Price Discrimination*

Simon P. Anderson[†] and Régis Renault[‡]

This version August 2008.

KEYWORDS ARBITRAGE, NONLINEAR PRICING, DISCRIMINATORY PRICING, RAMSEY-BOITEUX PRICING, MARGINAL COST OF PUBLIC FUNDS

1 Introduction

Whenever we take a trip by train or by plane, we are often well aware that the price we paid was quite different from that paid by our fellow passengers with whom we are sharing the carriage or cabin. We can bemoan this situation, if we booked late and do not qualify for an age discount and our ticket was one of the more expensive ones or perhaps be pleased at having gotten a good price. The different prices are illustrations of what economists call discriminatory pricing. This seems to be a textbook case where a service which is identical (same journey, same date, same time, same comfort class) is sold at different prices[1].

*We gratefully acknowledge travel funding from the CNRS and NSF under grants INT-9815703 and GA10273, and research funding under grant SES 0452864 . We thank Anita Anderson, Alain Béraud, Catherine de Fontenay, Robin Lindsey, André de Palma, Emile Quinet, and Sarah Tulman for their comments and suggestions. We would also like to thank Melbourne Business School and the Autoridade da Concorrencia in Lisbon for their hospitality.

†Department of Economics, University of Virginia, PO Box 400182, Charlottesville VA 22904-4128, USA. sa9w@virginia.edu

‡ThEMA, Université de Cergy-Pontoise, 33 Bd. du Port, 95011, Cergy Cedex, FRANCE and Institut Universitaire de France. regis.renault@u-cergy.fr

[1] Price discrimination is also common in other transportation services in addition to rail and air travel. Odlyzko (2004) provides examples from maritime transport, inland waterways,

Looking closer, it is a little oversimplified to claim that all travelers have actually received the same level of service. Less expensive tickets are often associated with numerous restrictions which clearly indicate a lower level of service. It is often necessary to buy the ticket a long time in advance, with restrictive conditions on cancellation and reimbursement. The traveler can then enjoy the same service as she would have had if she had paid full price – unless, of course, her plans change at the last minute. However, to get the cheap fare she has to accept some risk, had she been obliged to change or cancel her ticket, or indeed she might have had to put up with some inconvenience due to having not changed her ticket in order to not lose money. There are also reduced fares for those satisfying certain "demographic" considerations. Senior citizens and children often pay lower fares.

On May 5, for May 18, 2007 the prices for one-way second class travel between Charleroi Sud (Belgium) and Paris Nord (France), proposed on the internet site www.thalys.com were as follows: Librys 59 E; Mezzo 44.5 E; Mezzo+ 32.5 E; and Smilys 20.5 E. The rates for Mezzo+, Mezzo, and Smilys can be reserved only if sufficient seats are still available, and only if one buys a round trip ticket. It is less expensive to buy a round-trip Smilys ticket than to buy a one-way Librys ticket. However, Smilys can only be bought when reserving two weeks before the departure date and it is the only fare which is non-exchangeable and completely non-reimbursable. Mezzo and Mezzo+ are reimbursable up to fifty percent of the price of the ticket, up to the departure

and turnpikes, as well as railroads, to argue that price discrimination has been (and is still) prevalent in the development of these sectors.

date. The only difference between the two tickets is that the quota for Mezzo+ is sold out quicker than for Mezzo. It is also possible to get a Lys fare for 26 E if one has a membership in the program. This is an example of a two-part tariff.

For these examples, the menu of prices that are proposed must take into account the fact that travelers can personally arbitrage (or choose between) the different options. There are also fare categories which are not subject to such arbitrage, such as: Kid (15 E), for children under 12 years old; Kid & Co. (29.5 E), for adults accompanying a child under 12; Youth (29.5 E) for travelers under 26; and Senior (41.5 E) for travelers over 59. All of these classes, except for Smilys, are also available in first class – Librys, for example, costs 100 E.

Price discrimination arises when a firm sells different units of the same good at different prices. This applies perfectly to cases in which certain groups of customers benefit from special tariffs (i.e. students or senior citizens), or nonlinear pricing where the price per unit depends on the number of units brought (such as the price of a French Metro ticket when bought singly or in a pack of ten).[2]

Nevertheless, several practices that involve selling different services can be viewed as discriminatory. In such cases, price differences might also be explained by cost differences without necessarily invoking a discriminatory motive (see, for example, Lott and Roberts, 1991). To address this concern, some writers

[2] Price discrimination is often also based on time of travel, such as evening or summer tariffs. The simplest analysis assumes that the demands in each period are independent (see also the analysis of peak-load pricing in the Backhaul section below). Gerstner (1986) formulates a peak-load pricing model in which firms account for intertemporal demand shifting.

have proposed definitions based on the comparison of price differences relative to cost differences. Stigler (1987) proposed comparing the ratio of the prices of two services with the ratio of their marginal production costs. By this criterion, a situation is discriminatory if the two ratios are unequal. Phlips (1983) on the other hand proposes comparing absolute differences. Then prices are discriminatory if the difference in marginal costs is not equal to the difference in prices.

It is difficult to find a decisive argument for one definition over the other.[3] Both definitions indicate that prices can be discriminatory even if price differences are small, just as it can be discriminatory if price differences are large. Suppose an airline brings passengers to a Parisian airport from which its international flights leave, and has everyone pay the same price for a flight to New York. This pricing discriminates against travelers living near Paris (see Tirole, 1988, 1993, for a similar example).[4] The definitions do not say whether such discrimination harms economic efficiency: the airline's pricing scheme allows it to more effectively exploit its market power by bringing its transatlantic travelers to Paris.

A firm with some market power and proposing different services can set its prices to get the greatest profit, and its prices will not bear any simple relation (absolute or relative) to marginal costs. Exercising market power is channelled through the ability to price above marginal cost. Offering diverse services can be seen as a way of discriminating insofar as it allows the firm to adjust

[3] See Clerides (2004) for a definition in which there is no discrimination if the pricing structure is not subject to arbitrage by buyers.

[4] For further discussion, see Section 3.5.

the service proposed and its pricing to a demand that differs from customer to customer.[5]

Just as it is difficult to define discriminatory pricing, it is not easy to classify different discriminatory practices. The classic reference is Pigou (1938), who distinguishes between three possible degrees of discrimination, depending on the ability of the firm to distinguish between buyers who are prepared to pay a higher price and those inclined to pay less. Pigou defines first-degree discrimination as when consumers pay their maximal willingness to pay for each unit. This is also called *perfect price discrimination*.

Pigou recognized that this first form of price discrimination might not have great practical relevance. He notes that the firm is better able to segment the market between different groups of buyers who have different demands. Ideally, the firm would like to segment the market into groups with similar willingness to pay; such segments could be ranked from highest to lowest willingness to pay.

Such idealized segmentation constitutes second-degree price discrimination. As Pigou notes, however, in practice a firm can only imperfectly compartmentalize consumers according to their willingness to pay. The firm must use characteristics which it can directly monitor, such as the type of good that is being transported (for example, livestock or pig iron[6]) for a railroad that transports freight, or the location of the buyer. This latter practice is third-degree price discrimination.

Pigou's classification underlines the fact that discriminatory pricing is meant

[5] In this paper, we emphasize demand drivers and take marginal cost as constant. The proper attribution of costs is a complex problem in itself.

[6] See Leadbelly's "Rock Island Line".

for the firm to exercise its market power as well as possible. Rather than proposing a reference case corresponding to no discrimination, as did Stigler and Phlips, Pigou defined the benchmark where the ultimate objective of discrimination is attained, that is, where each unit is sold at the highest possible price.

From this viewpoint, first-degree price discrimination is the theoretical benchmark. We will see, nonetheless, that third-degree discrimination is very relevant to the understanding of the discriminatory practices which are based on directly observable characteristics such as age, departure location, and time and date of the journey.

On the other hand, second-degree discrimination only seems to be a particular case of third-degree discrimination which, as we will see, corresponds to a situation which is particularly favorable for the firm, where the verifiable information that allows the firm to discriminate is perfectly correlated with willingness to pay (for example, if older people are willing to pay more).

Nevertheless, numerous types of discriminatory pricing such as nonlinear pricing or offering different comfort classes within a train or plane do not rely on a verifiable criterion and allow the user to choose her preferred option. Such practices therefore cannot be viewed as third-degree price discrimination, and economists have gotten into the habit of calling such practices second-degree price discrimination (see Phlips, 1983, Tirole, 1988, 1993, chapter 3, Varian, 1989, Mougeot and Naegelen, 1994). This term is therefore currently used to cover practices that are quite different from those originally envisaged by Pigou.

The discussion above suggests that discriminatory pricing is tightly tied to the exercise of market power.[7] Under perfect competition, firms are constrained to sell their output at the price that is imposed by the market. It is then obviously impossible to sell different units of a good at different prices, or to try to affect prices by proposing a range of different services (at least as long as such services are sold in a competitive market). Even if the firm has some market power, its ability to discriminate between different buyers can be undone or mitigated by the buyers' ability to arbitrage between the different options proposed.[8]

This arbitrage can take two forms, depending on whether one or several buyers are involved. If it is possible to transfer the good, buyers can exchange the good or service between themselves and the firm cannot charge different prices, because those buyers who benefit from the lowest price will be able to buy in order to resell to those who would otherwise have to pay more. For example, if someone has a membership card which allows her to obtain her tickets at a cheaper price, she could buy a large number of them and resell them to those without a membership card. Similarly, if a low fare is offered under the condition that the ticket should be bought sufficiently far in advance, then entrepreneurial individuals could buy a large quantity of these cheap tickets in order to sell them just before the date when the tickets are valid. Even though this type of arbitrage can be limited by transactions costs, it nonetheless

[7] Nonetheless, McAfee, Mialon, and Mialon (2006) propose a simple model in which they show that the extent of price discrimination has no theoretical connection to the extent of market power.

[8] One other limit on the ability to price discriminate is that firms cannot effectively propose a price menu that is too complex. See Levinson and Odlyzko (2007) for a recent treatment.

represents a significant constraint on firms' pricing strategies, so much so that firms often put in place several techniques to stop it. They often require one to present the membership card during the trip, or, for airlines, present a piece of identification which has on it the name matching that of the ticket holder. This type of rule also allows the firm to circumvent the second type of arbitrage, in which one customer with several options does not choose the one which was designed for her. For example, if a firm wants to discriminate on the basis of age, by presenting a driver's license the customer verifies that she is paying the price that she should. When buyers can practice such arbitrage between the different pricing options that are offered, Pigou (1938) says that demand is transferrable.

In practice, firms can often discriminate without explicitly forbidding arbitrage. The firm then has to explicitly worry about potential arbitrage when it is setting up a discriminatory tariff structure. While the firm cannot force its customers to not arbitrage, it has to set up the right incentives in its pricing plan.

The analysis of arbitrage behavior across consumers is relatively complicated, and most of the literature on discriminatory pricing simply supposes that transactions costs are high enough to render it impossible. We, too, will implicitly invoke this assumption throughout the paper.[9] On the other hand, so-called personal arbitrage– by which a user can choose an option which is not intended for her purchase– has been the subject of numerous studies, especially

[9] See Alger (1999) for an analysis of the constraints that are imposed by the potential of arbitrage involving several buyers.

over the last three decades.

Section 2 presents a certain number of base concepts and general principles which are needed to understand the rest of the paper.[10] In Section 3 we will show how a firm can exploit information about customer demand in order to discriminate, by supposing that buyers do not indulge in personal arbitrage. We will distinguish between perfect discrimination (3.1), discrimination between several categories of buyers who are purchasing the same good (3.2), and market segmentation where the firm offers different services for which it can perfectly determine the buyers who are prone to buying each service. Section 4 looks at strategies which allow the firm to motivate buyers to not engage in personal arbitrage when the firm cannot stop it by using verifiable information. We first consider using nonlinear tariffs (4.1 and 4.2), and then the possibility of discriminating by offering different qualities (4.3). In Section 5, we conclude and also discuss the link between discriminatory pricing and competition. This discussion allows us to evaluate the robustness of the results obtained for monopolies.

2 Preliminary results and basic concepts

2.1 Uniform monopoly pricing

We start with the simplest pricing structure, uniform pricing, where all units of the same good are sold at the same price, such as a given journey where there is only one comfort class and all travelers pay the same price.

Let $D(p)$ be the quantity demanded at the price p. The simplest inter-

[10] For a simple and broader treatment of this subject, see Varian (2000)

pretation is to suppose that D describes the distribution of willingness to pay over different travelers. Each traveler only wants a single trip, and $D(p)$ then indicates the number of travelers willing to pay at least p.

It is also useful to define the inverse demand for each quantity q. This is the maximum price at which this number of trips can be sold, $P(q)$. When each traveler only wants a single trip, $P(q)$ is the willingness to pay of the marginal consumer: the individual who would not travel if the price were slightly higher. In order to simplify the analysis, suppose that the marginal cost of production is constant at rate c per unit, which is also therefore the average variable production cost (and is the cost generated by each traveler). Under perfect competition, the price would be given by this marginal cost, and the number of travelers would be $D(c)$. We assume that services are provided by a private monopolistic firm, whose objective is to maximize its profit. We start with the case where demand is linear,

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

The latter condition ensures that it is optimal to produce a strictly positive quantity. This situation is illustrated in Figure 1. The firm is constrained under uniform pricing to choose a single point on the demand curve. If it wants to carry q travelers, the highest uniform price that it could charge is:

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

Figure 1: Uniform Monopoly Pricing

The producer surplus, PS, which is profit gross of fixed costs, is represented by the rectangle above marginal cost (i.e. the mark-up per traveler), $p - c$, multiplied by the number of travelers q.[11] The optimal quantity must therefore maximize:

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

and the first order conditions for an optimal quantity can be written as:

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

The left hand side of this expression is marginal revenue, MR. This is a straight line with the same price intercept as inverse demand D, but its slope is twice as steep. It is easy to see from Figure 1 that the condition for equality between marginal cost and marginal revenue is satisfied for a quantity:

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

which is therefore half the quantity produced under perfect competition. The uniform price chosen is therefore:

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

[11] See Anderson and Engers (2007b) for further discussion of the concept of producer surplus.

and this generates a producer surplus, PS, equal to:

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

which corresponds to the rectangle PS in Figure 2.

In the general case, the equality of marginal revenue and marginal cost allows us to determine the optimal quantity.[12] When we describe discriminatory pricing, we will often use first order conditions for *prices* (rather than quantities). Under uniform pricing, the firm chooses a uniform price p^m in order to maximize $D(p^m)(p^m - c)$. The first order necessary condition for a maximum can be written in the form:

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

where $\eta(p^m) < 0$ is the elasticity of demand. The left hand side is the mark-up rate, also known as the Lerner Index. it is therefore higher when demand is inelastic (meaning an elasticity closer to zero). Note that the elasticity of demand at the uniform price chosen by the monopolist is always less than -1.

2.2 Welfare analysis and public policy

We have just seen that a private firm which maximizes its profit will choose a price that is higher than the perfectly competitive price, and can thus increase its profit. Obviously, this is done at the expense of the consumers (travelers), who face higher prices. We now show how this loss of consumer well-being can

[12] This quantity is strictly positive and unique if marginal revenue at zero is superior to marginal cost, and if the slope of marginal revenue is always less than that of marginal cost.

be measured in monetary terms in a way that can be compared with the extra profit extracted by the firm.

An individual's consumer surplus is the difference between the maximum price that the consumer is willing to pay for a trip and the price that she actually pays. The inverse demand curve is constructed by ranking willingness to pay in decreasing order, so that the q trips are sold to the q travelers willing to pay most. Aggregate consumer surplus, CS, generated by the sale of quantity $D(p)$ at price p, therefore corresponds to the area between inverse demand and the quantity from 0 to $D(p)$.[13] For the solution described in Figure 1, this is represented by the triangle CS in Figure 2. The loss in consumer surplus resulting from a change from pricing *at* marginal cost to pricing at the monopoly level is given by the monopoly producer surplus plus the area DWL in Figure 2. This loss of consumer surplus can be interpreted as the sum of what travelers would have been willing to pay, collectively, in order to be able to access tickets priced at marginal cost as opposed to the monopoly price. Insofar as this total is larger than the monopoly producer surplus, there is a beneficial exchange possibility for all market participants which has not been realized (the firm would be ready to cut its price down to marginal cost if, in exchange, it could receive compensation that is at least as large as its producer surplus).

Monopoly pricing therefore introduces an inefficiency which can be measured by that part of the loss of consumer surplus which is not offset by an increase in producer surplus. This *deadweight loss* is the area DWL in Figure 2. To

[13] See Anderson and Engers (2007a) for further discussion of the concept of consumer surplus.

understand this inefficiency more concisely, it is convenient to introduce the concept of *social surplus*. This is the sum of producer surplus and consumer surplus. Pricing at marginal cost enables the maximal social surplus to be attained. Deadweight loss is the reduction of social surplus caused by a higher price.[14]

One reasonable objective for public policy could be to minimize deadweight loss. This objective could be obtained by a public firm, or one subject to regulation, pricing at marginal cost. Such a solution is not generally very satisfactory. When there are increasing returns to scale, such pricing will not cover production costs. For example, when marginal cost is constant, marginal cost pricing will generate zero producer surplus, so the firm will not cover its fixed costs.

It follows that the firm must be partly financed by taxpayers. It may then be desirable to suffer some deadweight loss in order to avoid an overly large deficit. Pricing at marginal cost might then be replaced by pricing at average cost, so that the firm just covers its costs. Nevertheless, it seems rather arbitrary to impose the condition that the firm should not make losses. An alternative argument against pricing at marginal cost is that taxation induces inefficiency in the allocation of resources (see Meade, 1944). Optimal pricing must therefore strike a balance between maximizing social surplus in the markets served by the firm and the efficiency cost of raising tax revenue in the rest of the economy. The latter cost can be measured by the deadweight losses caused by the taxes in

[14] Any price below marginal cost causes a deadweight loss because it leads to the sale of some units for which consumers are willing to pay less than the extra social cost that their production would engender.

the markets where they are levied.[15] Under such circumstances, the firm will only be fully financed by its users if the cost of raising public funds is too large. This reasoning also implies that if the cost of raising public funds is too large, and if the government can appropriate the firm's profits (which is effectively the case if the firm is public), it may be desirable to earn a return exceeding production costs. This extra revenue will enable the government to reduce fiscal pressure on the rest of the economy.

We can now address the pricing problem of a public or regulated firm. Following Laffont and Tirole (1993), we introduce a parameter λ which is the marginal deadweight loss of raising public funds: an extra euro raised means a cost of $1+\lambda$ euros.[16] The firm's objective function can then be written as:

$$CS + (1+\lambda)\, PS.$$

When λ becomes large, the firm pays no heed to consumer surplus. This case correspond to a private firm maximizing its profits. Otherwise, the firm chooses p to maximize:

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

Using the relationship $CS'\,(p) = -D\,(p)$[17] gives the first order necessary condition of :

[15] For example, income taxes cause deadweight losses in the labor market.

[16] The size of λ has been the topic of several studies. One reasonable estimate for the US is 0.3 (see Ballard, Shoven, and Whalley, 1985, and Hausman and Poterba, 1987).

[17] Consumer surplus at price p is the integral of the demand for prices larger than p.

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

The price thus obtained is a special case of what are called Ramsey-Boîteux prices, which ensure maximization of social surplus under the constraint that the firm returns a particular level of profits, for example to cover its fixed costs. Under this interpretation, λ therefore indicates the severity of the budget constraint and is the marginal social surplus gain that could be obtained by reducing the profit level to be earned by one euro (see Ramsey, 1927, and Boîteux, 1956).

In the interpretations above, λ is necessarily positive, so that more weight is placed on producer surplus than on consumer surplus. A higher λ leads to a higher mark-up over marginal cost (with monopoly pricing resulting as λ goes to infinity).[18] This outcome is rather unsatisfying insofar as we might wish for public policy to respond not only to economic efficiency but also to redistribution. The current analysis does not need to take a stand on these issues.

Thus, if a higher price (of a train ticket) allows the government to efficiently collect revenues, these could be used to give lump sum transfers to consumers. Nevertheless, a system of monetary transfers creates difficulties in itself because of the perverse incentives it may induce, as well as perhaps for reasons of political viability. If indeed transfers are not to be made directly from Government revenues because it is intrinsically costly to do so, then the pricing scheme may be used directly for transfers, and it may therefore be reasonable to put a larger

[18] This follows from applying the Implicit Function Theorem to the first order condition to the firm's problem above.

weight on consumer surplus than on producer surplus. This translates in our formal analysis to $-1 < \lambda < 0$.[19] This allows us to understand why we might want to heavily subsidize certain services such as public transportation, even pricing below marginal cost. Finally, there may be other reasons for pricing below marginal cost, especially for transportation services. Up until now, we have not taken into account the possibility of positive or negative externalities such as pollution or congestion. This omission can easily be rectified by replacing, in our analysis, marginal private cost with marginal social cost, which would be greater or smaller depending on whether the externality were negative or positive.

3 Discrimination and verifiable consumer characteristics

3.1 Perfect discrimination

We first consider the case which is most advantageous to the seller. This arises when the seller has perfect information about the demand from each possible buyer. Perfect discrimination arises when the firm can use this information fully (first-degree discrimination, in Pigou's terminology). In order to use the information, the seller must be able to control the price and the characteristics of each unit sold to each buyer. For example, if an airline perfectly knew the needs and desires of all of its customers it could choose the price at which each

[19] A negative lambda might also be applied in a welfare analysis used to evaluate different market outcomes, and where the firms' losses do not contribute to (or need to be financed from) the public purse. Private firms' surplus would usually be weighted (much) less than consumer surplus. For example, a consumer surplus standard, as is arguably used in some antitrust circumstances, would put a weight on producer surplus of zero ($\lambda = -1$).

client would take the price, and dictate the date, time, and comfort class.

This principle can be illustrated very simply by supposing that the good sold is perfectly homogeneous, and that each buyer only wants a single unit. In this case, the tastes of the buyer are completely described by her willingness to pay. A profit-maximizing monopolist will then have each buyer pay her maximum willingness to pay, leaving the consumer no surplus. The firm will therefore want to sell to all those whose valuations exceed marginal cost. This situation is shown in Figure 1 for constant marginal cost and linear demand. We see here that this pricing policy leads to maximal social surplus, all of which is captured by the firm.

The above approach can easily be extended to deal with elastic individual demand (so the quantity demanded decreases as the price rises). For illustration, suppose that the demand curve in Figure 1 represented a single consumer. The inverse demand would then indicate the highest price that she would pay to consume one extra unit. If the quantity q were sold at the uniform price $P(q)$, the net consumer surplus would be $CS(P(q))$; we can then define the gross consumer surplus (for an individual[20]) as $V(q) = CS(P(q)) + P(q)\,q$. Optimal discriminatory pricing would then mean paying the willingness to pay for each unit, and selling units as long as this willingness to pay exceeds marginal cost. This then induces the quantity $q^* = D(c)$ sold at a tariff equal to the corresponding gross consumer surplus, $V(q^*)$ which is the surplus that the consumer would enjoy if she could consume the quantity q^* for free.

[20] The corresponding aggregate concept simply sums the surpluses over individuals.

Another method for getting to the same result would be to use a two-part tariff (see also Oi, 1971). Such a tariff would specify an entry, or membership, fee A that the consumer must pay in order to consume the good at all. If she joins, then she can buy as much as she wants at a price p. Setting $p = c$ ensures that the consumer will therefore choose q^*, and she will therefore enjoy a surplus of $V(q^*) - cq^*$. She will join as long as the entry fee is not larger than this, and so the seller will set a fee as large as possible subject to this individual constraint; namely it will set an entry fee of $A = V(q^*) - cq^*$. The total price paid by the consumer is then $A + cq^* = V(q^*)$, which means that her full surplus is extracted by the firm. The outcome is thus just the same as for the preceding pricing system. Even though a two-part tariff seems simpler, it still needs just as much information: while it may be easy to fix price at marginal cost, calculating the entry fee means knowing the consumer's surplus, and hence her full demand curve.[21]

The pricing solution for a public firm under the similar assumption of perfect knowledge and ability to discriminate is straightforward: it should choose exactly the same tariff structure. This is because all surplus is extracted from the consumer and earned by the firm for the public purse.

[21] No further complication is introduced for the preceding analysis when marginal costs are not constant. The optimal quantity, q^*, equates the demand price with marginal cost, and all consumer surplus is extracted. A single (all-or-nothing) tariff equal to gross consumer surplus at this quantity is optimal, and is equivalent to a two-part tariff with a per-unit price equal to the demand price $P(q^*)$ and an entry fee of $A = V(q^*) - P(q^*)\,q^*$ $(= CS(q^*))$.

3.2 One good and several groups of buyers

We next consider a situation in which a firm can observe a characteristic or characteristics of buyers– such as age, job, or residential address– and observing these characteristics allows the firm to infer something about demand. The firm can exploit some correlation between the observed variable and the individual demand in order to discriminate (perhaps such discrimination is not legal or socially acceptable, for example if it is based on gender or race).

The firm can then choose a price that depends on the observed characteristic, and the individual who does not have this characteristic can be excluded from prices not meant for her. However, discrimination is imperfect insofar as consumers are bundled together onto the same characteristics (e.g. the same age group), and individuals still may differ by willingness to pay within the group (within a group, selling different quantities with nonlinear pricing or introducing differentiation can be used to get potential buyers to reveal their tastes: see Section 4). The firm is therefore obliged to set a uniform price for each category (or group). Suppose, for example, there are two classes of buyers: youths under 26 and the rest of the population. Asking for identification (in the absence of fake IDs), the firm can know which category the buyer is in. It can therefore exclude arbitrage by which people in one group buy the good or service in order to sell it to people in the other group (for example, airline tickets with the traveler's name on them). In the absence of such arbitrage, demand from each group depends only on the price charged to members of the group.

Consider the case of a public firm: the specialization to a private firm will simply be given by letting λ become arbitrarily large. Let p_i be the price applied to group i, and let $D_i(p_i)$ be the resulting demand curve, and $CS_i(p_i)$ be the corresponding consumer surplus. We assume that marginal cost is constant at rate c, so that the firm can choose prices for groups independently of each other.[22] The firm therefore chooses p_i to maximize:

$$CS_i(p_i) + (1+\lambda) D_i(p_i)(p_i - c).$$

This leads to a first order condition which simply generalizes that, for a single market:

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

This type of discrimination benefits the firm because it faces groups of consumers whose elasticities differ. The less elastic demand is (the closer η is to zero), the higher the price charged to the group. Thus, youths enjoy lower train fares and airfares because they are more likely to reduce their demand if prices were higher. Clearly the firm benefits from discrimination because it can always opt to charge the same price across groups. However, only consumers for whom demand is quite elastic benefit, to the detriment of the other consumers.[23]

For a profit maximizing firm, we can determine whether discrimination is socially beneficial. Suppose for simplicity that $\lambda = 0$, so that social surplus

[22] When marginal costs are not constant, the optimal solution is obtained by equalizing marginal revenues across groups for any given total production level, and then choosing the total quantity which equalizes marginal cost with this common revenue.

[23] This discussion assumes that two demands are comparable in the sense that one is more elastic than the other for all prices. It is possible to construct examples for which prices under discrimination are higher or lower than under uniform pricing (see Nahata, Ostaszewski, and Sahoo, 1990)

is simply the sum of consumer surplus and producer surplus. Note first that, for a fixed total output, discrimination induces inefficiency in the allocation of this quantity between two groups. The price paid by the consumers in a group reflects what they are willing to pay in order to buy one extra unit. Social surplus can be increased by shifting consumption from those willing to pay less at the margin to those willing to pay more. Furthermore, since the output of a monopolist under uniform pricing is too low, discrimination can only improve welfare if it increases production. This necessary "output condition" was first proposed by Schmalensee (1981).

In general, the impact of discriminatory pricing on the total quantity produced is ambiguous. For example, suppose that the demand in each group i is linear and given by:

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

Applying the earlier analysis of linear demand, if both groups are served under uniform pricing the total output is:

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

while the quantities allocated to each group of buyers under discrimination are, respectively:

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

The total output is therefore identical[24], and discrimination is necessarily harmful to social welfare. This conclusion is invalid if only one group is served by the non-discriminating monopolist. Then, if marginal cost is constant, discriminatory pricing is clearly preferable because it allows more markets to be served. In this case, the group which is served under both pricing schemes is not worse off because it pays the same price under both, while the other group would not have been served at all under uniform pricing.

The above reasoning also gives us an unambiguous result when capacity is fixed. Discriminatory pricing cannot increase output, and is therefore clearly detrimental. If an airplane is full, it is therefore better that all passengers pay the same price. Nevertheless, this reasoning is only valid in the short run, because in the longer run the possibility of discrimination can motivate the airline to increase the number of flights. In this context, too, discrimination can engender a greater diversity amongst the traveling population.

As we argued in the introduction, second-degree discrimination as originally envisaged by Pigou (1938) can be seen as a special case of the above analysis of discrimination between different groups of buyers. If the observable characteristic used by the firm to discriminate can allow it to perfectly separate buyers into groups that can be ordered according to willingness to pay, then it can practice second-degree price discrimination à la Pigou.

To illustrate, consider the following example given in Anderson and Renault (2003b). Suppose that a traveler wishes to take a train trip and her willingness

[24] This is a special property of linear demand, that marginal revenue corresponding to the sum of demands is equal to the sum of the marginal revenues to the demands for each group.

to pay is perfectly positively correlated with age. Even though this would seem to allow the firm to practice perfect discrimination (because the client's age perfectly reveals her demand), it might in practice be costly to specify too many different prices. For example, it may be only possible to offer two prices. The railroad company must then set a threshold age above which people cannot get the low price.

Suppose, for example, that the inverse demand is given by $P(q) = 1 - q$. With zero marginal cost, the monopoly price and output are both one-half. It is then easy to see that the firm's optimal strategy is to charge full fare of 2/3 for the oldest third of the population, and to set a reduced fare of 1/3 for the younger ones. In this context, we can also determine the second best policy of choosing the critical age with the objective of maximizing social surplus, subject to the firm choosing its prices given this critical age. The second best solution is that the firm only offers the reduced fare to the younger half of the population. The firm would then choose a full fare of 1/2 for the older half of the population and a half-priced fare of 1/4 for the younger travelers. In this sense, the critical age chosen by the profit maximizing firm is too high.

3.3 Backhauling

We have assumed so far that the product is produced at constant marginal cost. In the case of transportation, services are provided and these depend upon the capacity offered. While it is out of our scope to cover the full range of issues associated with proper cost attribution, scheduling of service, and route network choice of passenger services, we nevertheless broach this issue with a simple example of capacity that is provided on an outbound trip: the train has to get back to the origin to make the next trip to the destination.

The analysis can be framed in terms of a monopoly passenger railway service: see Rietveld and Roson (2002) for a recent application in this vein. Many commuters might wish to make the trip to the Central Business District in the morning rush-hour commute, but few people want to go in the reverse direction soon afterwards.[25]

The analogue in the context of freight transportation is when a product is shipped to a destination and there is a relatively weak demand for shipping goods in the opposite direction. However, the trucks, ships, or freight trips must still make the return trip to pick up another load. This is termed the "backhaul problem".[26] We therefore frame the application to passenger transportation in the classic fronthaul and backhaul context (e.g., Mohring, 1976). This is well known in economic theory as a joint production problem, much as the textbook mutton and wool joint production in raising a sheep. Analogously, once an outbound trip to the final market is created (fronthaul), then a return trip is also created (backhaul).

The economic theory of pricing for competitive markets with backhaul is well understood. Suppose, as above, that the round trip costs c, and that the demand for such trips is given by a well-behaved downward-sloping demand,

[25] Of course, in the evening, the "backhaul" is stronger than the "fronthaul" in the sense that relatively empty trains come into the city and full ones go back out. For the present purposes we shall identify the stronger demand as the "fronthaul" even if it happens after the weaker demand (backhaul).

[26] See Anderson and Wilson (2008) for a treatment of the backhaul problem when a dominant firm faces a competitive fringe.

$D(p)$. Suppose too that the demand for trips back from the final market is $D_b(p)$. Denote the inverses of these demand curves as $P(Q)$ and $P_b(Q)$ respectively. For simplicity we assume that the incremental cost when carrying passengers for the trip back is zero. The relevant demand price for round trips is, therefore, the sum of the demand prices for the outbound and inbound trips, censored to be non-negative (because the transporter can always come back empty). That is, if Q is quantity, the demand price for the round trip is $P(Q)+P_b(Q)$ (where it is understood the demand prices are non-negative) and this sum is equal to p in equilibrium. Denote the solution as $\hat{Q}$: transport prices for each leg are then $P\left(\hat{Q}\right)$ and $P_b\left(\hat{Q}\right)$. Clearly, if the backhaul demand is weak then $P_b\left(\hat{Q}\right)$ can very well be zero. Some passengers can be carried, but backhaul demand is not contributing anything to reducing the price on the front haul, and effectively $\hat{Q}_b < \hat{Q}$ where $\hat{Q}_b$ is the number of travelers transported on the backhaul.

It is now simple enough to see how to introduce incremental costs for backhaul: they can be netted off the demand price for the backhaul. The same principle applies in what follows: it suffices to net the incremental costs off the demand price.

For a monopolist, say the commuter train where it has a large cost advantage and is unconstrained, the appropriate principle for determining the quantity to carry (and the corresponding prices) is that the sum of the marginal revenues equal the marginal cost. Here again, the monopolist is not obligated to carry as much back as it carries out, and so we truncate the marginal revenues at zero. Then the solution for the fronthaul quantity, Q^*, is given as the solution to $\max\{MR(Q), 0\} + \max\{MR_b(Q), 0\} = c$, where $MR(.)$ denotes the marginal revenue to the outbound demand curve (and $MR_b(Q)$ for the backhaul demand). In this case, with a weak backhaul demand the price charged will not be zero but rather the revenue maximizing point on the (net) backhaul demand curve. This "zero-cost monopoly price" (which is equivalently the revenue maximizing price) will prevail when some backhauls carry no passengers. Equivalently, any solution with $Q_b^* < Q^*$ involves $Q_b^* = MR_b^{-1}(0)$. For any solution to the monopoly problem, the prices on the two legs are given from the inverse demand curves as $P(Q^*)$ and $P_b(Q_b^*)$.

In the preceding analysis, the discriminating firm behaves just like a firm that offers different products with perfectly independent demands. In practice, discrimination founded on observable characteristics is often associated with multi-product production. The next section addresses the application of discriminatory pricing in such a context.

3.4 Yield Management

To set the stage, assume that a transportation vehicle (such as a train) has a single seat to fill.[27] The operator knows that one traveler will purchase the ticket, and that the traveler's reservation price for the trip is uniformly distributed on $[0, 1]$ (think linear demand curve, as in Section 2.1 and Figure 1). There is no cost associated with selling the ticket or filling the seat. The

[27] For references in this area, see Anderson and Schneider (2008), which also contains more analysis of competitive cases. See also Sinsou (1999) for a detailed presentation of the algorithmic methods firms can use for Yield and Revenue Management.

operator is to set a price to maximize its expected profit. Given that the associated demand curve is linear, the price the operator should set solves the classic monopoly problem, so $p = q = \frac{1}{2}$ and the associated profit is $\pi = \frac{1}{4}$.

Now consider the case where two travelers will arrive, one after the other, and their reservation prices are independently distributed on $[0, 1]$. The operator sets a single price for both. What price should it set? To find the answer, we first determine its profit. It is more transparent to write out the more general problem with $F(p)$ the probability an individual traveler's reservation price is below p (so $F(.)$ is the cumulative distribution). Then the profit is $\pi = p[1 - F(p) + F(p)(1 - F(p))]$, where the term in brackets is the probability that the first traveler buys, plus the probability that she does not $(F(p))$ times the probability that the second one does $(1 - F(p))$. In the case of linear demand, $F(p) = p$ for $p[0, 1]$ and the first-order condition implies that:

$$p = \frac{1}{\sqrt{3}} > \frac{1}{2}$$

The second order condition is readily seen to be satisfied; and profit is approximately 0.385. This is more profit because the operator can take a shot at a higher price given that the second customer provides some insurance in case the first one refuses. This logic is fine-tuned in the next example. Note that the price charged (in the current case of a single price for all) is increasing in the number of consumers, and the profit is too. As the number of travelers gets large, retaining a single seat, the price goes to the highest possible in the population (1) and the profit goes to 1 too since the probability of acceptance also tends to 1.

Profit is even higher if the operator can choose a separate price for each arriving guest. Suppose there are again two travelers, and they arrive sequentially. We want to find what price will the operator will set for the first one to arrive, and what it will charge the second one if the first traveler declines the seat. To find these prices, we work backwards. If the first traveler does not buy, we can simply use the answer from the monopoly case with one seat. This gives the expected profit of 1/4 if the room is not filled on the first shot. We can now determine the profit for the choice of a price p_1 to the first traveler. Noting that the probability that the first traveler does not buy is $F(p_1)$, this is $\pi = p_1 (1 - F(p_1)) + F(p_1) \frac{1}{4}$. The uniform distribution entails an optimal price of $\frac{5}{8}$. Plugging back into profit, this means a profit of $\pi = \frac{5}{8}\frac{3}{8} + \frac{5}{8}\frac{1}{4} = \frac{25}{64} = 0.39$. This discriminatory strategy is larger than when a single price must be set for both travelers, since the discriminatory problem subsumes the non-discriminatory one.

We next consider the case of three travelers and two places. Recall first that with simple linear demand, the monopoly price is $\frac{1}{2}$ and profit is $\frac{1}{4}$. These are indeed the price and expected profit that are earned on the last traveler if one place has already been sold. Likewise, if the very first sale is a failure, there are two places left and two travelers, so then the price is again $\frac{1}{2}$ in each period and the expected profit coming out of the failed first sale is $\frac{1}{2}$. If the first sale was a success, there is one place left and there are two remaining consumers. The continuation profit is then that derived just above, i.e., $\pi = \frac{25}{64}$.

We are now in a position to analyze the first period's price. This is given by the solution to

$$\max_p \pi = (1 - F(p)) \left[p + \frac{25}{64} \right] + F(p) \frac{1}{2}.$$

With a uniform distribution, the profit expression is $\pi = (1 - p) \left[p + \frac{25}{64} \right] + p\frac{1}{2}$, which generated a first order condition $(1-2p) + \frac{7}{64} = 0$, and hence the solution $p = \frac{71}{128}$. The corresponding sequence of prices is then $\frac{71}{128}$, followed by $\frac{5}{8}$(an increase) if the sale was a success or $\frac{1}{2}$ (a decrease) if it was a failure. The last price is $\frac{1}{2}$, if at least one previous sale foundered.

The type of exercise above can be applied to various other different pricing practices observed in transportation. For example, the SNCF sells a first batch of tickets at a lower price than a second tranche, which in turn is lower than the third tranche. The simplest set-up to analyze this practice is to assume that there are three travelers, and two seats. In contrast to what was just described, the pricing is now per seat, instead of the operator being able to condition per traveler. This implies that this is a special (constrained) case of the earlier analysis, and so leads to lower profits.[28]

The analysis can also be expanded to deal with uncertainty in the number of buyers, possibility of "strategic" buyers (or coming back after observing a first price), correlation in traveler values, etc. We pick up on this last topic because it is germane to transportation when firms may use the strength of early demand response to gauge later demand strength (for example, airlines

[28] The profit associated to the pricing per seat business model is $\pi = (1 - F(p_1)) \{ p_1 + p_2 (1 - F^2(p_2)) \} + F(p_1) \{ (1 - F(p_1)) [p_1 + p_2 (1 - F(p_2))] + F(p_1)(1 - F(p_1)) p_1 \}$.
The first term corresponds to the first traveler buying the first (lower-priced) seat, and then one of the other two buys the second seat. The second term corresponds to the first traveler declining the first seat, and then either the second buys it (and the third might buy the second seat), or else only the third might buy the first seat.

or passenger trains judging demand for special events like Olympic Games or tennis matches), and because it illustrates nicely the experimentation motive for varying prices.

Suppose then an operator has a single seat to fill, and there are two potential travellers. Both have the same valuation for traveling (perfect correlation), and this valuation is uniformly distributed on $[0, 1]$. If the operator charges p to the first traveler, it expects to sell the seat with probability $1-p$. If the first traveler does not buy, the operator knows the seat is worth less than p to the second. The updated valuation for the second traveler is uniformly distributed on $[0, p]$, and so the optimal price to set, conditional on a first refusal, is $p/2$. The second traveler accepts with probability $1/2$, so the expected profit, conditional on a first refusal, is $p/4$. This means the operator's profit problem as a function of the first price set, p, is

$$\max_{p\in[0,1]} \left\{ (1-p)\, p + p\frac{p}{4} \right\}.$$

This profit is maximized at $p = 2/3$. If the first traveler refuses, the price is dropped to $1/3$. The case of uncorrelated demand, given earlier, has less price dispersion, starting with a lower price $(5/8)$ and ending with a higher one $(1/2)$. With correlated demand, the operator starts with a higher price that includes an experimentation motive to find out more about the second traveler's valuation.

3.5 Discriminating with several products

When a firm sells several goods with perfectly independent demands, the situation is formally very similar to the one we have just discussed. The main difference is that marginal costs generally differ across products. With several products, there is a possibility of discrimination when different goods are sold at the same price or at prices that are close to each other. As we pointed out in the introduction, it can be quite difficult to determine whether a price difference is discriminatory.

The objective of this subsection is to briefly present some of the prominent examples that are frequently considered to represent price discrimination. In particular, we will discuss spatial discrimination and then look at discrimination over time and tied sales. In all cases that we consider, the client group which might buy one of the goods is clearly identifiable and there is no possibility of personal arbitrage. Using a multi-product offer to discriminate when arbitrage is possible will be studied in the next section. We first develop the optimal pricing strategy for a firm which sells to consumers located at different distances from its production point. We then show how this theory is relevant to a firm which must transport travelers over different distances, and we suggest other applications in transportation.

Suppose as a benchmark that consumers can access a transport service sold at a competitive price, and they can use this service to transport the good from its production point. Using this service, a consumer then pays for each unit of the good, an f.o.b. price set by the firm at its production point plus the transport cost.[29]

Without specific information on individual demand, the firm must charge a uniform tariff. This is the benchmark case of non-discrimination. However, if the firm can take care of delivery itself and if it can circumvent consumers' access to competing delivery services, it can extract more profit even though the delivery point does not reveal information about the demand of the buyers in question.

For illustration, suppose that demand is the same at every point in space. If the firm charges a price p_x to buyers located at a distance x away, the demand is $D(p_x)$ and the consumer surplus is $CS(p_x)$. Transport costs are linear and it costs tx to transport one unit a distance x, with $t > 0$. The firm chooses p_x to maximize

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

The price chosen will then be similar to that derived earlier, where we now simply use the full marginal cost $c + tx$. It can be shown that as long as the demand is more "concave"[30] than an exponential function, the rate of increase in the firm's price is less than t per unit distance (see Anderson, 1989). This means that the firm does not pass on all of the transportation costs into the price ("freight absorption"). In other words, the implicit f.o.b. price– the price net of transportation costs– is smaller the farther that buyers are from the

[29] The term f.o.b. stands for "free on board" or "freight on board" and means that the buyer must pay transportation costs for delivery. In spatial economics, the term f.o.b. price is synonymous with the term "mill price" to indicate the factory price (or the store price).

[30] See Anderson and Renault (2003a) for a precise definition of this concept. The condition is equivalent to Seade's (1987) condition on the elasticity of inverse demand.

production point.

Freight absorption is often used in practice[31]. Tirole (1988) notes that there are good reasons why there should be freight absorption even if demand curves do not satisfy the condition above. First, any pricing policy which over-bills transport costs might provoke arbitrage between consumers, with those who are closer to the factory perhaps buying the good and transporting it themselves, and reselling it to those farther away. Second, buyers far from the production site might also be more likely to be closer to a rival's production site, and this might render their demand more elastic. The firm may consequently be motivated to set a somewhat lower price.[32]

The social benefits from forcing a private firm to set a uniform price are ambiguous. The discriminating firm serves a larger geographic area (which, per se, improves social welfare) but it does so while imposing higher prices on nearby consumers. For linear demand, the social welfare is highest under spatial discrimination when $\lambda = 0$ (see Holahan, 1975), and therefore *a fortiori* is highest when $\lambda > 0$ because the producer surplus is always higher when the firm can discriminate.[33]

Now consider some applications in transportation. Consider first an airline selling long-distance flights between Paris and New York. Customers' points of origin differ: they start out from different regional French airports. Our

[31] Greenhut (1981) presents survey results on spatial pricing by firms.

[32] See Lederer and Hurter (1986) for a model of competitive spatial discrimination and Anderson, de Palma, and Thisse (1992, Chapters 8 and 9) for a review of the literature on spatial economics.

[33] The case of uniform pricing is a special case of discriminatory pricing. See Greenhut, Norman, and Hung (1987) for other developments with monopoly, and Anderson, de Palma, and Thisse (1989 and 1992b) for competition. Quinet (1998) describes in detail competition between two airports.

benchmark is when flights between different regional airports and Paris are sold on a competitive market and are therefore sold at marginal cost. Suppose that the airline can provide exclusive service to Paris, and that the demand for a trip to New York is the same irrespective of the starting point of the travelers (Paris, Bordeaux, Nice...), and that this demand is not "too convex". The freight absorption principle is manifested by more expensive tickets for tickets who originate farther from Paris (because of the higher cost of service), but price differences are less than the extra cost of serving these customers. In this sense, the more distant customers are subsidized by those closer to Paris.

The analysis also applies to railway pricing, or indeed an urban transportation network. Consider two trips of different length, which therefore involve different costs. Following Phlips' (1983) definition, pricing is discriminatory if the price difference between two trips is not the same as the cost difference. If the demand for the two trips is the same (so that we eliminate any price differences due to demand differences), and if the demand curve is not "too convex", then optimal prices will be closer than costs. In other words, the price should rise more slowly than the serving costs as a function of distance.

Of course, the analysis of urban transportation prices is much more complex than the sketch presented above. Demand typically depends on the distance traveled and the other transport modes available. Pricing should take congestion into account, etc., but this simplified framework at least serves to indicate the central role of discriminatory pricing.

There are numerous dimensions other than space over which buyers can be sorted. Time is one obvious example. Airlines often offer lower prices for return trips that include a Saturday night stay. This practice allows them to discriminate between leisure and business trips. Tourists' demand is relatively elastic, whereas business travelers face important timing restrictions. Similarly, prices which differ according to the time and date of departure can allow account to be taken of congestion (which is the reasoning behind peak hour pricing), but it can also be used to discriminate between clients with differing demand elasticities.[34]

Another practice which is often cited as an example of third-degree price discrimination is that of bundling or tying. Sales which combine two goods were recognized by Stigler (1963) to be an efficient means of discriminating. Such sales allow the firm to distinguish buyers interested in both goods from those interested in only one or the other. Package sales, combining transportation plus hotel, allow tourists to pay less for both services than if they were bought individually.

These examples indicate that segmenting the market can be substantially more subtle than just verifying IDs. However, strategies that do not rely on strictly verifiable information might be undone both by arbitrage across individuals or "within" the individual. The limits that personal arbitrage impose on discriminatory pricing were laid out in the 19th century by the French Ponts et Chaussées engineer Jules Dupuit. In his famous article on tolls (1849), he developed an example of price discrimination for a footbridge. He supposed

[34] There is a large literature on pricing under congestion, following Vickrey's (1969) seminal paper. De Palma and Lindsey (2000) treat the problem of two competing toll roads, and de Palma and Lindsey (1998) study the role of information acquisition in this context.

that many workers would like to use the bridge, but that a universal price of 1 centime per user would not suffice to cover construction costs.

In an attempt to impose a special workers' price that would be less than that paid by the rest of the population, Dupuit proposed a discriminatory scheme based on clothing: "for a crosser with a cap or a smock or jacket, the toll is reduced to 1 c." (p. 220) instead of 5 c. for the other travelers. However, he also notes that "it is quite likely that revenues will be reduced because some 5 c. crossers will benefit, by dint of their tenue, from the price reduction that was not meant for them" (p. 220). To staunch this potential arbitrage, he proposed only applying the price reduction at certain times of day (when workers were more likely to be present) or to require that worker present their pay stubs. In the next section we give a systematic analysis of strategies that allow the firm to prevent personal arbitrage.

4 Personal arbitrage

> *Travelers cannot be classed like merchandise by their appearance: they have to be allowed to sort themselves.* (Jules Dupuit, 1849)

Discrimination when there is personal arbitrage is effectively an information revelation problem. The firm knows that there is heterogeneity in the willingness to pay among buyers, but it doesn't have a means of knowing this information directly. In the example of Dupuit's (1849) footbridge, those who are willing to pay only one centime to cross may be relatively poor people who are not workers with a pay stub, or indeed they may be people who do not truly need to cross the bridge and who will not bother if it costs too much.

This problem of information revelation can only be resolved if individual demand varies with price, or if the firm can vary some characteristics of its product, which are differently evaluated by different consumers.

In the example of the footbridge, Dupuit noted that the second possibility could be to price differently according to the time of day. To see how the sale of different quantities can be exploited in the same example, suppose that the workers were sold a coupon, valid for six return trips per week, at a price of 12 c. If the other users only want, at most, one return trip, they will only pay 5 c. per trip rather than buying the ticket tailored for the workers. This type of "nonlinear pricing" is described in Subsections 1 and 2 below. Subsection 3 treats the use of multi-product offering in the presence of personal arbitrage.

4.1 Two-part pricing with heterogeneous customers

As we saw above, a two-part tariff specifies an entry fee A and a marginal price p, at which those who paid the entry fee can get the good. When the firm perfectly knows consumer tastes it can use this type of pricing to appropriate the entire consumer surplus, and thus perfectly discriminate. Although, in practice, transportation firms do not have such perfect information, they often use this type of pricing. The SNCF uses different passes for various periods of time. The French Metro (RATP) has the unlimited-ride Orange Card, which has a marginal price of zero. As we shall now see, such a two-part tariff can be useful when the firm does not know the consumers' tastes. (see also Oi, 1971)

Suppose there's a fraction α of consumers, corresponding to Type 1 consumers, whose demand at price p is $D_1(p)$ while the rest of the consumers demand $D_2(p)$. Suppose that $D_1(p) < D_2(p)$ for all $p \geq 0$ (the demand curves are shown in Figure 3), so that Type 1 consumers are low-demand and Type 2 consumers are high-demand. It follows that consumer surpluses satisfy $CS_1(p) < CS_2(p)$ for all $p \geq 0$.

When the firm has full information it would choose a marginal price $p = c$ and have Type 1 consumers pay an access fee $A_1 = CS_1(c)$, and $A_2 = CS_2(c)$ for the others. The number of trips would be $q_1^* = D_1(c)$ and $q_2^* = D_2(c)$ for consumers of Types 1 and 2, respectively. These are the first-best optimal quantities.[35] Under incomplete information, if the firm offered these two choices, clearly Type 2 consumers would pay the lower entry fee A_1, which would give them access to the same price per trip, and there would thus be personal arbitrage.

One simple solution to circumvent this type of arbitrage is to offer a single two-part tariff. If the firm offered the full-information prices tailored for Type 2 buyers, the consumers with the lower willingness to pay would not buy.

If there are enough Type 1 consumers (if α is large enough), this will not be the best solution. Instead, offering the first-best optimal price tailored for the Type 1 consumers, the firm can put in place an allocation which maximizes total surplus (because marginal price equals marginal cost and both types of buyers are active).

This will be an optimal solution for a public firm when there is no marginal

[35] They are the socially optimal quantities if the fiscal system allows lump-sum redistribution with a zero marginal cost of public funds.

cost of public funds ($\lambda = 0$). If $\lambda > 0$, the firm is also concerned with its profits. Even though the firm can extract the full surplus of Type 1 consumers, it must leave a strictly positive surplus to the others, as we shall now see.

The firm's objective can be split into two parts, the social surplus associated with Type 1 buyers:

$$CS_1(p) + (1+\lambda)(p-c) D_1(p) + \lambda A \qquad \text{(with weight } \alpha\text{)},$$

and the social surplus associated with Type 2 buyers:

$$CS_2(p) + (1+\lambda)(p-c) D_2(p) + \lambda A \qquad \text{(with weight } 1-\alpha\text{)}.$$

Each of these components includes consumer surplus and producer surplus plus the entry fee, which has a value of λ per euro because it is a transfer from consumers to the firm.

First, whatever the marginal price p, it is always desirable to extract the full consumer surplus from Type 1 consumers when $\lambda > 0$. If $A < CS_1(p) < CS_2(p)$, a small increase in the entry fee will have no impact on the quantities consumed and will raise revenues, so the firm is better off. It will therefore choose, as with full information, to set $A = CS_1(p)$, so that consumers with a low willingness to pay are just indifferent between joining and not joining. Nevertheless, in contrast to the solution with full information, the firm now wants to price above marginal cost. Even though the firm would maximize its objective over the Type 1 consumers by choosing $p = c$, this does not maximize social surplus with regard to the Type 2 consumers. The objective can then be written as:

$$CS_2(p) + (1+\lambda)(p-c) D_2(p) + \lambda CS_1(p),$$

where the last term takes into account that an increase in marginal price must be accompanied by a decrease in the entry fee to ensure that Type 1 consumers buy. Evaluating the derivative with respect to p when $p = c$ gives:

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

The firm can thus improve surplus by increasing the marginal price. As the effect of such an increase on social surplus for Type 1 buyers is negligible (because the derivative at $p = c$ is zero), this price hike is desirable. The optimal marginal price p_{mar} is strictly between marginal cost and the price which maximizes social surplus for Type 2 consumers. This latter price is below the Ramsey-Boiteux price corresponding to demand D_2, because of the negative effect of a higher marginal price on the membership fee.

It is straightforward to show that a fully nonlinear tariff, without the restriction that the second price be linear in quantity, performs better. Figure 3 shows that consumers with a high willingness to pay get a surplus (a rent) measured by area $KNLMJ$.

It is possible to reduce this rent without affecting quantities consumed. Suppose that instead of offering a two-part tariff, the firm gives each consumer the option of either the quantity $q_1 = D_1(p_{mar})$ at entry fee $T_1 = p_{mar}q_1 + CS_1(p_{mar})$, or else the quantity $q_2 = D_2(p_{mar})$ at entry fee $T_2 =$

$p_{mar}q_2 + CS_2(p_{mar})$, plus the area KNJ. Type 1 consumers will then prefer the combination (q_1, T_1) to (q_2, T_2). Type 2 consumers would enjoy a surplus of $JNLM$ if they chose (q_1, T_1), and they would get at least as much paying T_2 for quantity q_2. The firm then wants to leave the Type 2 consumers with as little rent as possible, that is exactly $JNLM$.

As we shall see, the pricing scheme that we have just described dominates the simple two-part tariff, but it is not generally the optimal tariff. It has a certain number of characteristics of the optimal tariff: Type 1 consumers have zero surplus, and consume less than their first-best optimal quantity (which would be consumed at marginal cost pricing); Type 2 consumers have a strictly positive surplus and are indifferent between the two choices offered. This latter indifference condition is at the heart of the incentive problem which must be solved to find the optimal nonlinear tariff.[36]

4.2 Optimal nonlinear pricing

Continuing the theme from above, suppose that there is a choice between two quantity-fee combinations (q_1, T_1) and (q_2, T_2). Up until now, we have not considered the problems with personal arbitrage from buyers with weak demand. Taking into account the possibility of such arbitrage can greatly complicate the general analysis, and we shall therefore side-step the issue.[37]

In what follows, because quantities are no longer determined by marginal prices, it will be more useful to write gross surplus as a function of the quantity

[36] For further advances on this topic, see Brown and Sibley (1986) and Wilson (1992).

[37] It is straightforward to show that the solution we shall describe is not subject to arbitrage by Type 1 buyers, insofar as it remains the optimal solution when we include an extra constraint to take such arbitrage into account.

consumed. To this end, let $V_i(q)$ denote the gross surplus of a buyer of type i consuming quantity q. If personal arbitrage only arises from buyers with high demand, it is clear that there should be no surplus left for consumers with low demand. If their surplus was positive, raising the tariff T_1 would raise revenues without affecting quantities because this would render Package 1 ((q_1, T_1)) less attractive to Type 2 consumers. This condition for zero surplus for Type 1 consumers can be written as:

$$T_1 = V_1(q_1).$$

On the other hand, a strictly positive surplus must be left to Type 2 buyers, because if the combination (q_1, T_1) is acceptable for Type 1 buyers then the Type 2 buyers would get a strictly positive surplus from this combination. To get them to buy Package 2 ((q_2, T_2)), they must be enticed with at least as much surplus. However, it is clear that it is not necessary to give them a strictly greater surplus because T_2 can be increased without changing anything else. This indifference condition can be written as:

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

or indeed, using the equality $T_1 = V_1(q_1)$:

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

The term in brackets is the "informational rent" which the high demand buyers must be guaranteed under personal arbitrage. If $\lambda > 0$, this rent represents a cost for the firm.

Figure 4 illustrates these constraints for arbitrary quantities q_1 and q_2. The informational rent R_1 is greater the higher the quantity consumed by Type 1 buyers. For the choice of q_2, it can be seen from Figure 4 that if $q_2 < q_2^*$, an increase of this quantity will raise social surplus for a Type 2 buyer as well as raising the firm's profits (with the informational rent unchanged) and it is not necessary to change the allocation destined for Type 1 buyers. By a symmetric argument, if $q_2 > q_2^*$, the firm can improve surplus by reducing the quantity addressed to Type 2 buyers.[38] The firm thus chooses to produce the first-best socially optimal quantity (obtained by ignoring the marginal cost of public funds) for buyers whose demand is high. This is also true for $\lambda > 0$. Profits that can be earned by selling to Type 2 buyers depend on social surplus and informational rent. Because the latter only depends on the quantities sold to Type 1 buyers, it is optimal to choose for Type 2 buyers the quantity that maximizes the corresponding social surplus.

If the firm could extract the full surplus of a Type 1 consumer, it would like this consumer to buy $q_1^* = D_1(c)$, because this would extract the maximal surplus. This would also yield the greatest fiscal revenue, which coincides with the social surplus. Because the firm is uncertain about the buyer type, it must take into account the impact the quantity q_1 has on the rent that would accrue were the demand high (as opposed to low).

Taking into account this possibility leads the firm to choose a quantity below

[38] Recall that the area between marginal cost and demand entails a negative surplus when the latter is below the former.

the first-best optimal one, q_1^*. At q_1^* the effect of a quantity reduction on social surplus for a Type 1 consumer is negligible, while it allows the surplus to a Type 2 consumer to be reduced. Such a reduction is therefore desirable.

More generally, a decrease of q_1 leads to a lower social surplus associated with Type 1 consumers and a decrease in the informational rents to Type 2 consumers. Figure 5 shows that the reduction in social surplus would be $P_1(q_1) - c$, which has a social value of $(1+\lambda)$ per euro, because this amount is entirely appropriated by the firm. The reduction in informational rent, which from the figure is $P_2(q_1) - P_1(q_1)$, is a transfer from Type 2 consumers to the firm and whose social value per euro is therefore λ. The optimal quantity can thus be seen from the figure: it is the value of q_1 for which the ratio of the vertical distance between the lower demand and marginal cost to the vertical distance between the two demand curves is equal to:

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

Formally, the firm chooses q_1 to maximize:

$$\begin{aligned} & \alpha\left[CS_1 + (1+\lambda) PS_1\right] - (1-\alpha)\lambda\left[V_2(q_1) - V_1(q_1)\right] \\ = \quad & \alpha\left[(1+\lambda)(V_1(q_1) - cq_1)\right] - (1-\alpha)\lambda\left[V_2(q_1) - V_1(q_1)\right], \end{aligned}$$

where the last term measures the impact of informational rent on the firm's objective when the buyer is of Type 2 (and a constant term involving $V_2(q_2)$ is omitted). The first order condition can be written as:

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

To interpret this result and compare it to the earlier pricing formulae, it is useful to rewrite it using the fact that the gross surplus derivative of type i is the price $P_i(q)$, given by the inverse demand curve, which is the demand price at which a consumer of type i would choose to consume q units. We can thus write the formula in terms of mark-ups as:

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

This formula allows a comparison with uniform Ramsey-Boiteux pricing. This latter pricing induces consumers to buy less than the first-best optimal quantity in order to generate a producer surplus and pricing above marginal cost. The possibility of generating such a surplus depends inversely on the price elasticity of demand: the more elastic demand is, the less feasible it is to raise price without causing too large of a drop in quantity. With nonlinear pricing, decreasing the quantity bought by low-demand consumers has a completely different role. It allows informational rent to high demand consumers to be reduced, while still motivating them to reveal their demand. The distortion is larger when the impact on rent is higher, as measured by $P_2(q_1) - P_1(q_1)$, and the relative share of high-demanders is higher, as measured by $\frac{1-\alpha}{\alpha}$. Because the informational rent is larger as q_1 increases, the high-demanders would prefer the distortion in q_1 to be as small as possible; their welfare is highest when they only constitute a small fraction of total buyers.

Although it is clear that the price per unit will differ between groups of buyers, nothing guarantees that it is decreasing in quantity. In practice, we often see that those who buy more benefit from larger discounts. For example, membership programs or frequent traveler programs allow those who travel more to pay less.

The current analysis does not immediately apply to this type of rebate, in part because we have not explicitly considered the possibility that large buyers could buy up several packages targeted to small buyers (see Alger, 1999, for more on this topic). More generally, the empirical implications of nonlinear pricing merit a deeper study (recent work includes Cohen, 2002, Ivaldi and Martimort, 1994, Leslie, 2003, and McManus, 2007).

4.3 Multiple qualities and discrimination

While there are certain types of nonlinear pricing in transportation, another frequent practice is to offer several classes of service on the same voyage. This practice goes back a long way, and Jules Dupuit gives an extremely perceptive analysis in his 1849 article.

> *It is clear that by increasing the number of classes indefinitely, consumers could be made to pay all of the utility that they get from using the railway. However, to do that requires distinguishing among consumers who get different utility from their transportation, and make them voluntarily sort themselves into one or another price category. However, this is a great difficulty, which gives rise to a whole*

host of measures which are generally quite poorly understood by the public.

Hence, a good many people, on seeing travelers in third class, traveling without a roof over the carriage, on poorly upholstered seats, they denounce the barbarity of the railways. It would cost very little, they say, to put some meters of leather and kilos of horse-hair [on the seats], and it is beyond greed to withhold them.

It is not because of the several thousand francs which they would have to spend to cover the third class wagons or to upholster the benches that a particular railway has uncovered carriages and wooden benches; it would happily sacrifice this for the sake of its popularity.

Its goal is to stop the traveler who can pay for the second class trip from going third class. It hurts the poor not because it wants them to personally suffer, but to scare the rich. The proof is that if today the State were to say to this railroad: here are one hundred thousand francs to improve your carriages, this subsidy would be certainly refused... improving the third class carriages could reduce revenues by two million francs and ruin the company.

Thus, it is for the same reason that companies, after being cruel to travelers in third class and miserly for those in second, become prodigious for those in first class. After having refused the poor some necessary comforts, they give the rich what is superfluous.

Walras (1875/1897) had a similar view of the logic which guided the pricing of French railways in the middle of the nineteenth century.

> *French railroads ask, respectively, for 10 c. in first class, 7.5 c. in second class, and 5.5 c. in third class; but they put 24 travelers in a first class carriage, 30 in a second class one, and 40 in third class. They also use less comfortable seats, etc. (...)*
>
> *As it happens, the railroads consider, rightly or wrongly, that the average price of 7.66 c., which is close enough to the 7.5 c. which is the second class fare, as being the price of maximal profit; but they do not want to miss the opportunity of getting more from travelers who are prepared to pay more, nor to refuse to get less from travelers who decided not to pay too much.*
>
> *When people earlier rejoiced at the rule of 1857-1858 that required the companies to put windows in third class, and when today they want heating in winter, and they complain about the harshness of the railroads, they are not understanding the key motivation.*
>
> *If the third class carriages were comfortable enough that many second class travelers and some of the first class ones would go there, net total product, as we understand it from the theory of monopoly, would fall. That is all there is to it.*
>
> *The railroads only have third class carriages to avoid missing out on a large number of travelers who, rather than pay the first or second class prices, would have continued to travel in stage coaches.*

Following Dupuit and Walras, the choice of the level of comfort in the different seating classes is effectively driven by the desire to make people pay a price corresponding to their willingness to pay, and to avoid personal arbitrage from those from whom the company wants to extract a high fare. Such arbitrage is discouraged by introducing sufficiently high comfort differences between classes. It is interesting to draw a parallel between the arguments of these authors and our results for nonlinear pricing. We have shown that potential personal arbitrage by high-demanders leads the firm to introduce a difference between the quantities it proposes, a difference which is larger than it would be under perfect discrimination. Dupuit and Walras suggest that railways use a similar logic when they choose the level of comfort. We now show that there is a formal equivalence between these two forms of price discrimination. To show the analogy between discrimination based on different qualities and nonlinear pricing of a homogeneous good, consider the following model due to Mussa and Rosen (1978). We also use an a graphical approach that is an alternative to the previous one that used demand curves.

Suppose that there are two types of users, l and h, differing in their willingness to pay for quality. Assume that their willingness to pay for an extra unit of quality is θ_h and θ_l, respectively, where $\theta_h > \theta_l$. Now let q be the quality level of the service proposed rather than the quantity sold, as it was under nonlinear pricing. The gross surplus of a buyer of type θ_i consuming services of quality q is then given by $V_i(q) = \theta_i q$. Each customer only wants one unit of the good, and her willingness to pay is her gross surplus. Her net surplus for a price T is

$\theta_i q - T$. The marginal cost of service $c(q)$ is increasing and strictly convex.

The combinations of quality q and tariff T that give an equal level of utility to consumer i are illustrated in Figure 6. These indifference curves are straight lines with slope θ_i (with quality on the horizontal axis and euro price on the vertical). The indifference curve through the origin corresponds to zero surplus and is indicated in the sequel by $IC_{i,0}$. Higher surplus levels correspond to indifference curves farther to the right. The producer surplus for one unit of service (one trip) from quality q sold at tariff T is given by $T - c(q)$. If the firm knew the value of θ_i for each traveler, it could perfectly discriminate and leave no surplus to the buyer. The corresponding producer surplus is $\theta_i q - c(q)$, and the firm will therefore choose the quality q_i^* that maximizes this (with social surplus being simply the producer surplus weighted by $1 + \lambda$).

Figure 6 shows the indifference curve for zero surplus and the marginal cost curve as a function of quality. The optimal quality is that which maximizes the distance between the indifference curve and marginal cost, which gives a value q_i^* where the curves have the same slope: formally, $\theta_i = c'(q_i^*)$ (the vertical distance between the two curves is always less than that between the indifference curve and the tangent to marginal cost at q_i^*). With perfect discrimination, the optimal quality will then optimize the willingness to pay for quality with marginal cost of increasing quality.

If the firm does not know the consumers' willingness to pay for quality then, if it were to offer the two first-best optimal qualities, it would necessarily leave some surplus to the high consumer type. Figure 7 shows that the combination

(q_h^*, T_h^*) is to the left of the high-type user's indifference curve, denoted by $IC_{h,l}$, which goes through (q_l^*, T_l^*), and she therefore prefers this latter combination. In order to motivate her not to choose the low quality, the firm must quote her a price such that she is on the indifference curve $IC_{h,l}$. Just as under nonlinear pricing, it is potential personal arbitrage by the high-demanders which constitutes the constraint for the firm. Again, because the firm is not concerned with arbitrage by the low-demanders, it can extract their full surplus so that the optimal combination will be on the indifference curve $IC_{l,0}$. High-demanders enjoy an informational rent of R, which is measured as the vertical difference between the indifference curve $IC_{h,0}$ and the indifference curve $IC_{h,l}$ (which is the difference between the perfectly discriminatory tariff and what they actually have to pay).

The social surplus associated with high-demanders is then the first-best social surplus. Although this rent does not depend on the high quality, the optimal high quality is q_h^*, as with perfect discrimination. On the other hand, as is seen from Figure 7, the firm wants to decrease the low quality below q_l^*. This permits the firm to reduce the informational rent because the indifference curve $IC_{h,l}$ shifts left while, for a small change, the loss of social surplus associated with low-demanders is negligible because q_l^* constitutes its maximum. This model thus confirms Dupuit's (1849) intuition. The comfort in third class is deliberately reduced to dissuade travelers who are ready to pay for higher levels of comfort from traveling at the cheaper fares. Today's economy class air travelers might sympathize. As Tirole (1988, 1993) points out, the comfort offered to the first class travelers actually is not "superfluous" because it is the level chosen under perfect discrimination; contrary to what Dupuit and Walras thought, it is only by adjusting the lower quality that the firm discourages personal arbitrage. There is therefore a perfect analogy between this model and that of nonlinear pricing, and the results are the same (under reinterpretation). Only passengers whose willingness to pay is high can retain some surplus, and this surplus increases as the fraction of high demanders decreases.

One variant of this model explored by Chander and Leruth (1989), and particularly relevant for transportation, supposes that the quality of comfort class decreases in the number of users as it becomes more congested. The firm then chooses two different prices, with the cheaper class having a lower quality just because it attracts more travelers. The two classes in the Parisian Metro until the 1980s give a striking illustration of this type of strategy.[39] Second class carriages only differed from first class by their color, but fares were lower.[40]

The modern day counterpart to the insights of Dupuit and Walras can be found without much difficulty in air travel. What was true in nineteenth century train carriages sometimes seems to be not very far off from what is found on modern economy class flights. Airlines could scarcely charge such premiums for first and business class travel if economy class were more comfortable.[41]

[39] Another illustration is that of a toll road with a parallel freeway to the same destination (see de Palma and Lindsey, 2000).

[40] This system was abolished by Charles Fiterman, Communist Transportation Minister in the first government of the Union of the Left.

[41] In practice, the problem faced by the airlines is very complicated because the information that it has about passengers is very sparse and evolves over time. See the sub-section on Yield Revenue Management above.

5 Conclusion

We have focused on discriminatory pricing for a single firm (without rivals) and suggested how this analysis is relevant for transportation. We analyzed a public firm under the assumption that public funds are valued more than consumer surplus, in order to account for inefficiencies in raising tax revenue. Profit maximization arises as a limit case when λ goes to infinity.

This theoretical framework enables us to highlight the similarity between the pricing problem of a public firm and that of a private firm. However, restricting this analysis to monopolies is more restrictive for the private sector. Although several transport modes are effectively monopolies, such as the railways and urban transportation systems, other sectors have several competing carriers, most notably the airlines. We now give a brief review of oligopoly competition[42]. For the discussion that follows, we consider private profit maximizing firms and we let $\lambda = 0$, so that consumer surplus and producer surplus are equally weighted.[43]

There is relatively little work on price discrimination under competition. As regards discrimination without personal arbitrage, standard oligopoly theory fairly quickly delivers some main conclusions. For one good sold to several groups the model most directly comparable with monopoly is Cournot's framework, where firms choose outputs and price equates aggregate output with the quantity demanded. If each firm has the same marginal costs c_i for serving

[42] For a review of the state of the art on price discrimination under oligopoly, see Stole (2007), and Armstrong and Vickers (2001).

[43] The Competition Authority of the European Union seems to put more weight on consumer surplus.

market i, Lerner's formula becomes:

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

where n_i is the number of active firms in market i. If firms have different costs, Lerner's formula becomes:

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

where

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

is the average cost for serving market i. This formula is directly comparable to the monopoly one. If two groups differ by elasticity of demand, we again find that the price is higher for inelastic demand. However, all prices go to marginal cost when the number of firms becomes large, and competition eliminates price differences. These results can also be applied when firms discriminate with different products, in which case the marginal cost differs from one market to another. A richer framework can be analyzed by supposing that there is also differentiation within each market. One common formulation for demand for differentiated products is founded on discrete choice models, which are described in greater detail in other chapters of this book. One interesting feature of these models is that it is easy to introduce new features, such as quality differences or congestion, into the consumer preferences which are at the heart of the model.[44]

[44] Anderson, de Palma, and Thisse (1992a) provides background for this model.

The multinomial logit model is a particularly useful formulation. If prices are the strategic variable, the equilibrium price of firm j in market i is

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

where D_{ij} is the equilibrium demand addressed to firm j in market i and depends on prices set by all active firms in market i. Anderson and de Palma (2001) show that this price equilibrium has several intuitive properties. For example, if consumers have a higher tendency to buy the product, then prices will be high and the differences will be large across variants. Furthermore, the more firms there are, or the more similar they are (from the point of view of a group of consumers), the lower prices will be and the more similar they will be across firms.

It can also be understood within this framework why introducing competition in a market can actually lead discriminating firms to exacerbate price differences across different services. Borenstein and Rose (1994) note that following airline deregulation in the United States, price differences increased. The theoretical explanation that they propose is that those travelers who are willing to pay more for a trip (hence, those for whom the reservation value is high) are also those who are the most loyal to a particular airline, which translates to a greater effective heterogeneity between products. When competition is introduced into the marketplace, the difference between the high business class fare and the low economy class fare is amplified by the differential intensity of competition at the two service levels. Competition is more intense for the economy segment.[45] However, more recent work by Gerardi and Shapiro (2007) finds that competition decreases the power of an airline to price discriminate. Price dispersion within the airline industry falls as competition increases, especially on routes where consumers have relatively heterogeneous elasticities of demand. Looking at patterns within different groups of air travelers, Bilotkach (2005) looks at across-airlines differences in economy class fares for the London-New York market, aimed at different consumer types. He finds that fares targeted to business fliers differ across airlines, while fares targeted to leisure fliers do not.

One issue in the economics of third-degree price discrimination under competition concerns the level of producer surplus when discrimination is not always possible (it could, for example, be illegal). Several authors came to the same conclusion: in contrast to monopoly, it may be that profits are lower when firms can discriminate. Hoover (1948, p. 57) anticipated this result in the context of spatial discrimination.

> *The difference between market competition under f.o.b. pricing (with strictly delineated market areas) and under discriminatory delivered pricing is something like the difference between trench warfare and guerrilla warfare. In the former case all the fighting takes place along a definite battle line; in the second case the opposing forces are intermingled over a broad area.*

This indicates that the implications of discrimination for profits can be very

[45] Borenstein and Rose base their theoretical arguments on Borenstein (1985) and Holmes (1989).

different, depending on the degree of competition. However, it remains true that welfare falls following the introduction of discrimination if total output does not rise.

In contrast to the analysis of third-degree discrimination, the study of discrimination under personal arbitrage in oligopoly is complicated and there are relatively few general contributions. Notable exceptions include Champsaur and Rochet (1989), Ivaldi and Martimort (1994), Stole (1995), Armstrong and Vickers (2001), and Rochet and Stole (2001). The latter two articles show that if duopolists offer relatively close services, nonlinear pricing can lead to two-part tariff with zero profits for each firm (an extension of the Bertrand Paradox). Rochet and Stole (2001) also show that if the services offered are more differentiated, equilibrium pricing resembles the monopoly case.

Theoretical progress on these issues is quite tricky, but it is important to continue to develop the theoretical framework which will allow empirical studies to examine discrimination with personal arbitrage under oligopoly.

References

[1] Alger, Ingela (1999). Consumer Strategies Limiting the Monopolist's Power: Multiple and Joint Purchases. *RAND Journal of Economics*, 30, 736-757.

[2] Anderson, Simon P. (1989). Socially Optimal Spatial Pricing. *Regional Science and Urban Economics*, 19, 60-86.

[3] Anderson, Simon P. and André de Palma (1989). Spatial Price Discrimination with Heterogeneous Products. *Review of Economic Studies*, 55, 573-592.

[4] Anderson, Simon P. and André de Palma (2001). Product Diversity in Asymmetric Oligopoly: Is the Quality of Consumer Goods Too Low? *Journal of Industrial Economics*, 49, 113-135.

[5] Anderson, Simon P., André de Palma and Brent Kreider (2001). Tax Incidence In Differentiated Products Oligopoly. *Journal of Public Economics*, 81, 173-192.

[6] Anderson, Simon P., André de Palma and Brent Kreider (2001). Efficiency of Indirect Taxes Under Imperfect Competition. *Journal of Public Economics*, 81, 231-251.

[7] Anderson, Simon P., André de Palma and Jacques-François Thisse (1989). Spatial Price Policies Reconsidered. *Journal of Industrial Economics*, 38, 1-18.

[8] Anderson, Simon P., André de Palma and Jacques-François Thisse (1992a). *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge, Mass.

[9] Anderson, Simon P., André de Palma and Jacques-François Thisse (1992b). Social Surplus and Profitability under Different Spatial Pricing Policies. *Southern Economic Journal*, 58, 934-949.

[10] Anderson, Simon P. and Maxim Engers (2007a). Consumer Surplus, Article for *International Encyclopedia of the Social Sciences*, 2nd Ed.

[11] Anderson, Simon P. and Maxim Engers (2007b). Producer Surplus, *Article for International Encyclopedia of the Social Sciences*, 2nd Ed.

[12] Anderson, Simon P. and Victor Ginsburgh (1999). International Pricing with Costly Consumer Arbitrage. *Review of International Economics*, 7, 126-139.

[13] Anderson, Simon P. and Luc Leruth (1993). Why Firms May Prefer not to Price Discriminate via Mixed Bundling. *International Journal of Industrial Organization*, 11, 49-61.

[14] Anderson, Simon P. and Régis Renault (2003a). Efficiency and Surplus Bounds in Cournot Competition. *Journal of Economic Theory*, 113, 253-264.

[15] Anderson, Simon P. and Régis Renault (2003b). Second degree price discrimination à la Pigou. Working Paper, Université de Cergy-Pontoise.

[16] Anderson, Simon P. and Yves Schneider (2008) Price Competition with Revenue Management and Costly Consumer Search. Mimeo, University of Virginia.

[17] Anderson, Simon P. and Jacques-François Thisse (1988). Price Discrimination in Spatial Competitive Markets. *European Economic Review*, 32 , 578-590.

[18] Anderson, Simon P. and Wesley W. Wilson [2008]. Spatial Competition, Pricing, and Market Power in Transportation: A Dominant Firm Model. *Journal of Regional Science*, 48(2), 367-397.

[19] Armstrong, Mark and John Vickers (2001). Competitive price discrimination. *RAND Journal of Economics*, 32, 579-605.

[20] Babu, Nahata, Krzysztof Ostaszewski, and P. K. Sahoo, (1990). Direction of price changes in third-degree price discrimination. *American Economic Review*, 80, 1254-1258.

[21] Ballard, Charles L., John B. Shoven and John Whalley (1985). General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States. *American Economic Review*, 75, 128-38.

[22] Bilotkach, Volodymyr (2005). *Understanding Price Dispersion in the Airline Industry: Capacity Constraints and Consumer Heterogeneity*. Advances in Airline Economics, volume 1 (ed. Darin Lee).

[23] Boîteux, Marcel (1956). Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica*, 24, 22-40.

[24] Borenstein, Severin (1985). Price discrimination in free entry markets. *RAND Journal of Economics*, 16, 380-397.

[25] Borenstein, Severin (1991). Selling costs and switching costs: explaining retail gasoline margins. *RAND Journal of Economics*, 22, 354-369.

[26] Borenstein, Severin and Nancy Rose (1994) Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, 102, 653-683.

[27] Brander, J. A. and A. Zhang (1990). "Market Conduct in the Airline Industry: An Empirical Investigation," *RAND Journal of Economics*, 21, 567-583. (Forthcoming in K. Button and P. Nijkamp (eds.), Classics in Transport Analysis, Edward Elgar Publishing Ltd.)

[28] Brown S. and Sibley, David (1986). *The theory of public utility pricing.* Cambridge University Press, Cambridge, England.

[29] Champsaur, Paul and Jean-Charles Rochet (1989). Multiproduct Duopolists. *Econometrica*, 57, 533-557.

[30] Chander, Parkash and Luc Leruth (1989). The Optimal Product Mix for a Monopolist in the Presence of Congestion Effects: A Model and Some Results. *International Journal of Industrial Organization*, 7, 437-449.

[31] Clerides, Sofronis (2004). Price discrimination with differentiated products: definition and identification. *Economic Inquiry*, 42, 402-412.

[32] Cohen, Andrew (2002) Package Size and Price Discrimination in the Paper Towel Market. Working Paper. University of Virginia.

[33] Corts, Kenneth S. (1998). Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment. *RAND Journal of Economics*, 29, 306-323.

[34] Cournot, Augustin (1838). *Recherches sur les principes mathématiques de la théorie des richesses.* L. Hachette, Paris.

[35] De Palma, André and Robin Lindsey (1998). Information and Usage of Congestible Facilities Under Different Pricing Regimes, *The Canadian Journal of Economics,* 31, 666-692.

[36] De Palma, André and Robin Lindsey (2000). Private toll roads : competition under various ownership regimes. *The Annals of Regional Science*, 34, 13-35.

[37] Dupuit, Jules (1849). De l'influence des péages sur l'utilité des voies de communication. *Annales des ponts et chaussées*, 17, mémoires et documents, 207.

[38] Gerardi, Kristopher and Shapiro, Adam Hale (2007). The Effects of Competition on Price Dispersion in the Airline Industry: A Panel Analysis. Federal Reserve Bank of Boston Working Paper 07-7.

[39] Gerstner, Eitan (1986). Peak Load Pricing in Competitive Markets. *Economic Inquiry*, 24, 349-361.

[40] Hausman, Jerry A. and James Poterba (1987). Household Behavior and the Tax Reform Act of 1986. *Journal of Economic Perspectives*, 1, 101-119.

[41] Holahan, William L. (1975). The welfare effects of spatial price discrimination. *American Economic Review*, 65, 498-503.

[42] Holmes, Thomas J. (1989). The effects of third degree price discrimination in oligopoly. *American Economic Review*, 79, 244-250.

[43] Hoover, E. M. (1937). Spatial price discrimination. *Review of Economic Studies*, 4, 182-191.

[44] Hoover, E. M. (1948). *The location of economic activities.* New York, McGraw-Hill.

[45] Hotelling, Harold (1938). The general welfare in relation to problems of taxation and of railway and utility rates. *Econometrica*, 6, 242-269.

[46] Ivaldi, Marc and David Martimort (1994). Competition under Nonlinear Pricing. *Annales d'Economie et de Statistique*, 34, 71-115.

[47] Laffont, Jean-Jacques, and Jean Tirole (1993). *A theory of incentives in procurement and regulation.* MIT Press, Cambridge, Mass.

[48] Lederer, Phillip J. and Arthur P. Hurter, (1986) Competition of firms: discriminatory pricing and location. *Econometrica,* 54, 623-40.

[49] Leslie, Phillip, (2003) Price Discrimination in Broadway Theatre. *RAND Journal of Economics*, 35, 520-541.

[50] Levinson, D. and A. Odlyzko (2007). Too expensive to meter: The influence of transaction costs in transportation and communication. http://rational.ce.umn.edu.

[51] Lott, John R. and R. D. Roberts (1991). A guide to the pitfalls of identifying price discrimination. *Economic Inquiry*, 29, 14-23.

[52] McAfee, R. Preston, Hugo M. Mialon, and Sue H. Mialon (2006). Does large price discrimination imply great market power? *Economic Letters*, 3, 360-367.

[53] McManus, Brian (2007) Nonlinear Pricing in an Oligopoly Market: the Case of Specialty Coffee. *RAND Journal of Economics*, 38(2), 512-532.

[54] Meade, James (1944). Price and output policy of state enterprise. *Economic Journal*, 54, 321-328.

[55] Mohring, Herbert [1976] *Transportation Economics.* Cambridge, MA, Ballinger.

[56] Mougeot, Michel and Florence Naegelen, (1994) La discrimination par les prix. *Economica*, Paris, France.

[57] Mussa, Michael, and Sherwin Rosen, (1978). Monopoly and product quality, *Journal of Economic Theory*, 18, 301-317.

[58] Nahata, Babu, Krzysztof Ostaszewski, and P. K. Sahoo (1990). Direction of Price Changes in Third-Degree Price Discrimination. *American Economic Review*, 80, 1254-1258.

[59] Oi, Walter, (1971). A Disneyland dilemma: Two-part tariffs for a Mickey Mouse monopoly, *Quarterly Journal of Economics*, 85, 77-90.

[60] Phlips, Louis (1983). *The economics of price discrimination.* Cambridge University Press, Cambridge, UK.

[61] Pigou, Arthur Cecil (1938). *The economics of welfare*. Macmillan and Co., London. 4th edition, 1932, reprinted 1938.

[62] Ramsey, Frank (1927). A contribution to the theory of taxation. *Economic Journal*, 37, 47-61.

[63] Rietveld, Piet, and Roberto Roson [2002] Direction dependent prices in public transport: A good idea? The back haul pricing problem for a monopolistic transport firm. *Transportation*, 29, 397-417.

[64] Rochet, Jean-Charles and Lars A. Stole (2001). Nonlinear pricing with random participation. *Review of Economic Studies*, 69, 277-311.

[65] Schmalensee, Richard (1981). Output and welfare implications of monopolistic third degree price discrimination. *American Economic Review*, 71, 242-247.

[66] Seade, J. E. (1985). Profitable cost increases and the shifting of taxation: Equilibrium responses of markets in oligopoly. University of Warwick Discussion paper no. 260.

[67] Shepard, Andrea (1991). Price discrimination and retail configuration. *Journal of Political Economy*, 99, 30-53.

[68] Sinsou, Jean-Paul (1999). *Yield & Revenue Management: Optimisation de la recette dans les transport de passagers.* Les presses de l'institut du transport aerien, Paris.

[69] Stigler, George (1963). A note on block booking. Supreme Court Review. Reprinted in *Organization of Industry* (University of Chicago Press).

[70] Stole, Lars A. (1995). Nonlinear pricing and oligopoly. *Journal of Economics and Management Studies*, 4, 529-562.

[71] Stole, Lars A. (2007). Price Discrimination in Competitive Environments. *Handbook of Industrial Organization*, Vol. 3, Chapter 34. Eds. Mark Armstrong and Robert H. Porter, North-Holland, Amsterdam.

[72] Tirole, Jean (1988). *The Theory of Industrial Organization.* MIT Press, Cambridge, Mass.

[73] Tirole, Jean (1993). *La théorie de l'organisation industrielle*, volume 1. Editions Economica, Paris, France.

[74] Thisse, Jacques-François and Xavier Vives (1988). On the strategic choice of spatial price policy. *American Economic Review*, 78, 122-137.

[75] Varian, Hal (1989). Price discrimination. *Handbook of Industrial Organization*, Vol. 1, Chapter 10. Eds. Richard Schmalensee and Robert Willig, North-Holland, Amsterdam.

[76] Varian, Hal (2000) *Introduction à la microéconomie.* DeBoeuck Université.

[77] Vickrey, William S. (1969). Congestion theory and transport investment. *American Economic Review* (Papers and Proceedings), 59, 251-261.

[78] Walras, Léon (1897). L'Etat et les chemins de fer. Manuscript written 1875. *Revue du Droit Public et de Science Politique*. Mai-juin et juillet-août 1897.

[79] Wilson, Robert (1992). *Non-linear pricing*. Oxford University Press, Oxford, England.

[80] Yin, Xiangkang (2004). Two part tariff competition in duopoly. *International Journal of Industrial Organization*, Elsevier, 22(6), 799-820.