CHAPITRE 5

Tarification discriminante

Simon P. ANDERSON et Régis RENAULT

Résumé Ce chapitre étudie un ensemble de pratiques de tarification permettant à une entreprise de transport de discriminer des utilisateurs dont les demandes diffèrent. L'objectif est de présenter les concepts et les modèles théoriques utilisés en science économique pour analyser de telles pratiques tout en soulignant la pertinence de ces cadres d'analyse en ce qui concerne la tarification dans les transports pour des entreprises publiques ou privées. La tarification discriminante prend des formes différentes selon que l'entreprise dispose ou non d'informations qui lui permettent de s'assurer que chaque formule tarifaire s'applique aux utilisateurs auxquels elle est destinée. Si de telles informations sont disponibles, la discrimination tarifaire peut prendre la forme de formules différenciées selon des caractéristiques observables du client telles que l'âge, le lieu de départ et d'arrivée (indépendamment de la distance), ou encore la plage temporelle pendant laquelle le transport est effectué. Dans le cas contraire, il n'est pas possible d'empêcher les utilisateurs d'arbitrer entre les différentes formules proposées. La discrimination prend alors la forme de formules d'abonnement, mais aussi de classes de confort différentes pour le transport des voyageurs. Le chapitre conclut par quelques considérations sur la relation entre la discrimination tarifaire et l'intensité de la concurrence sur un marché.

Mots clés arbitrage, tarification non linéaire, tarification discriminante, tarification de Ramsey-Boiteux, coût marginal des fonds publics.

1 Problématique

Lorsque nous nous déplaçons en avion ou en train, nous sommes bien conscients d'un fait : le prix de notre billet peut être très différent de ce qu'ont payé les autres voyageurs dont nous partageons la cabine ou le compartiment. Nous pouvons nous désoler de cette situation si notre billet est parmi les plus chers ou, au contraire, nous réjouir d'avoir fait une bonne affaire si nous bénéficions d'un tarif avantageux. Mais nous pouvons aussi, en prenant un peu de recul, voir dans ces circonstances une parfaite illustration de ce que les économistes appellent la tarification discriminante. C'est en effet un cas d'école où un service en tout point identique (même trajet, même date, même heure, mêmes conditions de confort) est vendu à des prix différents.

En y regardant de plus près, il est quelque peu simplificateur de considérer que tous les voyageurs ont bénéficié du même service.

Il existe certes des tarifs réduits, attribués d'après des critères « démographiques », qui permettent à certains groupes d'obtenir des billets bon marché dans des conditions semblables à celles s'appliquant au plein tarif. C'est par exemple le cas à la SNCF, pour les tarifs réservés aux personnes âgées de plus de 65 ans ou de moins de 26 ans, ou encore aux familles.

Mais pour les voyageurs qui ne satisfont à aucun de ces critères, les billets bon marché sont assortis de nombreuses restrictions qui impliquent clairement une moindre qualité de service. En général, il est nécessaire de prendre le billet longtemps à l'avance, avec des conditions restrictives sur l'annulation et le remboursement. Le voyageur peut alors jouir d'un service équivalent à celui qu'il aurait eu en payant plein tarif, s'il n'est pas empêché de partir en temps et en heure. Mais pour cela, il a dû accepter de supporter un certain risque au cas où il aurait été contraint de modifier ou annuler son billet ou encore, il a renoncé à un changement qui lui aurait convenu, pour échapper aux sanctions financières.

Les services ne sont donc pas en tout point identiques, mais seulement en apparence. Ceci est particulièrement évident lorsqu'il y a des classes de confort différentes ; mais c'est aussi le cas, par exemple, lorsque le prix est fonction de l'heure ou de la date du voyage. Si on ignore la capacité des entreprises de diversifier la gamme de leurs services, on ne peut rendre compte que d'un nombre limité de pratiques de tarifications discriminantes.

À la date du 16 juillet 2003, les tarifs de deuxième classe proposés sur le serveur internet du train Thalys (`www.thalys.com`) pour le trajet entre Charleroi Sud (Belgique) et Paris Nord étaient les suivants pour les diverses catégories : Librys 51,5 €, Mezzo 38,5 €, Mezzo+ 28,5 €, et Smilys 20,5 €. Les classes Mezzo+, Mezzo, et Smilys peuvent être obtenues dans la limite d'un contingent de places disponibles et à condition de réserver un aller-retour. Notons qu'il est meilleur marché d'acheter un billet Smilys aller-retour que d'acheter un aller simple Librys. Mais le Smilys ne peut être obtenu qu'en réservant deux semaines avant le départ et c'est le seul tarif qui est strictement non-échangeable et non-remboursable.

Le Mezzo et Mezzo+ sont remboursables à hauteur de 50 % du prix du billet jusqu'à la date du départ. La seule différence entre les deux tarifs est que le contingent Mezzo+ est épuisé plus rapidement que le contingent Mezzo. Il est aussi possible de bénéficier d'un tarif Lys de 26 €à condition de disposer de la carte d'abonnement Lys (c'est à dire un tarif binôme). Pour tous ces exemples, le menu de tarifs proposé doit tenir compte de la possibilité pour les voyageurs d'arbitrer entre les différentes options. Il existe aussi des catégories de tarif qui ne sont pas sujette à un tel arbitrage individuel tels que Kid (15 €) pour les enfants de moins de 12 ans, Kid&Co (26 €) pour les adultes accompagnant un enfant de moins de 12 ans, Youth (26 €) pour les jeunes de moins de 26 ans, et Senior pour les personnes âgées de plus de 59 ans. Toutes ces catégories, sauf Smilys, sont disponibles en Première Classe ; par exemple, pour Librys, à 89 €.

On peut définir simplement la tarification discriminante comme une situation dans laquelle une entreprise vend différentes unités d'un même bien à des prix différents. Une telle définition s'applique parfaitement aux cas ou certaines classes de clients bénéficient de tarifs spéciaux (étudiants, personnes âgées) ou à des modes de tarification dits non linéaires, où le prix par unité dépend de la quantité achetée (comme le prix d'un ticket de métro, à l'unité ou en carnet de dix).

Néanmoins, nous avons vu que de nombreuses pratiques considérées comme discriminatoires impliquent la vente de produits différents. Dans ce cas, les différences de prix pourraient s'expliquer par des différences de coûts sans qu'il soit besoin d'invoquer un motif de discrimination (voir Lott et Roberts, 1991).

Afin de remédier à cette difficulté, certains auteurs ont proposé des définitions basées sur la comparaison des variations de prix avec celles des coûts de production. Par exemple, Stigler (1987) a proposé de comparer le rapport des prix de deux biens au rapport des coûts marginaux de production[1]. On parle alors de discrimination si les deux rapports ne sont pas égalisés.

Phlips (1983), au contraire, suggère de comparer la différence de prix avec la différence de coûts marginaux et considère qu'il y a discrimination si la différence de coûts marginaux n'est pas exactement reflétée dans la différence de prix.

Il est à vrai dire difficile de trouver un argument décisif en faveur de l'une ou l'autre définition[2]. Elles ont toutes deux le mérite de montrer qu'on peut considérer comme discriminatoire une tarification qui implique de faibles différences de prix, mais aussi bien une tarification où les différences de prix sont importantes.

Ainsi, qu'une compagnie aérienne assure elle-même le transport de ses passagers vers un aéroport parisien d'où partent ses vols long-courriers, et fasse payer un prix identique à tous pour un vol de leur lieu de départ vers New York, ceci peut être considéré comme discriminatoire envers les voyageurs habitant près de Paris (voir Tirole, 1988, 1993, pour un exemple similaire). Mais ceci ne signifie en rien qu'une telle

1. Le coût marginal de production est le surcroît de coût qui résulterait de l'accroissement d'une unité de la quantité produite.
2. Voir Clerides (2001) pour une définition selon laquelle il n'y a pas de discrimination si le mode de tarification n'est pas remis en cause par l'arbitrage des acheteurs. Voir ci-dessous pour une discussion de l'arbitrage.

pratique est néfaste du point de vue de l'efficacité économique. Cela signifie simplement qu'avec une telle tarification la compagnie aérienne est à même d'exploiter plus avantageusement son pouvoir de marché que si elle n'assurait pas elle-même l'acheminement de ses voyageurs vers Paris.

De manière générale, une entreprise qui dispose d'un certain pouvoir de marché et propose différents services sera en mesure d'ajuster sa tarification pour obtenir un profit maximal sans que les prix ne reflètent des différences de coût marginal ; l'exercice d'un pouvoir de marché passe précisément par la capacité de tarifer au-dessus du coût marginal. L'offre de services diversifiés peut être vue comme un moyen de discriminer dans la mesure où elle permet d'adapter le service proposé et sa tarification à une demande qui diffère d'un client à l'autre[3].

De même qu'il est difficile de définir la tarification discriminante, il n'est pas très facile de définir des critères permettant de classer les différentes pratiques discriminatoires.

On renvoie traditionnellement à la classification de Pigou (1938) selon laquelle il y a trois degrés possibles de discrimination selon la capacité qu'a l'entreprise de discerner entre les acheteurs disposés à payer un prix élevé et ceux disposés à payer moins cher. Pigou définit la discrimination du premier degré comme une situation dans laquelle les consommateurs payent pour chaque unité le prix maximum qu'ils sont disposés à payer. Ceci correspond à la situation la plus favorable pour l'entreprise, mais nécessite qu'elle connaisse parfaitement la demande de chaque client et qu'elle puisse contrôler l'identité de l'acheteur pour chaque unité vendue. On parle en général ici de discrimination parfaite[4].

Pigou reconnaît que cette première forme de discrimination ne doit être vue que comme une référence théorique sans grande pertinence pratique[5]. Il souligne que l'entreprise est au mieux capable de segmenter le marché entre différents groupes d'acheteurs ayant des demandes différentes. Idéalement, l'entreprise souhaiterait être en mesure de constituer des segments composés d'acheteurs ayant des dispositions à payer voisines. Ces segments peuvent alors être classés, depuis la disposition la plus forte à la disposition la plus faible.

3. Dans ce chapitre, nous soulignons l'impact de demande, et nous traitons le coût marginal comme constant. En effet, l'attribution appropriée des coûts constitue un problème en soi.

4. Quinet (1998) cite un exemple ou la SNCF applique un pratique de tarification proche de la discrimination parfaite.

5. Pigou (1938, p. 280) suggère aussi que même si l'entreprise disposait d'une information suffisante pour discriminer parfaitement, elle serait réticente à le faire car cela l'obligerait à s'engager dans un marchandage bilatéral avec chaque client dans lequel son pouvoir de marchandage serait réduit.

Une telle segmentation idéale conduit à la discrimination du deuxième degré. Néanmoins, Pigou note qu'en pratique, une entreprise en est réduite à s'appuyer sur des segments qui ne peuvent pas être parfaitement classées selon la disposition à payer des consommateurs qui la composent. Elle doit utiliser des critères qui peuvent être directement vérifiés par elle, comme la nature du bien à transporter (par exemple, du charbon ou du cuivre) pour une compagnie de chemin de fer vendant du fret, ou le lieu de résidence de l'acheteur (par exemple dans le pays ou à l'étranger). Il s'agit alors de discrimination du troisième degré.

Le typologie de Pigou souligne le fait que la tarification discriminante a pour objectif premier de permettre à l'entreprise d'exercer son pouvoir de marché dans les meilleures conditions possibles. Plutôt que de proposer une situation de référence correspondant à l'absence de discrimination, comme c'est le cas chez Stigler et Phlips, Pigou définit la situation de référence où l'objectif ultime de la discrimination est atteint, c'est-à-dire lorsque chaque unité est vendue au prix le plus élevé possible.

De ce point de vue, la discrimination du premier degré apparaît comme une référence théorique incontournable. Nous verrons aussi que la discrimination du troisième degré est tout à fait pertinente pour comprendre les pratiques de discrimination qui reposent sur des critères directement vérifiables par l'entreprise tels que l'âge, le lieu de départ et d'arrivée ou l'heure et la date du voyage.

En revanche, la discrimination du deuxième degré n'apparaît que comme un cas particulier de celle du troisième degré qui, comme nous le verrons, correspond à une situation particulièrement favorable pour l'entreprise, où le critère vérifiable qui permet à l'entreprise de discriminer est parfaitement corrélé avec la disposition à payer (par exemple, plus on est âgé, plus on est disposé à payer cher).

Néanmoins, de nombreuses formes de tarification discriminante telles que la tarification non linéaire, ou le fait de proposer différentes classes de confort dans un train ou un avion, ne s'appuient pas sur un critère vérifiable par l'entreprise, et laissent les utilisateurs sélectionner la formule qui leur convient le mieux. Ces pratiques ne peuvent donc pas être considérées comme de la discrimination du troisième degré ; il est devenu habituel parmi les économistes de parler à leur sujet de discrimination du deuxième degré (voir Phlips, 1983, Tirole, 1988, 1993, chapitre 3, Varian, 1989, Mougeot et Naegelen, 1994), ce terme recouvrant donc des pratiques très différentes de celles envisagées initialement par Pigou.

La discussion ci-dessus suggère que la tarification discriminante est étroitement liée à l'exercice d'un pouvoir de marché. En situation de concurrence pure et parfaite, les entreprises sont astreintes à vendre leur production à un prix qui leur est imposé par le marché. Dans un tel contexte, il n'est évidemment pas possible de vendre différentes unités d'un bien à des prix différents ou de tenter d'affecter les prix en proposant une gamme de services différents (du moins tant que ces services sont vendus sur un marché concurrentiel). Même si l'entreprise dispose d'un certain pouvoir de marché, ses velléités de discriminer entre différents acheteurs peuvent être contrariées par la capacité de ces derniers à arbitrer entre les différentes options proposées[6].

Cet arbitrage peut prendre deux formes selon qu'il implique des échanges entre plusieurs acheteurs ou non. S'il est possible, selon l'expression de Pigou (1938), de transférer le bien, les acheteurs sont en mesure d'échanger le bien ou service entre eux et l'entreprise ne peut pratiquer des prix différents, car les clients qui bénéficient du tarif le plus bas achèteront à seule fin de revendre à ceux qui devraient autrement payer plus cher. Par exemple, si quelqu'un disposant d'un abonnement peut se procurer des billets à tarif réduit, il pourrait en acheter une grande quantité afin de les revendre à des non-abonnés. De même, si un tarif réduit est proposé sous condition que le billet soit acheté suffisamment tôt avant le voyage, il serait possible pour des personnes entreprenantes d'acheter ces billets bon marché en grande quantité pour ne les revendre que quelques jours avant la date à laquelle les billets peuvent être utilisés. Bien que cette forme d'arbitrage soit entravée par des coûts de transaction, elle constitue une contrainte suffisante sur la stratégie de tarification des entreprises pour que celles-ci mettent en place un certain nombre de pratiques visant à l'empêcher.

Elles exigent par exemple la présentation de cartes d'abonnement nominatives lors du voyage ou, dans le cas des compagnies aériennes, exigent la présentation de pièces d'identité correspondant au nom figurant sur le billet lors de l'enregistrement. Ce type de règle permet aussi d'éviter la seconde forme de l'arbitrage dans laquelle un client confronté à plusieurs options ne choisit pas celle qui lui est destinée. Par exemple si une entreprise souhaite discriminer sur la base de l'âge, la présentation d'une pièce d'identité permet de s'assurer que chacun paye le prix auquel il a droit[7]. Lorsque les acheteurs peuvent pratiquer un tel arbitrage entre les différentes formules proposées, Pigou (1938) dit que c'est la demande qui est transférable.

6. Une autre limite à la capacité de discriminer des entreprises est qu'elles ne peuvent pas proposer à leurs clients un système d'une trop grande complexité.

7. L'arbitrage non-autorisé (prendre la carte de son frère jumeau) est une fraude, mais n'est pas tout à fait semblable au non-paiement (et de ce fait, il est sanctionné moins sévèrement).

En pratique les entreprises parviennent à discriminer sans pour autant être en mesure d'obliger leurs clients à ne pas arbitrer. Cet arbitrage potentiel constitue alors une contrainte qui doit être prise en compte dans l'élaboration d'une tarification discriminante. Puisqu'on ne peut forcer les clients à renoncer à l'arbitrage, il faut les y inciter.

L'analyse du comportement d'arbitrage impliquant plusieurs acheteurs est relativement complexe et l'essentiel des travaux sur la tarification discriminante suppose simplement que les coûts de transaction sont assez élevés pour la rendre impossible. Nous adopterons cette hypothèse tout au long de ce chapitre[8]. En revanche, l'arbitrage dit « individuel », par lequel un usager peut choisir une option qui ne lui est pas destinée a fait l'objet de nombreux travaux, tout particulièrement dans les 25 dernières années.

Le problème qui se pose à l'entreprise est de proposer un ensemble de formules destinées à des clients ayant des demandes différentes, sans pour autant pouvoir observer directement à quel type de client elle a à faire. La tarification discriminante doit alors satisfaire des contraintes dites « incitatives », qui assurent qu'aucun client n'a intérêt à arbitrer en faveur d'une formule qui ne lui est pas destinée.

La section 2 présente un certain nombre de concepts de base et de principes généraux indispensables à la compréhension du reste du chapitre[9].

Dans la section 3, nous verrons comment une entreprise peut exploiter l'information dont elle dispose sur la demande de ses clients pour discriminer en supposant que les acheteurs ne se livrent à aucun arbitrage individuel. Nous distinguerons la discrimination parfaite (3.1), la discrimination entre plusieurs catégories d'acheteurs consommant le même bien (3.2) et la segmentation du marché où l'entreprise propose des services différents pour lesquels elle est en mesure d'identifier parfaitement les acheteurs susceptibles de consommer chaque service.

La section 4 est consacrée aux stratégies qui permettent à l'entreprise d'inciter les acheteurs à ne pas recourir à l'arbitrage individuel dans les situations où elles ne peuvent l'empêcher en utilisant de l'information vérifiable. Nous considérons d'abord l'utilisation d'une tarification non linéaire (4.1 et 4.2) puis la possibilité de discriminer en proposant des qualités différentes (4.3).

Dans la section 5, nous apportons quelques conclusions, et nous discutons plus particulièrement le lien entre la tarification discriminante

8. Voir Alger (1999) pour une analyse des contraintes imposées par un arbitrage potentiel impliquant plusieurs acheteurs.
9. Pour un traitement simple et plus complet de ce sujet, voir Varian (2000).

et la concurrence sur un marché dans la section 5. Ceci nous permettra de discuter la robustesse des résultats obtenus dans un cadre de monopole.

2 Résultats préliminaires et concepts de base

2.1 Tarification uniforme de monopole

Nous commençons par le mode de tarification le plus simple, dit uniforme, où toutes les unités d'un même bien sont vendues au même prix : par exemple, sur un trajet donné, il n'y a qu'un seul service offert et tous les voyageurs payent le même prix.

Soit $D(p)$ la quantité demandée au prix p. L'interprétation la plus simple consiste à supposer que D décrit la distribution des dispositions à payer des différents voyageurs, chacun ne souhaitant acheter qu'un seul trajet ($D(p)$ indiquant alors le nombre de voyageurs disposés à payer au moins p).

Il est aussi utile de définir la demande inverse qui pour chaque quantité q indique le prix maximal auquel elle peut être vendue, $P(q)$. Dans l'interprétation où chaque voyageur ne souhaite effectuer qu'un seul trajet, $P(q)$ est la disposition à payer du consommateur marginal (celui qui renoncerait à voyager si le prix était accru). Afin de simplifier l'analyse, supposons un coût marginal constant c qui correspond aussi au coût variable moyen de production (c'est le coût généré par chaque voyageur). En situation de concurrence parfaite, le prix serait donné par le coût marginal et le nombre de voyageurs serait $D(c)$.

Nous supposons ici que le service est assuré par une entreprise privée en situation de monopole dont l'objectif est de maximiser son profit. Commençons par considérer le cas où la demande est linéaire,

$$q = D(p) = a - bp, \quad a > 0, \quad b > 0, \quad \text{et} \quad \frac{a}{b} > c.$$

La dernière condition assure qu'il est optimal de produire une quantité strictement positive. Cette situation est illustrée sur la figure 5.1.

L'entreprise est contrainte par la tarification uniforme de choisir un point sur la courbe de demande ; si elle souhaite transporter q voyageurs, le prix uniforme le plus élevé qu'elle puisse pratiquer est

$$p = P(q) = \frac{a-q}{b}.$$

Le surplus du producteur SP (qui est le profit augmenté des coûts fixes) correspond à l'aire du rectangle au-dessus du coût marginal,

Figure 5.1 : Tarification uniforme de monopole

La recette marginale égale le coût marginal pour $q^m = (a - bc)/2$, et le prix correspondant est $p^m = (a + bc)/2b$ (voir texte).

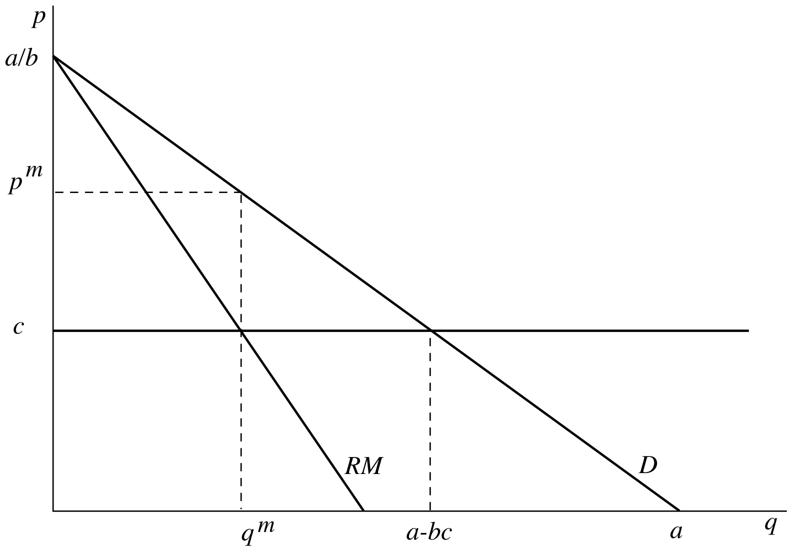
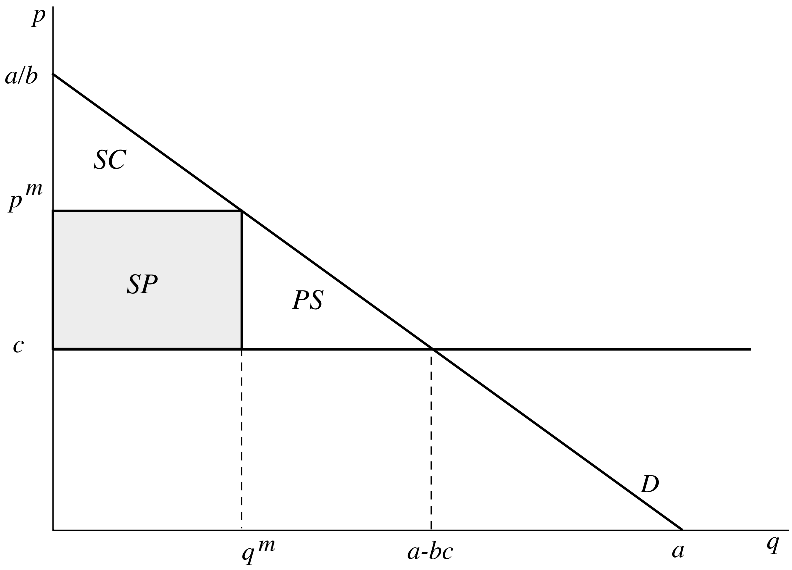


Figure 5.2 : Les deux surplus et la perte sèche

SC : surplus du consommateur (aire triangulaire comprise entre la courbe de demande inverse *D* et le prix p^m) ; *SP* : surplus du producteur (aire rectangulaire en grisé, au-dessus du coût marginal) ; *PS* : perte sèche (aire triangulaire à droite de *SP*).



c'est-à-dire la marge par voyageur $p - c$ multipliée par le nombre de voyageurs q. La quantité optimale doit donc maximiser

$$\left[\frac{a-q}{b} - c\right] q,$$

et les conditions du premier ordre pour une quantité optimale peuvent s'écrire

$$\frac{a-2q}{b} = c.$$

Le membre de gauche est la recette marginale RM ; c'est une droite de même ordonnée à l'origine que la demande inverse D, mais sa pente est deux fois plus forte. Il est aisé de voir sur la figure 5.1 que la condition d'égalité entre la recette marginale et le coût marginal est satisfaite pour une quantité

$$q^m = \frac{D(c)}{2} = \frac{a-bc}{2},$$

c'est-à-dire la moitié de la quantité produite en concurrence parfaite. Le prix uniforme pratiqué est donc

$$p^m = P(q^m) = \frac{a+bc}{2b},$$

et il en résulte un surplus du producteur SP égal à

$$(p^m - c)q^m = \frac{(a-bc)^2}{4b},$$

qui correspond au rectangle SP sur la figure 5.2.

Dans le cas général, l'égalité entre recette marginale et coût marginal permet de déterminer la quantité optimale[10]. Dans la présentation de la tarification discriminante nous nous appuierons fréquemment sur des conditions du premier ordre relatives au prix. En tarification uniforme l'entreprise choisit un prix uniforme p^m afin de maximiser

$$D(p^m)\ (p^m - c).$$

La condition nécessaire du premier ordre pour un maximum peut être écrite sous la forme

$$\frac{p^m - c}{p^m} = -\frac{1}{\eta(p^m)}, \qquad \text{avec} \quad \eta(p^m) = p^m \frac{D'(p^m)}{D(p^m)} < 0,$$

où $\eta(p^m)$ est l'élasticité de la demande. Le membre de gauche est le taux de marge, aussi appelé indice de Lerner. Il est donc d'autant plus élevé

10. Celle-ci est strictement positive et unique si, en zéro, la recette marginale est supérieure au coût marginal, et si la pente de la recette marginale est toujours inférieure à la pente du coût marginal.

que la demande est peu élastique, c'est-à-dire que l'élasticité de la demande est proche de zéro. En tout état de cause, l'élasticité au prix de monopole est inférieure à -1.

2.2 Analyse de bien-être et intervention publique

Nous venons de voir qu'une entreprise privée qui maximise son profit va choisir un prix plus élevé que le prix de concurrence parfaite et peut ainsi obtenir davantage de profit. Il est évident que ceci se fait au détriment des consommateurs (les voyageurs), qui doivent payer plus cher. Nous allons maintenant montrer comment cette perte de bien-être des consommateurs peut être mesurée en termes monétaires, de sorte qu'elle peut être comparée avec le profit supplémentaire réalisé par l'entreprise.

À cette fin, introduisons le *surplus du consommateur SC*, défini comme la différence entre le prix maximal qu'il est prêt à payer pour un trajet et le prix dont il doit effectivement s'acquitter. La courbe de demande inverse est construite en ordonnant les dispositions à payer par ordre décroissant, les q premiers trajets proposés étant vendus aux q voyageurs disposés à payer le plus cher.

Le surplus du consommateur SC, généré par la vente d'une quantité $D(p)$ au prix p, correspond donc à l'aire comprise entre la demande inverse et le prix pour des quantités comprises entre 0 et $D(p)$. Pour la solution décrite à la figure 5.1, c'est le triangle SC de la figure 5.2. La perte de surplus du consommateur résultant d'un passage d'une tarification au coût marginal à une tarification de monopole est donnée par le surplus du producteur en monopole augmenté de l'aire du triangle PS sur la figure 5.2. Or cette perte de surplus du consommateur peut être interprétée comme la somme que les voyageurs seraient prêts à verser collectivement pour pouvoir disposer de billets tarifés au coût marginal plutôt qu'au prix de monopole. Puisque cette somme est supérieure au surplus du producteur en monopole, il existe un échange bénéfique pour tous les participants qui n'est pas réalisé (l'entreprise serait disposée à diminuer son prix jusqu'au coût marginal si, en échange, elle percevait une compensation au moins égale à son surplus de monopole).

La tarification de monopole introduit donc une *inefficacité* qui peut être mesurée par la perte de surplus du consommateur qui n'est pas compensée par l'accroissement du surplus du producteur.

Cette *perte sèche* est l'aire PS sur la figure 5.2. Afin d'appréhender cette inefficacité de manière plus concise, il est commode d'introduire le *surplus social*, qui est la somme du surplus du producteur et du surplus du consommateur. La tarification au coût marginal permet d'obtenir le surplus social maximal. La perte sèche est la diminution du surplus social causée par un prix plus élevé[11].

Un objectif raisonnable de l'intervention publique pourrait donc être de minimiser la perte sèche.

Cet objectif peut être atteint par une entreprise publique ou réglementée qui adopte une *tarification au coût marginal*. Mais une telle solution n'est en général pas satisfaisante. En effet, en présence de rendements d'échelle croissants, l'entreprise ne rentre pas dans ses frais. Par exemple, dans le cas où le coût marginal est constant, la tarification au coût marginal génère un surplus du producteur nul, de sorte que l'entreprise n'est pas en mesure de couvrir ses coûts fixes. Il s'ensuit que l'activité de l'entreprise doit être en partie financée par les contribuables. Il peut donc sembler souhaitable d'accepter une perte sèche afin d'éviter que l'entreprise soit déficitaire. La tarification au coût marginal devrait alors être remplacée par une *tarification au coût moyen* de sorte que l'intégralité du coût soit supportée par les utilisateurs.

Néanmoins, il est quelque peu arbitraire d'imposer que l'entreprise ne doit pas faire de pertes. Une argumentation alternative contre la tarification au coût marginal est que le système fiscal souffre d'imperfections qui empêchent de prélever des fonds publics sans qu'il en résulte une perte d'efficacité dans l'allocation des ressources (Meade, 1944). La tarification optimale doit donc arbitrer entre l'objectif de maximisation du surplus social sur le marché servi par l'entreprise et le coût résultant de l'impact négatif des prélèvements fiscaux sur le reste de l'économie. Ce coût peut être mesuré par les pertes sèches engendrées par les différents impôts sur les marchés auxquels ils s'appliquent[12]. On ne choisira alors de faire financer l'intégralité de l'activité de la firme par les utilisateurs que si ce coût du prélèvement des fonds publics est trop important.

Mais ce raisonnement implique aussi que si le coût du prélèvement est trop important, et si le gouvernement est en mesure de s'approprier le profit de l'entreprise (ce qui est en particulier le cas si elle est publique), il peut être souhaitable de générer une recette supérieure aux coûts de production ; le revenu supplémentaire ainsi obtenu permet de réduire la pression fiscale sur le reste de l'économie.

Nous sommes maintenant en mesure de poser le problème de tarification d'une entreprise publique ou réglementée qui cherche à maximiser le surplus social, tout en prenant en compte le fait que toute re-

11. Une tarification en-dessous du coût marginal introduirait aussi une perte sèche, puisqu'elle conduirait à vendre des unités pour lesquelles la disposition à payer des acheteurs est plus faible que le coût supplémentaire qu'elles engendrent.

12. Par exemple, l'impôt sur le revenu génère des pertes sèches entre autres sur le marché du travail.

cette supplémentaire qu'elle peut générer permet de réduire les distorsions fiscales sur les autres marchés. Afin de formaliser cet arbitrage, à l'instar de Laffont et Tirole (1993), nous introduisons un paramètre mesurant le *coût marginal des fonds publics*. C'est-à-dire que nous supposons que tout euro supplémentaire de surplus du producteur pour la firme induit un accroissement du surplus dans le reste de l'économie dont la valeur est $\lambda > 0$[13]. La fonction objectif de l'entreprise peut donc s'écrire

$$SC + (1+\lambda)SP,$$

où SC et SP dénotent respectivement le surplus du consommateur et le surplus du producteur. En faisant tendre λ vers l'infini on parvient a une situation où la firme ignore entièrement le surplus du consommateur, ce qui correspond au comportement d'une firme privée maximisant son profit.

Soit p le prix et $D(p)$ la demande résultante. Le surplus du consommateur correspondant est dénoté $SC(p)$. Nous supposons aussi un coût marginal constant c. L'entreprise choisit donc p afin de maximiser

$$SC(p) + (1+\lambda)\ D(p)\ (p-c).$$

En utilisant $SC'(p) = -D(p)$[14], on obtient les conditions nécessaires du premier ordre pour un maximum sous la forme

$$\frac{p-c}{p} = -\frac{\lambda}{1+\lambda}\ \frac{1}{\eta(p)}. \tag{5.1}$$

Le prix ainsi obtenu est un cas particulier des prix dits de Ramsey-Boiteux, qui assurent la maximisation du surplus social, sous la contrainte que l'entreprise génère une recette minimale, par exemple pour couvrir des coûts fixes. Dans cette interprétation, λ est alors une mesure de la sévérité de la contrainte budgétaire : c'est le gain marginal de surplus social qui peut être obtenu en diminuant la recette à réaliser d'un euro. Dans notre interprétation, il mesure le gain de surplus qui peut être réalisé sur les autres marchés en accroissant la recette de l'entreprise d'un euro (voir Ramsey, 1927, et Boiteux, 1956).

13. La mesure de coût des fonds publics a fait l'objet de nombreuses études pour le cas de l'économie des Etats-Unis. On considère en général qu'une estimation raisonnable du coût marginal est 0,3 (voir Ballard, Shoven et Whalley, 1985, et Hausman et Poterba, 1987). À notre connaissance, il n'existe aucune étude sur le cas français.

14. En effet, le surplus du consommateur pour un prix p est l'intégrale de la demande pour les prix supérieurs à p.

Dans les deux interprétations ci-dessus, λ est nécessairement positif. Ceci suggère qu'on met davantage de poids sur le surplus du producteur que sur le surplus du consommateur. À la condition que la demande devient plus élastique pour un prix plus élevé (η strictement décroissante), plus λ est élevé plus le prix pratiqué s'élève au-dessus du coût marginal. En faisant tendre λ vers l'infini, on retrouve la formule habituelle du prix uniforme de monopole pour une firme qui maximise son profit. Ceci peut sembler insatisfaisant dans la mesure où on peut souhaiter que l'intervention publique réponde non seulement à des objectifs d'efficacité économique, mais aussi à des objectifs de redistribution. L'analyse présentée ici n'entre pas en conflit avec de tels objectifs. En effet, l'un des rôles du système fiscal est précisément d'assurer une répartition plus équitable des ressources.

Donc si une hausse du prix du bien considéré (hausse du prix du billet de train) peut permettre au gouvernement de générer efficacement des recettes fiscales, il sera d'autant plus à même d'effectuer des transferts au profit des plus démunis.

On objectera que parmi les utilisateurs il peut se trouver des personnes appartenant à de catégories particulièrement défavorisées. La théorie micro-économique nous enseigne qu'il est en général préférable d'apporter un soutien direct aux ménages à l'aide de transferts monétaires plutôt que de subventionner l'un ou l'autre des biens ou services qu'ils consomment.

Néanmoins, la mise en place d'un système de transferts monétaire pose elle-même des difficultés notamment pour des raisons incitatives mais aussi parce qu'un tel système peut ne pas être politiquement viable. Il peut alors être légitime de mettre un poids plus important sur le surplus du consommateur, ce qui dans notre formalisme se traduirait par une valeur de λ entre 0 et -1. Ceci permet de comprendre pourquoi on peut subventionner lourdement certains services comme les transports en commun, y compris en tarifant en-dessous du coût marginal.

Pour terminer cette discussion de l'intervention publique, notons que, particulièrement dans le cas des activités de transport, il existe une autre motivation importante pour choisir de tarifer en-dessous du coût marginal de production.

En effet, nous n'avons jusqu'à maintenant pas pris en compte les effets externes positifs ou négatifs (en matière de pollution ou de congestion) qu'implique le développement de certains types de transports. Cette omission peut facilement être corrigée en remplaçant dans notre analyse le coût marginal de production par un *coût marginal social* qui sera supérieur ou inférieur au coût marginal de production, selon que l'effet externe sera négatif ou positif.

3 Discrimination et critères vérifiables

3.1 Discrimination parfaite

Nous considérons tout d'abord le cas le plus favorable pour le vendeur, celui où il dispose de toute l'information sur la demande émanant de chaque acheteur potentiel. S'il est en mesure d'utiliser cette information pleinement, on parle de discrimination parfaite. Il s'agit de la discrimination au premier degré dans la typologie de Pigou. Pour utiliser l'information, le vendeur doit être en mesure de contrôler le prix et les caractéristiques de chaque unité vendue à chaque acheteur. Par exemple, si une compagnie aérienne connaissait parfaitement les goûts et besoins de chacun de ses clients, elle devrait, pour discriminer parfaitement, choisir le prix auquel chaque client effectuera chaque vol en spécifiant la date, l'heure et les conditions de confort.

Ce principe peut être illustré très simplement en supposant que le bien vendu est parfaitement homogène et que chaque acheteur ne souhaite se procurer qu'une unité. Dans ce cas les goûts de l'acheteur sont intégralement décrits en indiquant sa disposition à payer. Une entreprise privée en situation de monopole dont le but est de maximiser son profit choisira de faire payer à chaque acheteur sa disposition à payer, en ne lui laissant ainsi aucun surplus. Elle aura bien sûr intérêt à vendre à ceux qui sont disposés à payer le plus cher et elle acceptera de vendre à tous ceux qui sont prêts à payer plus que le coût marginal[15]. Cette situation est représentée graphiquement sur la figure 5.1 pour un coût marginal constant et une courbe de demande linéaire (traiter la quantité demandée comme une variable continue est une bonne approximation s'il y a beaucoup d'acheteurs, car en ce cas la demande de chacun est négligeable par comparaison avec la demande totale). La demande inverse décrit la distribution des prix de réserve par ordre décroissant. On voit que la politique de tarification de l'entreprise aboutit à la maximisation du surplus social qui lui revient en totalité.

L'approche ci-dessus peut être aisément étendue au cas où la demande individuelle est élastique (c'est-à-dire que la quantité demandée décroît si le prix augmente). À titre d'illustration, supposons que la courbe de demande de la figure 5.1 émane d'un unique consommateur. La demande inverse indique alors le prix le plus élevé qu'il est disposé à payer pour consommer une unité supplémentaire. Plus la quantité dont il dispose déjà, q, est élevée, plus ce prix est faible. La tarification discriminante optimale consisterait donc à faire payer chaque unité supplémentaire au prix donné par la demande inverse et d'accepter de vendre tant que ce prix est supérieur au coût marginal. Concrètement,

15. Dans le cas général où le coût marginal n'est pas nécessairement constant, la quantité vendue q est telle que la disposition à payer la plus faible parmi ceux qui achètent, $P(q)$, est égale au coût marginal en q.

ceci revient à vendre la quantité $q^* = D(c)$ à un tarif égal au surplus brut du consommateur correspondant, $V(q^*)$ (c'est-à-dire le surplus du consommateur s'il pouvait consommer la quantité q^* gratuitement)[16].

Une autre méthode pour parvenir au même résultat serait d'utiliser un tarif binôme (voir aussi Oi, 1971). Un tel tarif spécifie d'abord un prix d'abonnement, A, que le consommateur doit payer pour pouvoir consommer le bien. S'il s'acquitte de l'abonnement, le consommateur peut alors acheter n'importe quelle quantité au prix unitaire p. En fixant $p = c$, on s'assure que le consommateur choisira de consommer q^*. Il obtient ainsi un surplus du consommateur de $V(q^*) - cq^*$. Il acceptera de s'abonner tant que le tarif d'abonnement n'excède pas ce surplus. Si le vendeur choisit ce tarif d'abonnement, $A = V(q^*) - cq^*$, le prix total payé par le consommateur est $A + cq^* = V(q^*)$. Le résultat est strictement équivalent à celui obtenu avec le système de tarification précédent. Bien que la tarification binôme paraisse plus simple, elle nécessite tout autant d'information que la précédente. En effet, bien qu'il soit relativement facile pour l'entreprise de fixer le prix unitaire au coût marginal, la détermination du tarif d'abonnement optimal nécessite de connaître le surplus du consommateur et donc sa demande.

Pour terminer cette discussion de la discrimination parfaite, notons que la solution que nous avons décrite est celle qui serait choisie par une firme privée cherchant à maximiser son profit. En quoi diffère-t-elle de la solution qui serait choisie par une entreprise publique ? Nous avons noté que la solution choisie par l'entreprise privée maximise le surplus social et serait donc optimale en l'absence d'un coût marginal des fonds publics. La répartition du surplus en discrimination parfaite est en revanche très différente de ce qu'elle serait avec une tarification uniforme au coût marginal puisqu'elle ne laisse aucun surplus au consommateur. Si en revanche le coût marginal des fonds publics est positif, puisque l'entreprise publique en discriminant parfaitement peut générer des ressources pour l'Etat qui n'ont aucun coût en terme d'efficacité, il est souhaitable que l'entreprise publique adopte la même politique tarifaire qu'une entreprise privée. De toute évidence, la solution de discrimination parfaite ne peut être vue que comme une solution de référence théorique car sa mise en oeuvre nécessite une information parfaite sur la demande individuelle de chaque acheteur, qui n'est en général pas disponible. Mais il n'est pas rare qu'une entreprise dispose d'une information partiellement vérifiable sur la demande, qu'elle cherchera à utiliser à des fins de discrimination tari-

16. Notons que si on vendait la quantité q au prix uniforme $P(q)$, le surplus net du consommateur serait $SC(P(q))$; le surplus brut du consommateur est donc donné par $V(q) = SC(P(q)) + P(q)q$.

faires. Nous considérons maintenant ce type de comportement qui correspond à ce que Pigou appelle la discrimination du troisième degré.

3.2 Un seul bien et plusieurs groupes d'acheteurs

Nous considérons ici une situation dans laquelle une firme est en mesure d'observer une caractéristique de l'acheteur (âge, profession, adresse) qui lui procure de l'information sur sa demande. En d'autres termes, l'entreprise peut exploiter une certaine corrélation entre la variable observable et la demande de l'individu pour discriminer (il se peut néanmoins qu'une telle discrimination ne soit pas légale ou socialement acceptable, par exemple, si elle est basée sur le sexe ou l'appartenance ethnique).

L'entreprise pratique alors un prix qui dépend de cette caractéristique, laquelle peut être contrôlée pour éviter qu'un individu bénéficie d'un tarif qui ne lui est pas destiné. En revanche, elle n'est pas en mesure de discriminer entre les consommateurs de caractéristiques semblables (par exemple dans une même classe d'âge), et doit donc pratiquer un prix uniforme pour chaque groupe. Ceci peut être dû au fait qu'elle ne connaît que la demande agrégée pour chaque groupe, ou encore que les demandes individuelles sont unitaires et qu'il n'est pas possible d'introduire de différenciation dans le produit (la vente de quantités différentes avec une tarification non-linéaire ou l'introduction de différenciation pouvant être utilisées pour inciter les acheteurs à révéler leurs goûts ; voir la section 4). La firme doit donc pratiquer une tarification uniforme pour chaque catégorie.

Pour en revenir à la discrimination basée sur un critère vérifiable, supposons qu'il n'y a que deux catégories d'acheteurs (par exemple : les jeunes de moins de 26 ans et les autres). En demandant des justificatifs (et en l'absence de fraude), l'entreprise peut savoir si elle a affaire à un acheteur de l'une ou l'autre catégorie. On exclut par ailleurs toute possibilité d'arbitrage par lequel les acheteurs d'un groupe pourraient acheter le bien pour le revendre à l'autre groupe (par exemple, les billets d'avion sont nominatifs et ne peuvent être utilisés sans pièce d'identité). Du fait de cette absence d'arbitrage, la demande émanant de chaque groupe est fonction du seul prix appliqué à ce groupe. Soit p_i le prix appliqué au groupe i et $D_i(p_i)$ la demande qui en résulte. Le surplus du consommateur correspondant est dénoté $SC_i(p_i)$. Nous supposons aussi un coût marginal constant c, de sorte que la firme peut choisir les deux prix indépendamment l'un de l'autre[17]. L'entreprise choisit

17. Dans le cas général où les rendements ne sont pas nécessairement constants, la solution optimale s'obtient en égalisant les recettes marginales des deux groupes d'acheteurs pour une quantité totale produite donnée, puis en choisissant la quantité totale qui égalise le coût marginal à la recette marginale commune.

donc p_i afin de maximiser

$$SC_i(p_i) \; + \; (1+\lambda)\, D_i(p_i) \;\; (p_i - c).$$

On en déduit une condition du premier ordre analogue à la formule (5.1), page 103 :

$$\frac{p_i - c}{p_i} = -\frac{\lambda}{1+\lambda} \; \frac{1}{\eta_i(p_i)}.$$

On voit que ce type de discrimination est avantageux lorsque l'entreprise est confrontée à des groupes dont l'élasticité de la demande diffère. Moins la demande est élastique (plus η_i est proche de zéro) et plus le prix pratiqué est élevé. Ainsi les jeunes bénéficient de meilleurs tarifs de trains ou d'avions car ils sont davantage susceptibles de réduire leur demande si le prix augmente. Clairement, l'entreprise bénéficie de la discrimination, car elle a toujours la possibilité de ne pas la mettre en œuvre. En revanche, seuls les consommateurs dont la demande est assez élastique seront gagnants[18].

Dans le cas où l'entreprise est privée et maximise son profit, on peut demander quels seront les effets de l'introduction de la tarification discriminante sur le bien-être social. Nous supposons ici que le coût marginal des fonds publics est nul et que le surplus social est donc simplement la somme des surplus du producteur et du consommateur.

Notons d'abord que pour une quantité produite fixée, la tarification discriminante introduit une inefficacité dans l'allocation de cette quantité entre les deux groupes. En effet, le prix payé par les consommateurs d'une catégorie reflète ce qu'ils seraient prêts à payer pour consommer une unité supplémentaire. On pourrait donc obtenir un accroissement de surplus social en transférant une partie de la quantité consommée des acheteurs qui payent le moins cher, vers ceux qui payent le plus cher. Comme par ailleurs, la quantité produite par un monopoleur pratiquant une tarification uniforme est insuffisante, la discrimination ne peut améliorer le bien-être que si elle implique une production plus importante. Cette condition nécessaire pour que la discrimination conduise à plus d'efficacité a été mise en évidence par Schmalensee (1981).

D'une manière générale, l'impact de la tarification discriminante sur la quantité produite est ambigu. Par exemple, supposons que la demande de chaque catégorie i est linéaire et donnée par

$$D_i(p) = a_i - b_i\, p.$$

18. Cette discussion suppose que les deux demandes sont comparables dans le sens où l'une est plus élastique que l'autre pour tout niveau de prix. Il est possible de construire des exemples pour lesquels tous les prix en tarification discriminante seraient plus élevés ou plus faibles qu'en tarification uniforme (voir Nahata, Ostaszewski, et Sahoo, 1990).

Les résultats de la section 2.1 pour le cas linéaire montrent que si les deux catégories sont servies en régime de prix uniforme, la quantité produite est

$$\frac{a_1 + a_2 - (b_1 + b_2)c}{2},$$

alors que les quantités produites pour chaque groupe d'acheteurs en régime de discrimination sont respectivement

$$\frac{(a_1 - b_1 c)}{2} \qquad \text{et} \qquad \frac{(a_2 - b_2 c)}{2}.$$

La quantité totale est donc identique[19], et la discrimination est préjudiciable au bien-être social[20].

Notons que ce type de raisonnement permet aussi d'obtenir un résultat non-ambigu lorsqu'une capacité limitée est entièrement utilisée en tarification uniforme. La tarification discriminante ne peut alors pas conduire à une quantité plus importante et elle est donc nuisible. Par exemple si un avion voyage à plein, il est préférable que tous ses passagers payent le même prix. Néanmoins, ce raisonnement n'est correct qu'à court terme car à long terme, la possibilité de discriminer peut inciter la compagnie à accroître le nombre de ses vols. Dans le même ordre d'idée, il convient de souligner que le raisonnement selon lequel la discrimination introduit une distorsion qui n'est pas présente en tarification uniforme n'est correct que si cette dernière conduit à servir les deux catégories d'acheteurs. De fait, l'un des principaux bénéfices de l'introduction d'une tarification discriminante est de permettre que certains marchés soient servis alors qu'ils ne le seraient pas autrement.

Comme nous l'avons vu dans l'introduction, la discrimination du deuxième degré initialement envisagée par Pigou (1938) peut être vue comme un cas particulier de la discrimination entre différents groupes d'acheteurs, que nous venons de présenter. En effet, si le critère vérifiable utilisé par l'entreprise pour discriminer lui permet de séparer parfaitement un groupe d'acheteurs qui, pour toutes les unités achetées, est disposé a payer plus que ce que le reste de la clientèle est disposé à

19. C'est une propriété particulière des demandes linéaires que la recette marginale correspondant à la somme des demandes soit égale à la somme des recettes marginales des demandes de chaque groupe.

20. Cette conclusion n'est pas valide si un seul groupe est servi par un monopole non-discriminant. Dans ce cas, si le coût marginal est constant, la tarification discriminante est clairement préférable. En effet, le groupe qui est servi dans tous les cas a le même bien-être sous les deux formes de tarification parce qu'il paye le même prix, alors que l'autre groupe préfère clairement la tarification discriminant parce qu'il n'est pas servi en tarification uniforme.

payer pour n'importe quelle unité achetée, il est alors possible de pratiquer une discrimination du deuxième degré à la manière de Pigou.

Pour illustrer ce principe, considérons l'exemple suivant inspiré de Anderson et Renault (2003b). Supposons qu'un utilisateur souhaite effectuer un unique trajet en train et que sa disposition à payer est parfaitement et positivement corrélée avec son âge. Bien que cette situation permette en théorie de pratiquer une discrimination parfaite (puisque l'âge d'un client révèle parfaitement sa demande), il peut être en pratique coûteux de spécifier trop de tarifs différents. Par exemple il se peut qu'il ne soit possible d'offrir que deux tarifs. La compagnie de chemin de fer doit alors déterminer l'âge au-dessus duquel on ne peut plus bénéficier du tarif le moins cher. Supposons par exemple que la demande inverse soit linéaire et donnée par $P(q) = 1 - q$. Avec un coût marginal nul, le prix et la quantité de monopole seraient de 1/2. Il est alors facile de montrer que la stratégie optimale de l'entreprise serait de pratiquer un plein tarif de 2/3 pour le tiers le plus âgé de la population et de pratiquer un tarif réduit de 1/3 pour les plus jeunes. On peut aussi établir qu'une instance de réglementation de l'âge critique dont l'objectif serait de maximiser la somme du surplus du producteur et du consommateur (en l'absence de coût marginal des fonds publics) imposerait à l'entreprise de ne proposer le tarif réduit qu'à la moitié la plus jeune de la population. L'entreprise choisirait alors un plein tarif d'1/2 pour la moitié la plus âgée de la population, et un demi-tarif d'1/4 pour les plus jeunes. L'âge critique choisi par l'entreprise est donc trop élevé.

Dans l'analyse qui précède, l'entreprise qui discrimine se comporte comme une entreprise offrant différents produits dont les demandes seraient parfaitement indépendantes (les coûts pouvant être dépendants dans le cas où les rendements ne sont pas constants). En pratique, la discrimination basée sur des critères observables implique souvent une production multi-produit. La section suivante est consacrée aux applications de la tarification discriminante dans de tels contextes.

3.3 Discriminer avec plusieurs produits

Lorsqu'une entreprise vend plusieurs biens dont les demandes sont parfaitement indépendantes, la situation est formellement très similaire à celle qu'on vient de discuter. La principale différence est qu'en général le coût marginal pour chaque produit diffère, alors qu'il est le même pour toutes les catégories de consommateurs dans le cas d'un unique produit. Dans le cas de plusieurs produits, il y a un potentiel pour une discrimination à l'envers où des biens différents sont vendus à des prix identiques ou anormalement proches. Comme nous l'avons vu dans l'introduction, la question de savoir ce qu'est une différence de prix normale est difficile à trancher de manière générale.

L'objectif de cette sous-section est de discuter brièvement quelques applications habituellement considérées comme des cas de tarification discriminante et où nous pourrons définir clairement ce qu'on peut considérer comme de la discrimination. Nous étudierons plus particulièrement la discrimination spatiale ; puis nous discuterons brièvement d'autres possibilités comme la discrimination dans le temps ou les ventes liées. Dans toutes les situations que nous discutons ici, le groupe de clients susceptibles d'acheter l'un des biens offerts est clairement identifié et il n'y a donc pas d'arbitrage individuel. L'utilisation d'une offre multi-produit à des fins de discrimination en présence d'arbitrage individuel sera étudiée dans la section suivante. Nous développons d'abord la théorie de la tarification optimale d'une entreprise qui vend à des consommateurs situés à des distances différentes de son lieu de production. Puis nous montrerons en quoi cette théorie est pertinente pour une entreprise qui doit transporter des voyageurs sur des distances différentes, et nous suggérons quelques autres applications dans le domaine des transports.

Supposons comme situation de référence que les consommateurs ont accès à un service de transport à prix concurrentiel auquel ils peuvent avoir recours pour transporter le bien depuis le lieu de production. En utilisant ce service de transport, un consommateur paye pour chaque unité du bien, le prix f.o.b. pratiqué par l'entreprise sur le site de production, plus le coût de transport[21].

En l'absence d'informations spécifiques sur les demandes individuelles, l'entreprise est alors réduite à pratiquer une tarification uniforme. Ceci correspond à la situation de référence de non-discrimination. Néanmoins, si l'entreprise peut s'assurer elle-même la livraison et si elle peut éviter que les acheteurs aient accès à des services de livraison concurrents, elle peut atteindre un objectif plus élevé, quand bien même le lieu de livraison ne révèlerait pas d'information sur la demande des acheteurs concernés.

À titre d'illustration, supposons que la demande est la même en tout point de l'espace. Si l'entreprise facture p_x aux acheteurs situés à une distance x, la demande est $D(p_x)$, et le surplus du consommateur correspondant est $SC(p_x)$. Les coûts de transport sont linéaires, et il en coûte $tx, t>0$, de transporter une unité sur une distance x. L'entreprise choisit p_x pour maximiser

$$SC(p_x)+(1+\lambda)\,D(p_x)\,(p_x-(c+tx)).$$

21. Le sigle f.o.b. renvoie à « free on board » ou à « freight on board » ; il signifie que l'acheteur doit payer les frais de transport selon le lieu de livraison. Dans le contexte de l'économie spatiale, le terme « f.o.b. price » est synonyme avec le terme « mill price » pour dire le prix à l'usine (ou même au magasin).

Le prix choisi sera donc similaire à celui de l'équation (5.1), p. 103 avec c remplacé par $c + tx$. On peut montrer (Anderson, 1986) que sous condition d'une demande plus « concave »[22] qu'exponentielle, l'accroissement dans le prix p_x causé par un accroissement de la distance, est moindre que tx. Cela signifie que l'entreprise ne répercute pas l'intégralité des coûts de transports dans le prix (« amortissement des coûts de transport »). Autrement dit, le prix f.o.b. implicite (c'est-à-dire le prix net des coûts de transport) est d'autant plus faible que les acheteurs sont éloignés du lieu de production.

L'amortissement des coûts de transport est couramment utilisé par les entreprises[23]. Tirole (1988) souligne qu'il y a de bonnes raisons pour qu'il le soit, même pour des courbes de demande qui ne satisfont pas la condition ci-dessus. D'une part, toute politique de prix qui surfacturerait les coûts de transports risquerait d'entraîner un arbitrage entre consommateurs : ceux qui sont proches de l'usine ayant intérêt à acheter le bien et à le transporter eux-même pour le revendre aux acheteurs plus éloignés. D'autre part, il est peu vraisemblable que la demande soit indépendante de la distance. En effet, les acheteurs éloignés du site de production de l'entreprise ont d'autant plus de chance d'être proches de sites de production concurrents et auront donc une demande plus élastique. De ce fait l'entreprise sera incitée à pratiquer un prix plus faible pour ces consommateurs[24].

Notons pour finir que le bénéfice qu'il y aurait à forcer une entreprise privée à pratiquer un prix f.o.b. uniforme est ambigu. L'entreprise qui discrimine sert une zone géographique plus grande (ce qui accroît le bien-être social) mais elle le fait en imposant des prix plus élevés aux consommateurs proches. Pour le cas d'une demande linéaire, le bien-être social est plus élevé avec une tarification discriminante quand $\lambda = 0$ (voir Holahan, 1975) — et donc à plus forte raison $\lambda > 0$, puisque le surplus du producteur est toujours plus élevé lorsque l'entreprise peut discriminer (le cas d'un prix uniforme étant un cas particulier de la tarification discriminante)[25].

Considérons maintenant quelques applications dans le domaine des transports. Reprenons tout d'abord l'exemple donné dans l'intro-

22. Voir Anderson et Renault (2003a) pour une exposition précise de cette notion. La condition est aussi équivalente à celle mise à jour par Seade (1987) sur l'élasticité de la pente de la demande inverse.

23. Greenhut (1981) présente des données d'enquête sur la tarification spatiale des entreprises.

24. Voir Lederer et Hurter (1986) pour un modèle de discrimination spatiale en concurrence, et Anderson, de Palma et Thisse (1992, Ch. 8 et 9) pour une revue de la littérature sur l'économie spatiale.

25. Voir Greenhut, Norman et Hung (1987) pour d'autres développements dans le cas de monopole, et Anderson, de Palma et Thisse, (1989 et 1992b) pour le cas concurrentiel. Quinet (1998) décrit en détail l'exemple de la concurrence entre deux aéroports.

duction d'une compagnie aérienne qui vend des voyages sur un vol long-courrier, par exemple entre Paris et New York, à des clients qui sont susceptibles de rallier le point de départ sur des vols moyens-courriers plus ou moins longs (provenant par exemple des aéroports des différentes régions françaises). Notre situation de référence est celle où les vols entre les différents aéroports régionaux et Paris sont vendus sur un marché concurrentiel et donc tarifés au coût marginal. Supposons que la compagnie puisse assurer elle-même le transport exclusif jusqu'à Paris, que la demande pour un voyage à New York soit la même quelle que soit la provenance des voyageurs (Paris, Bordeaux, Nice…) et que cette demande ne soit pas « trop convexe ». Le principe d'amortissement des coûts de transport se traduit par des billets plus chers pour les clients qui sont plus loin de Paris (à cause du coût plus élevé du service) sans que pour autant les différences de prix reflètent intégralement les différences de coûts pour servir ces clients.

De ce point de vue, les clients les plus éloignés sont subventionnés par ceux qui sont plus proches.

Plus généralement, notre analyse s'applique à la tarification d'un réseau ferroviaire (ou encore d'un réseau de transport urbain). Considérons deux trajets de longueur différentes qui impliquent donc des coûts différents. Selon la définition de Phlips (1983), la tarification est discriminante si la différence de prix entre deux trajets n'est pas égale à la différence de coûts. Or, d'après les arguments théoriques présentés plus haut, si la demande pour ces deux trajets est la même (afin d'éliminer toute différence de tarification due à des différences dans la demande), et si la courbe de demande n'est pas trop convexe, on trouverait que dans la tarification optimale les prix doivent être moins différenciés que les coûts. Autrement dit, en fonction de la distance parcourue, il faut que les prix s'accroissent moins vite que les coûts de transport.

Bien sûr, l'analyse de la tarification des transports urbains est bien plus complexe que l'esquisse présentée ici — la demande dépend typiquement de la distance parcourue et des autres moyens de transport disponibles, et la tarification doit prendre en compte la congestion[26] — mais ce cadre simplifié permet de souligner le rôle central de la discrimination tarifaire.

Il existe de nombreuses dimensions autres que l'espace qui permettent à une entreprise de discriminer entre acheteurs. C'est notamment le cas de la dimension temporelle. Une pratique courante des compagnies aériennes est d'offrir des rabais sur des aller-retours à cheval sur un week-end. Il est bien connu que cette pratique leur permet

26. Cet amortissement des coûts de transport sera d'autant plus important que l'entreprise de transport cherchera à redistribuer en faveur des voyageurs qui doivent parcourir des distances importantes.

de discriminer entre des voyages d'agrément pour lesquels la demande est relativement élastique (car il est toujours possible d'y renoncer ou d'utiliser des moyens de transport plus lents) et les voyages professionnels pour lesquels la demande est peu élastique car ils impliquent d'importantes contraintes de temps. De même, les tarifs différenciés selon l'heure et la date de départ permettent de prendre en compte les différences de congestion au cours du temps (principe de la tarification d'heure de pointe), mais ils peuvent aussi être utilisés pour discriminer entre des clients dont l'élasticité de la demande diffère[27]. Dans ces deux cas, l'entreprise profite du fait qu'elle propose des services différents pour discriminer entre ses clients.

Pour finir, mentionnons une autre pratique qui est souvent citée comme un exemple de discrimination du troisième degré. Les ventes combinées de deux biens apparaissent comme un moyen efficace de discriminer (voir Stigler, 1963). Elles permettent de distinguer les acheteurs qui sont intéressés par les deux biens ou services vendus conjointement de ceux qui ne sont intéressés que par l'un des deux. Par exemple, la vente de forfaits « train + hôtel » permet de faire payer moins cher le trajet en train aux touristes intéressés par ce forfait qu'aux personnes qui voyagent pour affaires, qui ont leur propre moyen de se procurer un hébergement (ou qui font l'aller-retour dans la journée).

Les exemples ci-dessus montrent que la segmentation du marché d'une entreprise peut passer par des stratégies beaucoup plus élaborées que la simple vérification de justificatifs administratifs comme les pièces d'identité. Néanmoins, les stratégies qui ne reposent pas sur l'utilisation d'une information strictement vérifiable sont susceptibles d'être contrariées non seulement par l'arbitrage entre acheteurs mais aussi par l'arbitrage individuel : un acheteur bénéficiant du tarif qui ne lui est pas destiné.

Les limites qu'impose à la tarification discriminante l'arbitrage individuel ont été soulignées dès le milieu du dix-neuvième siècle par l'ingénieur des ponts et chaussées Jules Dupuit. Dans son célèbre article sur les péages appliqués aux voies de communication (1849), il développe l'exemple de ce qui peut être accompli en pratiquant une tarification discriminante pour le péage d'une passerelle piétonne.

Il suppose qu'une importante clientèle ouvrière trouverait avantage à utiliser la passerelle, mais à un prix de 1 centime par passage qui, s'il

27. Il y a de nombreux travaux qui traitent de l'ajustement de la tarification afin de gérer la congestion, à la suite de l'article fondateur de Vickrey (1969). Voir par exemple de Palma et Lindsey (2000) pour un traitement de ce problème dans une situation de concurrence entre deux routes à péage, et de Palma et Lindsey (1998) pour l'étude du rôle de l'acquisition d'information dans ce contexte.

était appliqué à l'ensemble des utilisateurs, serait insuffisant pour couvrir le coût de mise en place de la passerelle.

Afin de pratiquer un prix pour les ouvriers qui soit différent de celui payé par le reste de la population, Dupuit suggère de mettre en place une discrimination sur la base de la tenue vestimentaire en stipulant que « Pour le passant en casquette, en blouse ou en veste, le péage est réduit à 1 centime » (p. 220) au lieu de 5 centimes pour les autres passants. Mais il note aussitôt : « ...il est très possible que la recette (...) soit diminuée d'une certaine somme, parce qu'un certain nombre de passants à 5 centimes profiteront, grâce à leur costume, de la réduction qui ne leur est pas destinée » (p. 220). Afin de réduire cet arbitrage individuel potentiel, il propose de n'appliquer la réduction qu'à certaines heures (où les ouvriers sont susceptibles de passer) ou d'exiger que les ouvriers présentent leur livret. Dans la section suivante, nous présentons une analyse systématique des stratégies qui permettent d'éviter l'arbitrage individuel.

4 Effets de l'arbitrage individuel

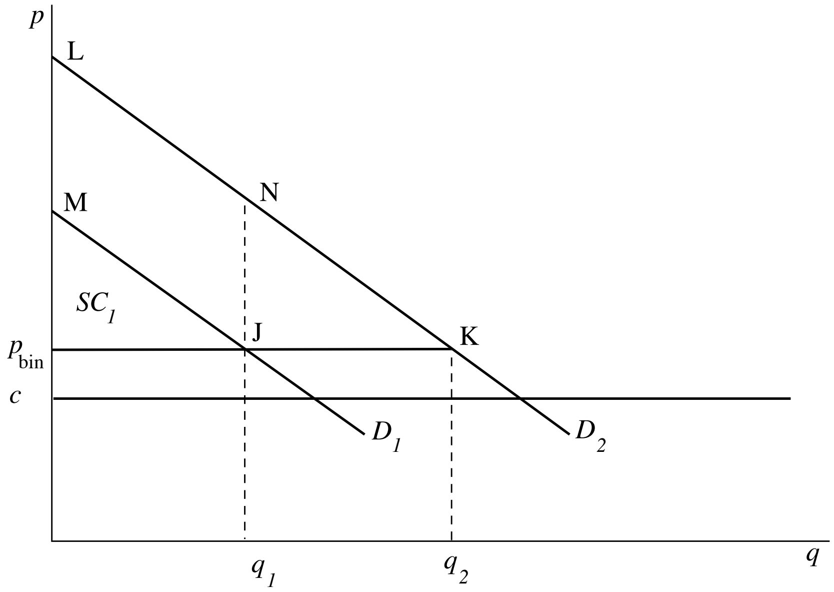
On ne peut pas classer les voyageurs comme des marchandises par leur caractère extérieur, on est obligé de les laisser se classer eux-mêmes (Jules Dupuit, 1849).

La discrimination en présence d'arbitrage individuel est en fait un problème de révélation d'information. L'entreprise sait qu'il y a une hétérogénéité dans la disposition à payer parmi les acheteurs, mais elle ne dispose pas d'un moyen de connaître cette information directement. Dans l'exemple de la passerelle de Dupuit (1849), ceux qui ne sont prêts à payer qu'un centime pour traverser peuvent être des personnes à revenu modeste qui ne sont pas des ouvriers disposant d'un livret, ou encore des personnes qui ont peu d'intérêt à traverser le fleuve et y renonceront s'il leur en coûte trop d'argent.

Ce problème de révélation d'information ne pourra être résolu que si la demande individuelle peut varier avec le prix, ou si l'entreprise est en mesure de moduler certaines caractéristiques de son produit, lesquelles sont appréciées différemment par différents types de consommateurs.

Dans l'exemple de la passerelle, Dupuit a déjà noté que la seconde possibilité pouvait être utilisée en tarifant différemment selon l'heure de la journée. Pour voir comment la vente de quantités différentes peut être exploitée dans ce même exemple, supposons que l'on vende aux ouvriers un ticket valable pour 6 aller-retours par semaine au prix de 12 centimes. Si les autres utilisateurs ne font pas plus d'un aller-retour, ils préfèreront continuer à payer 5 centimes par passage plutôt que d'acheter le ticket destiné aux ouvriers. Ce type de tarification, dit « non linéaire », est décrit dans les sous-sections 4.1 et 4.2 ; la sous-section 4.3

Figure 5.3: Caractère sous-optimal du tarif binôme
SC_1 : surplus du consommateur de type 1 (aire triangulaire comprise entre la courbe de demande inverse D_1 et le prix p_{bin}).



traite de l'utilisation d'une offre multi-produit en présence d'arbitrage individuel.

4.1 Tarif binôme avec clientèle hétérogène

Comme nous l'avons vu plus haut (3.1), un tarif binôme spécifie un droit d'accès A et un prix marginal p auquel ceux qui ont payé le droit d'accès peuvent se procurer le bien. Lorsque l'entreprise connaît parfaitement les goûts du consommateur, elle peut utiliser ce genre de tarification pour s'approprier l'intégralité du surplus du consommateur et ainsi discriminer parfaitement. Bien qu'en pratique les entreprises de transport ne disposent pas d'une information aussi parfaite, elles ont très largement recours à ce genre de tarification. C'est le cas par exemple de la SNCF qui propose différentes formules d'abonnement sur un trajet donné, ou de la RATP qui, avec la carte orange, propose un abonnement dont le prix marginal est nul. Nous allons voir maintenant dans quelle mesure un tarif binôme peut s'avérer utile lorsque l'entreprise est incertaine quant aux goûts du consommateur (voir aussi Oi, 1971).

Supposons qu'il existe une fraction α de consommateurs dont la demande au prix p est donnée par $D_1(p)$, tandis que le reste des consommateurs demande $D_2(p)$ au prix p. On suppose que $D_1(p) < D_2(p)$ pour tout $p \geq 0$ (les deux courbes de demande sont représentées sur la figure

5.3). Il s'ensuit que les surplus du consommateur correspondants sont tels que $SC_1(p) < SC_2(p)$ pour tout $p \geq 0$.

En situation d'information complète, l'entreprise choisirait un prix marginal $p = c$, et ferait payer des droit d'accès de $A_1 = SC_1(c)$ aux consommateurs de type 1 et $A_2 = SC_2(c)$ aux consommateurs de type 2. Les quantités consommées seraient alors $q_1^* = D_1(c)$ pour les consommateurs de type 1 et $q_2^* = D_2(c)$ pour les consommateurs de type 2. Les quantités consommées dans le cadre de cette solution de discrimination parfaite sont souvent appelées quantités optimales de premier rang[28]. En information incomplète, si l'entreprise proposait ces deux options, les acheteurs de type 2 choisiraient de payer l'abonnement A_1 puisqu'il est moins cher et donne accès aux mêmes droits : il y aurait donc arbitrage individuel.

Une solution simple pour éviter ce type d'arbitrage est de proposer un unique tarif binôme. Si l'entreprise propose uniquement la tarification d'information complète destinée aux acheteurs de type 2, les acheteurs dont la disposition à payer est faible n'achèteraient pas.

Si ces derniers sont assez nombreux (α suffisamment grand), ceci ne peut constituer la meilleure solution. En offrant au contraire le tarif de premier rang destiné aux acheteurs de type 1, l'entreprise met en œuvre une allocation qui maximise la somme des surplus du consommateur et du producteur (puisqu'ici le prix marginal est égal au coût marginal, et les deux types d'acheteurs participent).

Ceci constituerait une solution optimale pour une entreprise publique, en l'absence d'un coût marginal des fonds publics ($\lambda = 0$). Si le prélèvement de fonds publics est coûteux, ($\lambda > 0$), l'entreprise doit aussi se préoccuper de la recette qu'elle peut générer. Bien que, comme ce serait le cas en discrimination parfaite, l'entreprise s'approprie l'ensemble du surplus des consommateurs de type 1, elle doit laisser un surplus strictement positif aux consommateurs qui ont une disposition à payer élevée. Nous allons voir maintenant qu'il existe alors un tarif binôme qui permet d'atteindre un meilleur résultat.

L'objectif de l'entreprise peut être décomposé en deux composantes : le surplus social associé aux acheteurs de type 1

$$SC_1(p) + (1+\lambda)(p-c)D_1(p) + \lambda A \quad \text{(affecté du poids } \alpha\text{);}$$

et le surplus social associé aux consommateurs de type 2

$$SC_2(p) + (1+\lambda)(p-c)D_2(p) + \lambda A \quad \text{(affecté du poids } 1-\alpha\text{).}$$

28. Ce sont en effet les quantités socialement optimales si on suppose un système fiscal qui permet de redistribuer les ressources de l'économie de manière forfaitaire, auquel cas le coût marginal des fonds publics est nul.

Chacune de ces composantes comprend le surplus du consommateur et le surplus du producteur (tous deux avant paiement de l'abonnement), et l'abonnement, dont la valeur par euro est λ puisque c'est un transfert des consommateurs à l'entreprise.

Tout d'abord, quel que soit le prix marginal p, il est toujours souhaitable de prélever l'intégralité du surplus des consommateurs de type 1 (quand $\lambda > 0$). En effet, si $A < SC_1(p) < SC_2(p)$, un léger accroissement de l'abonnement n'aura aucun impact sur les quantités consommées et assurera une recette supplémentaire. L'objectif de l'entreprise s'en trouvera alors amélioré. Elle choisit donc, comme dans le cas en information complète, $A = SC_1(p)$, de sorte que les consommateurs ayant une faible disposition à payer sont juste indifférents entre s'abonner ou ne pas s'abonner. Néanmoins, nous allons voir que, contrairement à la solution en information complète, elle a intérêt à pratiquer un prix marginal au-dessus du coût marginal. Il est vrai que l'entreprise atteint un objectif maximum pour les consommateurs de type 1 en choisissant $p = c$ (puisqu'elle récupère l'ensemble du surplus social, elle a intérêt à ce qu'il soit maximisé). Mais la tarification au coût marginal ne maximise pas le surplus social associé aux consommateurs de type 2. Celui-ci peut s'écrire

$$SC_2(p) + (1+\lambda)\,(p-c)\,D_2(p) + \lambda SC_1(p),$$

où le dernier terme prend en compte le fait que, si on accroît le prix marginal, il faut baisser l'abonnement pour assurer la participation des acheteurs dont la demande est faible. La dérivée par rapport à p évaluée en $p = c$ est

$$\lambda\,(D_2(c) - D_1(c)) > 0.$$

L'entreprise pourrait donc accroître son objectif en augmentant le prix marginal. Comme l'impact d'un tel accroissement sur le surplus social des acheteurs de type 1 est négligeable, (la dérivée en $p = c$ étant nulle) cette augmentation de prix est souhaitable pour l'entreprise. Le prix marginal optimal p_{bin} est strictement compris entre le coût marginal et le prix qui maximise le surplus social pour les consommateurs de type 2. Ce dernier prix est inférieur au prix de Ramsey-Boiteux correspondant à la demande D_2, du fait de l'impact négatif d'une hausse du prix marginal sur le taux de l'abonnement.

Il est facile d'établir que si on exploite pleinement les possibilités d'une tarification non linéaire (sans se restreindre à une tarification affine) la solution décrite ci-dessus n'est pas optimale.

On voit sur la figure 5.3 que les consommateurs dont la disposition à payer est élevée bénéficient d'un surplus positif (une rente) mesuré par l'aire KNLMJ.

Nous allons voir qu'il est possible de réduire cette rente sans affecter les quantités consommées. Supposons qu'au lieu d'offrir un tarif binôme l'entreprise donne à chaque consommateur l'option de consommer soit la quantité $q_1 = D_1(p_{\text{bin}})$ au tarif $T_1 = p_{\text{bin}} q_1 + SC_1(p_{\text{bin}})$, soit la quantité $q_2 = D_2(p_{\text{bin}})$ au tarif $T_2 = p_{\text{bin}} q_1 + SC_1(p_{\text{bin}})$, plus l'aire KNJ. Les consommateurs de type 1 trouveront alors la combinaison (q_2, T_2) trop onéreuse et choisiront donc (q_1, T_1). Quant aux consommateurs de type 2, s'ils choisissent (q_1, T_1) leur surplus est donné par l'aire JNLM et ils auraient un surplus au moins aussi élevé s'ils payent T_2 pour la quantité q_2. Il est bien sûr intéressant pour l'entreprise de laisser aux consommateurs dont la disposition à payer est élevée la plus petite rente possible, c'est-à-dire exactement JNLM.

Comme nous allons le voir, la tarification que nous venons de décrire, bien qu'elle domine le simple tarif binôme, n'est pas en général la tarification optimale. Néanmoins, elle en a un certain nombre des caractéristiques : les consommateurs de type 1 ont un surplus nul et consomment une quantité inférieure à celle de premier rang (celle qui serait consommée avec une tarification au coût marginal) ; les consommateurs de type 2 ont un surplus strictement positif et sont indifférents entre les deux options offertes.

C'est cette dernière condition d'indifférence des acheteurs à forte disposition à payer qui est au centre du problème incitatif qui devra être résolu pour déterminer la tarification non-linéaire optimale[29].

4.2 Tarification non-linéaire optimale

Reprenons l'analyse là où nous l'avons laissée ci-dessus, et considérons une tarification qui spécifie un choix entre deux combinaisons quantité/prix (q_1, T_1) et (q_2, T_2). Jusqu'à présent, nous ne nous sommes jamais heurtés à des difficultés liées à un arbitrage individuel de la part des acheteurs dont la demande est faible. La prise en compte de la possibilité d'un tel arbitrage pourrait compliquer grandement l'analyse du cas général. Il s'avère que cette possibilité n'affecte pas la solution optimale dans le problème qui nous occupe et nous allons donc l'ignorer pour ne pas surcharger inutilement la présentation[30].

Dans la suite de l'analyse, puisque les quantités consommées ne sont plus déterminées par des prix marginaux, il sera plus commode d'écrire le surplus brut en fonction de la quantité consommée (cf. 3.1).

29. Pour plus de développements sur ce sujet voir Brown et Sibley (1986), ainsi que Wilson (1992).

30. Il est en fait facile d'établir que la solution que nous allons déterminer n'est pas sujette à l'arbitrage par les acheteurs de type 1, de sorte qu'elle resterait la solution optimale si on introduisait une contrainte supplémentaire pour prendre en compte un tel arbitrage.

À cette fin, nous dénoterons $V_i(q)$ le surplus brut d'un acheteur de type i qui consomme une quantité q.

Si l'arbitrage individuel est le fait des seuls acheteurs dont la demande est élevée, il est alors évident qu'il n'y a aucun intérêt à laisser un surplus positif à ceux dont la demande est faible. En effet, si ce surplus était positif, un accroissement du tarif T_1 permettrait d'accroître la recette sans affecter les quantités consommées, puisque cela rendrait l'achat de la quantité q_1 — et donc l'arbitrage individuel — d'autant moins attractif pour les acheteurs à forte demande. Cette condition de surplus nul pour les acheteurs à faible disposition à payer s'écrit

$$T_1 = V_1(q_1).$$

En revanche, on doit laisser un surplus strictement positif aux acheteurs à forte demande car, si la combinaison (q_1, T_1) est acceptable pour les acheteurs de type 1 (et leur procure donc un surplus d'au moins zéro), les acheteurs de type 2 peuvent s'assurer un surplus strictement positif en choisissant cette combinaison. Si on souhaite qu'ils choisissent une combinaison alternative (q_2, T_2) , elle devra leur procurer un surplus au moins égal. Mais il est aussi évident qu'il n'est pas nécessaire de leur assurer un surplus strictement supérieur car on pourrait alors accroître le tarif T_2 sans avoir à modifier le reste de la tarification. Cette condition d'indifférence s'écrit

$$V_2(q_2) - T_2 = V_2(q_1) - T_1,$$

ou encore, en utilisant l'égalité $T_1 = V_1(q_1)$,

$$T_2 = V_2(q_2) - \Big[V_2(q_1) - V_1(q_1)\Big].$$

Le terme entre crochets mesure la rente dite « informationnelle » que peuvent s'assurer les acheteurs à demande élevée grâce à l'arbitrage individuel. S'il y a un coût d'opportunité des fonds publics ($\lambda > 0$), cette rente représente un coût pour l'entreprise.

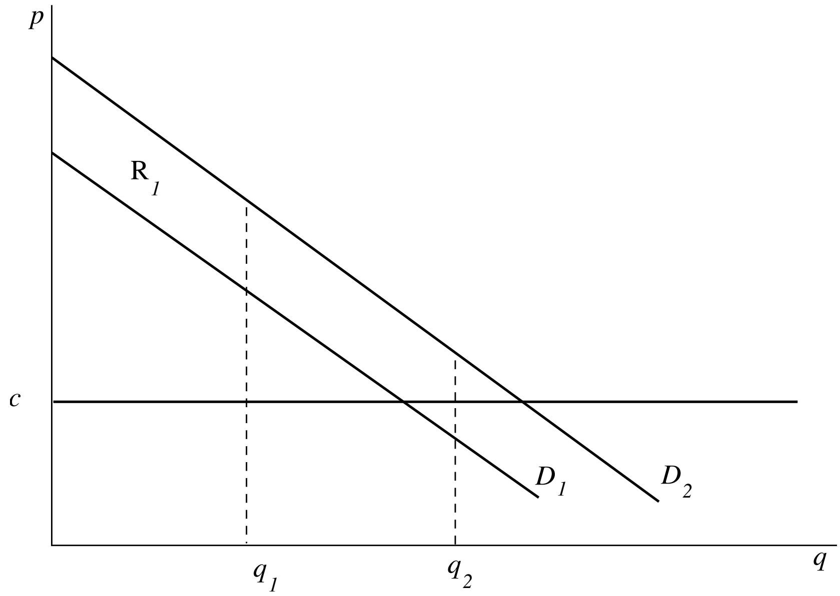
La figure 5.4 illustre ces contraintes pour des quantités q_1 et q_2 quelconques. La rente d'information R_1 est d'autant plus élevée que la quantité consommée par les acheteurs de type 1 est importante.

Pour ce qui est du choix de la quantité q_2, on voit sur cette figure[31] que si $q_2 < q_2^*$, un accroissement de cette quantité augmente le surplus social pour un acheteur de type 2 ainsi que la recette de l'entreprise (la rente d'information restant quant à elle inchangée) et il n'est pas nécessaire de modifier l'allocation destinée aux acheteurs de type 1. Par un

31. Nous choisissons une valeur de référence pour la variation de q_2, soit q_2^*, valeur de q pour laquelle $D_2 = c$.

Figure 5.4 : Tarifs et rente d'information

Le tarif T_1 payé par les acheteurs de type 1 est le surplus brut de ces acheteurs, c'est-à-dire l'aire comprise sous leur courbe de demande D_1. La rente d'information R_1 est donnée par l'aire située entre les deux courbes de demande pour des quantités q comprises entre 0 et q_1. Le tarif T_2 payé par les acheteurs de type 2 est leur surplus brut amputé de la rente informationnelle.

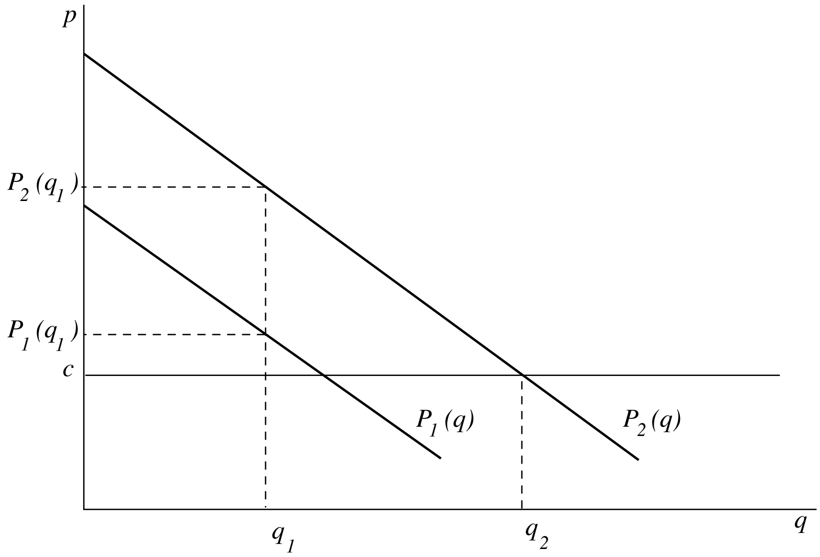


argument symétrique, si $q_2 > q_2^*$, l'entreprise peut améliorer son résultat en diminuant la quantité destinée aux acheteurs à forte demande[32]. L'entreprise choisit donc de produire la quantité socialement optimale de premier rang (obtenue en ignorant le coût marginal des fonds publics) pour les acheteurs dont la demande est élevée. Ceci est le cas en dépit d'un coût marginal des fonds publics positif. En effet, la recette qui peut être générée par la vente à des acheteurs à forte disposition à payer dépend du surplus social et de la rente informationnelle. Puisque cette dernière dépend uniquement de la quantité vendue aux acheteurs à faible demande, il est optimal de choisir la quantité pour les acheteurs de type 2 afin de maximiser le surplus social correspondant.

Puisque l'entreprise est en mesure de récupérer l'intégralité du surplus social lorsque l'acheteur a une faible disposition à payer, elle aurait intérêt à ce que ce type d'acheteur consomme $q_1^* = D_1(c)$. Elle obtient ainsi la recette fiscale maximale, qui coïncide avec le surplus social maximal. Néanmoins, puisqu'elle est incertaine sur le type de l'ache-

32. Rappelons que l'aire comprise entre le coût marginal et la courbe de demande correspond à un surplus négatif lorsque la seconde est située en-dessous du premier.

Figure 5.5: Tarification non-linéaire optimale
La distance verticale entre c et $P_1(q_1)$ est proportionnelle à $1/(1+\lambda)\ \alpha$, et la distance verticale entre $P_1(q_1)$ et $P_2(q_1)$ est proportionnelle à $1/\lambda(1-\alpha)$.



teur, elle doit aussi prendre en compte l'impact de la quantité q_1 sur la rente qu'elle devra consentir à l'acheteur, dans l'éventualité où sa demande est élevée. La prise en compte de cette éventualité la conduit à choisir une quantité q_1 plus faible que l'optimum de premier rang q_1^*. En q_1^*, l'impact d'une baisse de la quantité sur le surplus social pour un consommateur de type 1 est négligeable (la dérivée est nulle), tandis qu'une telle baisse permet de réduire la rente laissée à un consommateur de type 2. Une telle baisse est donc souhaitable.

Plus généralement, une diminution d'une unité de la quantité q_1 induit d'une part une baisse du surplus social associé aux consommateurs de type 1 et d'autre part une baisse de la rente d'information des consommateurs de type 2. On peut voir sur la figure 5.5 que la baisse de surplus social serait de $P(q_1) - c$ qui a une valeur sociale de $(1+\lambda)$ par euro, parce cette somme est entièrement récupérée par l'entreprise. La diminution de la rente d'information, qui d'après la graphique est de $P_2(q_1) - P_1(q_1)$, constitue un transfert des consommateurs de type 2 vers l'entreprise, et sa valeur sociale par euro est donc de λ. La quantité optimale peut donc être déduite du graphique : c'est la valeur de q_1 pour laquelle le rapport de la distance verticale entre la demande la plus faible et le coût marginal à la distance verticale entre les deux courbes de demande est égal à

$$\frac{\lambda}{1+\lambda}\ \frac{1-\alpha}{\alpha}.$$

Formellement, la firme choisit q_1 afin de maximiser

$$\begin{aligned}\alpha\Big[SC_1+(1+\lambda)SP_1\Big] \quad &- \quad (1-\alpha) \quad \lambda\Big[V_2(q_1)-V_1(q_1)\Big] = \\ \alpha\Big[(1+\lambda)(V_1(q_1)-cq_1)\Big] \quad &- \quad (1-\alpha) \quad \lambda\Big[V_2(q_1)-V_1(q_1)\Big],\end{aligned}$$

où le dernier terme mesure l'impact de la rente informationnelle sur l'objectif de l'entreprise lorsque l'acheteur est de type 2. Les conditions du premier ordre peuvent s'écrire

$$V_1'(q_1) = c + \frac{\lambda}{1+\lambda}\,\frac{1-\alpha}{\alpha}\,\Big[V_2'(q_1)-V_1'(q_1)\Big].$$

Afin d'interpréter ce résultat et de le comparer aux formules de tarification précédentes, il est utile de le réécrire en utilisant le fait que la dérivée du surplus brut de type i est le prix $P_i(q_1)$, donné par la demande inverse qui est le prix uniforme auquel un consommateur de type i choisirait de consommer q_1. On peut donc écrire la formule de taux de marge suivante :

$$\frac{P_1(q_1)-c}{P_1(q_1)} = \frac{\lambda}{1+\lambda}\,\frac{1-\alpha}{\alpha}\,\frac{\big[P_2(q_1)-P_1(q_1)\big]}{P_1(q_1)}.$$

Cette formule permet une comparaison avec la tarification uniforme de Ramsey-Boiteux. Cette dernière tarification induit les consommateurs à consommer moins que la quantité optimale de premier rang afin de pouvoir générer un surplus du producteur en tarifant au-dessus du coût marginal. La possibilité de générer un tel surplus dépend inversement de l'élasticité-prix de la demande : plus la demande est élastique, moins il sera possible d'accroître le prix sans induire une baisse de quantité trop importante. En tarification non linéaire, la diminution de la quantité consommée par les consommateurs à faible disposition à payer a un tout autre rôle. Elle permet de réduire la rente informationnelle que l'on doit consentir aux acheteurs dont la demande est élevée, afin de les inciter à révéler cette demande. La distorsion sera donc d'autant plus importante que sont élevés l'impact sur la rente, mesuré par $\big[P_2(q_1)-P_1(q_1)\big]$, et la part relative des acheteurs à forte demande, mesurée par $(1-\alpha)/\alpha$. Puisque la rente d'information est d'autant plus importante que la quantité basse est élevée, les acheteurs à forte disposition à payer ont intérêt à ce que la distorsion de la quantité basse soit la plus faible possible ; leur bien-être est donc d'autant plus élevé qu'ils représentent une faible proportion des clients potentiels.

Pour résumer, nous avons montré qu'il est optimal pour l'entreprise de produire la quantité optimale *en discrimination parfaite* à destination des acheteurs à *forte demande*, tandis que la quantité vendue aux acheteurs à *faible demande* est *plus faible qu'en discrimination parfaite.* Elle *récupère l'intégralité du surplus* de ces derniers, tandis qu'elle *laisse une rente* aux premiers. Enfin, cette rente est d'autant plus importante que la clientèle est constituée d'un nombre important d'acheteurs dont la demande est faible.

Bien qu'il soit clair qu'en général le prix unitaire diffère d'un groupe d'acheteurs à l'autre, rien ne permet d'affirmer qu'il est décroissant dans la quantité. En pratique, on observe souvent que ceux qui consomment davantage bénéficient de rabais. Par exemple, les formules d'abonnement ou les programmes de fidélité dans les transports permettent de faire payer moins cher à ceux qui voyagent fréquemment. L'analyse présentée ici ne procure pas d'explication immédiate à ce type de rabais — en partie parce qu'elle ignore la possibilité pour les gros acheteurs d'acheter plusieurs lots destinés aux petits acheteurs (sur ce sujet voir Alger, 1999). Plus généralement, les implications empiriques de la tarification non linéaire méritent une étude approfondie et rigoureuse (ceci a fait l'objet d'un certain nombre de travaux récents, cf. Cohen, 2002, Ivaldi et Martimort, 1994, Leslie, 2003, McManus, 2003).

4.3 Qualités multiples et discrimination

Bien qu'il existe dans le transport de voyageurs certaines formes de tarifications non-linéaires, une autre pratique très courante est d'offrir des qualités de service différentes sur un même trajet. Cette pratique est déjà ancienne et a fait l'objet d'une analyse extrêmement pénétrante de Jules Dupuit dans son article de 1849 :

> (...) il est clair (...) qu'en multipliant indéfiniment les classes, on pourrait faire payer aux consommateurs toute l'utilité qu'ils retirent du chemin.
>
> Mais pour cela il faut pouvoir distinguer les consommateurs qui attachent une utilité différente à leur transport et les obliger à se classer volontairement dans telle ou telle catégorie de tarif. Or c'est là une grande difficulté, qui donne lieu à une foule de mesures en général fort mal comprises du public.
>
> Ainsi, que de gens, en voyant les voyageurs de 3ème classe tantôt découverts, tantôt mal suspendus, toujours mal assis, ont crié à la barbarie des compagnies. Il en coûterait si peu, dit-on, pour mettre là quelques mètres de cuir, quelques kilogrammes de crin, qu'il y a plus que de l'avarice à les refuser.
>
> (...) ce n'est pas à cause des quelques milliers de francs qu'il serait nécessaire de dépenser pour couvrir les wagons de 3ème classe ou pour en rembourrer les banquettes, que telle compagnie a des wagons découverts et des banquettes de bois ; elle ferait volontiers ce sacrifice à sa popularité. Son but est d'empêcher le voyageur qui peut payer le wagon de 2ème classe d'aller dans celui des 3ème. On frappe sur le pauvre, non pas qu'on ait envie de le faire souffrir personnellement, mais pour faire peur au riche. La preuve, c'est que si aujourd'hui l'Etat disait à cette compagnie : voilà 100.000 Francs pour mettre des impériales sur vos Wagons (...), cette subvention serait certainement refusée. (...) améliorer les wagons de 3ème classe [peut] diminuer la recette de 2 millions et ruiner la compagnie.
>
> Enfin, c'est encore par ce même motif que les compagnies, après s'être montrées presque cruelles pour les voyageurs de troisième classe, avares pour ceux de se-

> conde, deviennent prodigues pour ceux de première. Après avoir refusé le nécessaire au pauvre, on donne le superflu au riche.

C'était aussi le point de vue de Walras (1875) sur la logique qui présidait à l'élaboration de la tarification des chemins de fers français au milieu du dix-neuvième siècle.

> Les compagnies françaises demandent respectivement 10 c. à ceux de 1ère classe, 7,5 c. à ceux de 2ème classe et 5,5 c. à ceux de 3ème ; mais elles mettent dans une voiture 24 voyageurs de 1ère classe, 30 de 2ème et 40 de 3ème. Elles ajoutent à cela des sièges plus ou moins rembourrés, etc. (...)
>
> En réalité les compagnies considèrent, à tort ou à raison, le prix moyen de 7,66 c, prix assez voisin de celui de 7,5 c. qui est le prix des secondes, comme étant le prix de bénéfice maximum ; mais elles ne veulent pas négliger d'accepter plus des voyageurs disposés à payer plus, ni même refuser d'accepter moins des voyageurs décidés à ne pas payer trop.
>
> Quand on réclamait jadis à si grand cri la fermeture des voitures de 3ème classe par des vitres tel que l'a stipulé le cahier des charges de 1857-58, quand on réclame aujourd'hui le chauffage en hiver et qu'on se plaint à ce propos de la dureté des compagnies, on ne saisit pas leur vrai mobile.
>
> Si les voitures de 3ème classe étaient assez confortables pour que beaucoup de voyageurs de 2ème et quelques-uns de 1ère y allassent, le produit net total, tel qu'il se compose d'après la théorie du monopole, sera abaissé. Et voilà tout.
>
> Les compagnies n'ont de voitures de 3ème classe que pour ne pas laisser échapper un grand nombre de voyageurs peu aisés qui, plutôt que de payer le prix de la 1ère ou de la 2ème, auraient continué à voyager en diligence.

D'après Dupuit et Walras le choix du niveau de confort dans les différentes classes de chemin de fer est pour l'essentiel guidé par la volonté de faire payer aux utilisateurs des prix en rapport avec leurs dispositions à payer et par le souci d'éviter un arbitrage individuel de la part de ceux pour lesquels on souhaite pratiquer un tarif élevé. Un tel arbitrage est découragé par l'introduction d'une différence suffisante entre les classes.

Il est intéressant de faire un parallèle entre les arguments de ces auteurs et nos résultats pour la tarification non linéaire.

En effet, nous avons montré que l'arbitrage individuel potentiel de la part des usagers à forte demande conduit l'entreprise à introduire une différence entre quantités proposées supérieure à ce qu'elles seraient en discrimination parfaite. Dupuit et Walras suggèrent que les compagnies de chemin de fer adoptent une logique similaire en jouant sur le différentiel de qualités plutôt que sur la différence de quantités. Nous allons voir maintenant qu'il y a effectivement une équivalence formelle entre ces deux pratiques de discrimination tarifaire. Afin d'illustrer l'analogie entre la discrimination basée sur l'offre de qualités différentes et la tarification non linéaire d'un bien homogène, considérons le modèle suivant dû à Mussa et Rosen (1978). La présentation de ce modèle est aussi l'occasion d'introduire une approche graphique alternative à celle s'appuyant sur les courbes de demande.

Supposons qu'il y ait deux types d'usagers qui diffèrent quant à leur disposition à payer pour une amélioration de la qualité du service proposé. Plus précisément, supposons que leurs dispositions à payer respectives pour une unité supplémentaire de qualité est donnée par Θ_1 et Θ_2, où $\Theta_2 > \Theta_1$. La lettre q dénote maintenant le niveau de qualité du service proposé plutôt que la quantité vendue comme c'était le cas en tarification non linéaire. Le surplus brut d'un acheteur de type Θ_i consommant un service de qualité q est alors donné par $V_i(q) = \Theta_i q$. Chaque client ne souhaite acheter qu'une seule unité et sa disposition à payer est donnée par son surplus brut. Son surplus net pour un tarif T est $\Theta_i q - T$. Le coût marginal du service proposé est d'autant plus important que la qualité est élevée ; il est dénoté $c(q)$ pour un niveau de qualité q, où c est strictement croissante et strictement convexe.

On peut représenter graphiquement les combinaisons de la qualité q et du tarif T donnant un surplus net identique à un acheteur de type i ; ces courbes sont appelées courbes d'indifférence. Il s'agit de droites croissantes de pente Θ_i (avec les qualités q en abscisse), la courbe d'indifférence passant par l'origine correspondant à un surplus nul (désignée dans la suite par $CI_{i,0}$) ; le niveau de surplus est d'autant plus important que la courbe d'indifférence est plus à droite. Le surplus du producteur pour une unité de service (un voyage) de qualité q vendue à un tarif T est donné par $T - c(q)$. Si l'entreprise connaît la valeur de Θ_i pour chaque utilisateur, elle est en mesure de discriminer parfaitement et donc, de ne laisser aucun surplus à l'acheteur. Le surplus du producteur correspondant est $\Theta_i q - c(q)$ et l'entreprise choisit donc la qualité q_i^* qui maximise cette grandeur (le surplus social étant simplement le surplus du producteur affecté d'un poids $1 + \lambda$).

La figure 5.6 représente la courbe d'indifférence correspondant à un surplus du consommateur nul et la courbe du coût marginal en fonction de la qualité. La qualité optimale est celle qui maximise la distance verticale entre la courbe d'indifférence et la courbe de coût marginal. Ceci correspond à la valeur q_i^*, pour laquelle les deux courbes ont la même pente, ce qui s'écrit formellement $\Theta_i = c'(q_i^*)$ (la distance verticale entre la courbe d'indifférence et le coût marginal est toujours inférieure à celle entre la courbe d'indifférence et la tangente au coût marginal en q_i^*).

En discrimination parfaite, la qualité optimale égalise donc la disposition à payer pour plus de qualité au coût marginal de l'accroissement de cette qualité.

Si l'entreprise ne connaît pas la disposition de chaque acheteur à payer pour la qualité, elle ne peut offrir les deux qualités optimales de

Figure 5.6 : Tarif et qualité en discrimination parfaite

$CI_{i,0}$ est la courbe d'indifférence de pente Θ_i et passant par l'origine, correspondant à un surplus nul ; $c(q)$ est la courbe du coût marginal. Les q sont ici des qualités.

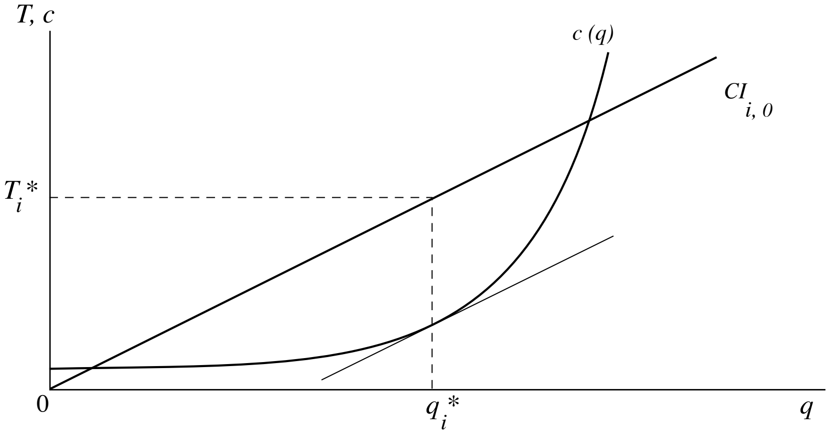
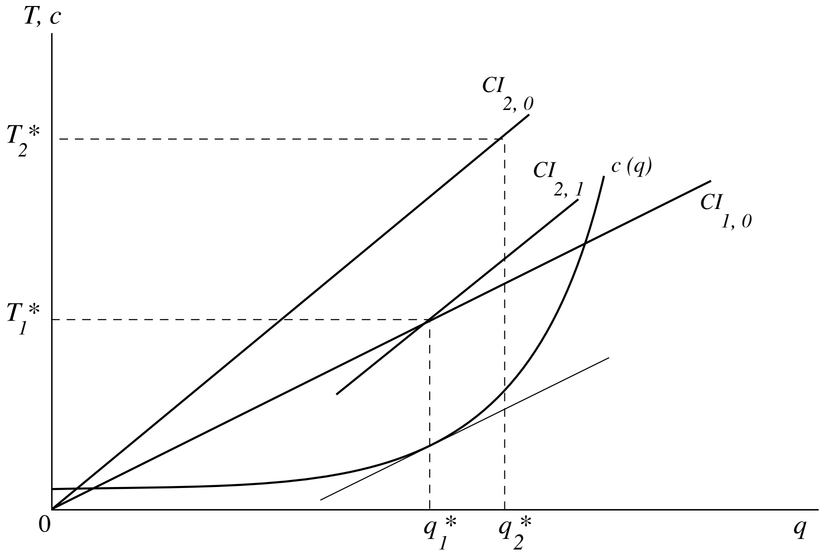


Figure 5.7 : Tarifs et qualités en présence d'arbitrage individuel

La rente d'information R correspond à la distance verticale entre la courbe d'indifférence $CI_{2,0}$ et la courbe d'indifférence $CI_{2,1}$.



premier rang q_1^* et q_2^* sans laisser de surplus à l'un ou l'autre type d'usager. On voit sur la figure 5.7 que la combinaison (q_2^*, T_2^*) se trouve à gauche de la courbe d'indifférence d'un usager de type 2 désignée par $CI_{2,1}$ passant par (q_1^*, T_1^*), et que cette dernière combinaison serait donc préférable pour lui. Afin de l'inciter à ne pas choisir la qualité basse, l'entreprise doit lui faire payer un tarif tel qu'il se retrouve au pire sur la courbe d'indifférence $CI_{2,1}$.

Comme c'était le cas en tarification non linéaire, c'est l'arbitrage individuel potentiel des acheteurs à forte disposition à payer qui constitue une contrainte pour l'entreprise. De nouveau, puisque l'entreprise n'est pas préoccupée par l'arbitrage individuel des usagers de type 1, elle peut les priver de l'intégralité de leur surplus de sorte que la combinaison optimale doit se situer sur la courbe d'indifférence désignée par $CI_{1,0}$. Quant aux acheteurs de type 2, ils bénéficient d'une rente d'information notée R, qui est mesurée par la distance verticale entre la courbe d'indifférence $CI_{2,0}$ et la courbe d'indifférence $CI_{2,1}$ (ce qui correspond à la différence entre le tarif de discrimination parfaite et ce qu'ils doivent effectivement payer).

Le surplus social associé aux consommateurs de type 2 est donc le surplus social de premier rang grevé de la rente d'information affectée du poids λ. Puisque cette rente ne dépend pas du niveau de la qualité élevée, le niveau optimal de celle-ci est q_2^*, comme en discrimination parfaite. En revanche, il est facile de voir que si, comme c'est le cas sur la figure 5.7, la qualité basse est q_1^*, l'entreprise a intérêt à la diminuer. En effet, ceci permet de réduire la rente d'information, puisque la courbe d'indifférence $CI_{2,1}$ se déplace vers la gauche, alors que, pour une diminution peu importante, l'impact sur le surplus social associé aux usagers de type 1 est négligeable puisque q_1^*, maximise celui-ci.

Ainsi, ce modèle confirme l'intuition de Dupuit (1849) : le confort des voitures de 3ème classe est délibérément détérioré afin de dissuader les voyageurs qui sont disposés à payer les tarifs des classes supérieures de chercher à voyager moins cher. En revanche, comme le souligne Tirole (1988, 1993), la qualité offerte aux voyageurs de première classe n'a rien de « superflu », puisqu'elle correspond à celle choisie en discrimination parfaite ; contrairement à ce que pensaient Dupuit et Walras, c'est uniquement en jouant sur la qualité basse que l'entreprise décourage l'arbitrage individuel. Du fait de l'analogie parfaite entre ce modèle et celui de la tarification non linéaire, on retrouve ici des résultats équivalents. Seuls les usagers dont la disposition à payer est élevée parviennent à préserver une partie de leur surplus, et celle-ci est d'autant plus importante qu'ils représentent une petite proportion de la clientèle totale.

Une variante de ce modèle, explorée par Leruth et Chander (1989) et particulièrement pertinente pour le secteur des transports, consiste à supposer que la qualité de chaque service proposé décroît lorsque le nombre d'utilisateurs augmente, ce qui se produit par exemple pour un bien sujet à congestion. L'entreprise choisit alors de proposer deux prix différents, le service le moins cher ayant une qualité plus basse du seul fait qu'il est choisi par davantage de gens. Les deux classes dans le métro parisien jusqu'au début des années 80 fournissent une illustration particulièrement frappante de ce type de stratégie[33]. Les voitures de seconde classe ne différaient de celles de première que par leur couleur, mais étaient accessibles à un prix plus faible[34].

Notons pour terminer que Dupuit et Walras pourraient trouver aujourd'hui au moins autant d'exemples susceptibles d'illustrer leurs arguments, aussi bien dans les transports ferroviaires que dans les transports aériens. La démocratisation de ces derniers a donné l'occasion à bon nombre de voyageurs de passer de longues heures de vol dans des conditions qui sont souvent considérées comme à la limite de l'acceptable. Nos commentateurs du dix-neuvième siècle seraient prompts à nous rappeler que c'est par ce moyen que les compagnies aériennes peuvent pratiquer des tarifs suffisamment élevés en première classe ou en classe business pour assurer la rentabilité du transport aérien[35].

Conclusions

Nous avons présenté dans ce chapitre la théorie de la tarification discriminante pour une seule entreprise (sans concurrents), et nous avons montré en quoi cette analyse est pertinente pour le secteur des transports.

Nous avons traité le cas où les fonds publics ont une valeur plus grande que le surplus du consommateur ($\lambda \geq 0$) pour prendre en compte l'inefficacité du prélèvement des fonds publics. Le poids plus important sur le surplus du producteur est justifié par le fait que les profits peuvent être récupérés par l'Etat, soit directement si l'entreprise est publique, soit parce qu'ils font l'objet d'une imposition. Le cas de la maximisation du profit par une entreprise privée apparaît comme un cas limite lorsque λ tend vers l'infini.

33. La présence sur un même itinéraire d'une voie routière à péage et d'une voie gratuite en est une autre illustration. Voir aussi de Palma et Lindsey (2000).

34. Ce système fut aboli par Charles Fiterman, ministre communiste des transports du premier gouvernement de l'Union de la Gauche.

35. En pratique, le problème auquel sont confrontées les entreprises de transport est potentiellement très complexe car l'information dont elle dispose sur la demande des différents voyageurs est très parcellaire et évolue au cours du temps. Pour élaborer leur tarification optimale, les entreprises ont recours à des méthodes algorithmiques, qui sont regroupées sous le terme de *Yield and revenue management*. Voir Sinsou (1999) pour une présentation détaillée de ces méthodes.

Ce cadre théorique nous permet de mettre en évidence la similarité entre le problème de tarification d'une entreprise publique et celui d'une entreprise privée et d'aborder des questions pertinentes pour les deux situations. Néanmoins, la restriction à une situation de monopole peut apparaître comme restrictive dans le cas des entreprises privées.

Bien que plusieurs modes de transport soient gérés par des entreprises en monopole (tels que le train, le bus, et le métro) il y a aussi des cas où la concurrence est présente (le transport aérien, par exemple) et l'introduction de plus de concurrence fait l'objet de nombreux débats. Nous passons maintenant à une brève revue du cas de la concurrence en oligopole. Pour la discussion qui suit, nous considérons des entreprises privées qui maximisent leur profit et les résultats sur le bien-être supposent $\lambda = 0$, qui veut dire que la valeur sociale d'un euro de surplus du producteur vaut un euro de surplus du consommateur[36].

Dans l'ensemble, il y a peu de travaux sur la tarification discriminante en situation de concurrence. Pour la discrimination en l'absence d'arbitrage individuel (cf. 3.2 et 3.3), la théorie de l'oligopole permet assez simplement d'obtenir des résultats. Dans le cas d'un seul produit vendu à plusieurs groupes, le modèle le plus directement comparable à celui du monopole est celui de Cournot (1838) où chaque firme choisit la quantité produite, et où le prix égalise la quantité totale à la demande. Dans ce modèle, chaque entreprise vend le même bien (hypothèse d'homogénéité du produit), et si chacune a le même coût marginal c_i de servir le marché i, la formule de Lerner devient

$$\frac{p_i - c_i}{p_i} = -\frac{1}{n_i}\,\frac{1}{\eta_i(p_i)},$$

où n_i est le nombre d'entreprises. Si les entreprises ont des coûts différents, la formule de Lerner devient

$$\frac{p_i - c^m}{p_i} = -\frac{1}{n_i}\,\frac{1}{\eta_i(p_i)},$$

où

$$c^m = \frac{1}{n_i}\sum_{j=1}^{n_i} c_j$$

est le coût moyen de servir le marché i. Cette formule est directement comparable à celle pour le monopole dans le cas où un marché correspond à un groupe de clients. Si les deux groupes diffèrent quant à l'élasticité de la demande, on trouve comme en monopole que le prix est plus

36. Il serait peut-être plus pertinent de considérer ici le problème qui se pose à une autorité chargée de contrôler les pratiques concurrentielles, comme le Conseil de la Concurrence ou la Commission européenne, qui semblent mettre plus de poids sur le surplus du consommateur.

élevé pour une demande élastique. Néanmoins, tous les prix tendent vers le coût marginal (qui est le même sur tous les marchés si le bien vendu est identique) quand le nombre d'entreprises devient très grand et la concurrence estompe les inégalités de prix. Ces résultats peuvent aussi s'appliquer lorsque les entreprises discriminent avec des produits différents (auquel cas le coût marginal c_i diffère d'un marché à l'autre). Mais un cadre d'analyse plus riche peut être obtenu en supposant qu'il y a aussi une différentiation sur chaque marché.

Une formulation courante de la demande pour des produits différenciés est basée sur les modèles de choix discrets, qui sont décrits plus en détail dans d'autres chapitres de ce livre. Un aspect intéressant de ces modèles est qu'il est facile d'introduire des éléments nouveaux (par exemple, les différences de qualités, la congestion) au niveau des préférences du consommateur qui sont au centre du modèle.

Le logit multinomial constitue une formulation particulièrement commode. Si on considère que la variable stratégique est le prix, le prix d'équilibre d'une entreprise j sur le marché i s'écrit

$$p_{ij} = c_{ij} + \frac{\mu_i}{1 - D_{ij}},$$

L'analyse d'Anderson et de Palma (2001) montre que cet équilibre possède plusieurs propriétés assez intuitives. Par exemple, s'il y a un marché où les consommateurs ont une préférence plus élevée pour ne pas acheter (peut-être les personnes âgées ont-elles moins souvent besoin d'aller en ville). Les prix d'équilibre sur ce marché seront alors plus bas, et les écarts entre ces prix seront plus faibles. De même, plus il y a d'entreprises ou plus elles sont homogènes (du point de vue d'un certain groupe de consommateurs), plus les prix seront bas et similaires.

On peut comprendre dans ce cadre pourquoi l'introduction de la concurrence sur un marché peut conduire des entreprises discriminantes à exacerber les différences de prix pour les différents services proposés. Borenstein et Rose (1994) observent que suite à la déréglementation du transport aérien aux Etats-Unis, on a vu s'accroître les disparités entre les prix pratiqués par les compagnies sur un trajet donné. L'explication théorique qu'ils proposent est que les clients prêts à payer plus cher pour un trajet (ceux dont la valeur de « ne pas acheter » est faible) sont aussi les plus fidèles à une compagnie, ce qui se traduit de leur point de vue par une plus grande hétérogénéité des produits. Lorsque la concurrence est introduite sur ce marché, la différence entre le prix fort de la classes business et le prix faible de la classe économique est amplifiée par la différence d'intensité de la concurrence pour les deux services, la concurrence étant plus intense pour celui qui est moins cher[37].

Une question qui a fait l'objet de plusieurs travaux sur la discrimination de troisième degré en présence de concurrence concerne le niveau du surplus du producteur selon que la discrimination est possible ou non (elle peut par exemple être illégale). Plusieurs auteurs sont parvenus à la même conclusion : contrairement au cas du monopole, il peut arriver que le profit total en oligopole soit plus faible quand les entreprises peuvent discriminer. Hoover avait déjà suggéré ce résultat dans le contexte de la discrimination spatiale (1948, p. 57) :

> ...la différence entre la concurrence en prix f.o.b. (avec les aires de marché strictement délimitées) et la concurrence avec des tarifications qui discriminent dans l'espace est un peu semblable à la différence entre la guerre de tranchées et la guérilla. Dans le premier cas tout le combat se fait le long d'une ligne de front bien définie ; dans le second, les forces opposées sont entremêlées sur un large espace.

Ces travaux montrent que les implications de la discrimination pour le surplus du producteur peuvent être très différentes selon qu'il y a ou non concurrence. En revanche, aucun contre-exemple n'a été trouvé au résultat selon lequel le bien-être social diminue suite à l'introduction de la discrimination si la quantité produite n'augmente pas.

Contrairement à l'analyse de la discrimination du troisième degré, celle de la discrimination en présence d'arbitrage personnel en oligopole est complexe, et il existe à ce jour peu de contributions significatives dans une formulation un tant soit peu générale. Des exceptions notables sont Champsaur et Rochet (1989), Ivaldi et Martimort (1994), Stole (1995), Armstrong et Vickers (2001), et Rochet et Stole (2001). Les deux derniers articles montrent en particulier que si deux entreprises en duopole proposent des services relativement proches, la tarification non linéaire d'équilibre peut se ramener à un tarif binôme qui génère un profit nul pour chaque entreprise. Rochet et Stole (2001) montrent aussi que si les services proposés sont plus différenciés, la tarification d'équilibre ressemble à celle de monopole.

Les progrès théoriques sur ce sujet semblent difficiles ; mais il est important que l'on continue à développer des cadres théoriques permettant d'étudier la discrimination avec arbitrage personnel en oligopole qui puissent servir de base à des travaux empiriques.

Remerciements

Nous remercions Anita Anderson, Catherine de Fontenay, André de Palma et Émile Quinet pour leurs commentaires et suggestions. Nous sommes redevables au programme de coopération entre le CNRS et

37. Borenstein et Rose appuient leur argumentation sur les travaux de Borenstein (1985) et Holmes (1989).