(b) Suppose demand for the product suddenly shifts to $D(p) = 200(12-p)$. In the short run, firms cannot change their production quantities at all. What is the new short-run equilibrium? In the intermediate run, firms in the industry can change their production quantities according to the total cost function just given. What is the intermediate-run equilibrium in this industry? In the long run, firms can enter and leave the industry. What is the new long-run equilibrium in this industry?

11.8 Suppose, in the industry of Problem 11.7, four firms have a superior production technology, which gives each the total cost function $\mathrm{TC}(x) = 50+x+0.04x^2$. An additional eight firms have the cost function from Problem 11.6. There are no other possible entrants into this industry.

If demand for the product is given by $D(p) = 200(10 - p)$, what is the equilibrium in this industry?

I have not told you whether the fixed costs of the 12 firms can be avoided. Does this matter to the answer to this problem? How?

11.9 Suppose that, in Problem 11.8, instead of eight firms with the cost function from Problem 11.7, an unlimited number of firms possess this cost function. Assume that all fixed costs can be avoided if a firm produces no output. If demand for the product is given by $D(p) = 200(10 - p)$, what is the equilibrium in this industry?

12. Market Efficiency

This chapter has two objectives:

- It presents the concepts of *consumer* and *producer surplus*, dollar-valued measures of the value that consumers and producers receive from participating in market exchange.
- Using these concepts, it shows why economists rhapsodize about markets and prices: Competitive markets are *efficient*, meaning they lead to the largest possible level of total surplus. But this statement must be qualified, and many of the qualifiers are discussed here as well.

This can be a frustrating chapter; the arguments given are more abstract than anything else in the book. If your skills of logical and mathematical proof are relatively weak, you may find it hard going. If so, at least understand what is being claimed, if not the arguments that support the claims. In the next chapter you get plenty of opportunity to see how these concepts and ideas are applied in more concrete settings.

If you surveyed the general populace, asking people to name phrases that economists use, *supply equals demand* would be the hands-down winner. But a contender for second place would probably be *the invisible hand*.

The invisible hand, a phrase coined by the father of economics, Adam Smith, refers to the role prices play in achieving a good allocation of resources in the economy. Economists, or at least economists who respect markets, are fond of rhapsodizing about the price mechanism. "Imagine," one of them might say, "many consumers and producers, with many conflicting preferences and capabilities, whose activities need to be coordinated. And the price mechanism does this so well, telling consumers the 'cost' of a particular item while telling producers the 'value' of the item in the marketplace." Warming up to her subject, our market-respecting economist explains, "Prices and the market mechanism are like an invisible hand that correctly and efficiently coordinates consumer desires and producer activities, achieving an unimprovable result."

Indeed, and with some justice, the invisible hand is sometimes cited as one reason that Soviet-style communism failed. The former Soviet Union was run as a centrally planned and administered economy. That is a pretty big operation to run from headquarters; and without in the least disparaging the talents of the commissariat and others who did the planning, the planning did not always work well.

In comparison, in price-driven economies, prices decentralize the planning process. To be very clear about this, decentralization is what gets the good results. Individual consumers have a lot of information and the time and incentive to process it, as they seek to maximize their utility. Firms have a lot of information and the time and incentive to process it, as they chase higher profits. Prices play an informational role: They sum up very concisely everything profit-maximizing firms and utility-maximizing consumers need to know about each other's desires and capabilities. Firms and consumers, guided by equilibrium prices and their own self-interest, reach an overall outcome that centrally planned and administered economies seemingly cannot touch.

Needless to say, there is a lot more to why Soviet style communism failed. The corrupt political system played a role, as did a lack of incentives for individuals to take risks and innovate. But there is little doubt that running by fiat an economy the size of the former Soviet Union's or even one the size of, say, Portugal is a daunting task. The central planners sitting in Moscow may have been able to decide how many shoes to make for sale in Minsk. But they probably had no clue about styles. And, if they got the styles wrong, the shoes might sit on the shelves of shoe stores in Minsk, as the consumers of Minsk made do with their somewhat worn old shoes. In comparison, the owner of a private shoe store in Zurich has a very strong incentive, and the time required, to learn what would be deemed a stylish pair of footwear in Zurich. Even if the bureaucratic planner sitting in Moscow had the incentive to follow the latest fashion trends in Minsk, which is doubtful, he still would have lacked the time and opportunity to glean the required information. Prices—or more properly, the decentralization of decision making and information gathering, allied with the incentives provided by the market system—get the job done.

What does the invisible hand accomplish, precisely? We discover in this chapter that the invisible hand, *if certain conditions are met*, produces an *efficient* outcome but not necessarily one that is *equitable*.

One more introductory thought, and we can begin. Economists attack this general issue in a number of different ways. One important dimension along which the attacks vary is the scope of the analysis: We can look at a single market, such as the market for shoes—what is called *partial equilibrium analysis*—or we can try to think about the whole economy, all at once—which is known in Economese as *general equilibrium analysis*. We have been developing tools appropriate for the first sort of attack, and so we look at these questions in the context of a single market. But, in the *Student's Companion*, following the solutions to problems for this chapter is a brief introduction to the general equilibrium approach to these questions.

12.1. Consumer and Producer Surplus

For most of this chapter, we look at a perfectly competitive market, where all supply comes from firms and all demand comes from consumers.

Figure 12.1 depicts the usual supply-equals-demand picture, with the equilibrium price and equilibrium quantity at the point where supply intersects demand. Note the two shaded regions. The darker region is bounded by the equilibrium price and the demand curve to the left of the equilibrium quantity. The area of this region is called the *consumer surplus*. The lighter region is bounded by the equilibrium price and the supply curve, to the left of the equilibrium quantity. The area of this region is called the *producer surplus*. I assert the following:

> Consumer surplus measures, in dollars, the benefits consumers obtain from trading in this market. Producer surplus measures, in dollars, the benefits producers obtain from trading in this market. Hence, these two quantities when summmed measure in dollars the value generated by the existence of this market.

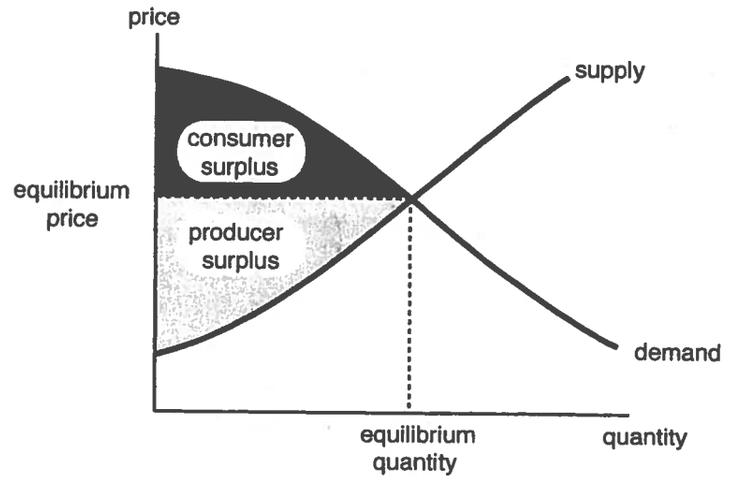


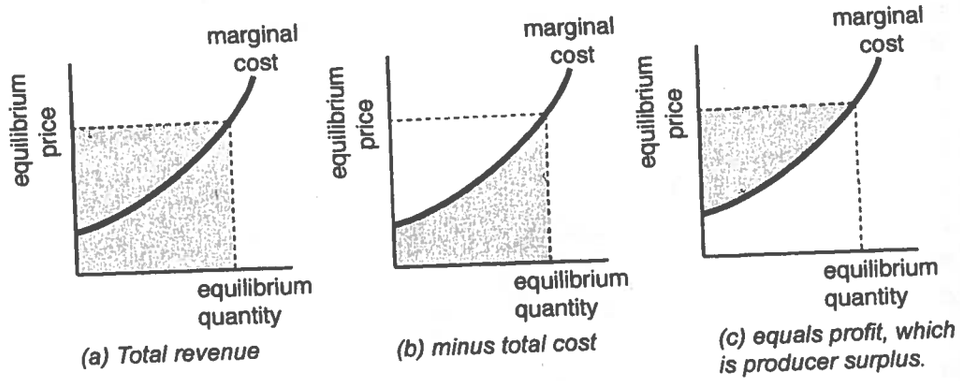
Figure 12.1. Consumer and producer surplus. Consumer surplus (the area of the darkly shaded region) measures in dollars the benefits that consumers obtain from the equilibrium outcome. Producer surplus (the area of the lightly shaded region) measures in dollars the benefits to producers of this market outcome.

Let me immediately qualify these assertions: they are not strictly true all the time. To understand them, we have to be a lot clearer about the term *benefits*. The next two sections indicate what these statements mean, when they are true, and why.

12.2. Producer Surplus

Producer surplus is, more or less, just a fancy way of saying *the profits of the firms in the industry*.

Assume for now that all firms in the industry have no fixed cost and rising marginal cost. Figure 12.2 graphs the marginal cost function of a single producer, which is also a graph of the producer's supply function. Suppose the equilibrium price and corresponding supply decision of the firm are as shown. In the first panel of the figure, the shaded rectangle's height is the price and its length is the firm's level of output; its area is the producer's total revenue. In the second panel, the shaded area under the marginal cost curve, up to the firm's level of output, is the integral of marginal cost, which is the total cost of the producer. So the difference between the areas of the two shaded regions, shaded in the third panel, is total revenue minus total cost, or profit.



(a) Total revenue (b) minus total cost (c) equals profit, which is producer surplus.

Figure 12.2. Producer surplus = profits for a single producer.

This is for a single producer. Now we extend to multiple producers. Suppose this industry has m firms, numbered 1, 2, ..., m. Write $s_j(p)$ as the supply function of firm number j, so that total industry supply is $S(p) = s_1(p) + \ldots + s_m(p)$.

1. The shaded area in the third panel of Figure 12.2 represents the profits of a single firm, say, firm number j. Remembering that supply is a function of price—so turn either your head or the page by 90°—this area is $\int_0^{p'} s_j(p)\,dp$, where p' is the equilibrium price and we correctly regard $s_j(p) = 0$ for prices p that are so low that the firm does not supply a positive amount at them.

2. The total industry producer surplus can likewise be written as $\int_0^{p'} S(p)\,dp$.

3. But then producer surplus, or $\int_0^{p'} S(p)\,dp$, is just

$$\int_0^{p'} [s_1(p) + \ldots + s_m(p)]\,dp = \int_0^{p'} s_1(p)\,dp + \ldots + \int_0^{p'} s_m(p)\,dp,$$

which is the sum, firm by firm, of the firms' profits.

Fixed Cost and Producer Surplus

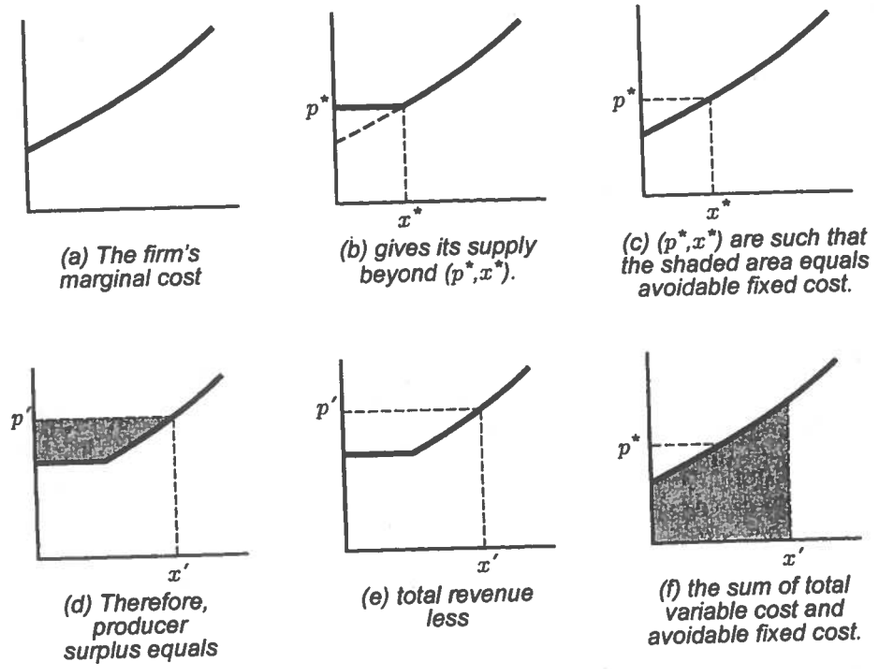
What if one or more of the firms in the industry has a more complex total-cost function? I am not going to be completely general, but let me at least state the result for firms with a rising marginal cost function and, possibly, a positive fixed cost. Divide the fixed cost of each firm into avoidable and unavoidable portions.

> Producer surplus for the industry—the area bounded by the industry supply function and the equilibrium price, out to the equilibrium quantity—is the sum of the profits of the firms in the industry gross of (not counting) their unavoidable fixed costs.

The part *gross of unavoidable fixed costs* is easy: Take one firm. Its unavoidable fixed cost, being unavoidable, has no impact on its supply decisions. Hence, it has no impact on the firm's supply curve. If we redid the argument of Figure 12.2, we would conclude that the shaded area in the third panel is the firm's profit gross of the firm's unavoidable fixed cost.

Suppose the firm can avoid some or all of its fixed cost. Figure 12.3 tells the tale. Panel a shows the firm's full marginal cost function. In panel b, the firm's supply function is shown; it traces marginal cost above a price level p^* and quantity x^*, but is zero for lower prices. Moreover, we know from Chapter 11 that the price–quantity pair p^* and x^* are the price and quantity at which the firm just covers the avoidable portion of its fixed cost. That is, the firm's variable profit gross of all fixed cost at p^* and x^* (the shaded region in panel c) just equals its avoidable fixed cost. So, in panels d, e, and f, for some price and production level (p', x') along the firm's supply curve, producer surplus (panel d) is the firm's total revenue (panel e) less the sum of its total variable cost (the heavily shaded region in panel f) and avoidable fixed cost (the lightly shaded region in panel f).

Putting this together, for a single firm with a rising marginal cost function and a positive fixed cost, its producer surplus area is its total revenue less the sum of its total variable cost and avoidable fixed cost, which is its profit gross of any unavoidable fixed cost.



(a) The firm's marginal cost
(b) gives its supply beyond (p^*,x^*).
(c) (p^*,x^*) are such that the shaded area equals avoidable fixed cost.
(d) Therefore, producer surplus equals
(e) total revenue less
(f) the sum of total variable cost and avoidable fixed cost.

Figure 12.3. Producer surplus with partially avoidable fixed cost.

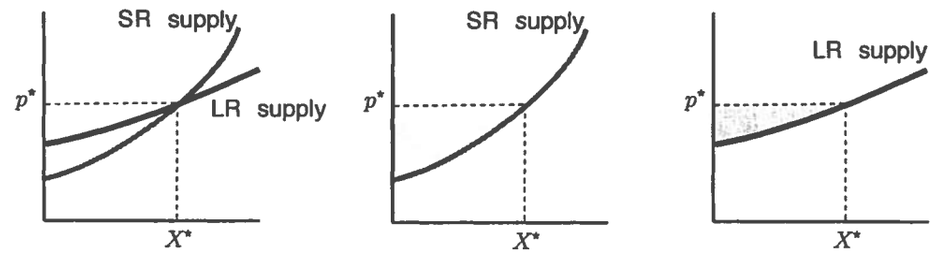
That is for a single firm. For several firms, simply repeat the argument with all those integral signs. As you do, remember that *profit* in the previous subsection becomes *profit gross of unavoidable fixed cost* in this one.

Unavoidable Fixed Cost and Different Periods

For most purposes, that producer surplus is gross of unavoidable fixed cost is unimportant. If the fixed costs of producers are unavoidable, then they are unavoidable in any institutional arrangement, and society simply bears them. But the presence of unavoidable fixed costs can cause confusion when we apply the idea of producer surplus in analyses of short- and longer-run responses to things like taxes and price ceilings. (I use *longer-run* instead of *long-run* to cover the case of intermediate-run analysis as well.)

Figure 12.4(a) presents the typical picture of short- and longer-run supply in a competitive industry: Short-run supply is less elastic than longer-run supply, and the two supply functions cross at the status-quo production level X^*.

In Figure 12.4(b) we reproduce short-run supply, and using the status-quo production level X^* and price p^*, we shade in the area that gives producer surplus. In Figure 12.4(c), we repeat this, but using the longer-run supply curve. So the shaded area in Figure 12.4(b) represents short-run



(a) (b) (c)

Figure 12.4. Short- and longer-run producer surplus. Panel a shows short- and longer-run supply; panel b the short-run producer surplus; and panel c the longer-run producer surplus, where in panels b and c, the quantity and price are the longer-run status-quo levels. Short-run producer surplus exceeds longer-run producer surplus. Does this mean that profits are higher in the short run than in the longer run? See the text for the explanation.

producer surplus = short-run producer profit, and the shaded area in Figure 12.4(c) gives longer-run producer surplus = longer-run producer profit. Because longer-run supply is flatter than short-run supply, the shaded area in Figure 12.4(b) is larger than the shaded area in Figure 12.4(c). So, it would seem that short-run producer profit is *greater* than longer-run producer profit at the status-quo level of production. In all our discussions of short-run and longer-run costs, short- and longer-run cost were equal at the status-quo level of production, so something seems amiss here.

Actually, nothing is amiss. The problem is that unavoidable fixed costs are higher in the short run than in the longer run, and the shaded areas represent profits gross of unavoidable fixed costs.

To keep things simple, assume all fixed costs are unavoidable. Then Figure 12.5, depicting a firm's short-run and longer-run *total costs*, tells the tale. The firm's short-run and longer-run total costs are equal at the its status-quo level of production, and longer-run total cost is otherwise less than short-run total cost. Clearly, then, the short-run total cost at production level 0—or the short-run fixed cost—exceeds the longer-run total cost at production level 0, which is just the longer-run fixed cost.

Indeed, Figures 12.4 and 12.5 show how to find the difference between short-run and longer-run unavoidable fixed costs for the industry as a whole, from the short-run and longer-run industry supply functions. The two total costs are equal at the status-quo production quantity, so the area between the short-run and longer-run industry supply functions, out to the status-quo level of production (the triangular area in Figure 12.4(a)) is the difference in fixed costs.

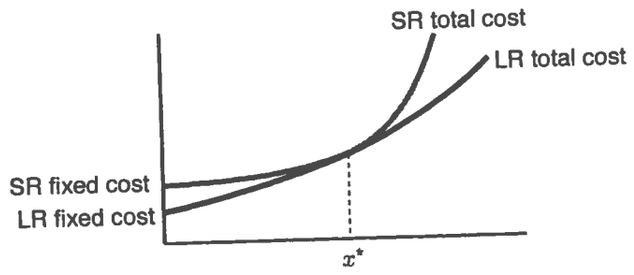


Figure 12.5. Short- and longer-run total cost functions.

In applications, questions arise such as these: If a tax is imposed on the industry or a price ceiling is enforced, how do producer profits change in the short run? How do they change in the long run? We cannot tell the absolute levels of fixed costs from the industry supply curves, so we cannot tell what is the absolute level of profits in any run. But, if we want to talk about the *change* in producer profits, fixed costs net out. The change in producer surplus equals the change in producer profits, whether the producer surplus measures net profits or profits gross of unavoidable fixed costs. Comparisons based on changes in producer surplus are completely legitimate, if you are careful: When you want to measure changes in profits in the short run, use the change in short-run producer surplus, based on the short-run supply curve. When you want to measure changes in profits in the longer run, use the change in longer-run producer surplus, based on the longer-run supply curve. In this way, any unavoidable fixed costs for the appropriate time frame will net out of the comparison. (Try Problem 12.3.)

Producer Surplus in Other Cases

We have been studying the concept of producer surplus exclusively in the context of a perfectly competitive industry, at the equilibrium outcome. Notwithstanding this single focus, the concept—essentially the sum of profits of producers, with some allowance made for various categories of fixed costs—is entirely general.

A nice thing about the context of a competitive market equilibrium is that producer surplus can be "seen" on the picture of supply and demand; it is the area of a recognizable region. In other contexts, this continues to be true. Most notably, in a monopoly industry, it is relatively easy to find producer profit gross of all fixed cost on the usual marginal cost–marginal revenue diagram, and you can find profit net of fixed cost if you have the average cost function on the graph. (See Problem 12.4 for where to find these.) But, in other contexts, it can be difficult to "see" producer surplus.

When we deal with producer surplus, we deal almost exclusively with competitive market equilibria or monopoly industries, so we nearly always can "see" producer surplus. But when you encounter other contexts, you will need to take considerable care in identifying producer surplus with some area on some graph.

Supplier Surplus When Suppliers Are Consumers

What if supply comes not from firms, but from consumers selling out of their endowment?

I do not take you through the details, but in this case, if we relabel producer surplus as *supplier surplus*, this area gives a dollar-valued measure of the benefits to suppliers of participating in the market, even if some or all the suppliers are consumers. The way to show this is (1) to figure out the supply function for an individual consumer who is selling out of his or her endowment, and (2) mimic the arguments we are about to make about consumer surplus. If you are worried about this, consult an advanced textbook in economics; but my advice is not to worry about it.

12.3. Consumer Surplus

Now we turn to consumer surplus. To begin with an admission, in general, the more heavily shaded area in Figure 12.1 is not precisely a dollar-valued measure of the benefits consumers take from consuming this good. Instead, this area gives an *approximation* of these benefits. I do not try to explain what this means: how a dollar–valued measure of benefits for a general utility is created, the nature of this approximation, when the approximation is particularly good. These are hard things to do, involving many hard-slogging pages of derivatives, which is in neither your interest nor mine. So you have to take my assertion on faith or consult a doctoral-level book on the subject.

You don't get off quite that easily, however. In a special case, the area in question provides a precise measure of these benefits. This is the case of *linear-money-left-over utility*, utility functions of the form $u(x_1, \ldots, x_k) = v_1(x_1) + \ldots + v_k(x_k) + m$, where m is money left over. You already read the argument for a single consumer back in Chapter 5 (consult pages 124–5) and taking the argument from one consumer to a set of consumers is one more application of the sum-of-integrals argument from pages 294–5.

Consumer Surplus for the Reservation-Demand Model

I complement the argument from Chapter 5 with a case that is, I hope, particularly transparent. This concerns a very special model of consumer demand,

known as the *reservation-price* model. In this model,

- For the good in question, each consumer wishes to consume either precisely one unit of the good or none at all.
- In dollar terms, for each consumer, consumption of 1 unit of the good is worth an amount $\$r$, called the *reservation price* of the good for the consumer. If the consumer pays $\$p$ for the good and consumes 1 unit of it, her utility, measured in dollar units, rises by $\$r - \$p$.
- Therefore, the consumer's decision whether to buy the good is simple. If the price of the good is less than her reservation price, buy. If the price is more than her reservation price, do not buy. If the price equals her reservation price, she does not care whether she buys the 1 unit or not.

For example, imagine 10 consumers, whose reservation prices for the good in question are, respectively, \$16, \$4.50, \$3, \$10, \$8, \$2, \$8, \$4, \$6, and \$5. If the price of the good is \$9, the first and fourth persons will buy units and the other eight will not. The utility gain to the first person, measured in dollars, will be \$16 – \$9 = \$7, and the gain to the fourth will be \$1. At a price of \$4.40, the 1st, 2nd, 4th, 5th, 7th, 9th, and 10th consumers will purchase, with gains, respectively, of \$11.60, \$0.10, \$5.60, \$3.60, \$3.60, \$1.60, and \$0.60. This gives us the demand function shown in Figure 12.6.

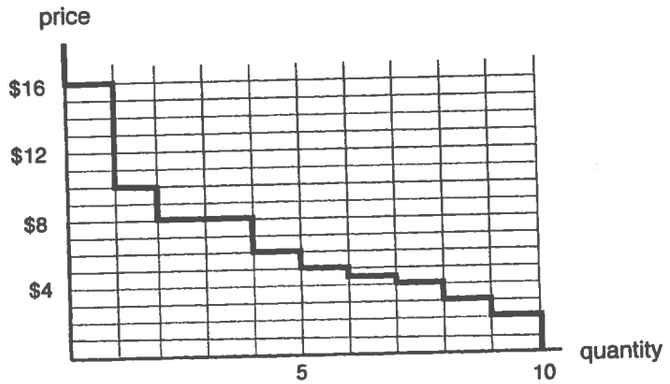


Figure 12.6. Demand in the case of reservation prices.

Usually we think of there being many more than 10 customers, with reservation prices scattered over a range. This gives a demand function that looks like a staircase with very tiny steps. Figure 12.7 shows such a staircase demand function, along with a supply function and three of the consumers whose reservation prices are \$9.50, \$8.50, and \$6.80. The equilibrium quantity is 24 and the equilibrium price is \$4.

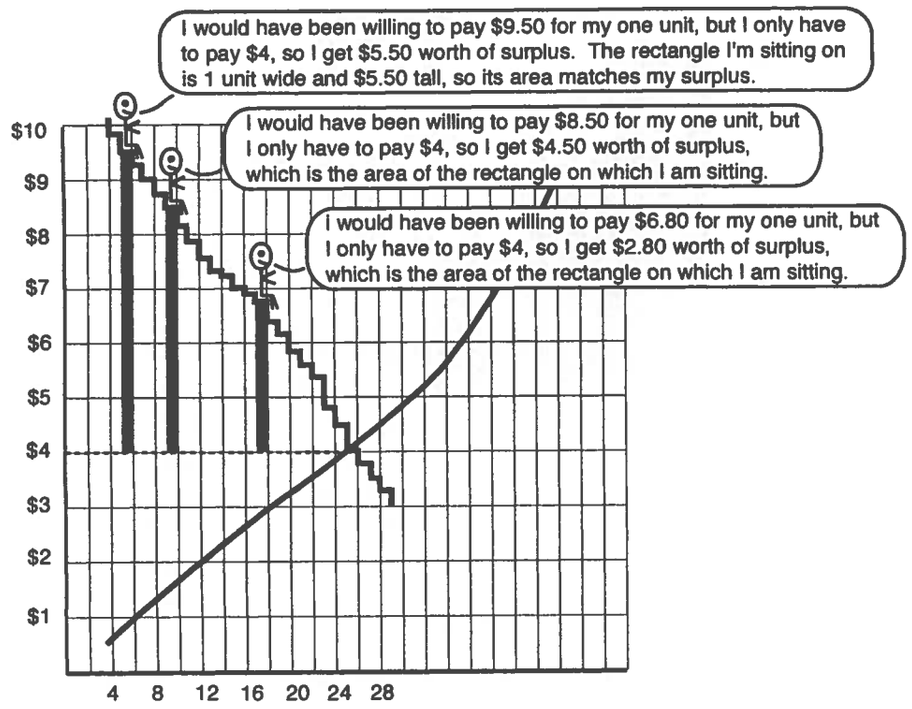


Figure 12.7. Supply, demand, and the surpluses enjoyed by three consumers.

Consider the consumer whose reservation price for the good is \$9.50. Since she has to pay only \$4, she is better off by the amount \$5.50. As she tells you in the figure, this *surplus* that she enjoys is the area of the shaded rectangle on which she sits. The other two consumers make similar statements.

In Figure 12.8, all the consumers are brought together. If we add all their utility gain or surplus rectangles, we get the shaded area that we called the *consumer surplus*.

Consumer Surplus in Nonequilibrium Situations

As with producer surplus, we sometimes want to apply the concept of consumer surplus in contexts other than the supply-equals-demand equilibrium of a competitive marketplace. The concept does not change in the least when applied to a monopoly industry, where the monopolist serves a given demand function, or indeed to any setting where a price for the good is taken as given by consumers, who then buy as much as they desire at that price. Slightly more complex are contexts in which the good is sold using a price-discriminating scheme, but usually these can be "figured out" if you are careful. For instance, in the ideal-for-the-seller world of first-degree price

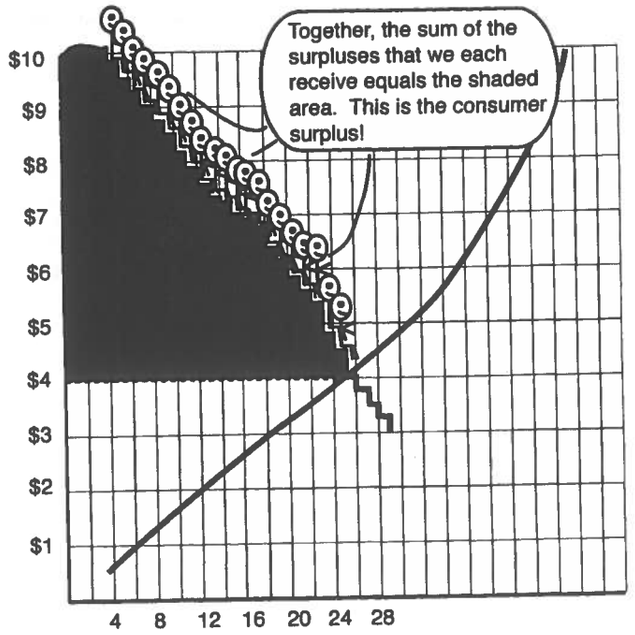


Figure 12.8. Consumer surplus is the surpluses enjoyed by all the consumers.

discrimination, consumer surplus is, by design, $0.

Other contexts can be hard to deal with. Suppose, for instance, the government nationalizes the shoe industry and produces 10 million pairs of shoes. Suppose further that the price per pair that would give demand of 10 million pairs is $50 per pair, but the Commissar for Shoes decides that a price of $10 per pair is more equitable. (Of course, a commissar would be dealing in a currency other than dollars.) Demand at $10 per pair is presumably significantly more than 10 million pairs, and so the 10 million pairs must be rationed among consumers. Queues form at the shoe department. Bribes are paid to shoe salespersons. A black market in shoes springs up, where people sell or barter shoes they do not want for things they do. And, as an economic analyst, you must work out the consumer surplus engendered by the 10 million pairs, sold—at least officially—for $10 per pair.

You cannot carry out this task without making a lot more assumptions. Even assuming linear-money-left-over utility functions, you need to know on whose feet those shoes will eventually be found, because you need to know how much benefit the shoe wearers get. As we are about to discover, a nice thing about competitive markets is that the folks who "value" the items for sale the most—who are willing to pay the most—get the items. (The quote marks around *value* will be explained later.) This allows for the computation of consumer surplus. But, when nonmarket allocation schemes

are used, unless we know the identity of the folks who get the goods and the dollar-valued utilities they receive from those goods, we are stumped. For more on this, see Problem 12.6.

Firms as Customers

In the story so far, demand has come entirely from consumers. For the remainder of this chapter and the next, we maintain this assumption. But in real life, demand arises as well from producers, who purchase inputs to their production process. What about them?

It takes too long to grind through the details, so I simply assert this: For a firm that purchases some good for use in its production, where the firm takes the price of this input as given, the area under the firm's demand-for-the-input curve, down to the price, out to the quantity it buys (the area of the "consumer surplus" region) is the gain in profits accruing to the producer from being able to purchase and use the factor input. When part of demand comes from firms, the term *consumer surplus* is no longer appropriate. But, if we call it *customer* or *purchaser surplus* instead, the heavily shaded area in Figure 12.1 measures just the right thing.

12.4. Competitive Markets Maximize Total Surplus

Now that we have standards by which to judge the benefits received by consumers and producers from particular market outcomes, we can see in what sense the equilibrium achieved by a perfectly competitive market is ideal. The result is simple to state: *A competitive market equilibrium maximizes the sum of consumer and producer surplus.* This result is true in substantial generality but requires some assumptions; in particular,

- Each consumer's utility from consumption depends only on his or her own level of consumption and not on how the goods are produced or what other consumers get to consume.
- Also, each firm's total cost of production depends only on what the firm itself produces.

In Chapter 14, we learn that this pair of assumptions is crucial if the invisible hand is to perform ideally; to use the language of Chapter 14, this assumes that there are no consumption or production *externalities*.

While the argument is true in substantial generality, proving it is true takes a somewhat abstract argument. So, instead of giving the general argument here, I provide a simple argument, based on the assumptions that firms have rising marginal cost functions and no fixed costs and consumers

have linear-money-left-over utility functions $v(x) + m$ where v moreover is concave.

With these assumptions, the argument runs as follows. In any plan for production and consumption, consumers consume amounts of the good in exchange (presumably) for cash, while firms take in cash in exchange for producing the good. The amount of cash received by firms must equal the amount of cash paid by consumers (monetary transfers must balance) and the amount of the stuff consumed must equal the amount produced. If consumer i gets x_i units of the good in return for m_i money, his consumer surplus is $v_i(x_i) - m_i - v_i(0)$. And if firm f gets m^f money for producing x^f goods, its surplus is $m^f - \mathrm{TC}_f(x^f)$. (I use subscripts on x such as x_i for consumer i's consumption level and superscripts on x such as x^f for the firm f's production quantity.) Therefore, the sum of everyone's surpluses is

$$\sum_i [v_i(x_i) - m_i - v_i(0)] \quad + \quad \sum_f [m^f - \mathrm{TC}_f(x^f)],$$

where the first sum is over all the consumers and the second sum is over all the firms. But the monetary transfers from consumers to firms must balance—that is, $\sum_i m_i = \sum_f m^f$—and so the sum of everyone's surpluses simplifies to

$$\sum_i [v_i(x_i) - v_i(0)] \quad - \quad \sum_f \mathrm{TC}_f(x^f). \tag{12.1}$$

Time for the first major conclusion:

> The problem of maximizing the sum of surpluses is the problem of finding the production levels for the firms and the consumption levels for the consumers that maximize the expression (12.1), subject to the requirement that what is consumed equals what is produced, or $\sum_i x_i = \sum_f x^f$.

What is the solution to this problem? This is a constrained maximization problem, very much like the problem solved in Section 9.3, in that the variables enter the constraint in a one-for-one tradeoff. The rule for this optimization problem, ignoring complications for x_i or x^f that are 0, is this:

> At the solution to this problem, all the marginal costs should be equal, all the marginal utilities should be equal, and they should all be equal to one other.

In brief, if one firm has a higher marginal cost than another, we could rearrange production and get the same amount of stuff at a lower total cost. If one consumer had higher marginal utility than another, we could rearrange their consumption and increase the utility of the first more than we decreased the utility of the second. And if Consumer A's marginal utility was, say, more than Firm B's marginal cost, we could have B make more, give it to A to consume, and raise A's utility more than we increased B's cost; the case where B's marginal cost exceeds A's marginal utility calls for decreasing A's consumption and B's production.

Such equating of margins is precisely what happens in a competitive market equilibrium. The key is the equilibrium price p. Recall that, in the introduction, I said that the role played by prices was informational. Now I can elaborate on this: The equilibrium price tells producers what their product is worth *on the margin* to consumers. It tells consumers what it costs *on the margin* to make the item. Of course, consumers and producers do not think in those terms: They maximize utility or profit. But as they do so, they cause the marginal cost of production to equal the marginal value of the good in consumption, and that is what maximizes total surplus. That is the invisible hand.

Breaking This Down: Production Efficiency; Consumption Efficiency; and the Right Total Quantity

I hope the argument just given is clear, but in case it is not, let me break this down. Step 0 in the argument is that, when we look at the total surplus, transfers of money do not matter. A dollar more or less in one party's pocket is a dollar less or more in another party's; these all net out of the total surplus calculation. What matters is the *physical goods outcome*, which consists of three things:

1. The total amount of the good, X, produced and consumed
2. The total cost incurred in producing X units of the good, which in turn depends on how this total production quantity is allocated among producers
3. The total utility generated for consumers who get to consume the good, which depends on how the X units are allocated among consumers

Then, with reference to these three:

- The cheapest way to produce X units is the way a competitive market does: Find the price p that gives supply X and let profit-maximizing firms choose how much to produce at price p.

- The greatest sum of consumers' utilities that X units of the good can produce is obtained by allocating the X units in a competitive market: Find the price p that gives demand X and let utility-maximizing consumers choose how much to consume at price p.
- The quantity X where supply equals demand maximizes the sum of utility generated in consumption less the total cost to the entire industry of production.

Or, paraphrasing, perfect competition minimizes the cost of production, maximizes the total utility obtained in consumption, and correctly sets the quantity produced and consumed.

Fairly general proofs of these propositions, more general than those given before, are provided in the *Student's Companion*.

This Does Not Work When Firms Have Market Power

What if the market is not perfectly competitive? Suppose it is served by a single profit-maximizing firm that faces a downward sloping demand function. Figure 12.9 presents demand, marginal revenue, and marginal cost. There is no supply function in this case, because a firm facing a downward sloping demand function sets the market price of its output rather than responding to an exogenously given market price.

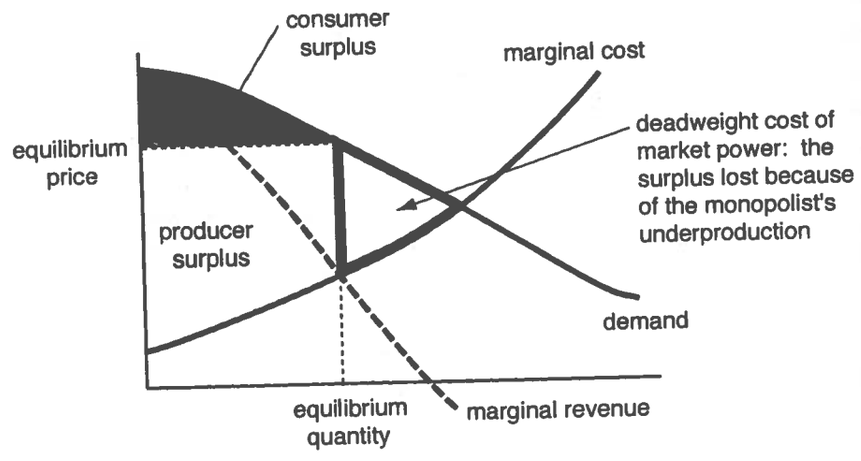


Figure 12.9. Surplus with a monopolist. A firm facing downward sloping demand produces where marginal cost equals marginal revenue, which is less than where marginal cost equals the marginal utility to the last consumer of the good. The loss in total surplus to society, relative to the best society could achieve, is the heavily outlined triangle, the so-called deadweight cost of market power.

The market equilibrium in this case is determined by the intersection of marginal cost and marginal revenue, giving the quantity and price marked on Figure 12.9. The area of the darkly shaded region is the amount of consumer surplus, and the area of the lightly shaded region is the firm's profit, gross of any fixed cost the firm faces. (See Problem 12.4.) Therefore, the combined shaded area is the total surplus generated in this market. It falls short of the maximal amount of surplus that could be generated by the area of the heavily outlined triangle. *The firm, in pursuit of maximal profits, produces at the level where its marginal cost equals its marginal revenue. This is less than the level that maximizes total surplus, where the firm's marginal cost equals the marginal utility or inverse demand.*

What About the Government's Share?

So far in this book, money that changes hands goes from firms to consumers or, more likely, from consumers to firms, with each dollar going out of one pocket winding up in another. But, in the next chapter, a third category of "player" is added to the story: There are firms, consumers, and the government, which may impose taxes or provide subsidies and the like. In the next chapter and Problem 12.7, when we evaluate outcomes that involve net inflows or outflows of cash from or to the government, we treat net government revenue on a dollar-for-dollar on-par basis with consumer and producer surplus. That is, the total surplus equals the net consumer surplus, plus the net producer surplus, plus the government's net revenue. In Chapter 14, we begin to question whether and why a dollar in the government's hands might be worth more or less than a dollar held by the private sector of the economy.

12.5. Efficiency versus Equity

So subject to some qualifications—that stuff about no externalities, which we explain in Chapter 14—competitive markets maximize total surplus. But is total surplus a good standard of comparison for social outcomes? What does it capture and what does it miss?

In any arrangement that causes the good to be produced and assigned to consumers, with money changing hands, two qualities to look for are efficiency and equity.

Efficiency measures whether the right amount of the good is produced in terms of the marginal costs and benefits of the good to society as a whole, whether it is produced by the low-marginal-cost producers, and whether it gets into the hands of the high-marginal-valuation consumers. *Equity*, on the other hand, concerns whether the combined transfers of goods and money

result in a fair distribution of the surplus that could be created by production and exchange.

The sum of consumer and producer surplus measures efficiency without paying heed to equity. There are many ways to make this point; here are four:

1. When it comes to total surplus, only the physical good transfers matter. Any transfer of money is a wash. If we took $100,000 from every other consumer and divided the money in equal shares among all the firms and the other half of the consumers, total surplus would be unchanged. I doubt that many people would believe that the equity of outcome would be unaffected.

2. In computing total surplus, we add consumer and producer surplus on equal terms. We presumably attach value to profits because they are eventually returned to shareholders of the firm. But, insofar as shareholders tend to be wealthier as a group than consumers, some would argue that firms' profits ought not to be weighted as much as consumer benefits. Institutional arrangements that, say, increase consumer benefits or surplus by the equivalent of $10 million and decrease firms' profits by $20 million may be judged to improve equity, even though this means a $10 million loss in total surplus.

3. When measuring total surplus, we measure each consumer's utility in dollar terms and exchange one consumer's utility for another's, dollar for dollar. Suppose we have two consumers, one quite rich, a second very poor. Suppose we have some item, a bag of groceries, say. We ask each consumer: How much *money*, on the margin, is this bag of groceries worth to you? The rich consumer, being quite rich, gets little value on the margin from her money; she's bought nearly everything she wants or needs. Still, flush with cash and nothing to spend it on, she is willing to pay $100 for the bag. The poor consumer has little cash and many needs—housing, clothing, and so forth—so he cannot afford to give up so much cash for the bag: He is willing to pay only $50 for the bag of groceries. Then, maximizing consumer surplus means giving the bag to the rich consumer. This is not very equitable. Put it this way. Many and perhaps even most people would say that taking $1000 from a rich consumer to distribute equally among 100 poor consumers increases equity, because $10 is worth more to the poor than it is to the rich. But, *by definition* in the measure of efficiency we are using (dollar-calibrated utility), $10 is worth the same "utility" to each consumer.

4. As a final demonstration that efficiency misses a lot that goes into judgments of equity, consider a monopoly firm interested, so it says, in maximizing total surplus: "It is true," it begins, "that when I set marginal revenue equal to marginal cost, I am underproducing relative to the amount required by total surplus maximization. This makes me feel rather bad. So permit me, if possible, to engage in first-degree price discrimination. In return for this right, I promise to produce at the quantity that maximizes total surplus." Our friendly and generous monoply firm can be counted on to do just what it says, if it is allowed to engage in first-degree price discrimination. But it does this *because it is going to take every bit of the surplus for itself*. It goes to each of the little stick consumers, perched on the demand curve of Figure 12.8, and offers to sell to them their unit of the good for precisely their reservation price, up to the last consumer from whom it can make money on such an arrangement, which is the last consumer whose value for the good exceeds its marginal cost. Very efficient, to be sure, but very inequitable, by almost any standard, and certainly far from generous.

Economists are not very happy when asked to evaluate the equity of a given institutional arrangement on the basis of economic principles. They may have their own notions of equity, but they tend to leave to philosophers the question of how to measure equity formally. Philosophers are ready to take up this challenge; John Rawls and Robert Nozick, two modern philosophers, staked out strong and fairly opposing positions on what constitutes equity. But this is economics, not philosophy, so in this book, we use total surplus as a formal measure of the efficiency of institutional arrangements, leaving trade-offs between efficiency and equity to informal judgment.

Do not misunderstand. Just because something is formalized does not make it important, nor those things that lack formality unimportant. We separate equity and efficiency, and, using total surplus, we speak formally about efficiency. But do not feel in the least bit restrained from sacrificing a bit or even a lot of efficiency to serve equity. What we do in this chapter is provide the tools that let us measure how much efficiency is being sacrificed.

12.6. Other Aspects of Welfare and Efficiency

Since we began this chapter speaking vaguely of the triumph of capitalism over state socialism, I should say a few more things regarding this topic.

This chapter concerns "static efficiency," the ability of of the price system to achieve an efficient level of production and distribution of goods and services in a "settled" economy. The main result is that if firms and consumers respond to prices by maximizing profit and utility—and if firms and con-

sumers are price takers and there are no externalities—then prices direct the economy to efficient production and distribution. This is not to say that a commissariat could not do just as well. But it is easy to believe that by decentralizing decisions on matters such as the number and styles of shoes to have available in shops in Minsk, markets (and profit and utility-maximization motives) outperform central planning.

But the story we've told in this chapter is inadequate for some other aspects of "efficient" economies. For instance, a large portion of what makes for material wealth is innovation and the creation of new products. A quite different story about incentives for innovation would have to be spun here. In particular, the story about static efficiency told in this chapter seems to imply that governments, to improve efficiency, should increase the level of competition. Antitrust activity seems to be clearly indicated. And things that stifle competition by protecting intellectual property, such as patents, would seem to be a bad idea. But innovation is spurred when innovators have the prospect of enjoying the fruits of their innovative activities. Some political economists would argue that the triumph of capitalism was not because the market-driven economies of the West resulted in a better assortment of shoes on the shelves of shoe stores in Los Angeles and Zurich than in Minsk. Instead, the triumph followed because capitalism and private enterprise (with protections built in for private property, including intellectual property) provide innovators the incentive to innovate and investors the incentive to seek out and finance worthwhile innovations.

We lack the tools needed to flesh out this sort of story at this point. We will be in reasonable shape to do so nearer the end of the book, although we will not do it. But it is worth noting that seeming implications of the story told here—about antitrust activity, patents, and the like—*may* be muted and even reversed when the process of innovation is added to the mix.

Another important aspect of economic efficiency concerns the ability of individuals within the society to transact with one another in a relatively low-cost manner. We take up this topic right at the end of the book, in Chapter 24. For now, it is probably useful to say that transactional efficiency may not be maximized by "Darwinian" economies, in which each firm and consumer is concerned solely and strictly with his, her, or its own welfare, so much so that an ends-justify-the-means mentality rules.

Finally, efficiency depends on what individuals value. We have used models of firm and consumer behavior in which self-interest is paramount. But, in real life, individuals have a positive taste for things like equity and providing for those less fortunate. In formal economic terms, when consumers have these sorts of tastes, we say that there are externalities in consumption. We deal with externalities in Chapter 14 and see there that they

can cause real problems for the rosy picture this chapter paints. For now, record that the clear division between equity and efficiency relied on in this chapter is not entirely realistic.

Executive Summary

- In a perfectly competitive market, producer surplus (the area between the price per unit and the industry supply curve out to the level of production) gives the sum of profits of suppliers of the good, gross of any unavaoidable fixed costs.
- Consumer surplus (the area between the demand curve and the price per unit out to the level of consumption) gives a dollar-valued measure of the benefits consumers take from purchasing and consuming the good. This measure is exact in cases of reservation-demand utility and money-left-over utility, where the money left over enters as $\ldots + m$. Otherwise, it is an approximate measure.
- In a perfectly competitive market, the market equilibrium maximizes the total surplus generated in the production, exchange, and consumption of the good in question.
- The good is produced in an overall cost-minimizing manner by the industry, with the marginal cost of the last unit produced equal to the equilibrium price.
- The good is consumed in an overall sum-of-benefits-maximizing manner by consumers, with the marginal utility of the last unit consumed equal to the equilibrium price.
- The cost of the marginal unit produced matches the utility its consumption engenders, because each is equal to the price: The right amount is being produced.
- The transfer of money from consumers to producers is a net wash in terms of total surplus.
- In Economese, this is phrased as, A competitive market equilibrium is efficient. This is how *the invisible hand* works. However, keep in mind the following three points.
- Efficiency is not equity. Economists are usually informal about measuring equity, but this does not mean that it is wrong to sacrifice some efficiency to achieve a more equitable outcome.
- A firm that faces downward sloping demand and sets its price rather than taking price as given, sets its marginal cost equal to its marginal revenue: Since marginal revenue is less than price (for downward sloping demand), which equals marginal utility, this means that too little is produced to achieve full efficiency.
- Externalities complicate the story and can affect the basic conclusion (see Chap-

ter 14), as can considerations of dynamic efficiency and the process of innovation (Chapter 23), transactional efficiency (Chapter 24), and consumer tastes for equity (Chapter 25).

Problems

12.1 Suppose the supply of a particular good is given by $S(p) = 1000(p-4)$, while demand is given by $D(p) = 3000(20-p)$. What is consumer surplus at the equilibrium of this market? What is producer surplus?

12.2 Problem 11.9 described a perfectly competitive industry with four firms having the total cost function $TC(x) = 50+x+0.04x^2$ and an unlimited supply of firms having the total cost function $TC(x) = 100+3x+0.04x^2$. In all cases, fixed costs are avoided by not producing. Suppose demand is given by $D(p) = 200(10-p)$. What are the consumer and producer surpluses at the competitive market equilibrium of this industry? How does producer surplus relate to the profit levels of the firms?

12.3 Refer to Problem 11.7. An industry has free entry and exit for an unlimited number of firms, each having total cost function $TC(x) = 100+3x+0.04x^2$. The industry demand is initially given by $D(p) = 200(10-p)$, and the long-run equilibrium has price \$7, total quantity 600, 12 active firms each producing 50, with \$0 profit per firm. In the short run, firms cannot change their production quantities. In the intermediate run, the 12 active firms can change their production quantities. In the long run, firms can enter or leave. In Problem 11.7, we found the short-, intermediate-, and long-run equilibria if demand shifts to $200(12-p)$. Compute the consumer and producer surpluses for the original equilibrium and then for the short-, intermediate-, and long-run equilibria following the shift in demand. Warning: To make meaningful comparisons with the status-quo situation, when you compute the status-quo level of producer surplus, do it *three times*. (Making sense of that last sentence is, more or less, the whole point of this problem.)

12.4 Figure 12.10 depicts the average and marginal cost functions for a firm with a fixed cost and rising marginal cost. It also gives a demand function and marginal revenue function for the firm, assuming this firm serves the market by itself and (therefore) has market power. Find the regions whose areas are the firm's profit gross of its fixed cost and net of its fixed cost, when it chooses price and quantity to maximize its profit. (Refer to Figure 12.9 if you need a hint for part of this.)

12.5 Imagine a monopoly whose marginal cost function is $MC(x) = 4 + x/1000$, facing a demand function given by $D(p) = 3000(20-p)$. What producer and consumer surpluses result if the monopoly maximizes its profit? Compare your answers with the answers to Problem 12.1.

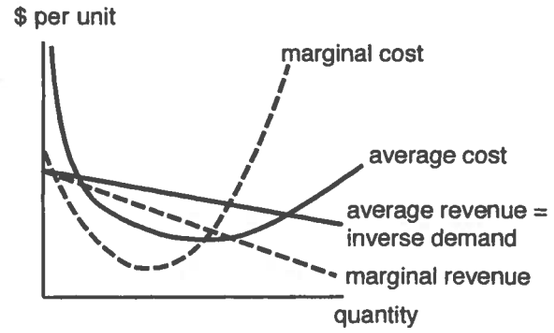


Figure 12.10. Problem 12.4: The situation facing a firm with market power.

12.6 Suppose the demand for pairs of shoes in the People's Republic of Slynavia is given by $D(p) = 250{,}000(90-p)$. (To keep the discussion simple, assume that each consumer wishes to buy at most a single pair of shoes.) The Commissar for Shoes can produce 10 million pairs of shoes to sell. A price of \$50 per pair would lead to demand for 10 million pairs. But the commissar decides that a price of \$10 per pair would be more in keeping with the government's ideology. At this price, 20 million pairs are demanded; hence, the shoes must be rationed.

(a) Suppose, in the rationing scheme employed, each of the 20 million consumers who wish to purchase a pair of shoes at \$10 per pair has an equal chance of getting 1 of the 10 million pairs available. That is, each has a 0.5 of getting a pair, *which he or she then wears*. What measure of consumer surplus is appropriate for this outcome?

(b) To explain the italicized *which he or she then wears*, the rationing outcome described has some consumers who value a pair of shoes at \$80 without shoes, while others who value a pair at \$20 wearing a pair. It seems likely in such cases that a black market in shoes will spring up. Suppose this is a well-functioning black market, where an equilibrium black-market price emerges and trades are made at that price. What is an appropriate measure of consumer surplus generated by the sale of shoes (by the state, for \$10 per pair) for the eventual outcome?

12.7 Imagine a market for a good in which demand is given by $D(p) = 10{,}000(10-p)$. Twenty-five identical firms supply this good, each of which has the total cost function $TC(x) = 4x + x^2/200$. All these firms are competitive; they act as if they have no impact on the prices they face.