer la productivité et la compétitivité dans le secteur*

agricole, Éditions OCDE.

SEN A. (2003), « Des famines et des libertés », in *L'économie est une science morale*, La Découverte.

Chapitre 4

- AKERLOF G. (1970), « The Market for “Lemons” : Quality Uncertainty and the Market Mechanism », in *Quarterly Journal of Economics*, traduction française dans BACACHE-BEAUVALLET M. et MONTOUSSÉ M. (2003), *Textes fondateurs en sciences économiques depuis 1970*, Bréal.
- AKERLOF G. et SHILLER R. (2009), *Les esprits animaux – Comment les forces psychologiques mènent la finance et l’économie*, Pearson.
- AKERLOF G. et SHILLER R. (2016), *Marchés de dupe – L’économie du mensonge et de la manipulation*, Éditions Odile Jacob.
- BEITONE A. et RODRIGUES C. (2017), *Économie monétaire – théories et politiques*, Armand Colin.
- COMBE E. (2016), *La politique de concurrence*, La Découverte, coll. « Repères ».
- COUPPEY-SOUBEYRAN J. (2015), *Monnaie, banques, finance*, PUF, coll. « Licence ».
- CROZET M., LAFOURCADE M. (2009), *La nouvelle économie géographique*, La Découverte, coll. « Repères ».
- EBER N. (2013), *Théorie des jeux*, Dunod, coll. « Topos ».
- GIRAUD G. (2009), *La théorie des jeux*, Flammarion, coll. « Champs ».
- MARTEAU D. (2016), *Les marchés de capitaux*, Armand Colin, coll. « Cursus ».
- MISHKIN F. S. (2013), *Monnaie, banques et marchés financiers*, Pearson Éducation.
- SCIALOM L. (2013), *Économie bancaire*, La Découverte, coll. « Repères ».

Chapitre 5

- BAUBY P., SIMILIE M. (2007), « Les services publics locaux : l'exemple de l'eau », *in* Cahiers français, *Les services publics*, La Documentation française, n° 339, juillet-août.
- CGDP – Commissariat général au développement durable (2010), « Services d'eau et d'assainissement : une inflexion des tendances ? », *Le point sur*, n° 67, décembre.
- BONTEMS P., ROTILLON G. (2013), *Économie de l'environnement*, La Découverte, coll. « Repères ».
- Conseil d'État (2010), *L'eau et son droit*, Rapport public, La Documentation française.
- CORNU M., ORSI F., ROCHFELD J. (2017), *Dictionnaire des biens communs*, PUF, Coll. « Quadrige ».
- GRIFFIN RONALD C., (2006) *Water Resource Economics*, MIT Press.
- JOB L. (2010), « Mission d'intérêt économique général et organisation industrielle dans l'Union européenne : étude du secteur français de l'eau potable et de l'assainissement », *in* POTVIN-SOLIS L. (éd.), *La libéralisation des services d'intérêt économique général en réseau en Europe*, Bruxelles, Bruylant, 2010.
- KONSTANTINOS C. (2006), « Brève histoire des compteurs d'eau à Paris 1880-1930 », *Terrains & travaux*, ENS Cachan, n° 11.
- LANDIER A., THESMAR D. (2010), *La société translucide : pour en finir avec le mythe de l'État bienveillant*, Fayard.
- LAURENT E., LE CACHEUX J. (2015), *Économie de l'environnement et économie écologique*, Armand Colin, coll. « Cursus ».
- LÉVÊQUE F. (2004), *Économie de la réglementation*, La Découverte, coll. « Repères ».
- MONTGINOUL M., RINAUDO J.-D. (2009), « Quels instruments pour gérer les prélèvements en eau souterraine ? Le cas du Roussillon », *Économie rurale*, n° 310, p. 40-56, 2009.
- OSTROM E. (2010), *Gouvernance des biens communs : pour une nouvelle approche des ressources naturelles*, De Boeck, éd. originale 1990.
- Regards Croisés sur l'Économie (2009), *Les économistes peuvent-ils sauver*

la planète ?, n° 6, novembre 2009.

ROTILLON G. (2010), *Économie des ressources naturelles*, La Découverte, coll. « Repères ».

TERREAUX J.-P. (éd.), 2005, *Économie des équipements pour l'eau et l'environnement*, Cemagref éditions.

TIROLE J. (2016), *Économie du bien commun*, PUF.

VALLÉE A. (2011), *Économie de l'environnement*, Points, coll. « Points Essais ».

Chapitre 6

- ACEMOGLU D., SHIMER R. (2000), « Productivity gains from unemployment insurance », *European Economic Review*, 44, p. 115-1125.
- ASKENAZY P., ERHEL C. (2017), *Qualité de l'emploi et productivité*, Éditions Rue d'Ulm, coll. du Cepremap.
- AMABLE B. (2005), *Les cinq capitalismes*, Le Seuil, coll. « Économie humaine ».
- ANNE D., L'HORTY P. (2013), *Économie de l'emploi et du chômage*, Armand Colin, coll. « Cursus ».
- BATARD P.-E., SAILLARD E. (2011), « Le chômage des jeunes : quel diagnostic ? », *Trésor-Éco*, n° 92.
- BAUDELLOT C., GOLLAC M. (1993) « Salaires et conditions de travail », *Économie et statistique*, n° 265, p. 65-84.
- BENHENDA A., ZUCMAN G. (dir.) (2013), « L'adieu au chômage », *Regards croisés sur l'économie*, n.° 13, La Découverte.
- CAHUC P. (2001) « Pourquoi y a-t-il des différences de salaires ? », *Reflets et perspectives de la vie économique*, 1, t. XL, p. 13-24.
- CAHUC P., KRAMARZ F. (2005), *De la précarité à la mobilité : vers une Sécurité sociale professionnelle*, Rapport au ministre de l'Économie, des Finances et de l'Industrie et au ministre de l'Emploi, du Travail et de la Cohésion Sociale, La Documentation française.
- CAHUC P., ZYLBERBERG A. (2015), *Les ennemis de l'emploi. Le chômage, fatalité ou nécessité ?*, Flammarion, coll. « Champs ».
- CARD D., KRUEGER A. (1994), « Minimum Wages and Employment : A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania », *American Economic Review*.
- CRÉPON B., DUFLO E., GURGAND M., RATHELOT R., ZAMORA P. (2011), « L'accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement : les enseignements d'une évaluation », *DARES Analyses*, n° 94.
- DOERINGER P., PIORE M. (1971), *Internal Labor Markets and Manpower Analysis*, Heath Lexington Books.
- DURU-BELLAT M., VAN ZANTEN A. (dir.), (2009), *Sociologie du système éducatif*

– les inégalités scolaires, PUF.

GODECHOT O. (2011), « Le capital humain et les incitations sont-ils les deux mamelles des salaires dans la finance ? », *Revue d'économie financière*, n° 104, p. 145-164.

GODECHOT O., GRUGAND M. (2000), « Quand les salariés jugent leur salaire », *Économie et statistique*, n° 331.

JOSEPH O., ROUAUD P. (2014), *Quand l'École est finie. Premiers pas dans la vie active*, éditions du Céreq.

LESUEUR J.-Y., SABATIER M. (2008), *Microéconomie de l'emploi*, De Boeck.

MORTENSEN D., PISSARIDES C. (1994), « Job Creation and Job Destruction in the Theory of Unemployment », *Review of Economic Studies*, vol. 61, n° 3, p. 397-415.

PICART C. (2008), « Flux d'emploi et de main-d'œuvre en France : un réexamen », *Économie et statistiques*, n° 412.

POLI A., LEFRANÇOIS C., CARADEC V. (2009), « Les discriminations liées à l'âge, entre exclusion et protection », *Mouvements*.

ROSEN S., (1986), « The theory of Equalizing Differences », in ASHENFELTER O. et LAYARD R., *Handbook of Labor Economics*, vol. 1.

VILLEVAL M.-C. (2016), *L'économie comportementale du marché du travail*, Presses de Sciences Po.

Chapitre 7

- AKERLOF G. (1978), « The Economics of “Tagging” as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning », *American Economic Review*, vol. 68, n° 1, p. 8-19.
- ARNSPERGER C., VAN PARIJS P. (2003), *Éthique économique et sociale*, La Découverte, coll. « Repères ».
- BOZIO A., GRENET J., (2017), *Économie des politiques publiques*, La Découverte, coll. « Repères ».
- FLEURBAEY M. (1996), *Théories économiques de la justice*, Economica, coll. « Économie et statistiques avancées ».
- FLEURBAEY M., MANIQUET F. (1999), « Optimal income taxation : An ordinal approach », THEMA Working Papers 99-43, THEMA (Théorie Économique, Modélisation et Applications), université de Cergy-Pontoise.
- JACQUET L., VAN DE GAER D. (2009), « A comparison of optimal tax policies when compensation or responsibility matter », Working Papers of Faculty of Economics and Business Administration, université de Gand, Belgique, 09/615.
- LE CLAINCHE C., REBOUD V. (2009), « Réduire les inégalités : pourquoi et comment ? », *Cahiers français*, n° 351, p. 71-78.
- MAGUAIN D. (2002), « Les théories de la justice distributive post-rawlsiennes. Une revue de la littérature », *Revue économique*, vol. 53, n° 2, p. 165-199.
- MIRRLEES J. (1971), « An Exploration in the Theory of Optimum Income Taxation », *Review of Economic Studies*, vol. 38, p. 175-208.
- SAEZ E. (2001), « Using Elasticities to Derive Optimal Income Tax Rates », *Review of Economic Studies*, n° 68, p. 205-229.
- SAEZ E. (2002), « Optimal Income Transfer Programs : Intensive Versus Extensive Labor Supply Responses », *Quarterly Journal of Economics*, vol. 117, n° 1, 1039-1073.
- SALANIÉ B. (2011), *The economics of taxation*, 2nd Revised edition, MIT Press (première édition disponible en français, *Théorie économique de la fiscalité*, 2002, Economica, coll. « Économie et statistiques avancées »).

- SIMULA L., TRANNOY A. (2010), « Optimal Income Tax under the Threat of Migration by Top-Income Earners », *Journal of Public Economics*, n° 94, p. 163-173.
- TRANNOY A. (2009), « RMI, RSA et cycle économique : faut-il déroger à la règle des 38 % ? », *Risques*, n° 77.
- TRANNOY A. (2010), « Peut-on encore taxer les hauts revenus ? », *Regards Croisés sur l'Économie*, n° 8, p. 158-165.
- VILLEVAL M-C. (2016), *L'économie comportementale du marché du travail*, Presses de Sciences po.

Index

accords horizontaux [1](#)
accords verticaux [1](#)
actifs spécifiques [1](#)
aléa moral [1](#), [2](#), [3](#)
allocation des ressources [1](#)
antisélection [1](#)
appariement [1](#)
asymétrie d'information [1](#), [2](#), [3](#)
atomicité [1](#)
autorités de la concurrence [1](#)
barrières à l'entrée [1](#), [2](#)
bénéfice externe [1](#)
bénéfice marginal social [1](#)
bénéfice privé [1](#)
bénéfice social [1](#)
biais cognitif [1](#)
biais d'ancrage [1](#)
biais systématique [1](#)
bien atypique [1](#)
bien de club [1](#)
bien collectif [1](#)

bien commun [1](#)
bien complémentaire [1](#)
bien de consommation [1](#)
bien de croyance [1](#)
bien d'expérience [1](#)
bien inférieur [1](#)
bien normal [1](#)
bien patrimonial [1](#)
bien privatif [1](#)
bien public local [1](#)
bien public mondial [1](#)
bien de recherche [1](#)
bien substituable [1](#)
bien supérieur [1](#), [2](#)
bien tutélaire [1](#), [2](#)
brevet [1](#)
budget des ménages [1](#), [2](#), [3](#), [4](#)
capital humain [1](#)
cartel [1](#)
chemin d'expansion de long terme [1](#)
choc de demande [1](#)
choc d'offre [1](#)
chômage frictionnel [1](#)
cohérence arbitraire [1](#)
coin socio-fiscal [1](#)

collusion [1](#)

combinaison productive [1](#)

commissaire-priseur walrasien [1](#)

concurrence imparfaite [1](#), [2](#)

concurrence monopolistique [1](#)

concurrence pure et parfaite [1](#), [2](#), [3](#)

consommateur [1](#)

contestabilité du marché [1](#)

contrainte budgétaire [1](#), [2](#)

contrat complet ou incomplet [1](#)

contrat incitatif [1](#)

convexité [1](#), [2](#), [3](#)

coopération [1](#)

coordination par la hiérarchie [1](#)

coordination par le marché [1](#)

courbe de beveridge [1](#)

courbe de consommation-revenu [1](#)

courbe de coût marginal [1](#), [2](#), [3](#)

courbe de coût moyen [1](#), [2](#)

courbe de demande [1](#), [2](#)

courbe d'offre [1](#)

courbe d'offre agrégée [1](#)

courbe d'offre de travail agrégée [1](#)

courbe d'offre de travail individuelle [1](#)

coût de catalogue [1](#)

coût de dommage [1](#)
coût d'opportunité [1](#), [2](#), [3](#)
coût d'usage du capital [1](#)
coût explicite [1](#)
coût externe [1](#)
coût fixe [1](#)
coût fixe moyen [1](#)
coût implicite [1](#)
coût irrécupérable [1](#)
coût privé [1](#)
coût de production [1](#), [2](#)
coût social [1](#), [2](#)
coût total [1](#), [2](#)
coût de transaction [1](#)
coût du travail [1](#)
coût variable [1](#)
coût variable moyen [1](#)
cycle de vie [1](#)
défaillances du marché [1](#)
délégation de service public [1](#)
demande agrégée [1](#)
demande captive [1](#)
demande de travail [1](#)
différenciation horizontale [1](#)
différenciation des produits [1](#), [2](#)

différenciation verticale [1](#)
dilemme du prisonnier [1](#)
discrimination par les prix [1](#)
discrimination statistique [1](#)
disposition à payer [1](#)
distorsion [1](#)
droite de budget [1](#)
droite d'isocoût [1](#)
duopole [1](#)
duopole de cournot [1](#)
école des choix publics [1](#)
économie du bien-être [1](#)
économie comportementale [1](#)
économie des conventions [1](#)
économie géographique [1](#)
économie normative [1](#)
économies d'échelle [1](#), [2](#), [3](#), [4](#)
économies de gamme [1](#)
éducation [1](#)
effet de dotation [1](#)
effet revenu [1](#), [2](#)
effet de substitution [1](#), [2](#)
efficacité allocative [1](#)
élasticité de la participation au marché du travail [1](#)
élasticité des heures travaillées au taux de salaire [1](#), [2](#)

élasticité de l'offre de travail au taux de salaire [1](#)

élasticité de participation de l'offre de travail [1](#)

élasticité-prix [1](#), [2](#)

élasticité-prix de court terme et de long terme [1](#)

élasticité-prix croisée de la demande [1](#)

élasticité-prix de la demande [1](#), [2](#)

élasticité-revenu de la demande [1](#)

élasticité-revenu de l'offre de travail [1](#)

entente illicite [1](#)

épargne [1](#)

équilibre à long terme [1](#)

équilibre du consommateur [1](#)

équilibre général [1](#)

équilibre de Nash [1](#)

évaluation contingente [1](#)

excluablilité [1](#)

exclusion du marché [1](#)

expérience contrôlée [1](#)

expérience de laboratoire [1](#)

expérience naturelle [1](#)

externalisation [1](#)

externalité [1](#), [2](#)

externalité positive [1](#)

externalité de réseau [1](#), [2](#), [3](#)

facteur complémentaire [1](#)

facteur de production [1](#), [2](#)
facteur substituable [1](#)
finance comportementale [1](#)
flexisécurité [1](#)
fluidité [1](#)
fonction de bien-être social [1](#)
fonction d'utilité [1](#)
fonction paramétrique des prix [1](#)
fonction de production [1](#)
frontière des possibilités de production [1](#)
frontière technologique [1](#)
groupe d'entreprises [1](#)
homogénéité [1](#)
horizon temporel [1](#)
impôt forfaitaire [1](#)
incidence fiscale [1](#), [2](#)
incitation [1](#), [2](#), [3](#), [4](#), [5](#)
incohérence temporelle [1](#)
indice d'Herfindahl-Hirschmann [1](#), [2](#)
industries de réseau [1](#)
innovation [1](#), [2](#)
insatiabilité [1](#), [2](#)
institution [1](#), [2](#), [3](#)
isoquante [1](#), [2](#)
jeux coopératifs [1](#)

jeux dynamiques [1](#)
jeux non coopératifs [1](#)
loi de Baumol [1](#)
loi de l'offre et de la demande [1](#)
loi du prix unique [1](#)
loi des rendements décroissants [1](#)
loi des rendements factoriels décroissants [1](#)
loisir [1](#)
marché concurrentiel [1](#)
marché interbancaire [1](#)
marché pertinent [1](#)
marché primaire du travail [1](#)
marché de quotas [1](#)
marché secondaire du travail [1](#)
marge extensive [1](#), [2](#)
marge intensive [1](#)
minima sociaux [1](#)
mobilité parfaite des facteurs de production [1](#)
modèle [1](#)
monopole [1](#)
monopole à frange [1](#)
monopole discriminant [1](#)
monopole légal [1](#)
monopole local [1](#)
monopole naturel [1](#), [2](#)

monopsone [1](#), [2](#)
négociation [1](#)
négociation salariale [1](#)
niveau optimal de pollution [1](#)
nouvelle économie institutionnelle [1](#)
offre de long terme [1](#)
offre de travail [1](#)
oligopole [1](#), [2](#), [3](#)
opportunisme [1](#), [2](#)
optimum de Pareto [1](#)
optimum de second rang [1](#)
optimum social [1](#)
panier de biens [1](#)
paradoxe de Bertrand [1](#)
passager clandestin [1](#)
paternalisme libertarien [1](#)
patrimoine [1](#), [2](#)
perte sèche [1](#), [2](#)
politique de la concurrence [1](#)
pouvoir d'achat [1](#)
pouvoir de marché [1](#), [2](#)
préférences [1](#)
préférences révélées [1](#)
preneur de prix [1](#), [2](#), [3](#)
principe du pollueur-payeur [1](#)

prix d'équilibre [1](#)
prix garanti [1](#)
prix hédonique [1](#)
prix plafond [1](#), [2](#), [3](#)
prix plancher [1](#)
production publique [1](#)
productivité [1](#)
productivité globale des facteurs [1](#)
productivité marginale d'un facteur [1](#)
productivité marginale du travail [1](#), [2](#)
profit [1](#)
protection de l'emploi [1](#)
quota [1](#), [2](#)
quota de prélèvements [1](#)
rationalité [1](#)
rationalité limitée [1](#)
rationnement [1](#)
rationnement du crédit [1](#)
recette marginale [1](#)
recette moyenne [1](#)
recette totale [1](#)
régie [1](#)
règle de consommation optimale [1](#)
réglementation [1](#), [2](#), [3](#), [4](#), [5](#)
régression [1](#)

régulateur public [1](#), [2](#)
relation d'agence [1](#)
rendement factoriel [1](#)
rendements d'échelle [1](#)
revenu arbitrage [1](#)
revenu disponible [1](#)
revenu permanent [1](#)
revenu primaire [1](#)
rivalité [1](#)
routine [1](#)
salaire [1](#)
salaire d'efficience [1](#)
salaire de réservation [1](#)
salaire de réserve [1](#)
sélection adverse [1](#), [2](#)
service public [1](#)
seuil de fermeture [1](#)
seuil de rentabilité [1](#)
soutenabilité faible [1](#)
soutenabilité forte [1](#)
spécificité des actifs [1](#)
subvention [1](#), [2](#)
surplus [1](#)
surplus des consommateurs [1](#)
surplus des producteurs [1](#)

surplus total [1](#)
talent [1](#)
tarification binôme [1](#)
taux de chômage [1](#)
taux d'activité [1](#)
taux d'emploi [1](#)
taux de dépréciation [1](#), [2](#)
taux de flexion [1](#)
taux marginal de substitution [1](#), [2](#)
taux marginal de substitution technique [1](#)
taux de rotation [1](#)
taux de salaire [1](#)
taux de salaire superbrut [1](#)
taxation [1](#), [2](#), [3](#)
taxation optimale [1](#)
taxe pigouvienne [1](#)
théorème de Coase [1](#)
théorème de l'économie du bien-être [1](#), [2](#)
théorie des différences compensatrices [1](#)
théorie évolutionniste [1](#)
théorie des incitations [1](#), [2](#)
théorie des jeux [1](#), [2](#)
tragédie des communs [1](#)
transfert monétaire [1](#)
transitivité des préférences [1](#)

transparence [1](#)

trappe à inactivité [1](#)

utilitarisme [1](#)

utilité marginale [1](#)

valeur cardinale [1](#)

valeur économique totale [1](#)

valeur ordinale [1](#)

valeur tutélaire du carbone [1](#)

viscosité des prix [1](#)

Listes des documents et graphiques

[Document 1.1 : Évolution du prix des logements et du revenu disponible des ménages](#)

[Graphique 1.1 : Les courbes d'offre et de demande](#)

[Graphique 1.2 : Un choc de demande positif](#)

[Graphique 1.3 : Le surplus des consommateurs et des producteurs](#)

[Graphique 1.4 : Surplus total et élasticité de l'offre et de la demande](#)

[Graphique 1.5 : L'exclusion du marché](#)

[Graphique 1.6 : L'effet d'une aide en équilibre partiel](#)

[Graphique 1.7 : L'effet d'une aide en équilibre général](#)

[Graphique 1.8 : Équilibre à court terme et à long terme sur le marché du logement](#)

[Graphique 1.9 : Les effets d'une subvention d'un montant fixe de la demande](#)

[Graphique 1.10 : Les effets d'une subvention proportionnelle de la demande](#)

[Graphique 1.11 : Les effets d'une subvention d'un montant fixe de l'offre](#)

[Graphique 1.12 : Les effets d'une taxe proportionnelle](#)

[Graphique 1.13 : Incidence fiscale et élasticité](#)

[Graphique 1.14 : Les effets du contrôle des prix](#)

[Graphique 1.15 : Les effets des quotas](#)

[Document 2.1 : Part au sein des dépenses culturelles et de loisirs en 2000 \(gris foncé\) et 2016 \(gris clair\)](#)

[Graphique 2.1 : La contrainte budgétaire](#)

[Graphique 2.2 : Effet d'une baisse du prix du cinéma](#)

[Graphique 2.3 : Effet d'une hausse du budget](#)

[Graphique 2.4 : Les préférences du consommateur](#)

[Graphique 2.5 : Une absence de goût pour le cinéma](#)

[Graphique 2.6 : Le taux marginal de substitution](#)

[Graphique 2.7 : L'équilibre du consommateur](#)

[Graphique 2.8 : Courbe de consommation-revenu et hausse du budget](#)

[Graphique 2.9 : L'effet d'une division par deux du prix des places de cinéma](#)

[Graphique 2.10 : Effet substitution \(ES\) et effet revenu \(ER\)](#)

[Document 2.2 : Revenu, consommation et patrimoine dans le modèle du cycle de vie](#)

[Document 2.3 : Estimation statistique de la courbe de demande de cinéma](#)

[Graphique 2.11 : L'élasticité-prix varie le long de la courbe de demande](#)

[Document 2.4 : Les différentes élasticités de la demande](#)

[Document 2.5 : Les caractéristiques des biens selon leurs élasticités](#)

[Document 3.1 : Coûts moyens de viande bovine en 2008](#)

[Graphique 3.1 : Coût total, coût moyen et coût marginal de court terme](#)

[Graphique 3.2 : Les courbes de coût moyen, coût fixe moyen et coût variable moyen](#)

[Graphique 3.3 : Fonctions de production à facteurs substituables et complémentaires](#)

[Graphique 3.4 : le choix de la combinaison productive d'un producteur de céréales](#)

[Graphique 3.5 : Les rendements d'échelle croissants](#)

[Graphique 3.6 : Le profit du producteur pour une quantité donnée](#)

[Graphique 3.7 : Seuil de rentabilité et de fermeture](#)

[Graphique 3.8 : La maximisation du profit](#)

[Graphique 3.9 : L'équilibre du marché européen sans subventions](#)

[Graphique 3.10 : Innovations et frontière des possibilités de production](#)

[Graphique 4.1 : Le comportement du monopole](#)

[Graphique 4.2 : Monopole *versus* concurrence](#)

[Graphique 4.3 : La perte de bien-être liée au manque de concurrence](#)

[Graphique 4.4 : L'effet d'un prix plafond](#)

[Graphique 4.5 : Le monopole naturel](#)

[Document 4.1 : Le résultat net des six principaux groupes bancaires français en 2016 et leur classement parmi les banques européennes](#)

[Graphique 4.6 : Le duopole de Cournot](#)

[Document 4.2 : Le dilemme du prisonnier](#)

[Document 5.1 : Typologie des biens économiques](#)

[Graphique 5.1 : Le cas d'une base de loisir, un bien de club à l'origine d'une perte sèche](#)

[Document 5.2 : Les externalités liées à l'eau](#)

[Graphique 5.2 : L'effet d'une externalité négative sur l'équilibre d'un marché agricole](#)

[Graphique 5.3 : Le niveau optimal de pollution](#)

[Graphique 5.4 : La taxe pigouvienne](#)

[Graphique 6.1 : L'arbitrage consommation-loisir](#)

[Graphique 6.2 : Les effets d'une augmentation du taux de salaire](#)

[Graphique 6.3 : Un exemple de courbe d'offre de travail individuelle](#)

[Document 6.1 : Part des travailleurs découragés en France](#)

[Graphique 6.4 : La détermination du taux de salaire d'équilibre](#)

[Graphique 6.5 : l'effet des cotisations sociales sur la détermination du taux de salaire d'équilibre](#)

[Document 6.2 : Situation au regard de l'activité par âge](#)

[Document 6.3 : L'évolution du taux de chômage et de la part du chômage par groupe d'âge](#)

[Graphique 6.6 : L'effet de l'instauration d'un salaire minimum](#)

[Graphique 6.7 : L'effet des cotisations sociales avec salaire minimum](#)

[Graphique 6.8 : L'effet du SMIC en situation de monopsonne](#)

[Graphique 6.9 : La courbe théorique de Beveridge](#)

[Document 7.1 : Évolution du salaire moyen de différents fractiles de salaires](#)

[Document 7.2 : Niveaux de salaires par sexe et par décile](#)

[Document 7.3 : Les prélèvements obligatoires en France en fonction du revenu](#)

[Graphique 7.1 : Les effets d'une augmentation des aides sociales sur l'offre de travail](#)

[Document 7.4 : Prime pour l'emploi et Revenu de solidarité active](#)

Listes des encadrés

Encadré 1.1 : La méthode des prix hédoniques, base de l'étude des caractéristiques d'un logement

Encadré 1.2 : Investir dans l'immobilier ?

Encadré 1.3 : Les effets des aides au logement

Encadré : le rationnement

Encadré 2.1 : Comment mesurer les dépenses des ménages ?

Encadré 2.2 : Une approche sociologique de l'incertitude sur la qualité des produits

Encadré 4.1 : Concurrence imparfaite et territoire

Encadré 4.2 : La politique de la concurrence

Encadré 5.1 : Soutenabilité forte, soutenabilité faible, et externalités