4

Efficiency and the Idealized Competitive Model

Collective action enables society to produce, distribute, and consume a great variety and abundance of goods. Most collective action arises from voluntary agreements among people—within families, private organizations, and exchange relationships. The policy analyst, however, deals primarily with collective action involving the legitimate coercive powers of government: public policy encourages, discourages, prohibits, or prescribes private actions. Beginning with the premise that individuals generally act in their own best interest, or at least what they perceive as their self-interest, policy analysts bear the burden of providing rationales for any governmental interference with private choice. The burden applies to the evaluation of existing policies as well as new initiatives. It is an essential element, if not the first step, in any analysis, and will often provide the best initial insight into complex situations.

Our approach to classifying rationales for public policy begins with the concept of a perfectly competitive economy. One of the fundamental bodies of theory in modern economics deals with the properties of idealized economies involving large numbers of profit-maximizing firms and utility-maximizing consumers. Under certain assumptions, the self-motivated behaviors of these economic actors lead to patterns of production and consumption that are efficient in the special sense that it would not be possible to change the patterns in such a way so as to make some person better off without making some other person worse off.

Economists recognize several commonly occurring circumstances of private choice, referred to as *market failures,* that violate the basic assumptions of the

idealized competitive economy and, therefore, interfere with efficiency in production or consumption. The traditional market failures, which we discuss in Chapter 5, provide widely accepted rationales for such public policies as the provision of goods and the regulation of markets by government agencies. Economists, until recently, have paid less attention, however, to the plausibility of some of the more fundamental assumptions about the behavior of consumers. For example, economic models usually treat the preferences of consumers as unchanging. Is this reasonable? Do consumers always make the right calculations when faced with decisions involving such complexities as risk? Negative answers to these questions, which we consider in Chapter 6, may also provide rationales for public policies.

Of course, efficiency is not the only social value. Human dignity, distributional equity, economic opportunity, and political participation are values that deserve consideration along with efficiency. On occasion, public decision makers or ourselves as members of society, may wish to give up some economic efficiency to protect human life, make the final distribution of goods more equitable, or promote fairness in the distribution process. As analysts, we have a responsibility to address these multiple values and the potential conflicts among them. We discuss distributional and other values as rationales for public policies in Chapter 7.

The Efficiency Benchmark: The Competitive Economy

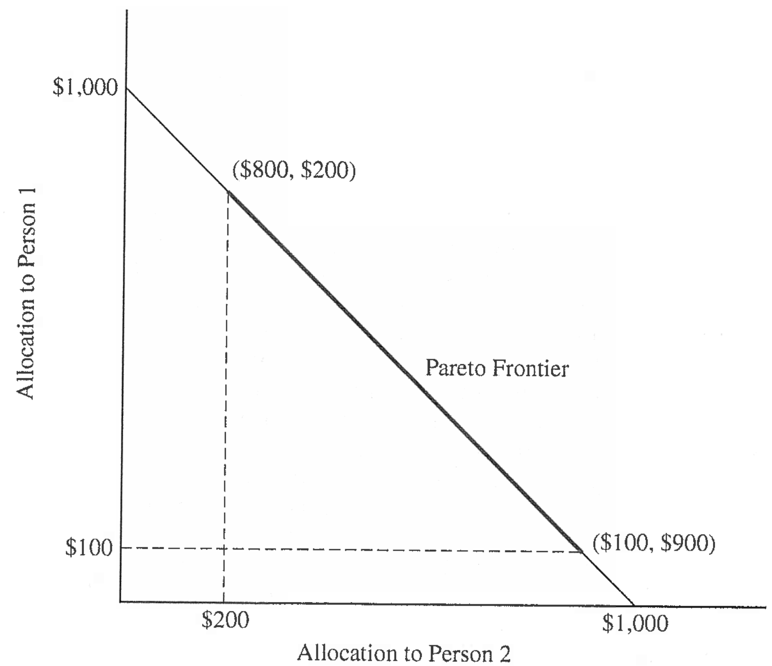
Imagine a world where each person derives *utility* (one's perception of one's own well-being) from personally consuming various quantities of all possible goods, including things, services, and leisure. We can think of each person as having a utility function that converts the list of quantities of the goods consumed into a utility index such that higher numbers imply greater well-being. We make several basic assumptions. First, other things equal, the more of any good a person has, the greater that person's utility. (We can incorporate unpleasant things such as pollution within this framework by thinking of them as "goods" that decrease utility or result in *disutility.*) And, second, additional units of the same good give ever-smaller increases in utility; in other words, they result in *declining marginal utility.*

Now make the following assumptions about the production of goods: Firms attempt to maximize profits by buying factor inputs (such as labor, land, capital, and materials) to produce goods for sale. The technology available to firms for converting factor inputs to final goods is such that, at best, an additional unit of output would require at least as many units of inputs to produce as the preceding unit; in other words, it becomes more costly in terms of resources to produce each additional unit of the good. Firms behave competitively in the sense that they believe that they cannot change the prices of factor inputs or products by their individual actions.

Each person has a budget derived from selling labor and his or her initial endowments of the other factor inputs, such as capital and land. People maximize their well-being by using their incomes to purchase the combinations of goods that give them the greatest utility.

In this simple world, a set of prices arises that distributes factor inputs to firms and goods to persons in such a way that it would not be possible for anyone to find a reallocation that would make at least one person better off without making at least one person worse off.[1] Economists refer to such a distribution as being *Pareto efficient*. It is a concept with great intuitive appeal: wouldn't, and shouldn't, we be dissatisfied with an existing distribution if we could find an alternative distribution that would make someone better off without making anyone else worse off? Although we would need other criteria for choosing between two distributions that were each Pareto efficient, we should, unless we are malevolent, always want to make Pareto-improving moves from inefficient distributions to efficient ones.

Figure 4.1 illustrates the concept of Pareto efficiency involving the allocation of $1,000 between two people. Assume that the two people will receive any mutually agreed-upon amounts of money that sum to no more than $1,000. The vertical axis represents the allocation to *person 1* and the horizontal axis represents the allocation to *person 2*. An allocation of all of the money to *person 1* would appear as the point on the vertical axis at $1,000; an allocation of all of the money to *person 2* would appear as the point on the horizontal axis at $1,000. The line connecting these two points, which we call the *potential Pareto frontier*, represents all the possible allocations to the two



Status quo: Point ($100, $200)
Potential Pareto Frontier: Line segment connecting ($1,000, $0) and ($0, $1,000)
Pareto Frontier: Line segment connecting ($800, $200) and ($100, $900)

Figure 4.1 Pareto and Potential Pareto Efficiency

[1]For the history of general equilibrium theory, see E. Roy Weintraub, "On the Existence of a Competitive Equilibrium: 1930–1954," *Journal of Economic Literature* 21(1) 1983, 1–39.

persons that sum exactly to $1,000. Any point on this line or inside the triangle it forms with the axes would be a technically feasible allocation because it gives shares that sum to no more than $1,000.

The potential Pareto frontier indicates all the points that fully allocate the $1,000. Any allocation that does not use up the entire $1,000 cannot be Pareto efficient, because it would be possible to make one of the persons better off by giving her the remaining amount without making the other person worse off. The actual Pareto frontier depends on the allocations that the two people receive if they reach no agreement. If they each receive nothing in the event they reach no agreement, then the potential Pareto frontier is the *actual Pareto frontier* in that any point on it would make at least one of the persons better off without making the other person worse off.

Now imagine that, if these two people reach no agreement about an allocation, *person 1* receives $100 and *person 2* receives $200. This point ($100, $200) can be thought of as the *status quo point*—it indicates how much each person gets in the absence of an agreement. The introduction of the status quo point reduces the Pareto frontier to the line segment between ($100, $900) and ($800, $200). Only moves to this segment of the potential Pareto frontier actually guarantee that each person is no worse off than under the status quo.

Note that whether a particular point on the potential Pareto frontier is actually Pareto efficient depends on the default allocations that comprise the status quo. More generally, Pareto efficiency in an economy depends on the initial endowments of resources to individuals.

The idealized competitive economy is an example of a *general equilibrium model*—it finds the prices of factor inputs and goods that clear all markets in the sense that the quantity demanded exactly equals the quantity supplied. Although general equilibrium models can sometimes be usefully applied to policy problems, limitations in data and problems of tractability usually lead policy analysts to evaluate policies in one market at a time.[2] Fortunately, a well-developed body of theory enables us to assess economic efficiency in the context of a single market.

Market Efficiency: The Meaning of Social Surplus

We need a yardstick for measuring changes in efficiency. Social surplus, which measures the net benefits consumers and producers receive from participation in markets, serves as an appropriate yardstick. In the context of the ideally competitive economy, the Pareto-efficient allocation of goods also maximizes social surplus. When we look across markets, the set of prices and quantities that give the greatest social surplus is usually Pareto efficient. Moreover, differences in social surplus between alternative market allocations approximate the corresponding sum of differences in individual welfares. As *social surplus is the sum of consumer surplus and producer surplus*, we consider each in turn.

[2]For a review of the use of general equilibrium models in education, see Thomas J. Nechyba, "What Can Be (and What Has Been) Learned from General Equilibrium Simulation Models of School Finance?" *National Tax Journal* 54(2) 2003, 387–414.

Consumer Surplus: Demand Schedules as Marginal Valuations

Imagine that you have the last five tickets to the World Cup Soccer final match. You walk into a room and announce to those present that you own all the remaining tickets to the event and that you will sell these tickets in the following manner: Starting with a high stated price, say, $500, you will keep lowering it until someone agrees to purchase a ticket. You will continue lowering the stated price until all five tickets are claimed. (An auction such as this with declining prices is known as a *Dutch auction;* in contrast, an auction with ascending prices is known as an *English auction.*) Each person in the room decides the maximum amount that he or she is willing to pay for a ticket. If this maximum amount must be paid, then each person will be indifferent between buying and not buying the ticket. Figure 4.2 displays the valuations for the persons in descending order from left to right. Although the stated prices begin at $500, the first acceptance is at $200. The purchaser obviously values this ticket at least at $200. Value in this context means the maximum amount the person is willing to pay, given his or her

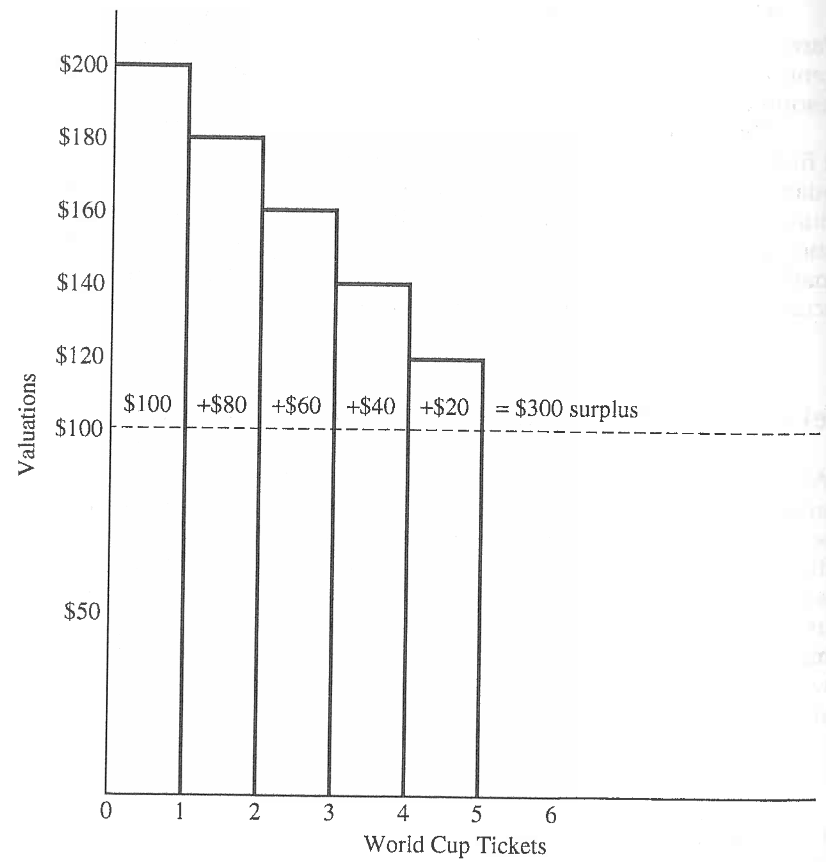


Figure 4.2 Consumer Values and Surpluses

budget and other consumption opportunities. You now continue offering successively lower stated prices until you sell a second ticket, which a second person accepts at $180, the second-highest valuation. Repeating this process, you sell the remaining three tickets at $160, $140, and $120, respectively.

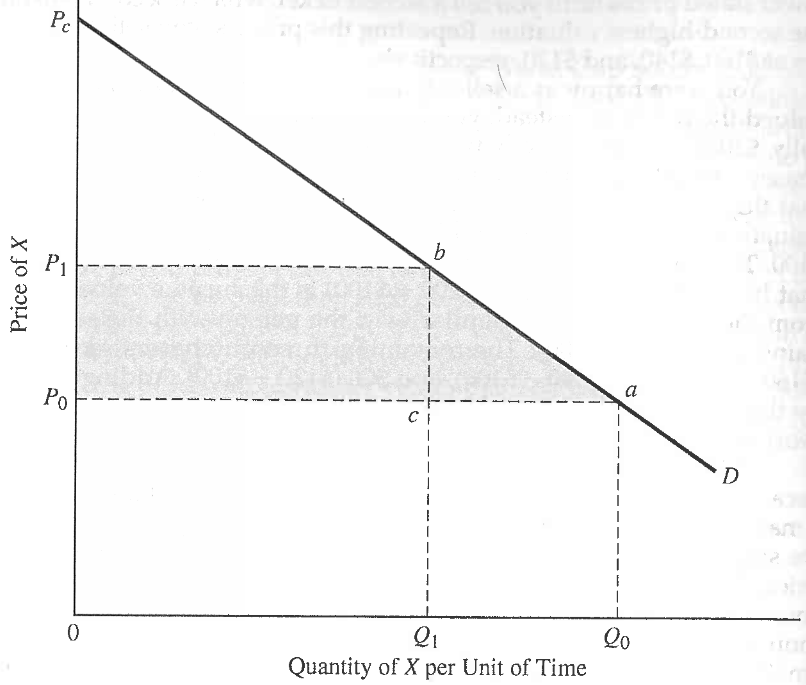
You were happy as a seller to get each person to pay the amount that he or she valued the ticket. If, instead, you had announced prices until you found one (specifically, $100) such that exactly five people wanted to buy tickets, then some of the purchasers would get tickets at a price substantially lower than the maximum amounts that they would have been willing to pay. For example, the person with the highest valuation would have been willing to pay $200 but only has to pay your set price of $100. The difference between the person's dollar valuation of the ticket and the price that he or she actually pays ($200 – $100) is the surplus value that the person gains from the transaction. In a similar way, the person with the second-highest surplus gains $80 ($180 – $100). The remaining three purchasers receive surpluses of $60 ($160 – $100), $40 ($140 – $100), and $20 ($120 – $100). Adding the surpluses realized by these five purchasers yields a measure of the *consumer surplus* in this market for World Cup tickets of $300.

The staircase in Figure 4.2 is sometimes called a *marginal valuation schedule* because it indicates how much successive units of a good are valued by consumers in a market. If, instead of seeing how much would be bid for successive units of the good, we stated various prices and observed how many units would be purchased at each price, then we would obtain the same staircase but recognize it as a *demand schedule.* Of course, we would also get a demand schedule by allowing individuals to purchase more than one unit at the stated prices. If we had been able to measure our good in small enough divisible units, or if demanded quantities were very large, then the staircase would smooth out to become a curve.

How do we move from this conceptualization of consumer surplus to its measurement in actual markets? We make use of demand schedules, which can be estimated from observing market behavior.

Line *D* in Figure 4.3 represents a person's demand schedule for some good, *X*. (Later we will interpret the curve as the demand schedule for all persons with access to the market.) Note that this consumer values all units less than *choke price,* P_c, the price that "chokes off" demand. The horizontal line drawn at P_0 indicates that she can purchase as many units of the good as she wishes at the constant price P_0. At price P_0 she purchases a quantity Q_0. Suppose, however, she purchased less than Q_0; then she would find that she could make herself better off by purchasing a little more because she would value additional consumption more than its cost to her. (The demand schedule lies above price for quantities less than Q_0.) Suppose, on the other hand, she purchased more than Q_0, then she would find that she could make herself better off by purchasing a little less because she would value the savings more than the lost consumption. At given price P_0, the quantity Q_0 is an *equilibrium* because the consumer does not want to move to an alternative quantity. The area of the triangle under the demand schedule but above the price line, $P_c a P_0$, represents her consumer surplus from purchasing Q_0 units of the good at price P_0.

Changes in consumer surplus are often the basis for measuring the relative efficiencies of alternative policies. For example, how does consumer surplus change in Figure 4.3 if a government policy increases price from P_0 to P_1? The new consumer surplus is the area of triangle $P_c b P_1$, which is smaller than the area of triangle $P_c a P_0$ by the area of the trapezoid inscribed by $P_1 b a P_0$ (the shaded region). We interpret the area of the rectangle $P_1 b c P_0$ as the additional amount the consumer must pay for the units



Loss in consumer surplus: P_1baP_0
Revenue collected by government: P_1bcP_0
Deadweight loss: abc

Figure 4.3 Changes in Consumer Surplus

of the good that she continues to purchase and the area of the triangle *abc* as the surplus the consumer gives up by reducing consumption from Q_0 to Q_1.

As an example of a government policy that raises price, imagine the imposition of an excise (commodity) tax on each unit of the good in the amount of the difference between P_1 and P_0. Then the area of rectangle P_1bcP_0 corresponds to the revenue raised by the tax, which, conceivably, could be rebated to the consumer to offset exactly that part of the consumer surplus lost. The consumer would still suffer the loss of the area of triangle *abc*. Because there are no revenues or benefits to offset this reduction in consumer surplus, economists define this loss in surplus due to a reduction in consumption as the *deadweight loss* of the tax. The deadweight loss indicates that the equilibrium price and quantity under the tax are not Pareto efficient—if it were possible, the consumer would be better off simply giving the tax collector a lump-sum payment equal to the area of P_1baP_0 in return for removal of the excise tax and its associated deadweight loss.

The loss of consumer surplus shown in Figure 4.3 approximates the most commonly used theoretical measure of changes in individual welfare: *compensating variation*. The compensating variation of a price change is the amount by which the consumer's budget would have to be changed so that he or she would have the same

utility after the price change as before. It thus serves as a dollar measure, or *money metric*, for changes in welfare. If the demand schedule represented in Figure 4.3 were derived by observing how the consumer varied purchases as a function of price, holding utility constant at its initial level (it thus would be what we call a *constant-utility demand schedule*), then the consumer surplus change would exactly equal the compensating variation.

Figure 4.4 illustrates how compensating variation can be interpreted as a money metric, or proxy, for utility. The vertical axis measures expenditures by a person on all goods other than good X; the horizontal axis measures how many units of X she consumes. Suppose initially that she has a budget, *B*, but that she is not allowed to purchase any units of X, say, because X is manufactured in another country and imports of it are prohibited. She will, therefore, spend all of *B* on other goods. The *indifference curve* I_0 indicates all the combinations of expenditures on other goods and units of X that would give her the same utility as spending *B* on other goods and consuming no units of X. Now imagine that the import ban is lifted so that she can purchase units of X at price P_x. She can now choose any point on the line connecting *B* with the point on the horizontal axis indicating how many units of X she could purchase if she spent zero on other goods, B/P_x. Once this new budget line is available, she will maximize her utility by selecting a point on the highest feasible indifference I_1, by purchasing x_1 units of X and spending her remaining budget on other goods. Once X is available, it would be possible to return her to her initial level of utility by reducing her initial budget by the distance between *B* and C on the

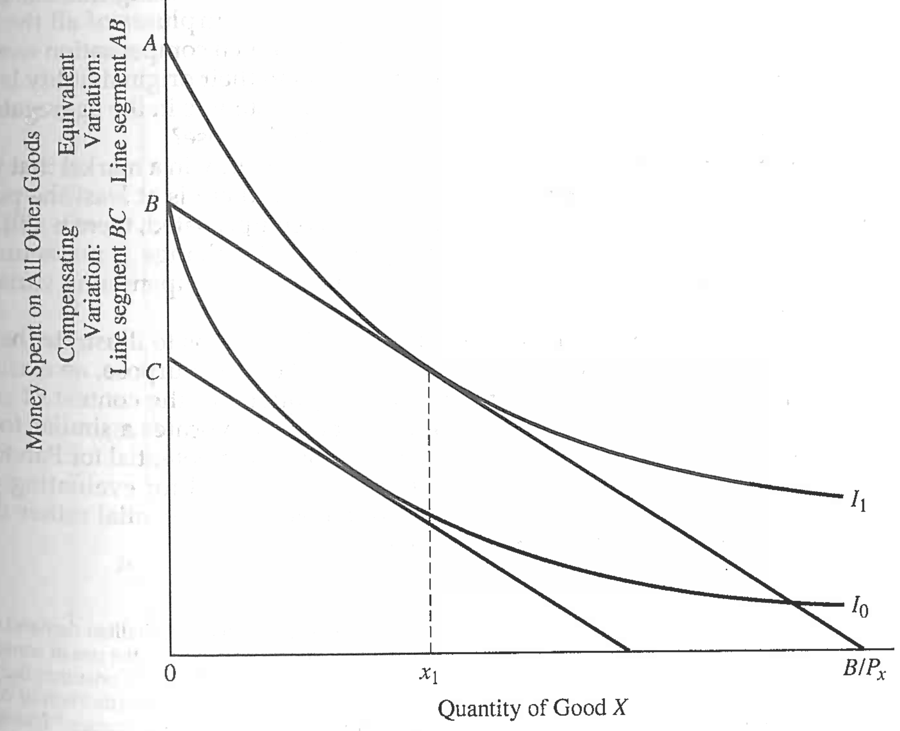


Figure 4.4 Money Metrics for Utility

vertical axis. This amount is her compensating variation associated with the availability of X at price P_x. It is a dollar value, or money metric, for how much value she places on being able to consume X.

Instead of asking how much money we could take away to make the person as well off after introduction of imports of X as before, we could ask how much money we would have to give her if X were not available so that she is as well off as she would be with imports of X. This amount, called *equivalent variation*, is shown as the distance between A and B on the vertical axis—if her budget were increased from B to A, then she could reach indifference curve I_1 without consuming X.

In practice, we usually work with empirically estimated demand schedules that hold constant consumers' *income* (rather than utility) and all other prices. This *constant income*, or *Marshallian, demand schedule* involves decreases in utility as price rises (and total consumption falls) and increases in utility as price falls (and total consumption rises). In comparison with a demand schedule holding utility constant at the initial level, the Marshallian demand schedule is lower for price increases and higher for price reductions. Fortunately, as long as either the price change is small or expenditures on the good make up a small part of the consumer's budget, the two schedules are close together and estimates of consumer surplus changes using the Marshallian demand schedule are close to the compensating variations.[3]

Now, moving from the individual to society, consider the relationship between price and the quantity demanded by all consumers. We derive this *market demand schedule* by summing the amounts demanded by each of the consumers at each price. Graphically, this is equivalent to adding horizontally the demand schedules for all the individual consumers. The consumer surplus we measure using this market demand schedule would just equal the sum of the consumer surpluses of all the individual consumers. It would answer these questions: How much compensation would have to be given in aggregate to restore all the consumers to their original utility levels after a price increase? How much could be taken from consumers in the aggregate to restore them all to their original utility levels after a price decrease?

Thus, if we can identify a change in price or quantity in a market that would produce a net positive increase in social surplus, then there is at least the potential for making a Pareto improvement. After everyone is compensated, there is still something left over to make someone better off. Of course, the change is not actually Pareto improving unless everyone is given at least his or her compensating variations from the surplus gain.

Our primary use of consumer surplus in Chapter 5 is to illustrate the inefficiencies associated with the various market failures. For this purpose, an exclusive focus on the potential for Pareto improvement is adequate. In the context of cost-benefit analysis, the Kaldor-Hicks compensation principle advocates a similar focus on net positive changes in social surplus as an indication of the potential for Pareto improvements. When we consider cost-benefit analysis as a tool for evaluating policies in Chapter 16, we discuss the implications of focusing on potential rather than actual improvements in the welfare of individuals.

[3]In any event, the consumer surplus change measured under the Marshallian demand curve will lie between compensating variation and equivalent variation. For a discussion of the use of consumer surplus changes as a measure of changes in individual welfare, see Robert D. Willig, "Consumer Surplus without Apology," *American Economic Review* 66(4) 1976, 589–97. For a more intuitive justification of consumer surplus, see Arnold C. Harberger, "Three Basic Postulates for Applied Welfare Economics," *Journal of Economic Literature* 9(3) 1971, 785–97.

Producer Surplus: Background on Pricing

In the ideal competitive model, it is standard to assume that marginal cost of production for each firm rises with increases in output beyond equilibrium levels. Because firms have some fixed costs that must be paid before any production can occur, the average cost of producing output first falls as the fixed costs are spread over a larger number of units, then rises as the increasing marginal cost begins to dominate, because the use of some input, such as labor, becomes less efficient. Consequently, some output level minimizes the average cost of the firm. The curve marked AC in Figure 4.5 represents a U-shaped *average cost* curve for the firm. The *marginal cost* curve is labeled MC. Note that the marginal cost curve crosses the average cost curve at the latter's lowest point. When marginal cost is lower than average cost, the latter must be falling. When marginal cost is higher than average cost, the latter must be rising. Only when marginal cost equals average cost does average cost remain unchanged. This is easily grasped by thinking about your average score for a series of examinations—only a new score above your current average can raise your average.

The total cost of producing some level of output, say, Q_S, can be calculated in one of two ways. First, because average cost is simply total cost divided by the quantity of output, multiplying average cost at this quantity (AC_S in Figure 4.5) by Q_S yields total cost as given by the area of rectangle AC_SbQ_S0. Second, recalling that marginal cost tells us how much it costs to increase output by one additional unit, we can approximate total cost by adding up the marginal costs of producing successive units from the first all the way up to Q_S. The smaller our units of measure, the

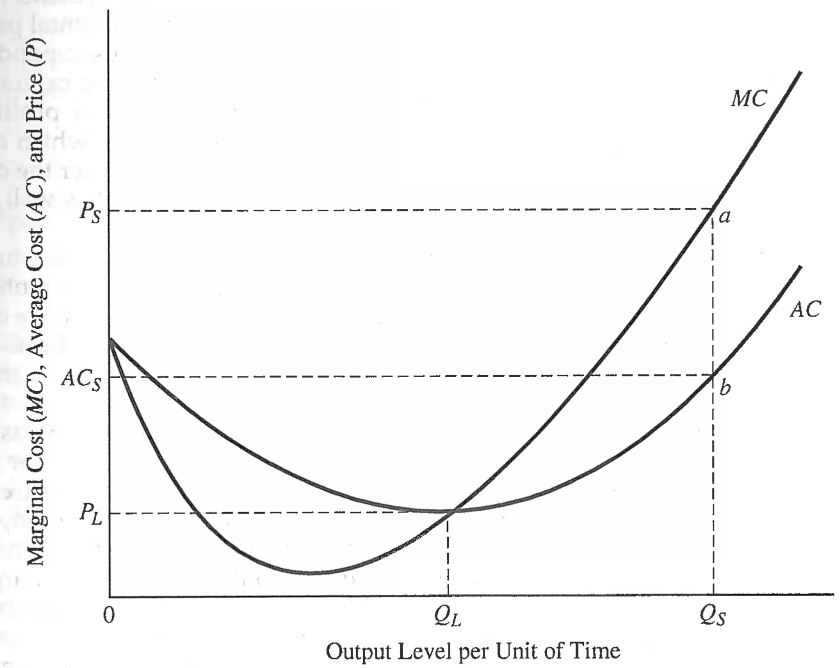


Figure 4.5 Average and Marginal Cost Curves

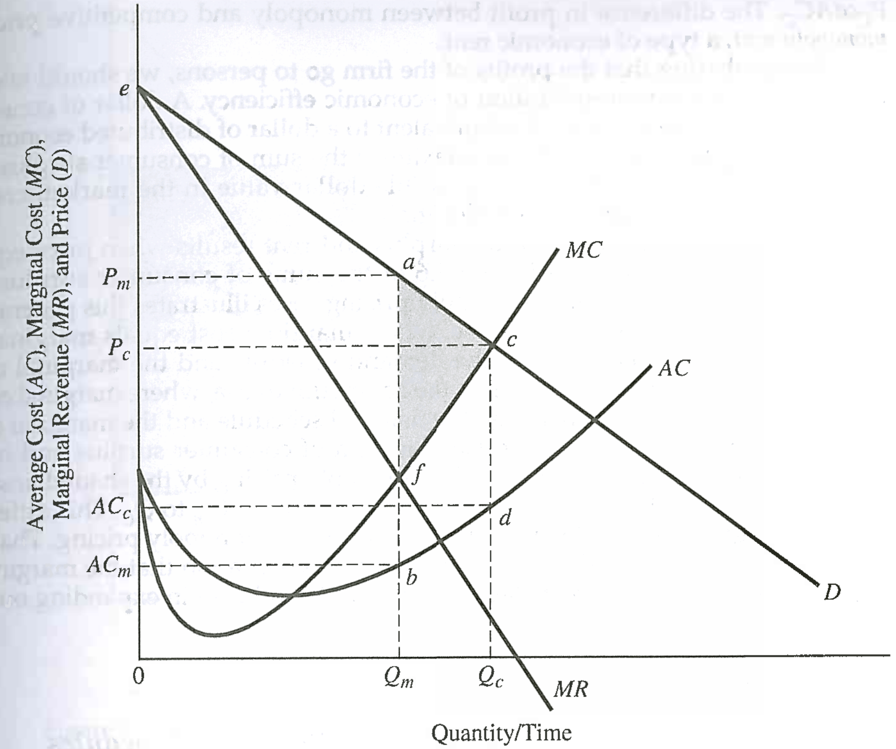
closer will be the sum of their associated marginal costs to the total cost of producing Q_S. In the limiting case of infinitesimally small units, the area under the marginal cost curve (*MC* in Figure 4.5) from zero to Q_S exactly equals total cost. (Those familiar with calculus will recognize marginal cost as the derivative of total cost, so that integrating marginal cost over the output range yields total cost; this integration is equivalent to taking the area under the marginal cost curve between zero and the output level.)

Now imagine that the market price for the good produced by the firm is P_S. The firm would maximize its profits by producing Q_S, the quantity at which marginal cost equals price. Because average cost is less than price at output level Q_S, the firm would enjoy a profit equal to the area of rectangle $P_S abAC_S$. Profit equals total revenue minus total cost. (Total revenue equals price, P_S, times quantity, Q_S, or the area of rectangle $P_S aQ_S 0$; the total cost of producing output level Q_S is the area of rectangle $AC_S bQ_S 0$.) In the competitive model, profit would be distributed to persons according to their initial endowments of ownership. But these shares of profits would signal to others that profits could be made simply by copying the technology and inputs used by the firm. As more firms enter the industry, however, total output of the good would rise and, therefore, price would fall. At the same time the new firms would bid up the price of inputs so that the entire marginal and average cost curves of all the firms would shift up. Eventually, price would fall to P_L, the level at which the new marginal cost equals the new average cost. At P_L profits fall to zero, thus removing any incentive to enter the industry.

With no constraints on the number of identical firms that can arise to produce each good, the Pareto-efficient equilibrium in the idealized competitive model is characterized by zero profits for all firms. (Note that we are referring to economic profits, not accounting profits. *Economic profit* is total revenue minus payments at competitive market prices to all factors of production, including an implicit rental price for capital owned by the firm. *Accounting profit* is simply revenue minus expenditures.) If the firm does not make an explicit payment to shareholders for the capital it uses, then accounting profits may be greater than zero even when economic profits are zero. To avoid confusion, economists refer to economic profits as *rents*, which are defined as any payments in excess of the minimum amounts needed to cover the cost of supply. Rents can occur in markets for inputs such as land and capital as well as in product markets.

In the real economic world, unlike our ideal competitive model, firms cannot be instantaneously replicated; at any time some firms may thus enjoy rents. These rents, however, attract new firms to the industry, so that over the long run we expect rents to disappear. Only if some circumstance prevents the entry of new firms will the rents persist. Therefore, we expect the dynamic process of profit seeking to move the economy toward the competitive ideal.

To understand better the concept of rent, it is useful to contrast pricing in a monopolistic industry with one that is competitive. To begin, consider the case of an industry with a single firm that does not have to worry about future competition. This monopoly firm sees the entire demand schedule for the good, labeled *D* in Figure 4.6. It also sees a *marginal revenue curve (MR)*, which indicates how much revenue increases for each additional unit offered to the market. The marginal revenue curve lies below the demand schedule because each additional unit offered lowers the equilibrium price not just for the last but for all units sold. For example, imagine that increasing supply from 9 to 10 units decreases the market price from $100/unit to



	Competitive Pricing	Monopoly Pricing
Consumer Surplus:	P_cce	P_mae
Total Revenue:	P_ccQ_c0	P_maQ_m0
Total Cost:	AC_cdQ_c0	AC_mbQ_m0
Rent:	P_ccdAC_c	P_mabAC_m
Deadweight Loss:	0	*acf*

Figure 4.6 Monopoly Pricing, Rents, and Deadweight Loss

$99/unit. Revenue increases by $90 (10 times $99 minus 9 times $100). The height of the marginal revenue curve above 10 units is thus $90, which is less than the height of the demand schedule, $99. As long as marginal revenue exceeds marginal cost (*MC*), profits can be increased by expanding output. The profit-maximizing level of output occurs when marginal cost equals marginal revenue (where *MC* intersects *MR*). In Figure 4.6, this output level for the monopoly firm, Q_m, results in a market price, P_m, and profits equal to the area of rectangle $P_m abAC_m$: total revenue (P_m times Q_m) minus total cost (AC_m times Q_m).

In contrast to the case of monopoly, consider the production decisions of one of the firms in a competitive industry. Because it provides a small part of the total industry supply, it ignores the effects of its supply on market price and, therefore, equates marginal cost with price (the intersection of *MC* and *D*), yielding price P_C and profits

$P_C cdAC_C$. The difference in profit between monopoly and competitive pricing is the *monopoly rent*, a type of economic rent.

Remembering that the profits of the firm go to persons, we should take account of these rents in our consideration of economic efficiency. A dollar of consumer surplus (compensating variation) is equivalent to a dollar of distributed economic profit. If we set the price and quantity to maximize the sum of consumer surplus and rent, then we will generate the largest possible dollar value in the market, creating the prerequisite for a Pareto-efficient allocation.

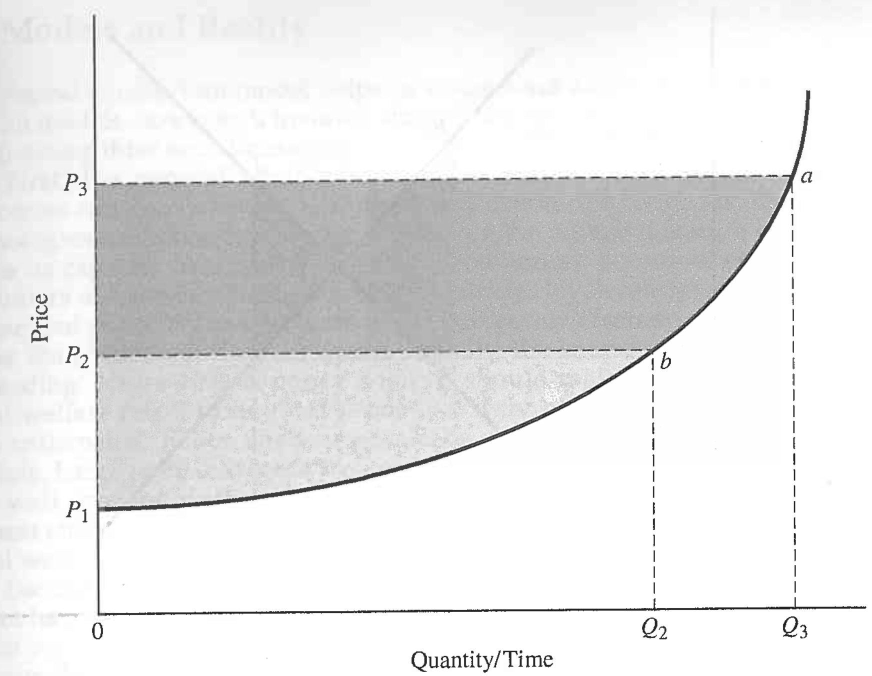
The largest sum of consumer surplus and rent results when price equals marginal cost. A comparison in Figure 4.6 of the sums of consumer surplus and rent between the competitive and monopoly pricing cases illustrates this general proposition. The sum in the monopoly case, where marginal cost equals marginal revenue $(MC = MR)$, is the area between the demand schedule and the marginal cost curve from quantity zero to Q_m. The sum in the competitive case, where marginal cost equals price $(MC = P)$, is the area between the demand schedule and the marginal cost curve from quantity zero to Q_C. Obviously, the sum of consumer surplus and rent under competitive pricing exceeds that under monopoly pricing by the shaded area between the demand schedule and the marginal cost curve from Q_m to Q_C. This difference, the area of triangle *acf*, is the deadweight loss caused by monopoly pricing. That this area is the deadweight loss follows directly from the observation that the marginal benefit *(D)* exceeds the marginal cost *(MC)* for each unit produced in expanding output from Q_m to Q_C.

Producer Surplus: Measurement with Supply Schedules

We usually deal with markets in which many firms offer supply. Therefore, we desire some way of conveniently summing the rents that accrue to all the firms supplying the market. Our approach parallels the one we used to estimate compensating variations. First, we introduce the concept of a supply schedule. Second, we show how the supply schedule can be used to measure the sum of rents to all firms supplying the market.

Imagine constructing a schedule indicating the number of units of a good that firms offer at each of various prices. Figure 4.7 shows the common case of firms facing increasing marginal costs. Firms offer a total quantity Q_2 at price P_2. As price increases, firms offer successively greater quantities, yielding an upward-sloping *supply schedule.* The schedule results from the horizontal summation of the marginal cost curves of the firms. (For example, refer *MC* in Figure 4.5.) Each point on the supply curve tells us how much it would cost to produce another unit of the good. If we add up these marginal amounts one unit at a time, beginning with quantity equal to zero and ending at the quantity supplied, then we arrive at the total cost of producing that quantity. Graphically, this total cost equals the area under the supply curve from quantity zero to the quantity supplied.

Suppose the market price is P_3 so that the quantity supplied is Q_3. Then the total cost of producing Q_3 is the area $P_1\, aQ_3 0$. The total revenue to the firms, however, equals price times quantity, given by the area of rectangle $P_3 aQ_3\, 0$. The difference between total revenue and total cost equals the total rent accruing to the firms. This difference, called *producer surplus,* is measured by the total shaded area in Figure 4.7 inscribed by $P_3 a\, P_1$.



Loss in producer surplus resulting from a fall in price from P_3 to P_2: P_3abP_2

Figure 4.7 A Supply Schedule and Producer Surplus

Producer surplus need not be divided equally among firms. Some firms may have unique advantages that allow them to produce at lower cost than other firms, even though all firms must sell at the same price. For instance, a fortunate farmer with very productive land may be able to realize a rent at the market price, while another farmer on marginal land just covers total cost. Because the quantity of very productive land is limited, both farmers face rising marginal costs that they equate with market price to determine output levels. The general point is that unique resources—such as especially productive land, exceptional athletic talent, easy-to-extract minerals—can earn rents even in competitive markets. Excess payments to such unique resources are usually referred to as *scarcity rents.* Unlike monopoly, *scarcity rents do not necessarily imply economic inefficiency.*

Changes in producer surplus represent changes in rents. For example, if we want to know the reduction in rents that would result from a fall in price from P_3 to P_2, then we compute the change in producer surplus in the market. In Figure 4.7 the reduction in rents equals the dark shaded area P_3abP_2, the reduction in producer surplus.

Social Surplus

We now have the basic tools for analyzing efficiency in specific markets. A necessary condition for Pareto efficiency is that it should not be possible to increase the sum of compensating variations and rents through any reallocation of factor inputs or final products. We have shown how changes in consumer surplus measure the sum of

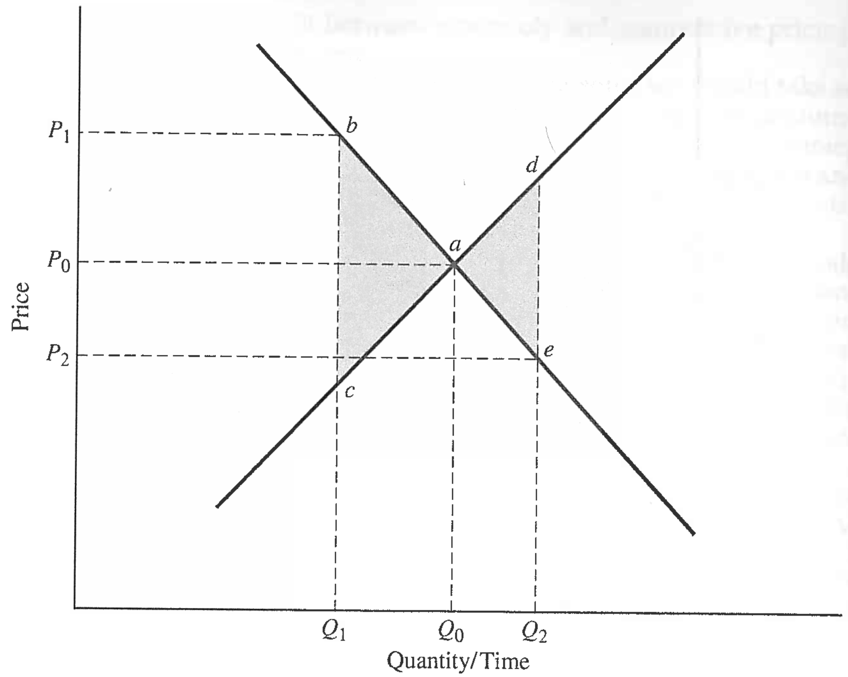


Figure 4.8 Inefficiencies Resulting from Deviations from the Competitive Equilibrium

compensating variations and how changes in producer surplus measure changes in rents. The sum of consumer and producer surpluses in all markets is defined as *social surplus*. Changes in social surplus, therefore, measure changes in the sums of compensating variations and rents. For evaluating the efficiency implications of small changes in the price and quantity of any one good, it is usually reasonable to limit analysis to changes in social surplus in its market alone.

Figure 4.8 reviews the inefficiencies associated with deviations from the equilibrium price and quantity in a competitive market in terms of losses of social surplus. The efficient competitive equilibrium occurs at price P_0 and quantity Q_0, the point of intersection of the supply *(S)* and demand *(D)* schedules. A policy that increases price to P_1 involves a loss of social surplus given by the area of triangle *abc*—each forgone unit between Q_1 and Q_0 yields marginal value (as given by the height of the demand schedule) in excess of marginal cost (as given by the height of the supply schedule). Consequently, social surplus can be increased by lowering price so that the quantity supplied and demanded moves closer to Q_0. A policy that decreases price to P_2 involves a loss of social surplus given by the area of triangle *ade*—each additional unit supplied and demanded between Q_0 and Q_2 yields marginal cost (as given by the height of the supply schedule) that is in excess of marginal benefit (as given by the height of the demand schedule). Consequently, raising price to move the quantity supplied and demanded closer to Q_0 increases social surplus.

Caveats: Models and Reality

The general equilibrium model helps us understand a complex world. As is the case with all models, however, it involves simplifications that may limit its usefulness. It is worth noting three such limitations.

First, the general equilibrium model is static rather than dynamic. Real economies constantly evolve with the introduction of new goods, improvement in technologies, and changes in consumer tastes. An amazing feature of the price system is its capacity for conveying information among decentralized producers and consumers about such changes, which Friedrich Hayek refers to as "the particulars of time and place."[4] The equilibria of the competitive framework give us snapshots rather than videos of the real world. Usually, the snapshot is helpful and not too misleading. Nevertheless, policy analysts should realize that substantial gains in social welfare result from innovations that were not, and perhaps could not, have been anticipated. Policy analysts should be careful not to take too static a view of markets. Large rents that seem well protected by barriers to entry, for example, may very well spur innovative substitutes. In discussing each of the market failures in the next chapter, we consider some of the market responses that may arise to reduce social welfare losses.

Second, the general equilibrium model can never be complete: modelers simply do not have enough information to incorporate all goods and services. If they could, then it would be unlikely they could solve the model for its equilibrium. Our switching from the general equilibrium model to models of individual markets is a purposeful restriction of the model so that it can be usefully applied. In most applications it is a reasonable approach, though sometimes goods are such strong complements or substitutes that it is not reasonable to look at them separately.[5]

Third, the assumptions of the general equilibrium model are often violated in the real world. In the two chapters that follow, we consider the most important of these violations of assumptions. We do so in the context of specific markets, acknowledging that doing so may not fully capture all their implications in the wider economy.[6] Nonetheless, we see this analysis as highly valuable in helping policy analysts get started in understanding the complexity of the world in which they work.

Conclusion

The idealized competitive economy provides a useful conceptual framework for thinking about efficiency. The tools of applied welfare economics, consumer and producer surplus, give us a way of investigating efficiency within specific markets. In the next chapter, we explicate four situations, the traditional market failures, in which equilibrium market behavior fails to maximize social surplus.

[4]F. A. Hayek, "The Use of Knowledge in Society," *American Economic Review* 35(4) 1945, 519–30 at 522.
[5]See Anthony E. Boardman, David H. Greenberg, Aidan R. Vining, and David L. Weimer, *Cost–Benefit Analysis: Concepts and Practice*, 3rd ed. (Upper Saddle River, NJ: Pearson Prentice Hall, 2006), Chapter 5.
[6]R. G. Lipsey and Kelvin Lancaster, "General Theory of the Second Best," *Review of Economic Studies* 24(1) 1956–1957, 11–32.

For Discussion

1. Assume that the world market for crude oil is competitive, with an upward-sloping supply schedule and a downward-sloping demand schedule. Draw a diagram that shows the equilibrium price and quantity. Now imagine that one of the major oil exporting countries undergoes a revolution that shuts down its oil fields. Draw a new supply schedule and show the loss in consumer surplus in the world oil market resulting from the loss of supply. What assumptions are you making about the demand for crude oil in your measurement of consumer surplus?
2. Now assume that the United States is a net importer of crude oil. Show the impact of the price increase resulting from the loss of supply to the world market on social surplus in the U.S. market.

5

Rationales for Public Policy

Market Failures

The idealized competitive model produces a Pareto-efficient allocation of goods. That is, the utility-maximizing behavior of persons and the profit-maximizing behavior of firms will, through the "invisible hand," distribute goods in such a way that no one could be better off without making anyone else worse off. Pareto efficiency thus arises through voluntary actions without any need for public policy. Economic reality, however, rarely corresponds perfectly to the assumptions of the idealized competitive model. In the following sections we discuss violations of the assumptions that underlie the competitive model. These violations constitute market failures, that is, situations in which decentralized behavior does not lead to Pareto efficiency. Traditional market failures are shown as circumstances in which social surplus is larger under some alternative allocation to that resulting under the market equilibrium. Public goods, externalities, natural monopolies, and information asymmetries are the four commonly recognized market failures. They provide the traditional economic rationales for public participation in private affairs.