

As the scale economy index S is defined as the ratio of average cost to marginal cost, it follows that: $S = 1/\eta_C$.

RAY AVERAGE COST AND MULTIPRODUCT SCALE ECONOMIES

Scale economies are indicated by declining average cost, or declining Ray Average Cost (RAC) for a multiproduct firm. If a firm has 2 products so that its cost function is $C(q_1, q_2) + F$ we may implicitly define total output q by the equations $q_1 = \lambda_1 q$, and $q_2 = \lambda_2 q$, where λ_1 and λ_2 sum to unity. Then ray average cost is:

$$RAC(q) = \frac{C(\lambda_1 q, \lambda_2 q) + F}{q} \quad (4.A4)$$

Multiproduct scale economies imply that RAC declines as output expands or that

$$\frac{dRAC(q)}{dq} = \frac{q(\lambda_1 MC_1 + \lambda_2 MC_2) - C(\lambda_1 q, \lambda_2 q) - F}{q^2} < 0 \quad (4.A5)$$

where MC_i is the marginal cost of producing good i . It follows immediately that the sign of $dRAC(q)/dq$ is determined by the sign of the numerator of this expression. In other words, if $q_1 MC_1 + q_2 MC_2 > C(q_1, q_2) + F$ then $dRAC(q)/dq > 0$, while if $q_1 MC_1 + q_2 MC_2 < C(q_1, q_2) + F$ then $dRAC(q)/dq < 0$. Now define the ratio

$$S^M = \frac{C(q_1, q_2) + F}{q_1 MC_1 + q_2 MC_2} \quad (4.A6)$$

If $S^M > 1$ ray average cost decreases with output and so exhibits multiproduct increasing returns to scale. If $S < 1$, ray average cost is increasing, and so exhibits multiproduct decreasing returns to scale. If $S = 1$, neither scale economies nor diseconomies exist for the multiproduct firm.

FORMAL COST FUNCTION ANALYSIS AND EMPIRICAL ESTIMATION: THE COBB-DOUGLAS CASE

Assume that production is generated from capital K and labor L inputs via a Cobb-Douglas production function:

$$Q = K^\alpha L^\beta \quad (4.A7)$$

If r is the rental price of capital and w is the wage rate of labor, then total cost $C = rK + wL$. Cost minimization requires choosing the capital and labor inputs that minimize cost for achieving a given level of output. If we denote the target output level as \bar{Q} , the firm's problem then becomes:

$$\text{Minimize } C = rK + wL \text{ subject to } \bar{Q} = K^\alpha L^\beta$$

Since the production relation implies: $L = \bar{Q}^{\frac{1}{\beta}} K^{-\frac{\alpha}{\beta}}$, the cost function becomes $C = rK + w\bar{Q}^{\frac{1}{\beta}} K^{-\frac{\alpha}{\beta}}$. Minimizing this with respect to capital and solving for K yields:

$$K = \left(\frac{\alpha}{\beta} \frac{w}{r} \right)^{\frac{\beta}{\alpha+\beta}} \bar{Q}^{\frac{1}{\alpha+\beta}} \quad (4.A8)$$

Substituting this value into the labor input implied by the production constraint, the cost-minimizing labor input is:

$$L = \left(\frac{\beta}{\alpha} \frac{r}{w} \right)^{\frac{\beta}{\alpha+\beta}} \bar{Q}^{\frac{1}{\alpha+\beta}} \quad (4.A9)$$

Together, the optimal K and L equations imply that the minimal cost for producing any given level of output \bar{Q} is:

$$C = w \left(\frac{\beta}{\alpha} \frac{r}{w} \right)^{\frac{\beta}{\alpha+\beta}} \bar{Q}^{\frac{1}{\alpha+\beta}} + r \left(\frac{\alpha}{\beta} \frac{w}{r} \right)^{\frac{\beta}{\alpha+\beta}} \bar{Q}^{\frac{1}{\alpha+\beta}} = \left[\left(\frac{\alpha}{\beta} \right)^{\frac{\beta}{\alpha+\beta}} + \left(\frac{\beta}{\alpha} \right)^{\frac{\alpha}{\alpha+\beta}} \right] r^{\frac{\alpha}{\alpha+\beta}} w^{\frac{\beta}{\alpha+\beta}} \bar{Q}^{\frac{1}{\alpha+\beta}} \quad (4.A10)$$

The exponents for each factor price, r and w , sum to unity so that the cost function is homogeneous of degree one in factor prices. Since the exponent on the output level is $1/(\alpha + \beta)$ there will be constant returns to scale if $\alpha + \beta = 1$; scale economies if $\alpha + \beta > 1$; and scale diseconomies if $\alpha + \beta < 1$.

$$\ln C = \ln \left[\left(\frac{\alpha}{\beta} \right)^{\frac{\beta}{\alpha+\beta}} + \left(\frac{\beta}{\alpha} \right)^{\frac{\alpha}{\alpha+\beta}} \right] + \left(\frac{\alpha}{\alpha+\beta} \right) \ln r + \left(\frac{\beta}{\alpha+\beta} \right) \ln w + \left(\frac{1}{\alpha+\beta} \right) \ln \bar{Q} \quad (4.A11)$$

For estimation purposes, this can easily be translated into: $\ln C = \text{Constant} + \delta_1 \ln r + \delta_2 \ln w + \delta_3 \ln Q$. With observations on input prices and output levels, estimation of the coefficients δ_i allow us to recover the underlying production parameters, α and β .

If the assumption of Cobb-Douglas technology is incorrect, the estimated coefficients will not satisfy theoretical restrictions. Hence, empirical analysis often employs a more flexible specification that includes Cobb-Douglas as a special case. One such specification is the translog cost function. For the two-input case above, this function is:

$$\begin{aligned} \ln C = & \text{Constant} + \delta_1 \ln r + \delta_2 \ln w + 0.5[\delta_{11}(\ln r)^2 + \delta_{12}(\ln w)(\ln r) \\ & + \delta_{21}(\ln w)(\ln r) + \delta_{22}(\ln w)^2] + \delta_3 \ln Q \\ & + \delta_{31}(\ln Q)(\ln r) + \delta_{32}(\ln Q)(\ln w) + 0.5\delta_{33}(\ln Q)^2 \end{aligned} \quad (4.A12)$$

As before, we expect δ_1 and δ_2 to be positive fractions that sum to unity. However, the measure of scale economies $S = 1/\frac{\partial \ln C}{\partial \ln Q}$ now depends on the level of output. That is we now have: $S = 1/\frac{\partial \ln C}{\partial \ln Q} = 1/(\delta_3 + \delta_{33} \ln Q + \delta_{31} \ln r + \delta_{32} \ln w)$. Only if $\delta_{31} = \delta_{32} = \delta_{33} = 0$, will the index of scale economies be independent of the level of Q .

Part Two

Monopoly Power in Theory and Practice

In Part Two, we consider the pure monopoly case in much more detail than the simple, textbook case presented in Chapter 2. In particular, we consider a single firm facing a downward-sloping demand curve and the price and non-price tactics that it may use. While focusing on a single firm omits much strategic interaction, there are some cases in which this is realistic. In many regions, for example, there is just one ski lift operator within a radius of fifty miles and only one amusement park serving an even greater area. Second, and more importantly, the tactics discussed such as quantity discounts and bundling are also available in a setting of multiple rivals. However, it is much easier to understand the role of such measures in this more competitive environment *after* seeing them used by a single firm with no strategically-linked rival.

Chapter 5 begins with an examination of linear price discrimination tactics, such as market segmentation. It also includes an analysis of the 2003 study of price discrimination in new car sales by Scott-Morton, Zettelmeyer, and Silva-Risso. This is followed by a discussion of nonlinear price discrimination techniques, e.g., two-part tariffs, in Chapter 6. That chapter also includes a discussion of the evidence on airline price discrimination presented by Stavins (2001).

In Chapter 7, we explore the choice of product quality. This allows us to introduce the two concepts of horizontal differentiation, in which consumers disagree about what makes a high-quality product (but may agree on the value of higher quality), and vertical differentiation, in which consumers agree about the quality ranking of different goods (but typically disagree about quality's marginal value). Both of these concepts, but especially that of horizontal differentiation, will be used extensively in subsequent chapters. We also present the empirical analysis of Berry and Waldfogel (2010) linking newspaper quality to market size.

Chapter 8 reviews the classic techniques of bundling and tie-in sales. Here we consider not only the surplus-extracting aspects of such tactics, but also their potential for leveraging market power from one market to another. This permits a discussion of the related (and extensive) developments in antitrust policy. The chapter concludes with an illustration of the basic, profit-enhancing effects that bundling can have based on the empirical study by Crawford (2008) of bundling in cable TV.

5

Price Discrimination and Monopoly: Linear Pricing

The standard definition of price discrimination is that a seller sells the same product to different buyers at different prices. Examples of price discrimination abound. Consider, for example, the market for prescription and generic drugs. Both casual and formal empirical evidence finds that brand name prescription drug prices are on average lower in Canada than in the United States. Table 5.1 demonstrates this with evidence from two recent studies, one by Skinner and Rovere (2008) and the other by Quon, Firszt, and Eisenberg (2005). While the precise drugs covered by each study differ somewhat, the basic conclusion in each is the same. US consumers pay between 24 percent and 57 percent more for prescription drugs on average than do their Canadian neighbors.

The case of prescription drugs is far from an isolated example. Passenger airline companies and hotel chains are past masters at what they euphemistically call “yield management,” as any frequent traveler can readily attest. In the great majority of business-to-business transactions, prices are arrived at through prolonged negotiation. Sophisticated travelers who visit the Grand Bazaar in Istanbul know that all prices are negotiable—and are typically lower if the traveler happens to be accompanied by a native of Istanbul who does the negotiating! Nearer to home, purchasing a pre-owned automobile usually involves equally intensive negotiation. In each of these cases of negotiated prices, there is no reason to believe that the price offered to one buyer will be the same as the price offered to another.

The extensive use of price discrimination by sellers raises two sets of questions. First, what market conditions make price discrimination feasible? Second, what makes price discrimination profitable? It is, after all, difficult to believe that discriminatory pricing is extensively employed but unprofitable! Essentially, our task in this and upcoming chapters is to analyze the methods firms can use to implement price discrimination in a way that increases profits relative to charging the same price to everyone.

This raises another important issue. Any increased profit that price discrimination generates must come either from a reduction in consumer surplus, improved market efficiency, or some combination of the two. From a policy perspective, it matters a great deal which of these is the case. As a result, we also want to explore the welfare implications of price discrimination. Finally, it is worthwhile noting that discriminatory prices can also affect market competition. This occurs when the buyers are not final consumers but instead, retailers such as drug stores. If large drug store chains are charged different wholesale prices than are small, independent pharmacies, then retail competition between these two

Table 5.1 Comparison of US and Canadian prescription prices, selected drugs

<i>Brand Name</i>	<i>Percent Canadian Reduction from US Price [Skinner & Rove (2008)]</i>	<i>Percent Canadian Reduction from US Price [Quon, Firszt, and Eisenberg (2005)]</i>
Accupril	43%	17%
Altace	54%	40%
Arthrotec	67%	—
Ativan	88%	—
Avandia	—	27%
Celebrex	62%	53%
Coversyl	57%	—
Crestor	57%	21%
Diovan HCT	54%	27%
Flomax	—	40%
Fosomax	—	30%
Lipitor	40%	34%
Lopressor	96%	—
Neurontin	—	14%
Prozac	—	50%
Vasotec	21%	—
Viagra	—	-37%
Zestril	46%	—
Average	57.08%	26.33%



groups will not be conducted on a level playing field. We address these issues in the next group of chapters.

5.1 FEASIBILITY OF PRICE DISCRIMINATION

A firm with market power faces a downward sloping demand curve. So if the firm charges the same price to each consumer—the standard case of non-discriminatory pricing—the marginal revenue it gets from selling an additional unit of output is less than the price charged. In order to sell the additional unit, the firm must lower its price not only to the consumer who buys the additional unit, but to all its other consumers as well. Having to lower the price for all its customers in order to gain an additional consumer weakens the monopolist's incentive to serve more consumers. As a result, the textbook monopoly undersupplies its product relative to the efficient outcome.

Non-discriminatory pricing by a monopolist is not just a source of potential inefficiency. It is also a constraint on the firm's ability to convert consumer surplus into profit, particularly from those consumers willing to pay a lot for its product. Price discrimination is a powerful technique that can greatly increase firm profits. In some cases, moreover, price discrimination may induce the monopolist to sell more output thereby coming closer to the competitive outcome and enhancing market efficiency.

While a monopolist can increase profit through price discrimination, it is important to realize that price discrimination is not always easily accomplished. There is a reason that

the standard textbook case assumes that each customer pays the same price. To discriminate successfully, the monopolist must overcome two main obstacles. The first of these is the *identification problem*: the firm needs to be able to identify who is who on its market demand curve. The second is the problem of *arbitrage*: being able to prevent those who are offered a low price from reselling to those charged a high price.

In considering the identification problem, it is useful to recall a common assumption in the textbook monopoly model: that the firm knows the quantity demanded at each price. Without this knowledge the firm would not know its marginal revenue curve and, hence, would not be able to determine the profit-maximizing output. Let's examine more carefully what this assumption means in practice.

For some products such as bicycles, TVs, DVD players, or haircuts, each consumer purchases at most one unit of the good over a given period of time. The firm's market demand curve is then an explicit ordering of consumers by their reservation prices—the top price each is willing to pay. For these goods, knowledge of the demand curve means that the monopolist knows that the top part of the demand curve is made up of those consumers willing to pay a relatively large amount for the one unit they will purchase, whereas the bottom part of the demand curve is made up of those willing to pay only a little.

For other products, such as movies, CDs, refreshments, and tennis lessons, the construction of the market demand curve is slightly more complex. This is because each individual consumer can be expected to purchase an increasing (decreasing) quantity of the good as the price is reduced (increased). As a result, for these goods the market demand curve reflects not only differences in the willingness to pay across consumers, but also differences in the willingness to pay as any one consumer buys more of the product.

When the monopolist practices uniform pricing, these distinctions are not relevant. In that case, the assumption that the firm knows its demand curve means only that it knows how willingness to pay for the good *in the overall market* varies with the quantity of the good sold. To be able to practice price discrimination, the monopolist must be able to acquire and exploit more information about consumers than is assumed in the standard model. The monopolist must know how the market demand curve has been constructed from the individual consumer demand curves. In other words, the monopolist must know how different kinds of consumers differ in their demands for its good.

This is easier for some sellers than for others. For example, tax accountants effectively sell one unit of their services to each client in any given year. Further, they know exactly how much their clients earn and, more importantly, how much they save their clients by way of reduced tax liabilities. They can certainly use this information to identify their clients' willingness-to-pay. Similarly, a car dealer typically sells one car to a customer. The dealer may be able to identify those buyers with the greatest or least willingness to pay by asking potential buyers where they live, work, or shop and more generally through the process of negotiation. The same is often true for realtors, dentists, and lawyers.

Sellers of retail merchandise, however, face a more anonymous market. Various schemes such as varying the price depending on time of purchase—"early-bird" specials or Saturday morning sales—or offering coupons that take time to collect, can help retailers identify "who's who" on their demand curve. Nevertheless, the identification problem is still difficult to overcome. Moreover, even if weekend sales or coupon schemes do successfully identify the firm's different consumers, such schemes may be too costly to implement.

Even when a monopolist can solve the identification problem, there is still the second obstacle to price discrimination: arbitrage. As was noted previously, to discriminate successfully the monopolist must be able to prevent those consumers who are offered a low

price from reselling their purchases to other consumers to whom the monopolist wants to charge a high price. Again, this is more easily accomplished for some goods and services than for others. Medical, legal, and educational services are not easily resold. One consumer can't sell her appendectomy to another! Similarly, a senior citizen cannot easily resell a discounted movie theater ticket to a teenager. For other markets, particularly consumer durables such as bicycles and automobiles, resale—or sale across different markets—is difficult to prevent. This is an important part of the drug pricing story noted at the start of this chapter. Pharmaceutical companies can only price discriminate successfully if they can keep the American and Canadian markets separate, in other words, only if they can prevent arbitrage.

To sum up, we expect firms with monopoly power to try to price discriminate. In turn, this implies that we should expect these firms to develop techniques by which they can identify the different types of consumers who buy their goods and prevent resale or consumer arbitrage among them. The ability to do this and the best strategy for achieving price discrimination will vary from firm to firm and from market to market.

We now turn to the practice of price discrimination and investigate some of the more popularly practiced techniques. The tradition in economics has been to classify these techniques into three broad classes: first degree, second degree, and third degree price discrimination.¹ More recently, these types of pricing schemes have been referred to respectively as personalized pricing, menu pricing, and group pricing.² In this chapter, we focus on third-degree price discrimination, or group pricing.

5.2 THIRD-DEGREE PRICE DISCRIMINATION OR GROUP PRICING

Third-degree price discrimination, or group pricing, is defined by three key features. First, there is some easily observable characteristic such as age, income, geographic location, or education status by which the monopolist can group consumers in terms of their willingness to pay for its product. Second, the monopolist can prevent arbitrage across the different groups. In the prescription drug case with which we started this chapter, this implies that it is possible to prevent the re-importation to the United States of prescription drugs initially exported from the United States to Canada. Finally, a central feature of third-degree price discrimination is that, while the monopolist quotes different unit prices to different groups, all consumers *within* a particular group are quoted the same unit price. Consumers in each group then decide how much to purchase at their quoted price.

Group pricing is an example of price discrimination precisely because the price quoted to one group of consumers is not the same as the price quoted to another group *for the same good*. This type of pricing policy is the one most commonly found in economics textbooks and is referred to in the industrial organization literature as *linear pricing*—hence the title of this chapter. Consumers within a group are free to buy as much as they like at the quoted price, so that the average price per unit paid by each consumer is the same as the marginal price for the last unit bought.

¹ Price discrimination is a fascinating topic and its interest to economists goes well beyond the field of industrial organization. The distinction between first, second, and third degree discrimination follows the work of Pigou (1920). A more modern treatment appears in Phlips (1983).

² These terms were first coined by Shapiro and Varian (1999).

Table 5.2 Annual membership dues for the American Economic Association

Annual Income	Subscription Price
Less than \$70,000	\$20
\$70,000 to \$105,000	\$30
Above \$105,000	\$40

The world is full of examples of third-degree price discrimination. Senior discounts and “kids are free” programs are both examples. An interesting case that is familiar to economists is the fee schedule for membership in the *American Economic Association*, the major professional organization for economists in the United States. Payment of the fee entitles a member to receive professional announcements, newsletters, and three very important professional journals, *The American Economic Review*, *The Journal of Economic Perspectives*, and *The Journal of Economic Literature*, each of which is published quarterly.

The 2012 fee schedule is shown in Table 5.2. As can be seen, the aim is to price discriminate on the basis of income. A particularly interesting feature of this scheme is that the Association makes no attempt to check the veracity of the income declared by a prospective member. What they appear to rely upon is that economists will be either honest or even boastful in reporting their income. In addition, the Association must also hope to avoid the arbitrage problem whereby junior faculty members who pay a low subscription fee resell to senior faculty members who pay a high one. Casual observation suggests such reselling is rare.

The practice of the American Economic Association is not unique. Many academic journals charge a different price to institutions such as university libraries than to individuals. The print and online 2012 subscription rate to the *Journal of Economics and Management Strategy*, for example, is \$59 for an individual but \$446 for an institution.

Airlines are particularly adept at applying third-degree price discrimination. It has sometimes been suggested that the number of different fares charged to Economy class passengers on a particular flight is approximately equal to the number of passengers! A common feature of this type of price discrimination is that it is implemented by restrictions on the characteristics of the ticket. These include constraints upon the time in advance by which the flight must be booked, whether flights can be changed, the number of days between departure and return, whether the trip involves staying over a Saturday night, and so on. We return to the airline case later in this chapter.

Other examples of third-degree price discrimination are restaurant “early bird specials” and supermarket discounts to shoppers who clip coupons. Similarly, department stores that lower their apparel prices at the end of the season are attempting to charge a different price based on the observable characteristic of the time of purchase.³ Segmenting consumers by time of purchase is also evident in other markets. Consumers typically pay more to see a film at a first-run theater when the film is newly released than to see it at a later date at a second-run cinema or, still later, as a downloaded film at home.

³ Discounting over time in a systematic fashion runs the risk that if consumers know prices will fall in the future, they will delay their purchases. If the number of customers that postpone is “too” large, seasonal discounts will not be a very good strategy.

An essential feature of all third-degree price discrimination schemes is that the monopolist has some easily observed characteristic that serves as a good proxy for differences in consumer willingness to pay. This characteristic can be used effectively to divide the market into two or more groups, each of which will be charged a different price. The monopolist must next be able to ensure that resale of the product by those who are offered a low price to those who are offered a high one is not feasible. Consider the airline example again. The requirement to stay over a Saturday night is designed to discriminate between those consumers who are traveling on business and those who are not. Senior discounts typically require proof of age but this in itself is not necessarily sufficient to prevent arbitrage. Suppose, for example, that a local supermarket were to offer a senior discount. Forty- and fifty-somethings who have a senior parent living nearby would then have an incentive to have their parent purchase the groceries. This is why most senior discounts require both proof of age *and* proof of consumption (e.g., theater or movie tickets).

Once the different consumer groups have been identified and separated, the general rule that characterizes third-degree price discrimination is easily stated. *Consumers for whom the elasticity of demand is low should be charged a higher price than consumers for whom the elasticity of demand is relatively high.*

5.3 IMPLEMENTING THIRD-DEGREE PRICE DISCRIMINATION OR GROUP PRICING

The logic underlying the pricing rule is fairly straightforward. Here, we illustrate it with a simple example and defer formal presentation to the Appendix at the end of this chapter.

Suppose that the publishers of J.K Rowling's final volume in the Harry Potter series, *Harry Potter and the Deathly Hallows*, estimate that inverse demand for this book in the United States is $P_U = 36 - 4Q_U$ and in Europe is $P_E = 24 - 4Q_E$. In each case, prices are measured in dollars and quantities in millions of books sold at publication of the first edition of the book. Marginal cost is assumed to be the same in each market and equal to \$4 per book. The publisher also incurs other costs such as cover design and promotion, but we treat these as fixed and independent of sales volume and so ignore them in our analysis.

As a benchmark, suppose that the publisher treats the two markets as a single, integrated market and so quotes the same, nondiscriminatory price to consumers in the United States and Europe. To work out the profit maximizing price, the publisher first needs to calculate aggregate market demand at any price P . This means that they need to add the two market demand curves *horizontally*. In the United States, we have $P = 36 - 4Q_U$, which can be inverted to give $Q_U = 9 - P/4$ provided, of course, that $P \leq \$36$. In Europe we have $P = 24 - 4Q_E$ so that $Q_E = 6 - P/4$ provided $P \leq \$24$. This gives us the aggregate demand equation:

$$\begin{aligned} Q &= Q_U + Q_E = 9 - P/4 && \text{for } \$36 \geq P \geq \$24 \\ Q &= Q_U + Q_E = 15 - P/2 && \text{for } P < \$24 \end{aligned} \tag{5.1}$$

We can write this in the more normal inverse form as:

$$\begin{aligned} P &= 36 - 4Q && \text{for } \$36 \geq P \geq \$24 \\ P &= 30 - 2Q && \text{for } P < \$24 \end{aligned} \tag{5.2}$$

This demand relationship is illustrated in Figure 5.1. The kink in the aggregate demand function at a price of \$24 and a quantity of 3 million occurs because at any price above \$24 books will be sold only in the United States whereas once the price drops below \$24 both markets are active. The marginal revenue function associated with this demand function satisfies the usual “twice as steep” rule:

$$\begin{aligned} MR &= 36 - 8Q && \text{for } Q < 3 \\ MR &= 30 - 4Q && \text{for } Q > 3 \end{aligned} \quad (5.3)$$

This is also illustrated in Figure 5.1. The jump in the marginal revenue function at a quantity of 3 million arises because when price falls from just above \$24 to just below \$24, the inactive European market becomes active. That is, when the price falls to just below \$24, it brings in a new set of consumers.

We can use equations (5.1)–(5.3) to calculate the profit maximizing price, aggregate quantity, and quantity in each market. Equating marginal revenue with marginal cost, assuming that both markets are active, we have $30 - 4Q = 4$ so that $Q^* = 6.5$ million. From the aggregate demand curve (5.2) this gives a price of $P^* = \$17$. It follows from the individual market demands (5.1) that 4.75 million books will be sold in the United States and 1.75 million books in Europe. Aggregate profit (ignoring all the fixed and other set-up costs) is $(17 - 4)*6.5 = \$84.5$ million.

That this pricing strategy is not the best that the monopolist can adopt is actually clear from Figure 5.1. At the equilibrium we have just calculated, the marginal revenue on the last book sold in Europe is greater than marginal cost whereas marginal revenue on the last book sold in the United States is less than marginal cost (it is actually negative in this example!). Transferring some of the books sold in the United States to the European market will, therefore, lead to an increase in profit.

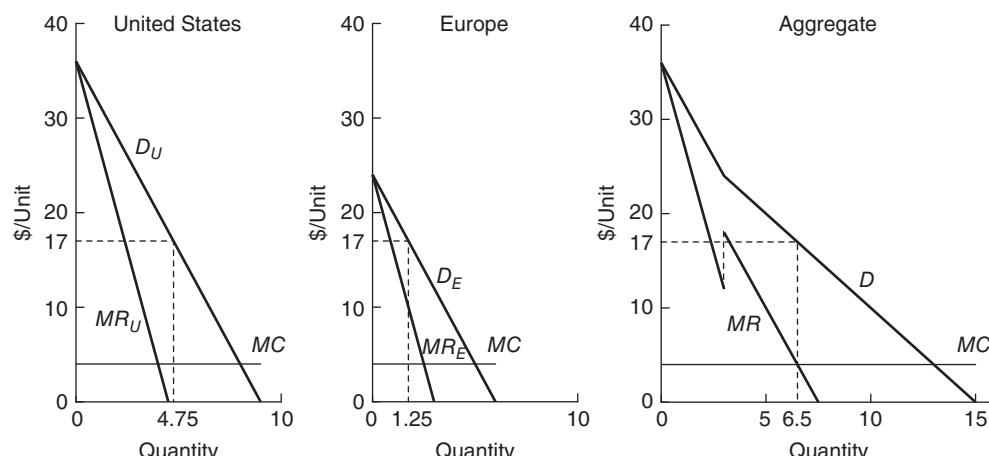


Figure 5.1 Non-discriminatory pricing—constant marginal cost

The firm identifies aggregate demand and the associated marginal revenue. It chooses total output where marginal revenue equals marginal cost and the non-discriminatory price from the aggregate demand function. Output in each market is the market clearing output.

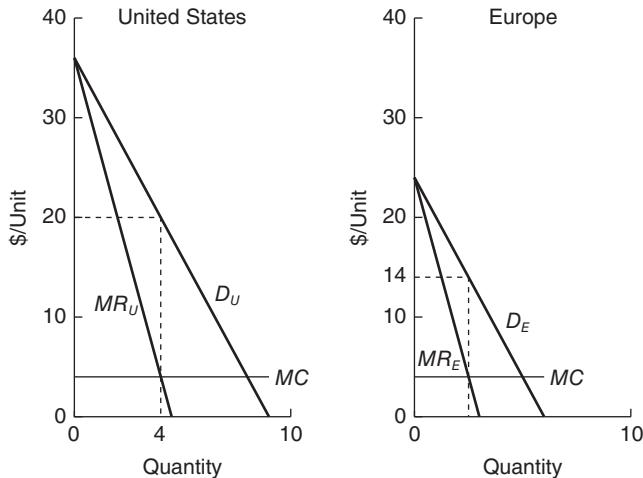


Figure 5.2 Third-degree price discrimination or group pricing—constant marginal cost
The firm sets output where marginal revenue equals marginal cost in each market and sets the market clearing price in each market.

We can make this more explicit. A necessary condition for profit maximization under third-degree price discrimination is that marginal revenue must equal marginal cost in *each* market that the monopolist serves. If this were not the case in a particular market, then the last unit sold in that market would be generating either more or less in cost than it is earning in revenue. Cutting back or increasing total production in that market would therefore raise profits. If marginal cost in serving each market is identical, as in our case, then this condition implies that marginal revenue be the same on the last unit sold in each market. If this condition does not hold, the monopolist can raise revenue and profit with no increase in production (and hence, no increase in costs), simply by shifting sales from the low marginal revenue market to the high one.

The application of these principles to our example is illustrated in Figure 5.2. Recall that demand in the United States market is $P_U = 36 - 4Q_U$, and in Europe it is $P_E = 24 - 4Q_E$. This means that marginal revenue in the United States is $MR_U = 36 - 8Q_U$ and in Europe is $MR_E = 24 - 8Q_E$. Now apply the rule that marginal revenue equals marginal cost in each market. This gives a profit-maximizing output in the United States of $Q_U^* = 4$ million books at a price of $P_U^* = \$20$, and in Europe a profit-maximizing output of $Q_E^* = 2.5$ million books at a price of $P_E^* = \$14$. Profit from sales in the United States is \$64 million and in Europe is \$25 million, giving aggregate profit (again ignoring all the fixed and other set-up costs) of \$89 million, an increase of \$4.5 million over the nondiscriminatory profit.

How does this outcome relate to the elasticity rule that we presented above? An important property of linear demand curves is that the elasticity of demand falls smoothly from infinity to zero as we move down the demand curve. This means that, for any price less than \$24 (and greater than zero) the elasticity of demand in the United States market is lower than in the European market. (You can check this by evaluating the demand elasticity in the two markets at any price for which both markets are active.) Our rule then states that we should find a higher price in the United States than in Europe, precisely as in our example.

How is our analysis affected if marginal cost is not constant? The same basic principles apply with one important change. If marginal production costs are not constant, we cannot

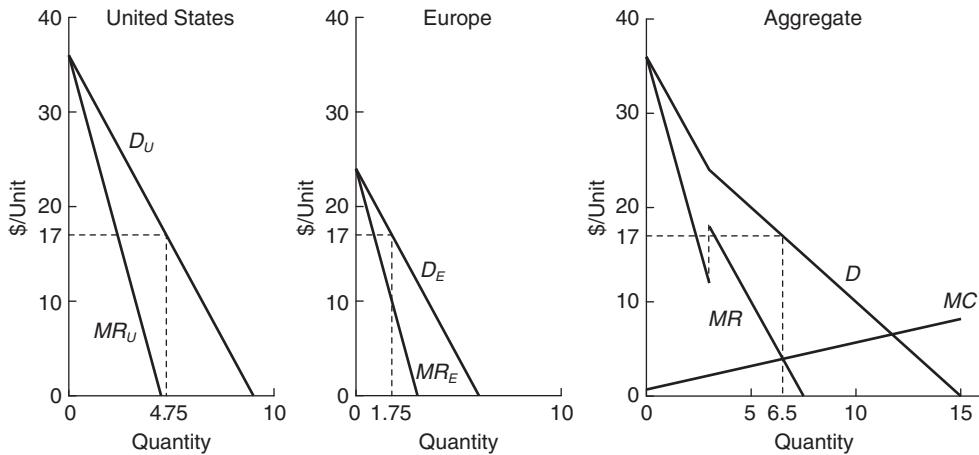


Figure 5.3 Non-discriminatory pricing with non-constant marginal cost

The firm identifies aggregate demand and the associated marginal revenue. It chooses total output where marginal revenue equals marginal cost and the non-discriminatory price from the aggregate demand function. Output in each market is the market clearing output

treat the two markets independently. Whatever output the monopolist chooses to supply to the United States, for example, affects the marginal cost of supplying Europe. As a result we need to look at the different markets together. Nevertheless, we still have simple rules that guide the monopolist's pricing decisions in these markets.

To illustrate this point, suppose that the publisher of *Harry Potter and the Deathly Hallows* has a single printing facility that produces books for both the United States and European markets and that marginal cost is given by $MC = 0.75 + Q/2$, where Q is the total number of books printed.

Figure 5.3 illustrates the profit-maximizing behavior if the monopolist chooses not to price discriminate. The basic analytical steps in this process are as follows:

1. Calculate aggregate market demand as above.
2. Identify the marginal revenue function for this aggregate demand function. From our example, if $Q > 3$ so that both markets are active, this is $MR = 30 - 4Q$.
3. Equate marginal revenue with marginal cost to determine aggregate output. Thus we have $0.75 + Q/2 = 30 - 4Q$ giving $Q^* = 6.5$ million books.
4. Identify the equilibrium price from the aggregate demand function. Because both markets are active, the relevant part of the aggregate demand function is $P = 30 - 2Q$, giving an equilibrium price of $P^* = \$17$.
5. Calculate demand in each market at this price: 4.75 million books in the United States and 1.75 million books in Europe.

Now suppose that the monopolist chooses to price discriminate. This outcome is illustrated in Figure 5.4. The underlying process is clearly different, and the steps in implementing profit maximizing price discrimination are as follows:

1. Derive marginal revenue in each market and add these *horizontally* to give aggregate marginal revenue. Marginal revenue in the United States is $MR = 36 - 8Q_U$ for any

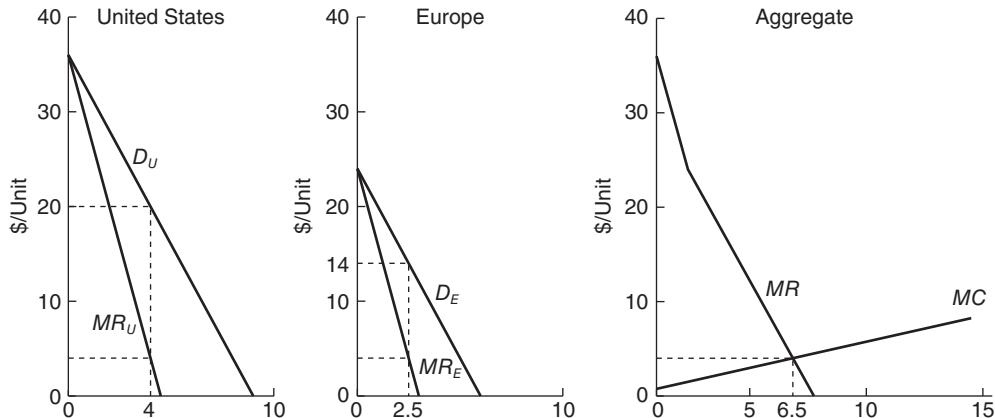


Figure 5.4 Third-degree price discrimination or group pricing with non-constant marginal cost
The firm calculates aggregate marginal revenue and equates this with marginal cost. Output in each market equates marginal revenue with aggregate marginal cost. Price in each market is the market-clearing price.

marginal revenue under \$36. In Europe, $MR = 24 - 8Q_E$ for marginal revenue below \$24. Inverting these gives $Q_U = 4.5 - MR/8$ and $Q_E = 3 - MR/8$. Summing these gives an aggregate marginal revenue:

$$\begin{aligned} Q &= Q_U + Q_E = 4.5 - MR/8 \quad \text{for } Q \leq 1.5 \\ Q &= Q_U + Q_E = 7.5 - MR/4 \quad \text{for } Q > 1.5 \end{aligned} \quad (5.4)$$

This can be inverted to give aggregate marginal revenue its more usual form

$$\begin{aligned} MR &= 36 - 8Q \quad \text{for } Q \leq 1.5 \\ MR &= 30 - 4Q \quad \text{for } Q > 1.5 \end{aligned} \quad (5.5)$$

Note how this step differs from the non-discriminatory case. In the latter both markets are treated as one, so we start with aggregate demand and derive its associated marginal revenue. In the discriminatory pricing case, by contrast, the markets are supplied separately, with the profit maximizing condition that $MC = MR$ in both markets so we need aggregate marginal revenue, not aggregate demand.

2. Equate aggregate marginal revenue with marginal cost to identify the equilibrium aggregate quantity *and* marginal revenue. So we have $30 - 4Q = 0.75 + 2Q$ giving $Q^* = 6.5$. As a result, the equilibrium marginal revenue is \$4, which is equal to the marginal cost of the last unit produced.
3. Identify the equilibrium quantities in each market by equating individual market marginal revenue with the equilibrium marginal revenue and marginal cost. In the United States this gives $36 - 8Q_U = 4$ or $Q_U^* = 4$ million books and in Europe $24 - 8Q_E = 4$ or $Q_E^* = 2.5$ million books.
4. Identify the equilibrium price in each market from the individual market demand functions, giving a price of \$20 in the United States and \$14 in Europe.

The foregoing procedure is again derived from two simple rules that guide the monopolist's pricing decisions with third-degree price discrimination. These rules apply no matter the shape of the monopolist's marginal cost function. The rules are:

1. Marginal revenue must be equal for the last unit sold in each market.
2. Marginal revenue must equal marginal cost, where marginal cost is measured at the *aggregate* output level.

There is one further interesting point that is worth noting regarding the contrast between uniform pricing (no price discrimination) and third-degree price discrimination. When demand is linear and both markets are active under both pricing schemes, *aggregate demand is identical with the two pricing policies*. This is proved formally in the chapter Appendix. The intuition is simple to see. When both markets are active, aggregate marginal revenue is identical with the two pricing policies (we are below the discontinuity in MR in Figure 5.3). So equating aggregate marginal revenue with aggregate marginal cost must give the same aggregate output. The reason that third-degree price discrimination is more profitable in this case is that the aggregate output is allocated more profitably across the two markets—to ensure that marginal revenue on the last unit sold in each market is equal.

We complete our discussion of third-degree price discrimination in this section by making explicit the relationship between the price set and the elasticity of demand in any specific market segment. Our review of monopoly and market power in Chapters 2 and 3 explained how we could express the firm's marginal revenue in any market in terms of price and the point elasticity of demand at that price. Specifically, marginal revenue in market i is given by $MR_i = P_i \left(1 - \frac{1}{\eta_i}\right)$ where η_i is (the absolute value of) the elasticity of demand. The larger is η_i the more elastic is demand in this market. Now recall that third-degree price discrimination requires that the profit-maximizing aggregate output must be allocated such that marginal revenue is equalized across each market (and, of course, equal to marginal cost). For example, if there are two markets this says that $MR_1 = MR_2$. Substituting from the equations above, we then know that $MR_1 = P_1 \left(1 - \frac{1}{\eta_1}\right) = MR_2 = P_2 \left(1 - \frac{1}{\eta_2}\right)$. We can solve this for the ratio of the two prices to give:

$$\frac{P_1}{P_2} = \frac{\left(1 - \frac{1}{\eta_2}\right)}{\left(1 - \frac{1}{\eta_1}\right)} = \frac{\eta_1 \eta_2 - \eta_1}{\eta_1 \eta_2 - \eta_2}. \quad (5.6)$$

From this, it is clear that price will indeed be lower in the market with the higher elasticity of demand. The intuition is that prices must be lower in those markets in which consumers are sensitive to price. Such price sensitivity means that raising the price will lose too many customers, and this loss more than offsets any gain in surplus per customer. To put it differently, when consumers are price sensitive, the strategy of lowering price can actually raise the monopolist's total surplus because it brings in many additional purchases. We encourage you to reinterpret the various examples with which we motivated our analysis in terms of demand elasticities. For example, is it reasonable to think that business travelers will have a lower elasticity of demand for air travel at a particular time than vacation travelers?

5.1

Practice Problem

The manager of a local movie theater believes that demand for a film depends on when the movie is shown. Early moviegoers who go to films before 5 pm are more sensitive to price than are evening moviegoers. With some market research, the manager discovers that the demand curves for daytime (D) and evening (E) moviegoers are $Q_D = 100 - 10P_D$ and $Q_E = 140 - 10P_E$, respectively. The marginal cost of showing a movie is constant and equal to \$3 per customer no matter when the movie is shown. This includes the costs of ticketing and cleaning.

- a. What is the profit maximizing pricing policy if the manager charges the same price for daytime and evening attendance? What is attendance in each showing and what is aggregate profit per day?
- b. Now suppose that the manager adopts a third-degree price discrimination scheme, setting a different day and evening price. What are the profit maximizing prices? What is attendance at each session? Confirm that aggregate attendance is as in (a). What is aggregate profit per day?

5.4 PRODUCT VARIETY AND THIRD-DEGREE PRICE DISCRIMINATION OR GROUP PRICING

We have thus far defined price discrimination as occurring whenever a firm sells an identical product to two or more buyers at different prices. But what if the products are not identical? Ford, for example, offers several hundred (perhaps even several thousand) varieties of the Ford *Taurus* with slightly different features. Procter and Gamble offers a wide range of toothpastes in different tastes, colors, and claimed medicinal qualities. Kellogg's offers dozens of breakfast cereals that vary in terms of grain, taste, consistency, and color.

Many examples of what looks like third-degree price discrimination or group pricing arise when the seller offers such *differentiated* products. For example, books are first released as expensive hardcover editions and only later as cheap paperbacks. Hotels in a ski area are more expensive in winter than in summer. First class air travel costs more than coach. The common theme of these examples is that they all involve variations of a basic product. This is a phenomenon that we meet every day in buying restaurant meals, refrigerators, electronic goods, and many other goods and services. In each of these situations, what we observe is a firm selling different varieties of the same good—distinguished by color, material, or design. As a brief reflection on the typical restaurant menu will reveal, what we also usually observe is that the different varieties are aimed at different groups and sold at different prices.

In considering these as applications of price discrimination we have to be careful. After all, the cost incurred in producing goods of different types, such as hardback and paperback books, or first class versus coach flights, is different. Philips (1983) provides perhaps the best definition of third-degree price discrimination or group pricing once we allow for product differentiation: “Price discrimination should be defined as implying that two varieties of a commodity are sold (by the same seller) to two buyers at different *net* prices, the net price being the price (paid by the buyer) corrected for the cost associated with the product differentiation.”

Using this definition, it would not be discriminatory to charge \$750 extra for a car with antilock brakes if it costs \$750 extra to assemble a car with such brakes. By contrast, the difference in price between a coach class fare of \$450 and a first class fare of \$8,000 for service between Boston and London must be seen as almost entirely reflecting price discrimination because the additional cost of providing first class service per passenger is well below the \$7,550 difference in price. In other words, price discrimination among different versions of the same good exists only if the difference in prices is not justified by differences in underlying costs: this is what Philips means by the *net price*.

Consideration of product variety leads to a very important question. Does offering different varieties of a product enhance the monopolist's ability to charge different net prices? That is, does a firm with market power increase its ability to price discriminate by offering different versions of its product? As we shall see, the general answer is yes.

We can obtain at least some insight into this issue by recalling the two problems that successful discrimination must overcome, namely, identification and arbitrage. In order to price discriminate, the firm must determine who is who on its demand curve and then be able to prevent resale between separate consumers. By offering different versions or models of its product, the monopolist may be able to solve these two problems. Different consumers may buy different versions of a good and therefore reveal who they are through their purchase decisions. Moreover, because different customers are purchasing different varieties, the problem of resale is considerably reduced.

As an example of the potential for product differentiation to enhance profit, consider an airline that we will call Delta Airlines (DA), operating direct passenger flights between Boston and Amsterdam. DA knows that there are three types of customers for these flights: those who prefer to travel first class, those who wish to travel business class, and those who are reconciled to having to travel coach. One part of the arbitrage problem is, of course, easily solved: in order to sit in a first class seat you need a first class ticket. However, there is another aspect to this problem. If the difference in price is great enough relative to the value a consumer places on a higher class of travel, a business class traveler, for example, might choose to fly coach. For simplicity, we assume that this arbitrage, or self-selection problem does not arise. That is, we assume that first class passengers prefer not to travel rather than sit in business or coach, and business class passengers similarly will not consider coach travel—they place sufficiently high values on the differences in quality between the types of seat that they will not trade down. (See end-of-chapter problem 5 for an example of this case).⁴

DA's market research indicates that daily demand for first class travel on this route is $P_F = 18,500 - 1,000Q_F$, for business class travel is $P_B = 9,200 - 250Q_B$, and for Coach travel is $P_C = 1500 - 5Q_C$. The marginal cost is estimated to be \$100 for a coach passenger, \$200 for a business class passenger, and \$500 for a first class passenger.

The profit maximizing third-degree price discrimination scheme for differentiated products of this type satisfies essentially the same rules as for homogeneous products. Simply put, DA should identify the quantity that equates marginal revenue with marginal cost for each class of seat and then identify the equilibrium price from the relevant demand function. For first class passengers this requires $MR_F = 18,500 - 2,000Q_F = 500$, or $Q_F^* = 9$. The resulting first class fare is $P_F^* = \$9,500$. In business class, we have

⁴ There is also the possibility that coach or business travelers would want to trade up. The equilibrium prices that we derive in the example preclude such a possibility.

Reality Checkpoint

Variations on a Theme—Broadway Ticket Prices

In New York, over 20,000 people attend Broadway shows each night. As avid theatergoers know, prices for these tickets have been rising inexorably. The top price for Broadway shows has risen 31 percent between 1998 and 2004 and has more than doubled since then. However, discounts offered through coupons, two-for-one deals, special student prices, and then TKTS booth in Times Square, significantly reduce this price impact.

Why so much discounting? The value of a seat in a theater, like a seat on an airplane, is highly perishable. Once the show starts or the plane takes off, a seat is worth next to nothing. So, it's better to fill the seat at a low price than not fill it at all.

Stanford economist Phillip Leslie investigated Broadway ticket price discrimination using detailed data for a 1996 Broadway play, *Seven Guitars*. Over 140,000 people saw this play, and they bought tickets in seventeen price categories. While some of the difference was due to seat quality—opera versus mezzanine versus balcony, a large amount of price differentials remained even after quality adjustments. The average difference of two

tickets chosen at random on a given night was about 40 percent of the average price. This is comparable to the price variation in airline tickets.

Leslie used advanced econometric techniques to estimate the values that different income groups put on the various categories of tickets. He found that Broadway producers do a pretty good job, in general, at maximizing revenue. He found the average price set for *Seven Guitars* was about \$55 while, according to Mr. Leslie's estimates, the value that would maximize profit was a very close \$60. His data also indicated that the optimal uniform price would be a little over \$50. Again, price discrimination is less about the average price charged and more about varying the price in line with the consumer's willingness to pay. In this connection, Leslie found that optimal price discrimination drew in over 6 percent more patrons than would optimal uniform pricing.

Source: P. Leslie. "Price Discrimination in Broadway Theatre," *Rand Journal of Economics* 35 (Autumn, 2004): 520–41.

$MR_B = 9,200 - 500Q_B = 200$, or $Q_B^* = 18$ and $P_B^* = \$4,700$. Finally, in coach we have $1500 - 10Q_C = 100$, giving $Q_C^* = 140$ and $P_C^* = \$800$.

The example we have just presented resolved the arbitrage problem by assuming that different types of travelers are committed to particular classes of travel. Of course, this may not always be the case. For example, the downturn in economic activity through 2011 has encouraged many businesses to seek ways to cut costs. In particular, business travelers increasingly are required by their companies to fly coach. It remains the case that these types of travelers are willing to pay more (though not as much more as before) for air travel than casual or vacation travelers. Now, however, the airline's ability to exploit the difference in willingness to pay faces a potentially severe arbitrage problem.

To see this more clearly, let's simplify the problem and suppose that the airline has just two types of customers, business people and vacationers. Business people are known to have a high reservation price, or willingness to pay, for a return ticket, which we will denote as V^B . Vacationers, by contrast, have a low reservation price, denoted as V^V . By assumption, $V^B > V^V$, and the airline would obviously like to exploit this difference by charging business customers a high price and vacationers a low one. However, the airline

cannot simply impose this distinction. A policy of explicitly charging business customers more than vacationers would quickly lead to every customer claiming to be on holiday and not on business. To be sure, the airline could try to identify which passengers really are on holiday, but this would be costly and likely to alienate customers.

If this were the end of our story, it would appear that the airline has no choice but to sell its tickets at a single, uniform price. It would then face the usual textbook monopoly dilemma. A high price will earn a large surplus from every customer that buys a ticket but clearly leads to a smaller, mostly business set of passengers. By contrast, a low price will encourage many more people to fly but, unfortunately, leave the company with little surplus from any one consumer.

Suppose, however, that business and holiday travelers differ in another respect as well as in their motives for flying. To be specific, suppose that business travelers want to complete their trip and return home within three days, whereas vacationers want to be away for at least one week. Suppose also that the airline learns (through surveys and other market research) that business travelers are willing to pay a premium beyond a normal ticket price if they can be guaranteed a return flight within their preferred three-day span. In this case, product differentiation by means of offering two differentiated tickets—one with a minimum time away of a week and another with no minimum stay—will enable the airline to extract considerable surplus from each type of consumer.

The complete strategy would be as follows: First, set a low price of V^V for tickets requiring a minimum of one week before returning. Because holiday travelers do not mind staying away seven days, and because the ticket price does not exceed their reservation price, they will willingly purchase this ticket. Because such travelers are paying their reservation price, the airline has extracted their entire consumer surplus and converted it into profit for itself.⁵

Second, the airline should set a price as close to V^B as possible for flights with no minimum stay. The limit on its ability to do this will be such factors as the cost of paying for a hotel for extra nights, the price of alternative transportation capable of returning individuals in three days, and related considerations. Denote the dollar value of these other factors as M . Business people wanting to return quickly will gladly pay a premium over the one week price V^V up to the value of M , so long as their total fare is less than V^B . (The precise condition is $V^V + M < V^B$.) Using such a scheme enables the airline to extract considerable surplus from business customers, while simultaneously extracting the entire surplus from vacationers.

In short, even if the airline cannot squeeze out the entire consumer surplus from the market, it can nevertheless improve its profits greatly by offering two kinds of tickets. This is undoubtedly the reason that the practice just described is so common among airlines and other transportation companies. (See Reality Checkpoint.) Such companies offer different varieties of their product as a means of having their customers self-select into different groups. Automobile and appliance manufacturers us a similar strategy—offering different product lines meant to appeal to consumers of different incomes or otherwise different willingness to pay. Stiglitz (1977) labels such mechanisms as *screening devices* because

⁵ An alternative and frequently used distinction is to require that the traveler stay over a Saturday night in order to qualify for a cheap fare. Presumably a corporation will not want to finance the lodgings of its employees when they are not on company business. Further, business travelers will typically want to spend weekends with family and loved ones. On both counts, the Saturday night requirement works as a self-separating device.

Reality Checkpoint

You Can't Go Before You Come Back

It is not uncommon to find that a coach fare to fly out on Tuesday and return quickly on Thursday costs well over twice the coach fare to fly out on Tuesday and return a week later. So, for travelers wanting to return in two days, an obvious strategy is to buy two round trip tickets—one, say, that departs on Thursday the 10th and returns on Thursday the 17th, and another that departs on Tuesday the 15th and returns on Tuesday the 22nd. The passenger can use the outgoing half of the second ticket on Tuesday the 15th and then fly back on the return flight of the first ticket that flies on the 17th. Unfortunately for such savvy travelers—and for the students and other needy consumers who could use the unused portions of each flight—airlines are alert to such practices. In particular, when a passenger checks in for a flight, the airline checks to see if the passenger has an unused portion of a return flight. If so, the fee is automatically adjusted to the higher fare. The airlines have a great incentive to make sure that those who are willing to pay a substantial premium to return in two days really do pay it.

Los Angeles resident Peter Szabo found this out the hard way. He had planned a trip that would take him first to Boston, then to New York, and then to Philadelphia. He paid \$435 dollars for a US Airways flight to Boston with a return flight from Philadelphia, with an intermediate bus trip from Boston to Philadelphia that stopped in Manhattan. A week or so before his departure, however, a business firm in New York that was interested in making a deal with Szabo asked to meet with him and offered to fly him to New York for free. So, Szabo amended his itinerary to go to New York first, then Boston, and then return from Philadelphia using the second half of the original ticket. He soon discovered that this was not possible. Because he had not used the first half, he could not use the second half. If he wanted to fly from Philadelphia now, he would have to purchase a new, one-way ticket *plus* a \$150 ticket-change fee.

Sources: "Why It Doesn't Pay to Change Planes or Plans," *London Daily Telegraph* 11 March, 2000, p. 27; D. Lazarus. "Using Just Half of a Round Trip Ticket Can Be Costly," *Los Angeles Times* 11 January, p. B3.

they screen or separate customers precisely along the relevant dimension of willingness to pay.

A rather curious kind of screening is illustrated by Wolfram Research, manufacturers of the *Mathematica*® software package. In making its student version of the software, Wolfram disables a number of functions that are available in the full academic or commercial versions. In 2011, Wolfram offered the full version of *Mathematica*® at around \$2,495, the academic version at around \$1,095, and the student version at around \$140. There is little doubt that this is a case involving substantial differences in net prices.

The motivation behind this screening by means of product differentiation seems equally clear. Wolfram realizes that some customers do not need—or at least do not want to pay very much for—the full version of their software. Wolfram markets the low-priced version of *Mathematica*® for these consumers, and then sells the extended version to customers with a high willingness to pay for the improved product. Note that the two products must really differ in some important respect (to consumers at least). If Wolfram did not reduce

the capabilities of the student version, it would have to worry about arbitrage between the two customer groups, with students buying for their professors!

The Wolfram example just described is a type of screening referred to by marketing experts as “crimping the product.” Deneckere and McAfee (1996) argue that crimping, or deliberately damaging a product to enhance the ability to price discriminate, has been a frequent practice of manufacturers throughout history. Among the examples that they cite are (1) IBM’s *Laser Printer E*, an intentionally slower version of the company’s higher-priced top-of-the-line laser printer and (2) simple cooking wine, which is ordinary table wine with so much salt added that it is undrinkable. Some people have even argued that the US Post Office deliberately reduces the quality of its standard, first class service so as to raise demand for its two-day priority and overnight mail services.

Each of these examples is a clear case of a difference in net prices. The lower-quality product sells for a lower price, yet—because it starts as a high-quality product and then requires the further cost of crimping—the lower-quality product is actually more expensive to make. Why do firms crimp a high-quality product to produce a low-quality one instead of simply producing a low-quality one in the first place? The most obvious answer relates to costs of production. Given that a firm with monopoly power such as Wolfram knows that there are consumers of different types willing to buy different varieties of its product, the firm must decide how these consumer types can be supplied with products “close” to those that they most want at least cost. It may well be cheaper to produce the student version of *Mathematica*® by crimping the full version rather than to set up a separate production line dedicated to manufacturing different versions of the software package.

The final type of product differentiation that we consider in this chapter is differentiation by *location of sale*.⁶ In many cases, a product for sale in one location is not the same as the otherwise identical product for sale in another location. A prescription drug such as *Lipitor*® for sale in Wisconsin is not identical to the same prescription drug for sale in New York state. Even with the advent of sophisticated Internet search engines, a new automobile for sale in one state is not identical to the same new automobile for sale in another state.

To illustrate why this type of product differentiation can lead to price discrimination, suppose that there is a company, Boston Sea Foods (BSF), which sells a proprietary brand of clam chowder. BSF knows that demand for its chowder in Boston is $P_B = A - BQ_B$ and in Manhattan is $P_M = A - BQ_M$, where quantities are measured in thousands of pints. In other words, the firm believes that these two markets have identical demands. BSF has constant marginal costs of c per thousand pints of chowder. Transport costs to reach the Boston market are negligible but it costs BSF an amount t to transport a thousand pints of chowder to Manhattan.

How does BSF maximize its profits from these two markets, given that BSF employs linear pricing? BSF should apply the rules that we have already developed. It should equate marginal revenue with marginal cost in each market. In the Boston market this requires that $A - 2BQ_B = c$, so that $Q_B^* = (A - c)/2B$ and the Boston price is $P_B^* = (A + c)/2$. In the Manhattan market we have, by contrast, $A - 2BQ_M = c + t$, so that $Q_M^* = (A - c - t)/2B$ and the Manhattan price is $P_M^* = (A + c + t)/2$.

Why is this outcome an example of third-degree price discrimination? Recall our definition of price discrimination with differentiated products. For there to be *no* such discrimination, any difference in price should be equal to the difference in the costs of

⁶ We return to spatial differentiation in more detail in Chapter 7.

product differentiation. In our BSF example, it costs BSF t per thousand pints to send chowder from Boston to Manhattan but the difference in price in the two markets is only $t/2$. In other words, BSF is price discriminating by absorbing 50 percent of the transport costs of sending its chowder to Manhattan.

What about the arbitrage problem in the BSF example? Manhattanites might want to buy their chowder directly in the Boston market, but it is economic for them to do so only if they have access to a transport technology that is at least 50 percent cheaper than that employed by BSF, a very tall order other than for those who choose to vacation in Boston.

Returning to our prescription drug example in Table 5.1, one possible explanation for the difference in prices in the three United States regions might be differences in costs of supplying these three regions. Another, of course, would be differences in demands in the three regions arising from differences in these regions' demographics or incomes.

5.2

Practice Problem

NonLegal Seafoods (NS) sells its excellent clam chowder in Boston, New York, and Washington. NS has estimated that the demands in these three markets are respectively $Q_B = 10,000 - 1,000P_B$, $Q_{NY} = 20,000 - 2,000P_{NY}$ and $Q_W = 15,000 - 1,500P_W$, where quantities are pints of clam chowder per day. The marginal cost of making a pint of clam chowder in their Boston facility is \$1. In addition, it costs \$1 per pint to ship the chowder to New York and \$2 per pint to ship to Washington.

- What are the profit maximizing prices that NS should set in these three markets? How much chowder is sold per day in each market?
- What profit does NS make in each market?

5.5 THIRD-DEGREE PRICE DISCRIMINATION OR GROUP PRICING AND SOCIAL WELFARE

The term “price discrimination” suggests inequity and, from a social perspective, sounds like a “bad thing.” Is it? To answer this question, we must recall the economist’s approach to social welfare and the problem raised by the standard monopoly model. Economists view arrangements as less than socially optimal whenever there are potential trades that could make both parties better off. This is the reason that a standard monopoly is sub-optimal. The textbook monopolist practicing uniform pricing restricts output. At the margin, consumers value the product *more* than it costs the monopolist to produce it. A potentially mutually beneficial trade exists but under uniform pricing such a trade will not occur.

The question that we consider in this section is whether third-degree price discrimination worsens or reduces this monopoly distortion. The intuitive reason that third-degree discrimination may reduce efficiency relative to the uniform pricing case is essentially that such a policy amounts to uniform pricing within two or more separate markets. It thus runs the risks of compounding the output-reducing effects of monopoly power.

We can be more specific regarding the welfare effects of third-degree price discrimination by drawing on the work of Schmalensee (1981). This is illustrated for the case of two

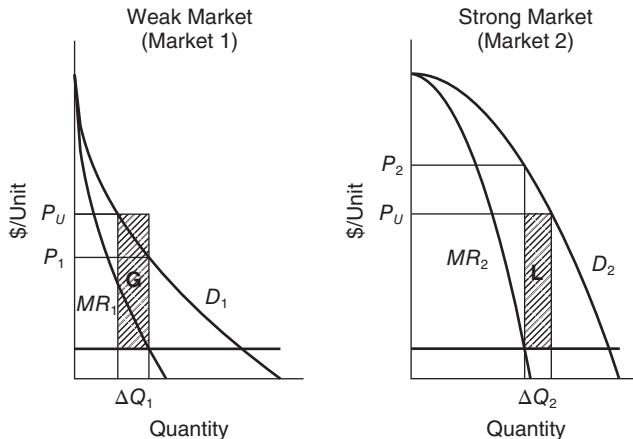


Figure 5.5 Welfare effects of third-degree price discrimination or group pricing

The upper limit on the welfare gain is area G and the lower limit on welfare loss is area L . The upper limit of the net welfare impact is $G - L$ and is positive only if aggregate output is greater with discriminatory pricing than with nondiscriminatory pricing.

markets in Figure 5.5. In this figure, P_1 and P_2 are the profit maximizing discriminatory prices—obtained by equating marginal revenue with marginal cost in each market—while P_U is the optimal nondiscriminatory price. Market 2 is referred to as the strong market because the discriminatory price is higher than the uniform price while market 1 is the weak market. ΔQ_1 and ΔQ_2 are respectively the difference between the discriminatory output and the nondiscriminatory output in the weak and the strong market. It follows, of course, that $\Delta Q_1 > 0$ and $\Delta Q_2 < 0$.

Our normal definition of welfare is the sum of consumer plus producer surplus. Using this definition, an upper limit on the increase in surplus that follows from third-degree price discrimination in Figure 5.5 is the area G minus the area L . This gives us the following equation. (In writing equation (5.7) we have used the property that $\Delta Q_2 < 0$.)

$$\Delta W \leq G - L = (P_U - MC)\Delta Q_1 + (P_U - MC)\Delta Q_2 = (P_U - MC)(\Delta Q_1 + \Delta Q_2) \quad (5.7)$$

Extending this analysis to n markets, we have

$$\Delta W \leq (P_U - MC) \sum_{i=1}^n \Delta Q_i \quad (5.8)$$

It follows from equation (5.8) that for $\Delta W \geq 0$ it is necessary that $\sum_{i=1}^n \Delta Q_i \geq 0$. In other words, a necessary condition for third-degree price discrimination to increase welfare is that it increases total output.

We know from the Harry Potter example (and from the more general presentation in the Appendix) that when demands in the various markets are linear, total output is identical with discriminatory and non-discriminatory pricing. It follows that with linear demands

third-degree price discrimination must reduce total welfare. The increase in profit is more than offset by the reduction in consumer surplus. Schmalensee states:

“If one thinks that demand curves are about as likely to be concave as convex ... (this) ... might lead one to the conclusion that monopolistic third-degree price discrimination should be outlawed.” (Schmalensee 1981, 246) See also Shih, Mai and Lui (1988).

However, before jumping to the suggested conclusion, we need to note an important caveat. Our analysis implicitly assumes that the same markets are served with and without price discrimination. This may very well not be the case. In particular, one property of price discrimination that we have not yet considered is that it can make it profitable to serve markets that will not be served with non-discriminatory prices. If this is the case, then the additional welfare from the new markets that third-degree price discrimination introduces more than offsets any loss of welfare in the markets that were previously being served.

A simple example serves to make this point. Suppose that monthly demand for a patented AIDS drug treatment in North America is $P_N = 100 - Q_N$ and in Sub-Saharan Africa is $P_S = \alpha 100 - Q_S$ with $\alpha < 1$ reflecting the assumption that African consumers have a lower demand because their income is so much smaller. Further assume that the marginal cost of producing each month's treatment is constant at $c = 20$ per unit and that transport costs to the African market are negligible.

Now assume that the patent holder is constrained from price discriminating across the two markets. As before, we start by inverting the demand functions to give $Q_N = 100 - P$ and $Q_S = \alpha 100 - P$. If the price is low enough to attract buyers in both markets then aggregate demand is: $Q = (1 + \alpha)100 - 2P$ or $P = (1 + \alpha)50 - Q/2$, and marginal revenue is $MR = (1 + \alpha)50 - Q$. Equating marginal revenue with marginal cost $c = 20$, gives the equilibrium output, $Q = (1 + \alpha)50 - 20 = 30 + \alpha 50$, and price $P = 35 + 25\alpha$.

Now recall our assumption that both markets are active without price discrimination. For this assumption to hold it must be that the equilibrium price when there is no discrimination is less than the maximum price— $\alpha 100$ —that Sub-Saharan African consumers are willing and able to pay. That is, for our assumption to hold it must be the case that $35 + 25\alpha < \alpha 100$. In turn, this implies that for both markets to be active with no price discrimination it is necessary that $\alpha > 35/75$ or $\alpha > 0.466$. In other words, for the Sub-Saharan African market to be served it is necessary that the maximum willingness to pay for AIDS drugs in that market be at least 47 percent of the maximum willingness to pay in North America.

Moreover, even if $\alpha > 0.466$, the Sub-Saharan African market may not be served. From the patent-holding firm's perspective, it is not quite enough that the maximum willingness to pay exceeds the price charged if it serves both markets. This is because the monopolist always has the option of choosing a higher price and serving only the North American market. In the end-of-chapter problem 6, you are asked to show that $\alpha > 0.531$ for it to be profitable for the firm to serve both markets when price discrimination is for some reason prohibited.

Why might allowing the firm to price discriminate in this example lead to an increase in total welfare? Whether or not the monopolist is allowed to price discriminate, the firm will set the monopoly price in the United States so there is no change on total surplus in the United States. What about the Sub-Saharan market? For the monopolist to be willing to supply this market all that is needed now is that the reservation price 100α be greater than marginal cost of \$20, or $\alpha > 0.2$. If this condition is satisfied, total surplus is increased by

third-degree price discrimination because a market is opened up that would otherwise not be served.

5.3

Return to practice problem 5.1 and confirm that total welfare is greater with nondiscriminatory pricing than with third-degree price discrimination.

Practice Problem

5.6 EMPIRICAL APPLICATION: PRICE DISCRIMINATION IN THE NEW CAR MARKET

As we have seen, a firm with market power has a very strong interest in discriminating in the price it charges to different consumers—getting the most from those with the greatest willingness to pay and concluding a sale to those with lower valuations of the good at a lower price. As we have seen, an explicit pricing rule such as time-of-day pricing can sometimes be used to this effect. However, in a number of cases, the actual price paid is not the result of an explicit pricing schedule but, instead, reflects the outcome of a bargaining process between seller and buyer. New car purchases are a good example.

Automobiles come with a Manufacturer's Suggested Retail Price (MSRP) but the actual price paid by any consumer is typically the outcome of a bargaining process between the buyer and the car dealer. As a result, different consumers will end up paying different prices. In part, this will depend on factors that affect the dealer's willingness to reduce price, such as an end-of-the-month sales bonus or the number of rival dealerships nearby. Controlling for these factors, however, variations in the price paid should reflect variations in consumer demand elasticity. Consumers with incomes well above the average are likely to pay more while consumers who are well-informed about car prices at other dealers, i.e., those with low search costs, will be more price sensitive and negotiate a lower price. There may also be overt discrimination based on other factors such as gender and race.

Scott Morton, Zettelmeyer, and Silva-Risso (2003) (hereafter SZS) examine actual new car prices for over 671,000 transactions at over 3,500 dealers during a 14 month period beginning on January 1, 1999 and ending on February 28, 2000. Working with a market research firm, the authors were able to collect data for each of these transactions regarding the price paid, the make, model, and trim level of the car, the financing and any trade-in values that accompanied the deal, the dealership that made the sale, the name, address, and age of the buyer and, very importantly, whether the buyer used the online car-referral service, Autobytel.

SZS can match the name and address of each buyer to a specific census block group, which is a precise neighborhood of 1,100 people that comprises about one-fourth of the geographic tracts compiled by the US Census Bureau. For each such census block, the authors have data on the percentage of the population that is black, Hispanic, and Asian (and hence, the percentage that is in none of these racial groups). They also have data on other socioeconomic variables for the block such as median household income, the percentage that own their homes, the percentage that completed college, and the occupational breakdown between professionals, business executives, blue collar workers, and technicians.

SZS then identify 834 specific “cars” depending on the make, model, and trim involved. For each such car, they measure the net price paid by the buyer as the transaction price less

any manufacturer's rebate and also less any excess value given for a trade-in as measured by the difference between the value assigned the trade-in vehicle and that car's book value. Because they cannot control for all option features, they include a variable that should indicate additional special features called DVehCost. This is measured as the difference between the vehicle's recorded invoice (dealer cost) and the average cost for that type of car. As noted, Szs also control for the number of nearby rival dealers, the "hotness" of the car (how long since it was first introduced), end-of-month sales, and various other dealer characteristics. The first of the main results of the Szs research for our purposes are shown in Table 5.3. For ease of interpretation, all coefficient estimates have been multiplied by 100.

Because the dependent variable is the natural log of the net purchase price, the coefficients are naturally interpreted as percentage price effects. In this light, first note that the dealer control variables generally have the right sign and are statistically significant. For example, the inclusion of unobserved features that raise the vehicle cost above the dealer's average also raise the final price, as does the presence of a trade-in as dealers try to recover some of the trade-in value paid. Competition as measured by the number of rival dealers nearby significantly lowers the price.

Table 5.3 Determinants of new car prices actually paid

Variable	Estimated Coefficient	Standard Error
%Black	0.015	(0.00054)*
%Hispanic	0.011	(0.001)*
%Asian	-0.004	(0.00096)*
Female	0.210	(0.01)*
Age	0.005	(0.00063)*
Age > 64	-0.170	(0.03)*
Median Income	-0.00002	(1.39e-06)*
(Median Income) ²	1.3e-10	(7.58e-12)*
%College Grad	-0.0031	(0.00095)*
%<High School	0.0039	(0.0013)*
%HomeOwner	-0.0027	(0.00045)*
%Professional	0.0046	(0.0014)*
%Executives	-0.00013	(0.0015)
%Blue Collar	0.00018	(0.001)
%Technicians	0.0046	(0.0035)
Median Home Value	2.7e-06	(1.28e-07)*
End of Month	-0.3500	(0.015)*
Weekend	0.1100	(0.016)*
DVehCost	88	(0.13)*
Competition	-0.02	(0.0035)*
Any Trade	0.31	(0.01)*
Constant	1,001	(0.13)*
R ²	0.97	

*Significant at 1% level.