Chapter 9

Monopoly

As you will recall from intermediate micro, monopoly is the situation where there is a single seller of a good. Because of this, it has the power to set both the price and quantity of the good that will be sold. We begin our study of monopoly by considering the price that the monopolist should charge.[1]

9.1 Simple Monopoly Pricing

The object of the firm is to maximize profit. However, the price that the monopolist charges affects the quantity it sells. The relationship between the quantity sold and the price charged is governed by the (aggregate) demand curve $q(p)$. Note, in order to focus on the relationship between q and p, we suppress the wealth arguments in the aggregate demand function.

We can thus state the monopolist's problem as follows:

$$\max_p pq(p) - c(q(p)).$$

Note, however, that there is a one-to-one correspondence between the price charged and the quantity the monopolist sells. Thus we can rewrite the problem in terms of quantity sold instead of the price charged. Let $p(q)$ be the inverse demand function. That is, $p(q(p)) = p$. The firm's profit maximization problem can then be written as

$$\max_q p(q)q - c(q).$$

It turns out that it is usually easier to look at the problem in terms of setting quantity and letting price be determined by the market. For this reason, we will use the quantity-setting approach.

---

[1]References: Tirole, Chapter 1; MWG, Chapter 12; Bulow, "Durable-Goods Monopolists," *JPE* 90(2) 314-332.
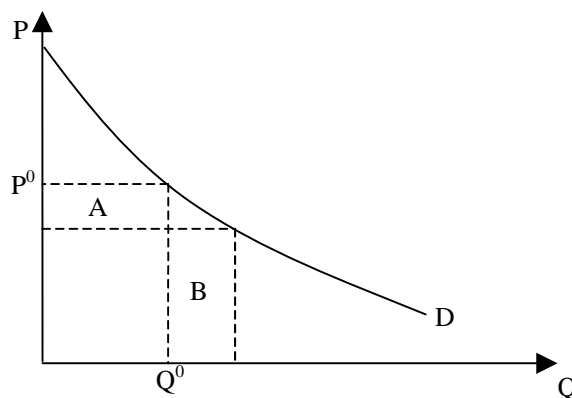
Figure 9.1: The Monopolist's Marginal Revenue

In order for the solution to be unique, we need the objective function to be strictly concave (i.e. $\frac{d^2\pi}{dq^2} < 0$). The second derivative of profit with respect to q is given by

$$\frac{d^2}{dq^2}\left(p\left(q\right)q - c\left(q\right)\right) = p''\left(q\right)q + 2p'\left(q\right) - c''\left(q\right).$$

If cost is strictly convex, $c''\left(q\right) > 0$, and since demand slopes downward, $p'\left(q\right) < 0$. Hence the second and third terms are negative. Because of this, we don't need inverse demand to be concave. However, it can't be "too convex." Generally speaking, we'll just assume that the objective function is concave without making additional assumptions on $p\left(\right)$. Actually, to make sure the maximizing quantity is finite, we need to assume that eventually costs get large enough relative to demand. This will always be satisfied if, for example, the demand and marginal cost curves cross.

The objective function is maximized by looking at the first derivative. At the optimal quantity, q^*,

$$p'\left(q^*\right)q^* + p\left(q^*\right) = c'\left(q^*\right)$$

On the left-hand side of the expression is the marginal revenue of increasing output a little bit. This has two parts - the additional revenue due to selling one more unit, $p\left(q^*\right)$ (area B in Figure 9.1), and the decrease in revenue due to the fact that the firm receives a lower price on all units it sells (area A in Figure 9.1). Hence the monopolist's optimal quantity is where marginal revenue is equal to marginal cost, and price is defined by the demand curve $p(q^*)$.[2] See Figure 9.2 for the graphical depiction of the optimum.

---

[2]This is also true for the competitive firm. However, since competitive firms are price takers, their marginal revenue is equal to price.
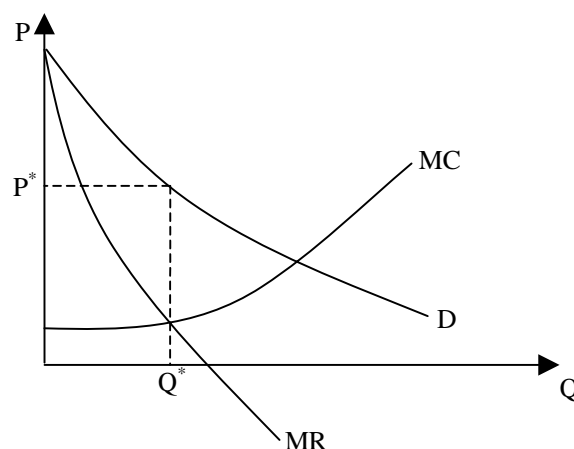
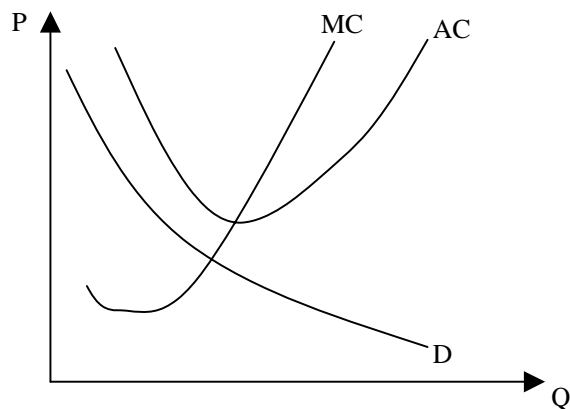Figure 9.2: Monopolist's Optimal Price and Quantity

Figure 9.3: The Monopolist Cannot Make a Profit

If the monopolist's profit is maximized at $q = 0$, it must be that $p(0) \leq c'(0)$. This corresponds to the case where the cost of producing even the first unit is more than consumers are willing to pay. We will generally assume that $p(0) > c'(0)$ to focus on the interesting case where the monopolist wants to produce a positive output. However, even if we assume that $p(0) > c'(0)$, the monopolist may not want to choose a positive output. It may be that shutting down is still preferable to producing a positive output. That is, the monopolist may have fixed costs that are so large that it would rather exit the industry. Such a situation is illustrated in Figure 9.3. Thus we interpret the condition that $p(0) > c'(0)$ (along with the appropriate second-order conditions) as saying that if the monopolist does not exit the industry, it will produce a positive output.

If there is to be a maximum at a positive level of output, it must be that the first derivative equals zero, or:

$$p'(q^*)q^* + p(q^*) = c'(q^*)$$

Note that one way to rewrite the left side is:

$$p'(q^*)q^* + p(q^*) = p(q^*)\left(\frac{dp}{dq}\frac{q^*}{p^*} + 1\right) = p(q^*)\left(1 - \frac{1}{|\varepsilon_p^*|}\right)$$

where ε_p^* is the price elasticity of demand evaluated at (q^*, p^*). Now we can rewrite the monopolist's first-order condition as:

$$\frac{p(q^*) - c'(q^*)}{p(q^*)} = p'(q^*)\frac{-q^*}{p(q^*)} = \frac{1}{|\varepsilon_p^*|}.$$

The left-most quantity in this expression, $(p - mc)/p$ is the "markup" of price over marginal cost, expressed as a fraction of the price. This quantity, called the Lerner index, is frequently used to measure the degree of market power in an industry.

Note that at the quantity where $MR = MC$, $p > MC$ (since $p'(q)q$ is negative). Thus the monopolist charges more than marginal cost.[3] The social optimum would be for the monopolist to sell output as long as consumers are willing to pay more for the last unit produced than it costs to produce it. That is, produce up until the point where $p = MC$. But, the monopolist cuts back on production because it cares about profit, not social optimality. And, it is willing to reduce q in order to increase the amount it makes per unit. This results in what is known as the **deadweight loss of monopoly,** and it is equal to the area between the demand curve and the marginal cost curve, and to the right of the optimal quantity, as in Figure 9.4. This area represents social surplus that could be generated but is not in the monopoly outcome.

9.2   Non-Simple Pricing

The fact that the monopolist sells less than the societally optimal amount of the output arises from the requirement that the monopolist must sell all goods at the same price. Thus if it wants to earn higher profits on any particular item, it must raise the price on all items, which lowers the quantity sold. If the monopolist could raise the price on some items but not others, it could earn higher profits and still sell the efficient quantity. We now consider two examples of more complicated pricing mechanisms along these lines, non-linear pricing and two-part tariffs.

---

[3]Also, from these two equations we can show that the monopolist will always choose a quantity such that price is elastic, i.e. $|\varepsilon_p^*| > 1$.
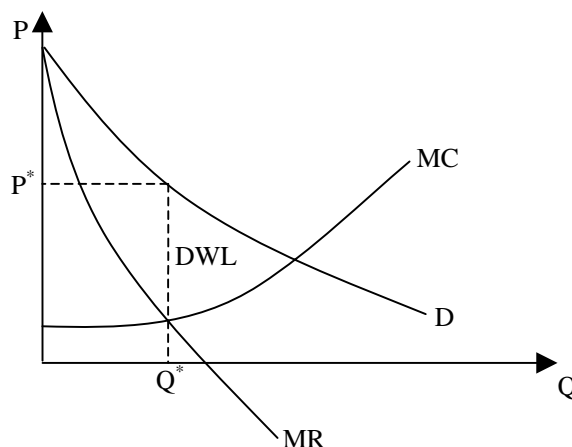
Figure 9.4: Deadweight Loss of Monopoly

9.2.1   Non-Linear Pricing

Consider the case where the monopolist charges a price scheme where each unit is sold for a different price.

For the moment, we assume that all consumers are identical and that consumers are not able to resell items once they buy them. In this case, the monopolist solves its profit-maximization problem by designing a scheme that maximizes the profit earned on any one consumer and then applying this scheme to all consumers.

In the context of the quasilinear model, we know that the height of a consumer's demand function at a particular quantity represents his marginal utility for that unit of output. In other words, the height of the demand curve represents the maximum a consumer will pay for that unit of output.

With this in mind, if the monopolist is going to charge a different price for each unit of output, how should it set that price?   Obviously, it wants to set the price of each unit equal to the consumer's maximum willingness to pay for that unit, i.e. equal to the height of the demand curve at that q.   If the monopolist employs a declining price scheme, where $p(q) = D^{-1}(q)$, up to the point where demand and marginal cost cross, the monopolist can actually extract all of the social surplus. Note: if we are worried that output can only be sold in whole units, then the price of unit q should be given by:

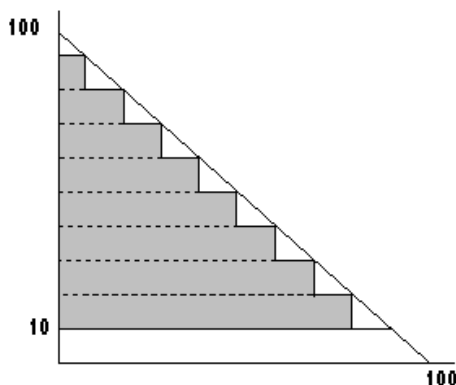$$p_q = \int_{q-1}^{q} D^{-1}(q)\, dq.$$

Figure 9.5: Non-Linear Pricing

To illustrate non-linear pricing, consider a consumer who has demand curve $P = 100 - Q$, and suppose the monopolist's marginal cost is equal to 10. At a price of 90, the consumer demands 10 units of output. While the consumer would not purchase any more output at a price of 90, it would purchase more output if the price were lower. So, suppose the monopolist sells the first 10 units of output at a price of 90, and the second 10 units at a price of 80. Similarly, suppose the monopolist sells units 21-30 at a price of 70, 31-40 at a price of 60, 41-50 at a price of 50, etc., all the way up to units 71-80, which are sold at a price of 20. This yields Figure 9.5.

The monopolist's producer surplus is equal to the shaded region. Thus by decreasing the price as the number of units purchased increases, the monopolist can appropriate much of the consumer surplus. In fact, as the number of "steps" in the pricing scheme increases, the producer surplus approaches the entire triangle bounded by the demand curve and marginal cost. Thus if the monopolist were able to charge the consumer $\int_{q-1}^{q} D^{-1}(q)\, dq$ for the block of output consisting of unit q, it would appropriate the entire consumer surplus.

In practice, declining block pricing is found most often in utility pricing, where a large buyer may be charged a high price for the first X units of output, a lower price for the next X units, and so on.

Recall that we said that a monopolist charging a single price can make a positive profit only if the demand curve is above the firm's average total cost (ATC) curve at some point. If the firm can engage in non-linear pricing, this is no longer the case. Since the firm's revenue is the entire shaded region in the above diagram, it may be possible for the firm to earn a positive profit even if the demand curve is entirely below the ATC curve. For example, if ATC in the previous example
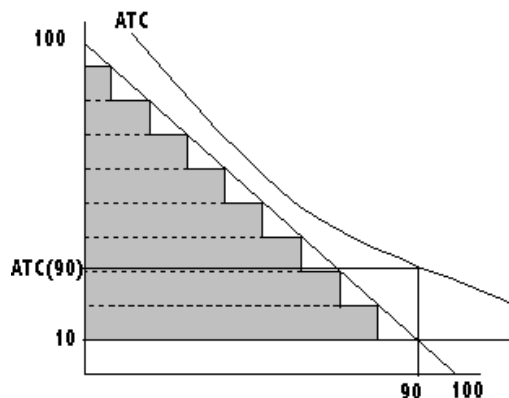
Figure 9.6: Profit Even With Demand Below ATC

is as in Figure 9.6, then the monopolist can make a profit whenever the area of the shaded region is larger that $TC(90) = ATC(90) \times 90$.

9.2.2   Two-Part Tariffs

Another form of non-simple pricing that is frequently employed is the two-part tariff. A two-part tariff consists of a fixed fee and a price per unit consumed. For example, an amusement park may charge an admission fee and a price for each ride, or a country club may charge a membership fee and a fee for each round of golf the member plays. The question we want to address is how, in this context, the fixed fee (which we'll call F) and the use fee (which we'll call p) should be set.

As in the non-linear pricing example, we continue to assume that all consumers are identical and that the monopolist produces output at constant marginal cost. For simplicity, assume that $MC = 0$, but note that a positive marginal cost is easily incorporated into the model. Further, we continue to think of the consumer as having quasilinear utility for output q and a numeraire good. In this case, we know that the consumer's inverse demand curve, $p(q)$, gives the consumer's marginal utility from consuming unit q, and that the consumer's net benefit from consuming q units is given by the consumer surplus at this quantity, $CS(q)$.

The firm's profit from two-part tariff (F, p) is given by $F + p * q(p)$. We break our analysis into two steps. First, for any p, how should F be set? Second, which p should be chosen? So, fix a p. How should F be set. Since the monopolist's profit is increasing in F, the monopolist wants to set F as large as possible while still inducing the consumer to participate. That is, the consumer's net benefit must be at least zero.
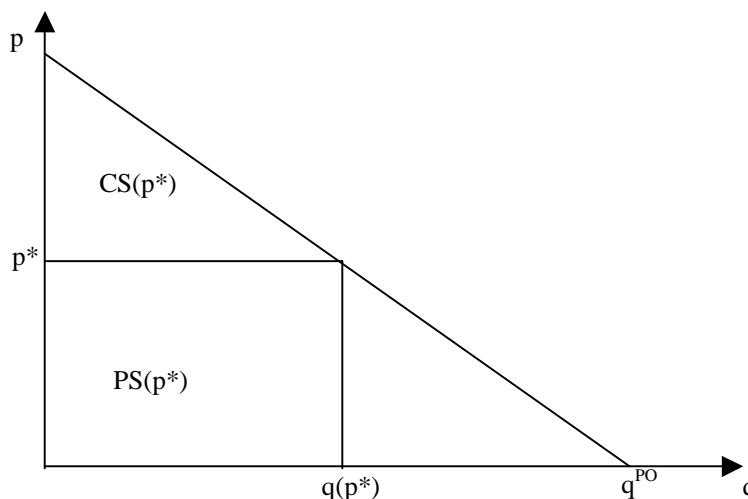
Figure 9.7: The Two-Part Tariff

At price p, the consumer chooses to consume $q(p)$ units of output, and earns surplus $CS(q(p))$. Net surplus is therefore $CS(q(p)) - F$, and we want this to be non-negative: $CS(q(p)) - F \geq 0$. As argued above, the firm wants to set F as large as possible. Therefore, for any p, setting $F(p) = CS(q(p))$ maximizes profit. This makes sense, since the firm wants to set the membership fee in order to extract all of the consumer's surplus from using its product.

Next, how should p be chosen? Given $F(p)$ as defined above, the monopolist's profit from the two-part tariff is given by: $CS(q(p)) + p*q(p)$. Note that this is the sum of consumer and producer surplus (see Figure 9.7). As is easily seen from the diagram, this expression is maximized by setting the use fee equal to zero (marginal cost) and the fixed fee equal to $CS(q(0))$. Thus the optimal two-part tariff involves $p = 0$ and $F = CS(0)$.[4]

We can understand the optimal two-part tariff in another way. When the monopolist chooses F optimally, it claims the entire surplus created from the market. Thus the monopolist has an incentive to maximize the surplus created in this market. By the first welfare theorem, we know that total surplus is maximized by the perfectly competitive outcome. That is, when p is such that $p^* = MC$. Thus, in order to maximize total surplus, the firm should set $p = MC$. If it does so, "producer surplus" is zero. However, the firm sets $F = CS(q(p^*))$, and extracts the entire social surplus in the form of the fixed fee.

---

[4] More generally, if the firm has positive marginal cost, then p should be set such that $p = MC$ at $q(p)$, and F should be set equal to $CS(q(p))$.

Interestingly, there seems to be a trend toward this sort of two-part pricing in recent years. For example, at Walt Disney World, the standard admission package charges for entry into one of the several theme parks but charges nothing for rides once inside the park. This makes sense, since the marginal cost of an additional ride is very low.[5]

9.3   Price Discrimination

Previously we considered the case where the monopolist faces one or many identical consumers, and we investigated various pricing schemes the monopolist might pursue. However, typically, the monopolist faces another problem, which is that some consumers will have a high willingness to pay for the product and others will have a low willingness to pay. For example, business travellers will be willing to pay more for airline tickets than leisure travellers, and working-age adults may be more willing to pay for a movie ticket than senior citizens. We refer to the monopolist's attempts to charge different prices to these different groups of people as **price discrimination.** Frequently, three types of price discrimination are identified, although the distinctions are, at least to some extent, arbitrary. They are called first-degree, second-degree, and third-degree price discrimination.[6]

9.3.1   First-Degree Price Discrimination

First-degree price discrimination - also called "perfect price discrimination" - refers to the situation where the monopolist is able to sell each unit of output to the person that values it the most for their maximum willingness to pay. For example, think of 100 consumers, each of whom have demand for one unit of a good. Let $p(q)$ be the qth consumer's willingness to pay for the good (note, this is just the inverse demand curve). A perfectly discriminating monopolist is able to charge each consumer $p(q)$ for the good: Each consumer is charged her maximum willingness-to-pay. In this case, the monopolist's marginal revenue is equal to $p(q)$, and so the monopolist sets quantity q so that $p(q) = c'(q)$, which is exactly the condition for Pareto optimality. Figure 9.8 depicts the price-discrimination optimum, which is also the Pareto optimal quantity.

We can also think of first-degree price discrimination where each individual has a downward

---

[5]This contrasts with the scheme Disney used when I was younger, which involved both an admission fee and positive prices for each ride. In fact, they even charged higher prices for the most popular rides.

[6]There are many different definitions of the various kinds of price discrimination. The ones I use are based on Varian's *Intermediate Microeconomics.*
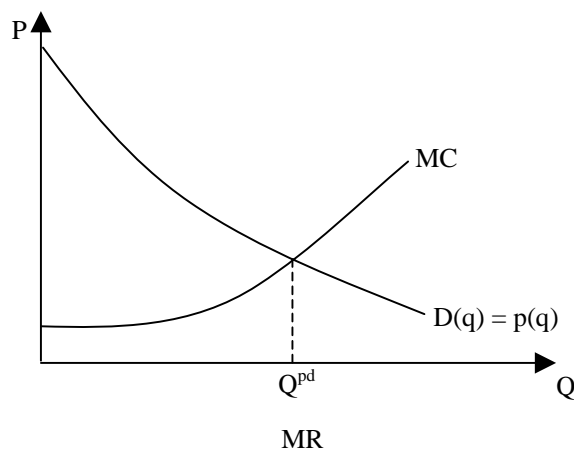
Figure 9.8: First-Degree Price Discrimination is Efficient

sloping demand for the good. In this case, the monopolist will sell a quantity such that $p^* = p(q) = c'(q)$. However, each individual will be sold quantity q_i^*, where $p^* = p_i(q_i^*)$, and $p_i(q_i)$ is individual i's inverse demand curve. The monopolist will charge the consumer the total consumer surplus associated with q_i^* units of output,

$$t_i^* = \int_0^{q_i^*} p_i(s)\, ds.$$

Thus the monopolist's selling scheme is given by a quantity and total charge (q_i, t_i) for each consumer. This scenario is depicted in Figure 9.9: In the left panel of the diagram, we see the aggregate demand curve, and the price-discriminating monopolist sets total quantity equal to where marginal cost equals aggregate demand (unlike the non-discriminating monopolist, who sets quantity where marginal cost equals marginal revenue). The panel on the right shows that the demand side of the market consists of two different types of consumers (D_1 and D_2), and the monopolist maximizes profits by dividing the quantity between the two buyers such that the marginal willingness-to-pay (WTP) is equal across the two buyers and equals marginal cost: $WTP_1 = WTP_2 = MC$. And the price is set for each consumer as specified above for t_i^*, to capture the full surplus.

In order for first-degree price discrimination to be possible, let alone successful, the monopolist must be able to identify the consumer's willingness to pay (or demand curve) and charge a different price to each consumer. There are two problems with this. First, it is difficult to identify a consumer's willingness to pay (they're not just going to tell you); and second, it is often impractical or illegal to tailor your pricing to each individual. Because of this, first-degree price discrimination
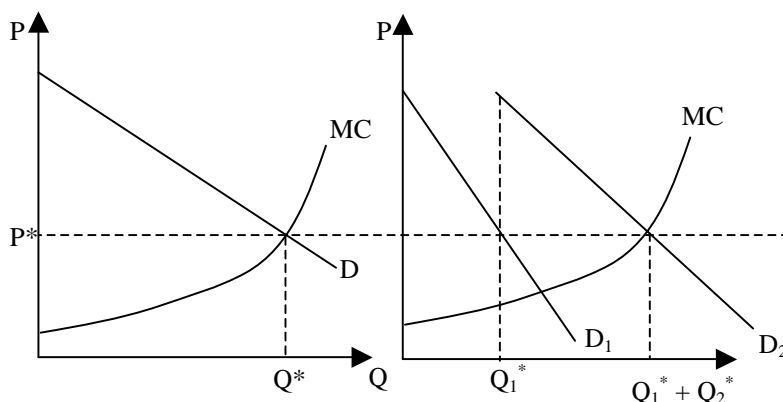
Figure 9.9: First-Degree Price Discrimination

is perhaps best thought of as an extreme example of the maximum (but rarely attainable) profit the monopolist can achieve.

9.3.2 Second-Degree Price Discrimination

Second-degree price discrimination refers to the case where the monopolist cannot perfectly identify the consumers. For example, consider a golf course. Some people are going to use a golf course every week and are willing to pay a lot for each use, while others are only going to use the course once or twice a season and place a relatively low value on a round of golf. The owner of the golf course would like to charge a high price to the high-valued users, and a low price to the low-valued users. What's the problem with this? On any given day, people will have an incentive to act like low-valued users and pay the lower price. So, in order to be able to charge the high-valued people a high price and the low-valued people a low price, the monopolist needs to design a pricing scheme such that the high-valued people do not want to pretend to be low-valued people. This practice goes by various names, including second-degree price discrimination, non-linear pricing, and screening.

Let's think about a simple example.[7] Suppose the monopolist has zero marginal cost, and the demand curves for the high- and low-valued consumers are given by $D_H(p)$ and $D_L(p)$. Assume for the moment that there are equal numbers of high- and low-valued buyers. This scenario is depicted in Figure 9.10.

---

[7]This example is based on the analysis in Varian's *Intermediate Microeconomics*.
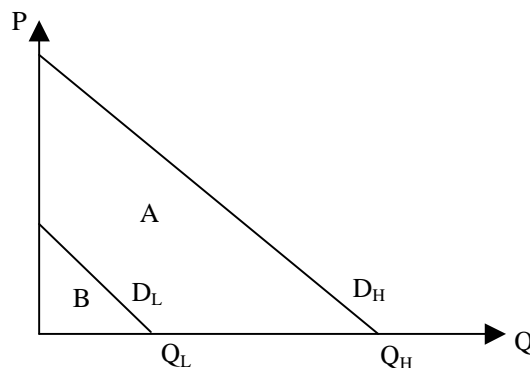
Figure 9.10: Second-Degree Price Discrimination

If the monopolist could perfectly price discriminate, it would charge the low type a total price equal to the area of region B and sell the low type Q_L units, and charge the high type total price $A + B$ for Q_H units. Notice that under perfect price discrimination, consumers earn zero surplus. The consumer surplus from consuming the object is exactly offset by the total cost of consuming the good.

Now consider what would happen if the firm cannot identify the low and high types and charge them different prices. One thing it could do is simply offer (Q_L, B) and $(Q_H, A + B)$ and let the consumers choose which offer they want. Clearly, all of the low-valued buyers would choose (Q_L, B) (why?). But, what about the high-valued buyers? If they choose $(Q_H, A + B)$, their net surplus is zero. But, if they choose (Q_L, B), they earn a positive net surplus equal to the area between D_H and D_L to the left of Q_L (again, why?). Thus, given the opportunity, all of the high-valued buyers will self-select and choose the bundle intended for the low-valued buyers.

The monopolist can get around this self-selection problem by changing the bundles. Specifically, if the bundles intended for the high-valued and low-valued consumers offered the same net surplus to the high-valued consumers, the high-valued consumers would choose the bundle intended for them.[8] If the monopolist lowered the price charged to the high types, they would earn more surplus from taking the bundle intended for them, and this may induce them to actually select that bundle. For example, consider Figure 9.11.

---

[8] Here we make the common assumption that if an agent is indifferent between two actions, he chooses the one the economist wants. This is pretty innocuous, since we could always change the offer slightly to make the agent strictly prefer on of the options. Technically, the reason for this assumption is to avoid an "open set" problem: The set of bundles that the high type strictly prefers to a particular bundle is open, and thus there may not be a profit-maximizing bundle for the monopolist, if we don't assume that indifference yields the desired outcome.
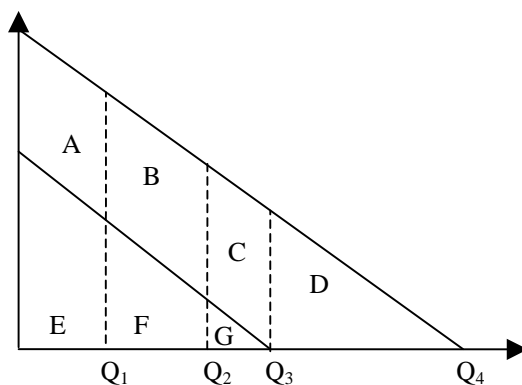
Figure 9.11: Monopolistic Screening

The optimal **first-degree** pricing scheme is to offer $\beta_1 = (Q_3, E + F + G)$ to the low type and $\beta_2 = (Q_4, A + B + C + D + E + F + G)$ to the high type. However, high prefers β_1, which offers him surplus $A + B + C$ to β_2. In order to get the high-valued type to accept a bundle offering Q_4, the monopolist can charge no more than $E + F + G + D$. Let's call this bundle $\beta_3 = (Q_4, E + F + G + D)$. If the monopolist has zero production costs, selling β_3 to the highs and β_1 to the lows will be better than selling β_1 to all consumers.

The key point in the previous paragraph is that the monopolist must design the bundles so that the buyers self-select the proper bundle. If you want the highs to accept β_3, it had better be that there is no other bundle that offers them higher surplus.

The monopolist can do even better than offering β_1 and β_3. Suppose the monopolist offers $\beta_4 = (Q_2, E + F)$ to low and $\beta_5 = (Q_4, C + D + E + F + G)$ to high. Since high earns surplus $A + B$ from either bundle, the self-selection constraint is satisfied. The monopolist earns profit $C + D + 2E + 2F + G$, as compared to $2E + 2F + 2G + D$ from offering β_1 and β_3. So, as long as area C is larger than area G, the monopolist earns more profit from offering β_4 and β_5 than from offering β_1 and β_3.

What is the optimal menu of bundles? In the previous paragraph, G is the revenue given up by making less from the lows, while C is the revenue gained by making more from the highs. The optimal bundle will be where these two effects just offset each other. In the diagram, they are shown by $\beta_6 = (Q_1, E)$ and $\beta_7 = (Q_4, B + C + D + E + F + G)$.

Figure 9.12 gives a clean version of the diagrams we have been looking at above. The optimal second-degree pricing scheme is illustrated. The monopolist should offer menu (Q_L^*, A) and
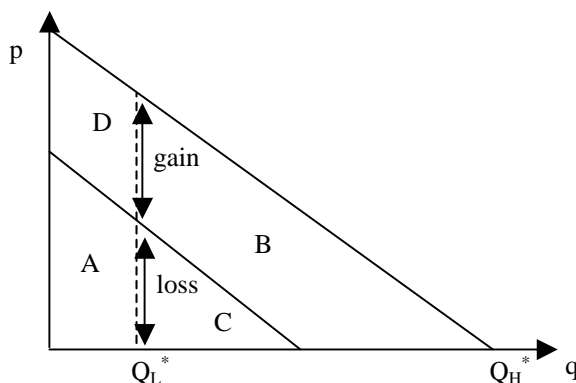
Figure 9.12: Optimal Second-Degree Pricing Scheme

$(Q_H^*, A + B + C)$. Under this scheme, the low-valued consumers choose (Q_L^*, A) and earn no surplus. Notice that Q_L^* is less than the Pareto optimal quantity for the low-valued consumers. The high-valued consumers are indifferent between the two bundles, and so choose the one we want them to choose: $(Q_H^*, A + B + C)$. Notice that under this scheme the high-valued consumers are offered the Pareto optimal quantity, Q_H^*, but earn a positive surplus.   Overall, as the quantity offered to the low-valued consumers decreases, the monopolist gains on the margin the area marked "gain" in the diagram and loses the area market "loss." The optimal level of Q_L^* is where these lengths are just equal.

Once we have derived the optimal second-degree pricing scheme, there is still one thing we have to check. The monopolist could always decide it is not worth it to separate the two types of consumers. Rather it could just offer Q_H^* at price $A + B + C + D$ in Figure 9.12 and sell only to the high-valued consumers. Hence, after deriving the optimal self-selection scheme, we must still make sure that the profit under this scheme, $2A + C + B$, is greater than the profit from selling only to the high types, $A + B + C + D$. Clearly, this depends on the relative size of A (the profit earned by selling to the low-types under the optimal self-selection mechanism) and D (the rent that must be given to the high type to get them to buy when offer (Q_L^*, A) is also on the menu).[9]

The previous graphical analysis is different in style than you are used to. I show it to you for a couple of reasons. First, it is a nice illustration of a type of problem that we will see over and over again, called the monopolistic screening (or hidden-information principal-agent) problem.[10]   One

---

[9] If there were different numbers of high- and low- type consumers, we would have to weight the size of the gains and losses by the size of their respective markets.

[10] Since you haven't seen these other problems yet, this paragraph may not make sense to you. That's okay. We'll

party (call her the principal) offers a menu of contracts to another party (call him the agent) about which she does not know some information. The contracts are designed in such a way that the agents self-select themselves, revealing that information to the principal. The menu of contracts is distorted away from what would be offered if the principal knew the agent's private information. Further, it is done in a specific way. The high type is offered the full-information quantity, but a lower price (phenomena like these are often referred to as "no distortion at the top" results), while the low type is offered less than the full-information quantity at a lower price. Distorting the quantity offered to the low type allows the principal to extract more profit from the high type, but in the end the high type earns a positive surplus. This is known as the informational rent – the extra payoff high gets due to the fact that the principal does not know his type.[11] As I said, this is a theme we will return to over and over again.

9.3.3   Third-Degree Price Discrimination

Third-degree price discrimination refers to a situation where the monopolist sells to different buyers at different prices based on some observable characteristic of the buyers. For example, senior citizens may be sold movie tickets at one price, while adults pay another price, and children pay a third price. Self-selection is not a problem here, since the characteristic defining the groups is observable and verifiable, at least in principle. For example, you can always check a driver's license to see if someone is eligible for the senior citizen price or not.

The basics of third-degree price discrimination are simple. In fact, you really don't need to know any economics to figure it out. Let $p_1(q_1)$ and $p_2(q_2)$ be the aggregate demand curves for the two groups. Let $c(q_1 + q_2)$ be the firm's (strictly convex) cost function, based on the total quantity produced. The monopolist's problem is:

$$\max_{q_1, q_2} p_1(q_1) q_1 + p_2(q_2) q_2 - c(q_1 + q_2).$$

come back to this point later.

[11] This is only one type of principal-agent problem. Often, when people refer to "the" principal-agent problem, they are referring to a situation where the agent makes an unobservable effort choice and the principal must design an incentive contract that induces him to choose the correct effort level. To be precise, such models should be referred to as "hidden-action principal-agent" models, to differentiate them from "hidden-information principal-agent" problems, such as the monopolistic screening or second-degree price discrimination problems. See MWG Chapter 14 for a discussion of the different types of principal-agent models.

The first-order conditions for an interior maximum are:

$$p_1'(q_1^*) q_1^* + p_1(q_1^*) = c'(q_1^* + q_2^*)$$
$$p_2'(q_2^*) q_2^* + p_2(q_2^*) = c'(q_1^* + q_2^*).$$

In general, you would need to check second-order conditions, but let's just assume they hold. The first-order conditions just say the monopolist should set marginal revenue in each market equal to total marginal cost. This makes sense. Consider the last unit of output. If the marginal revenue in market 1 is greater than the marginal revenue in market 2, you should sell it in market 1, and vice versa. Hence any optimal selling scheme must set MR equal in the two markets. And, we already know that MR should equal MC at the optimum.

So, third-degree price discrimination is basically simple. But, it can have interesting implications. For example, suppose there are two types of demand for Harvard football tickets. Alumni have demand $p_a(q_q) = 100 - q_a$, while students have demand $p_s(q_s) = 20 - 0.1q_s$. Suppose the marginal cost of an additional ticket is zero. How many tickets of each type should Harvard sell? Set MR = MC in each market:

$$20 - 0.2q_s = 0$$
$$100 - 2q_a = 0,$$

So, $q_a^* = 50$ and $q_s^* = 100$ as well. Alumni tickets are sold at $p_a = 50$, while student tickets are sold at price $p_s = 10$. Total profit is $50 * 50 + 100 * 10 = 2500 + 1000 = 3500$.

Now, suppose that the stadium capacity is 151 seats (for the sake of argument), and that all seats must be sold. Should the remaining seat be sold to alumni or students? Think of your answer before going on.

If the additional seat is sold to alumni, $q_a = 51$, $p_a = 49$, and total profit on alumni sales is 2499, yielding total profit 3499. On the other hand, if the additional seat is sold to students, $q_s = 101$, $p_s = 20 - 0.1(101) = 9.9$, total profit on student sales is 999.9, and so total profit is 3499.9. Hence it is better to sell the extra ticket to the students, even though the price paid by the alumni is higher. Why? The answer has to do with marginal revenue. We know that in either case selling another unit of output decreases total profit (why?). By selling an additional alumni ticket, you have to lower the price more than when selling another student ticket. Hence it is better to sell the ticket to a student, not because more is made on the additional ticket, but because less is lost due to lowering price on the tickets sold to all other buyers in that market.

9.4   Natural Monopoly and Ramsey Pricing

From the point of view of efficiency, monopolies are a bad thing because they impose a deadweight loss.[12]   The government takes a number of steps to prevent monopolies. For example, patent law grants a firm exclusive rights to an invention for a number of years in exchange for making the design of the item available, and allowing other firms to license the use of the technology after a period of time.   Other ways the government opposes monopolies is through anti-trust legislation, under which the government may break up a firm deemed to exercise too much monopoly power (such as AT&T) or prevent mergers between competitors that would create monopolies..

However, there are some monopolies that the government chooses not to break up. Instead the government allows the monopoly to operate (and even sanctions its operation) but regulates the prices it can charge. Why would the government do such a thing? The government allows monopolies to exist when they are so-called natural monopolies. A **natural monopoly** is an industry where there are high fixed costs and relatively small variable costs.   This implies that AC will be decreasing in output.   Because of this, it makes sense to have only one firm.[13]   The best examples of natural monopolies are utilities such as the electric, gas, water, and (formerly) telephone companies. Take the water company: In order to provide households with water, you need to purify and filter the water and pass it through a network of pipes leading from the filtration plant to the consumer's house. The filtration plants are expensive to build, and the network of pipes is expensive to install and maintain. Further, they are inconvenient, since laying and maintaining pipe can disrupt traffic, businesses, etc. Think of how inefficient it would be if there were four or five different companies all trying to run pipes into a person's house!

Because of the inefficiency involved in having multiple providers in an industry that is a natural monopoly, the government will allow a single firm to be the monopoly provider of that product, but regulate the price that it is allowed to charge. What price does the government choose? Consider the diagram of a natural monopolist in Figure 9.13.

---

[12] Of course, there are also critical distribution issues with monopolies, and these issues can persist even in situations in which the DWL of monopoly has been eliminated.   For example, with perfect price discrimination, there is no DWL - the efficient quantity is sold - but the distribution of surplus (all to the monopolist, zero to consumers) is clearly a cause for societal concern.

[13] The technical definition of a natural monopoly is that the industry cost function is subadditive.   That is, $c(q_1 + q_2) \leq c(q_1) + c(q_2)$.   Hence it is always cheaper to produce $q_1 + q_2$ units of output using a single firm than using two (or more) firms.   This is a slightly weaker definition of a natural monopoly than decreasing average cost, especially in multiple dimensions.
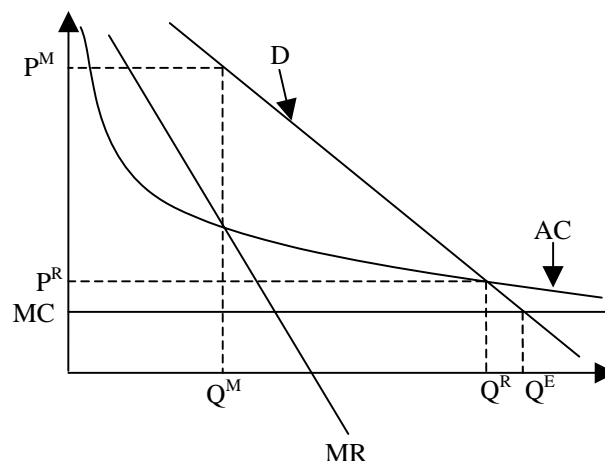
Figure 9.13: Ramsey Pricing

The monopolist has constant marginal cost, $c'(q) = MC$, and positive fixed cost. Note that this implies that $AC > MC$ for all q, the defining feature of a natural monopolist.

If the monopoly is not regulated, it will charge price p^M. The Pareto optimal quantity to sell is where inverse demand equals marginal cost, labeled Q^E in the diagram. The corresponding price is $P^E = MC$. However, since this price is below AC, the monopolist will not be able to cover its costs if the government forces it to charge P^E. In order for the monopolist to cover its production costs, it must be allowed to charge a higher price. However, as it increases price above MC, the quantity drops below Q^E, and there is a corresponding deadweight loss. Thus the government's task is to strike a balance between allowing the monopolist to cover its costs and keeping prices (and deadweight loss) low. The price that does this is the smallest price at which the monopolist is able to cover its cost. This price is labeled P^R in the diagram. The R stands for Ramsey, and the practice of finding the prices that balance deadweight loss and allow the monopolist to cover its costs is called Ramsey pricing. Finding the Ramsey price is easy in this example, but when the monopolist produces a large number of products (such as electricity at different times of the day and year, and electricity for different kinds of customers), the Ramsey pricing problem becomes much harder.[14] Ramsey pricing (and other related pricing practices) is one of the major topics in the economics of regulation (or at least it was for a long time).[15]

---

[14]Ramsey pricing is really a topic for monopolists that produce multiple outputs. The Ramsey prices are then the prices that maximize consumer surplus subject to the constraint that the firm break even overall.

[15]Another type of regulatory mechanism is the regulatory-constraint mechanism, where the monopolist is allowed

In recent years, technology has progressed to the point where many of the industries that were traditionally thought to be natural monopolies are being deregulated. Most of these have been "network" industries such as electric or phone utilities. Technological advances have made it possible for a number of providers to use the same network. For example, a firm can generate electricity and put it on the "power grid" where it can then be sold, or competing local exchange carriers (phone companies) can provide phone service by purchasing access to Ameritech's network at prespecified rates.[16] Because it is now possible for multiple firms to use the same network, these industries are being deregulated, at least in part. Generally, the "network" remains a regulated monopoly, with specified rates and terms for allowing access to the network. Provision of services such as electricity generation or connecting phone calls is opened up to competition.

9.4.1   Regulation and Incentives

When a monopolist is regulated, the government chooses the output price so that the monopolist just covers its costs.[17]   Because the monopolist knows that it will cover its costs, it will not have an incentive to keep its costs low. As a result, it may incur costs that it would not incur if it were subjected to market discipline. For example, it could purchase fancy office equipment, and decorate the corporate headquarters. This phenomenon is sometimes known as gold plating. Gold-plating is something regulators look out for when they are determining the cost base for the regulated firm.

An interesting phenomenon occurs when a firm operates both in a regulated and an unregulated industry. For example, consider local phone providers, which are regulated monopolists on local service (the "loop" from the switchbox to your house) but one of many competitors on "local toll" calls (calls over a certain distance – like 15 miles). Since the firm knows it will cover its costs on the local service, it may try to classify come of the costs of operating its competitive service as costs of local service, thereby gaining a competitive advantage in the local toll market.

The extent to which the problems I mentioned here are real problems depend on the industry. However, they are things that regulators worry about. Whenever a regulated monopoly goes before a rate commission to ask for a rate increase, the regulators ask whether the costs are appropriate

---

to do whatever it wants, subject to a constraint such as its return on assets can be no larger than a pre-specified number.   If you are interested in such things, see Berg and Tschirhart, *Natural Monopoly Regulation* or Spulber, *Regulation and Markets.*

[16]Ameritech is the local phone company in the Chicago area, where I started writing these notes.   In the Boston area (where I'm adding this note), the relevant company is Verizon.

[17]Actual regulation usually allows the firm to cover its costs and earn a specified rate of return on its assets.

and whether they should be allocated to the regulated or competitive sector of the monopolist's business. The question of how regulated firms respond to regulatory mechanisms, i.e. regulatory incentives, is another major subject in the economics of monopolies.

9.5   Further Topics in Monopoly Pricing

9.5.1   Multi-Product Monopoly

When a monopolist sells more than one product, it must take into account that the price it charges for one of its products may affect the demand for its other products.[18] This is true in a wide variety of contexts, but let's start with a simple example, that of a monopolist that sells goods that are perfect complements. For example, think of a firm that sells vacation packages that consist of a plane trip and a hotel stay. Consumers care only about the total cost of the vacation. The higher the price of a hotel room, the less people will be willing to pay for the airline ticket, and vice versa.

Demand for vacations is given by:

$$q\left(p_V\right) = 100 - p_V,$$

where p_V is the price of a vacation, $p_V = p_A + p_H$, and p_A and p_H are the prices of airline travel and hotel travel respectively. Each airline trip costs the firm c_H, and each hotel stay costs c_H.

To begin, consider the case where the firm realizes that consumers care only about the price of a vacation, and so it chooses p_V in order to maximize profit.

$$\max_{p_V} \left(100 - p_V\right)\left(p_V - \left(c_H + c_A\right)\right),$$

since the cost of a vacation is $c_H + c_A$. The first-order condition for this problem is:

$$
\begin{aligned}
\frac{d}{dp_V}\left(\left(100 - p_V\right)\left(p_V - \left(c_H + c_A\right)\right)\right) &= 0 \\
p_V^* &= \frac{100 + c_H + c_A}{2} = 50 + \frac{c_H + c_A}{2}.
\end{aligned}
$$

Thus if the firm is interested in maximizing total profit, it should set p_V^* as above, and divide the cost among the plane ticket and hotel any way it wants. In fact, if you've ever bought a tour, you

---

[18]See Tirole, *Theory of Industrial Organization*, starting at p. 70.

know that you usually don't get separate prices for the various components. Profit is given by:

$$\left(100 - \frac{100 + c_H + c_A}{2}\right)\left(\frac{100 + c_H + c_A}{2} - c_H - c_A\right)$$
$$= \frac{1}{4}\left(100 - (c_H + c_A)\right)^2.$$

Now consider the case where the price of airline seats and hotel stays are set by separate divisions, each of which cares only about its own profits. In this case, the hotel division takes the price of the airline division as given and chooses p_H in order to maximize its profit:

$$(100 - p_H - p_A)(p_H - c_H)$$

which implies that the optimal choice of p_H responds to the airline price p_A according to the "reaction curve" or "best-response function" :

$$p_H = \frac{100 + c_H - p_A}{2}.$$

Similarly, the airline division sets p_A in order to maximize its profit, yielding reaction curve:

$$p_A = \frac{100 + c_A - p_H}{2}.$$

At equilibrium, the optimal prices set by the two divisions must satisfy the two equation system:[19]

$$p_A = \frac{100 + c_A - p_H}{2}$$
$$p_H = \frac{100 + c_H - p_A}{2}$$

which has solution:

$$p_A^* = \frac{100 + 2c_A - c_H}{3}$$
$$p_H^* = \frac{100 - c_A + 2c_H}{3}$$

The total price of a tour is thus:

$$p_H^* + p_A^* = \frac{200 + c_A + c_H}{3} \simeq 66.7 + \frac{c_A + c_H}{3}.$$

Unless $c_H + c_A > 100$ (in which case the cost of production is greater than consumers' maximum willingness to pay), the total price of a tour when the hotel and airline prices are set separately is greater than the total price when they are set jointly, $p_V^* = 50 + \frac{c_H + c_A}{2}$.

---

[19] This kind of best-response equilibrium is known in game theory as a Nash equilibrium - more on that in the game theory chapters.

When prices are set separately, the values of p_H^* and p_A^* above imply quantity:

$$100 - \left(\frac{200}{3} + \frac{1}{3}c_A + \frac{1}{3}c_H \right)$$
$$= \frac{100}{3} - \frac{1}{3}c_A - \frac{1}{3}c_H$$

and total profit:

$$\left(\frac{100}{3} - \frac{1}{3}c_A - \frac{1}{3}c_H \right) \left(\frac{200}{3} + \frac{1}{3}c_A + \frac{1}{3}c_H - c_H - c_A \right)$$
$$\frac{2}{9} \left(100 - c_H - c_A \right)^2$$

Finally, since $\frac{2}{9} < \frac{1}{4}$, the firm earns higher profit when it sets both prices jointly than when the prices are set independently by separate divisions.

The previous example shows that when the firm's divisions set prices separately, they set the total price too high relative to the prices that maximize joint profits. The firm as a whole would be better off lowering prices – the increased demand would more than make up for the decrease in price.

What is going on here? Begin with the case where the firm is charging p_V^*, and suppose for the sake of simplicity that the hotel price and airline price are equal. At p_V^*, the marginal revenue to the entire firm is equal to its marginal cost. That is,

$$MR_{Firm} = 100 - 2p_V^* = c_H + c_A.$$

Now think about the incentives for the hotel manager in this situation. If the hotel increases its price by a small amount, beginning from $\frac{p_V^*}{2}$, its marginal revenue is:

$$MR_{Hotel} = 100 - 2p_H^* - p_A^* = 100 - 2\frac{p_V^*}{2} - \frac{p_V^*}{2} > 100 - 2p_V^* = MR_{Firm}.$$

Thus the hotel's marginal gain in revenue due to raising its price is greater than the marginal gain in revenue to the entire firm. Why? Raising the price reduces the quantity demanded, but the increase in price all goes to the hotel, while the decrease in quantity is split between the hotel and airline. But, the hotel manager doesn't care about this latter effect on the airline. The same logic holds for the airline manager's choice of the airline price. Because of this, each division will charge a price that is too high.

What about if the goods were substitutes instead of complements? Think about a car company pricing its product line. If its lines are priced separately, each division head has an incentive to

lower the price and steal some business from the other divisions. Because all division heads have this incentive, the prices for the cars are lower when prices are set separately then they would be if the firm set all prices centrally. This is sometimes known as cannibalization. The firm must worry that by lowering the price on one line it is really just stealing business from the other product lines.

We will return to these types of issues when we study oligopoly. But, for now, we just want to motivate the idea that even the monopolist has to worry about issues of strategy.[20]

9.5.2   Intertemporal Pricing

Consider the following scenario. A monopolist produces a single good that is sold in two consecutive periods, 1 and 2. Using the quantity-based approach again, let $p_1(q_1)$ be the inverse demand curve in the first period and $p_2(q_2, q_1)$ be the inverse demand in the second period. Note that this formulation indicates that the first-period demand does not depend on the second-period quantity. This implies that first-period consumers do not plan ahead in their purchase decisions.

Let $c_1(q_1)$ and $c_2(q_2)$ be the cost functions in the two periods, with a discount factor $\delta = \frac{1}{1+r}$. The monopolist maximizes:

$$p_1(q_1) \cdot q_1 - c_1(q_1) + \delta\left(p_2(q_2, q_1) \cdot q_2 - c_2(q_2)\right)$$

The first-order conditions are:

$$p_1'(q_1) \cdot q_1 + p_1(q_1) + \delta\left(\frac{\partial p_2(q_2, q_1)}{\partial q_1} \cdot q_2\right) = c_1'(q_1)$$

$$\frac{\partial p_2(q_2, q_1)}{\partial q_2} \cdot q_2 + p_2(q_2, q_1) = c_2'(q_2)$$

Thus the monopolist sets marginal revenue equal to marginal cost in the second period. But, what about the first period? In this period, the monopolist must take into account the effect of the quantity sold in the first period on the quantity it will be able to sell in the second period. There are two possible ways this effect could go:

- Goodwill: $\frac{\partial p_2(q_2, q_1)}{\partial q_1} > 0$. Selling more in the first period generates "goodwill" that increases demand in the second period, perhaps through reputation effects or good word-of-mouth. In this case, the monopolist will produce more in the first period than it would if there were only one period in the model (or if demand in the two periods were independent).

---

[20]Strategic issues are the main subject of the rest of these notes, starting with game theory, economists' tool for studying strategic interactions.

- Fixed Pie: $\frac{\partial p_2(q_2,q_1)}{\partial q_1} < 0$. Selling more in the first period means that demand is lower in the second period. This would be true if the market for the product is fixed. In this case the additional sales in period 1 are cannibalized from period 2, leaving nobody to buy in period 2. In this case, the monopolist will want to sell less than the amount it would in the case where demand was independent in order to keep demand up in the second period.

Note - the stories told here are not quite rigorous enough. In Industrial Organization (IO) economics, people work on models to account for these things explicitly. Our object here is to illustrate that the firm's strategy in the first period will depend on what the firm is going to do in the second period.

9.5.3   Durable Goods Monopoly

Another way in which a monopolist can compete against itself is if it produces a product that is durable. For example, think about a company that produces refrigerators. Substitute products include not only refrigerators produced today by other firms, but refrigerators produced yesterday as well. Put another way, the refrigerator is competing not only against other refrigerator makers, it is also competing against past and future versions of its own refrigerator.

Consider the following simple model of a durable goods monopoly. A monopolist sells durable goods that last forever. Each unit of the good costs c dollars. There are N consumers numbered 1,2,...,N. Consumer n values the durable good at n dollars. That is, she will pay up to n dollars, but no more. Each day the monopolist quotes a price, p_t, and agrees to sell the good to whomever wants to purchase it at that price. How should the monopolist choose the path of prices? Assuming no discounting, the monopolist is willing to wait, so it should charge price N on day 1, $N - 1$ on day 2, etc. Each day, it skims off the highest value customers that remain. If there is a positive discount rate, then the optimal price path will balance the extra revenue gained by skimming against the cost of putting off the revenue of the lower valued customers into the future. Either way, the monopolist is able to garner almost all the social surplus as profit. Can you think of examples of this type of behavior? What about hard cover vs. soft cover books?

But, there is the problem with this model. If customers know that the price will be lower tomorrow, they will wait to buy. Of course, how willing they are to wait will depend on the length of the period. The longer they have to wait for the price to fall, the more likely they are to buy today. We can turn the previous result on its head by asking what will happen as the length of

the period gets very short. In this case, consumers will know that by waiting a very short time, they can get the product at an even lower price. Because of this they will tend to wait to buy. But, knowing that consumers will wait to buy, the monopolist will have an incentive to lower the price even faster. The limit of this argument is that the monopolist is driven to charge $p = MC$ immediately, when the period is very short.

So, we have seen that the durable goods monopolist will make zero profit when it has the opportunity to lower its price as fast as it wants to. This phenomenon is known as the Coase Conjecture, so called because Coase believed it but didn't prove it. It has since been proven. Notice that the monopolist's flexibility to charge different prices over time actually hurts the firm. It would be better off if it could commit to charging the declining price schedule we mentioned earlier. In fact, the monopolist would be better off if it could commit never to lower prices and just charging the monopoly price forever. In this case, the firm wouldn't sell any units in the second period, but it would also face no pressures to lower the price in the first period.

How can a monopolist commit to never lowering its prices? The market gives us many examples:

- Print the price on the package

- Some products never go on sale by reputation, such as Tumi luggage

- Third-party commitment, such as using a retailer who is contractually prohibited from putting items on sale

- "Destroy" the factory after producing in the first period, by limiting production runs (such as "limited edition" collectibles or the "retirement" of Beanie Babies)

- Money back guarantee – if the price is lowered in the future, the monopolist will refund the difference to all purchasers.

- Planned obsolescence – if the goods aren't that durable, there isn't a problem. In the absence of an ability to commit to keeping prices high, firms should produce goods that aren't particularly durable.

- Leasing goods instead of selling them

This same phenomenon happens even in a two-period model. The monopolist would like to charge a high price in both period. But after the first period has passed it has an incentive to

lower the price in order to increase sales. But the customers, knowing that the monopolist will lower prices in the second period, will not buy in the first period. Thus, the monopolist will have to charge a lower price in the first period as well. The monopolist would be better off if it could somehow commit to keeping prices high in both periods in this model.

Consider the following scenario. Instead of selling the durable good, the monopolist will lease it for a year. In this case, units of the good rented in period 2 are no longer substitutes for units of the good rented in period 1. In fact, somebody who likes the good will rent a unit in each year.

What should the monopolist do in this case? The monopolist should charge the monopoly rental price (high price) in each period to rent the good. Thus if the monopolist can commit to rental instead of selling, it can earn higher profits. Renting is similar to the other tools listed above, in that it is an example of the firm's intentionally restricting its own flexibility. Thus the monopolist, by eliminating its opportunity to cut prices on sales later, is able to do better. In this context, choice is not always a good thing.

Formal Model of Renting vs. Selling[21]   There are two periods, and goods produced in period 1 may be used in period 2 as well (i.e. it is durable). After period 2 the good becomes obsolete. Assume that the cost of production is zero to make things simple, and the monopolist and consumers have discount factor δ. Demand in each period is given by $q(p) = 1 - p$. The monopolist can either lease the good for each period or sell it for both periods. If the monopolist decides to sell the good, then consumers who purchased in the first period can resell it during the second period.

Suppose the monopolist decides to lease. The optimal price for the monopolist to charge in each period is $1/2$. This yields quantity $1/2$ in the first period, which can be leased again in period 2. No additional quantity is produced in period 2. Thus discounted profits are given by

$$\pi_{lease} = \frac{1}{4} (1 + \delta)$$

Suppose the monopolist decides to sell. In this case, the quantity offered for sale in period 1 is reoffered by the resale market in period 2. Thus the residual demand in period 2 is given by $p_2 = 1 - q_2 - q_1$. Thus in period 2 the monopolist chooses q_2 to solve:

$$\max q_2 (1 - q_1 - q_2)$$

which implies that the monopolist sells $q_2 = \frac{1-q_1}{2}$ and earns profit $\left(\frac{1-q_1}{2}\right)^2$.

---

[21] From Tirole, pp. 81-84.

Now, what should the monopolist do in period 1? The price that consumers are willing to pay is given by the willingness to pay in period 1 plus the discounted price in period 2, since the consumer can always lease the object (or resell it) in period 2 for the market price. Thus consumers are willing to pay

$$(1 - q_1) + \delta p_2^a$$

where p_2^a is their belief about what the price will be in period 2.

We suppose that consumers correctly anticipate the price in the second period.[22] That is, $p_2^a = p_2 = \frac{1-q_1}{2}$. Thus as a function of q_1 the maximum price the monopolist can charge is given by:

$$p_1 = (1 - q_1) + \delta \frac{1 - q_1}{2} = (1 - q_1)\left(1 + \frac{\delta}{2}\right)$$

Thus in the first period the monopolist chooses q_1 to maximize:

$$\pi_{sales} = \pi_1 + \delta \pi_2 = (1 - q_1)\left(1 + \frac{\delta}{2}\right) q_1 + \delta \left(\frac{1 - q_1}{2}\right)^2$$

The first-order condition yields:

$$\frac{d}{dq_1}\left((1 - q_1)\left(1 + \frac{\delta}{2}\right) q_1 + \delta \left(\frac{1 - q_1}{2}\right)^2\right) = 0$$

$$(1 - q_1)(1 + \frac{\delta}{2}) - q_1(1 + \frac{\delta}{2}) - \delta(\frac{1 - q_1}{2}) = 0$$

$$q_1^* = \frac{2}{4 + \delta} < \frac{1}{2}$$

$$q_2^* = \frac{1 - q_1}{2} = \frac{1 - \frac{2}{4+\delta}}{2} = \frac{1}{2} \cdot \frac{2 + \delta}{4 + \delta}$$

$$q_1^* + q_2^* = \frac{2}{4 + \delta} + \frac{1}{2} \cdot \frac{2 + \delta}{4 + \delta} = \frac{1}{2} \cdot \frac{6 + \delta}{4 + \delta} > \frac{1}{2}$$

In terms of prices,

$$p_2^* = 1 - (q_1^* + q_2^*)$$

$$p_2^* = \left(1 - \frac{1}{2} \cdot \frac{6 + \delta}{4 + \delta}\right) = \frac{1}{2} \cdot \frac{2 + \delta}{4 + \delta} < \frac{1}{2}$$

$$p_1^* = (1 - q_1)\left(1 + \frac{\delta}{2}\right) = \left(1 - \frac{2}{4 + \delta}\right)\left(1 + \frac{\delta}{2}\right) = \frac{1}{2} \frac{(2 + \delta)^2}{4 + \delta} < \frac{1 + \delta}{2}$$

---

[22] In game theory terms, this is a Perfect Bayesian equilibrium.

Note that $\frac{1+\delta}{2}$ is the monopoly price if the second period is ignored ($\delta = 0$). Going back to the expression for total profit:

$$
\begin{aligned}
\pi_{sales} &= (1 - q_1)\left(1 + \frac{\delta}{2}\right)q_1 + \delta\left(\frac{1 - q_1}{2}\right)^2 \\
&= \frac{1}{2}\frac{(2+\delta)^2}{4+\delta} \cdot \frac{2}{4+\delta} + \delta\left(\frac{1}{2} \cdot \frac{2+\delta}{4+\delta}\right)^2 \\
&= \left(\frac{2+\delta}{4+\delta}\right)^2 + \delta\left(\frac{1}{2} \cdot \frac{2+\delta}{4+\delta}\right)^2 \\
&= (1 + \frac{\delta}{4})\left(\frac{2+\delta}{4+\delta}\right)^2
\end{aligned}
$$

Comparing the sales profit, $(1 + \frac{\delta}{4})\left(\frac{2+\delta}{4+\delta}\right)^2$, to the leasing profit, $\frac{1}{4}(1+\delta)$, we can see that $\pi_{lease} = \pi_{sales}$ only when $\delta = 0$. It turns out that for all $\delta > 0$ (meaning the second period matters, at least somewhat), then $\pi_{lease} > \pi_{sales}$. Leasing is more profitable. Why? When selling the good, the monopolist cannot resist the temptation to lower the price in the second period and sell more; thus, it cannot sell as much in the first period ($q_1^* < \frac{1}{2}$) as would be optimal, and this reduces overall profits. The monopolist would be better off if it could commit to leasing rather than selling.