# 3. The compensation principle and the welfare function

In Chapter 2, emphasis was given to the Pareto criterion as a means for selecting among alternative policies. Results show that many ‘first-best’ bundles or many Pareto-optimal points usually exist for an economy but, unfortunately, the Pareto principle does not give a basis for selecting among them. Narrowing the range of possibilities to a single first-best bundle (which essentially requires determining the ideal income distribution) requires a more complete criterion. One such criterion, which was introduced much later than the Pareto principle in the hope that it would be a more powerful device for choosing among policies, is the *compensation principle*, sometimes called the *Kaldor–Hicks compensation test* after the two economists to which it is attributed (Kaldor 1939; Hicks 1939). The development of the compensation principle was thus an attempt to broaden the states of the world that could be compared using an accepted welfare criterion. Simply stated, state $B$ is preferred to state $A$ if at least one individual could be made better off without making anyone worse off at state $B$ – not that all individuals are *actually* no worse off – by some feasible redistribution following the change. Unlike the Pareto principle, the compensation criterion does not require the actual payment of compensation.

The issue of compensation payments is at the heart of many policy discussions. Some argue that compensation should be paid in certain cases. According to Lester Thurow (1980, p. 208), ‘If we want a world with more rapid economic change, a good system of transitional aid to individuals that does not lock us into current actions or current institutions would be desirable.’ However, most policies that have been introduced have not entailed compensation. For example, bans on DDT and other pesticides have in many cases resulted in producer losses, but producers have not been compensated for their losses in revenue.

Although the compensation principle does, in fact, expand the set of comparable alternatives (at the expense of additional controversy), some states remain noncomparable. The latter part of this chapter considers the necessary features of a criterion that ranks all possible states of an economy. However, empirical possibilities for the resulting more general theoretical constructs appear bleak.

## 3.1 THE COMPENSATION PRINCIPLE

*According to the compensation principle, state B is preferred to state A if, in making the move from state A to state B, the gainers can compensate the losers such that at least one person is better off and no one is worse off.* Such states are sometimes called *potentially Pareto preferred* states. The principle is stated in terms of *potential* compensation rather than *actual* compensation because, according to those who developed the principle, the payment of compensation involves a value judgment. That is, to say that society should move to state *B* and compensate losers is a clearly subjective matter, just as recommendation for change on the basis of the Pareto criterion is a subjective matter. For example, if a Pareto improvement is undertaken, then, as demonstrated in Section 2.6, the possibilities that represent further Pareto improvements may be more restricted. Conceivably, the true optimum state of society may not be reachable by further applications of the criterion if the wrong initial Pareto improvement is undertaken. Similarly, to say that society should move to state *B* without compensating losers is also a subjective matter of perhaps a more serious nature. Thus, nonpayment of compensation also involves a value judgment. In terms of objective practice, one can only point out the potential superiority of some state *B* without actually making a recommendation that the move be made.

### The Pure Consumption Case[1]

Consider the application of the compensation principle to comparing different distributions of a given bundle of goods, again using the basic model of two goods and two individuals developed in Chapter 2. In Figure 3.1, point *a* is preferred to point *b* on the basis of the Pareto principle. But how does one compare point *b* with a point such as *c*, where *c* is not inside the lens-shaped area? The compensation principle offers one possibility. For example, suppose that one redistributes the bundle such that, instead of being at point *b*, individual *A* is at point *d* and individual *B* is at point *e*. Note that the welfare of each is unchanged. However, at these points there is an excess of $q_2$ equal to $q_2^3 - q_2^2$ and an excess of $q_1$ equal to $q_1^3 - q_1^2$. Now, if the move *actually* takes place to point *c*,

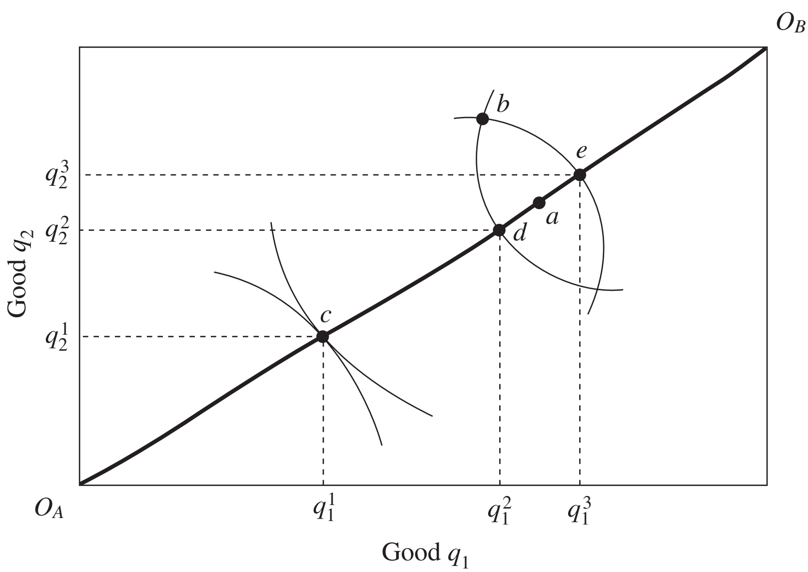


*Figure 3.1*

1. This section is largely based on Bailey (1954).

individual $A$ is clearly worse off while individual $B$ is better off. Individual $A$ loses $q_2^2 - q_2^1$ of $q_2$ and $q_1^2 - q_1^1$ of $q_1$, but individual $B$ gains $q_2^3 - q_2^1$ of $q_2$ and $q_1^3 - q_1^1$ of $q_1$. The amount individual $B$ gains in physical amounts of $q_1$ and $q_2$ is greater than the loss to individual $A$. Hence, point $c$ is potentially preferred to point $b$. *By the compensation principle, everyone is potentially better off by moving from point b to point c because the amount individual B gains is greater than the amount individual A loses.* This result holds even though compensation is not actually paid. If compensation is paid in terms of $q_1$ and $q_2$, both parties would, in effect, not agree to move to point $c$. Instead, a move would take place from point $b$ only to somewhere within the lens-shaped area. But points within the lens-shaped area are comparable with point $b$ by the Pareto principle. Thus, the application of the compensation criterion does not increase the ability to make statements about *actual* increases in welfare.

To view the problem in a different way, consider to what extent individual $B$ would have to bribe individual $A$ in order to make the move from point $b$ to point $c$. The minimum amount is $q_2^2 - q_2^1$ of $q_2$ and $q_1^2 - q_1^1$ of $q_1$. Hence, in equilibrium, one would move from point $b$ to point $d$ only if compensation were paid. Individual $B$ would gain $q_2^3 - q_2^2$ of $q_2$ and $q_1^3 - q_1^2$ of $q_1$ in the move if the minimum bribe is paid. Thus, point $c$ is never actually reached if compensation is paid.

## Distribution of Different Bundles

Consider next how the compensation principle can be used to compare different distributions of *different output bundles*. Recall from the preceding case that potential gains can be made in a move from point $b$ to point $c$ if, in the actual move to point $c$, the amount one individual loses is less than the amount the other individual gains. With this in mind, consider Figure 3.2 where the indifference curve $C$ corresponds to production at $O_B$ and to distribution at point $a$. Similarly, with production at $O_B^*$, the Scitovsky curve corresponding to distribution at point $b$ is $C^*$. At point $b$, one individual is worse off than at point $a$, and the other individual is better off. However, potential gains are possible in the move from point $a$ to point $b$ because the amount the loser loses is less than what the gainer gains. Potential gains are clear because production at $O_B^*$ can be distributed to keep welfare the same as at point $a$ by moving along the Scitovsky indifference curve $C$ to point $f$. By so doing, $fh$ of $q_2$ and $fg$ of $q_1$ are left over. Thus, if the compensation principle is used as a policy criterion, the move would be made (even though at point $b$ one of the individuals may be actually worse off than at point $a$).

At this point, a comparison and contrast can be drawn between the compensation principle and the Pareto principle. Using the compensation principle with initial production bundle at $O_B$ and distribution at point $a$, a move to the production bundle at $O_B^*$ is supported regardless of the way it is actually distributed. Using the Pareto principle, however, the move is supported only if the actual distribution corresponds to moving along the Scitovsky curve $C$ to point $f$, keeping the welfare of each individual constant and then dividing the excess of $fg$ of good $q_1$ and $fh$ of good $q_2$ among the two individuals in some way so that neither is worse off.

The reason that production at $O_B^*$ is preferred to production at $O_B$, in either case, is that the starting point with distribution at point $a$ is a second-best state. The corresponding Scitovsky curve $C$ is not tangent to the production possibility frontier $PP$. Like the

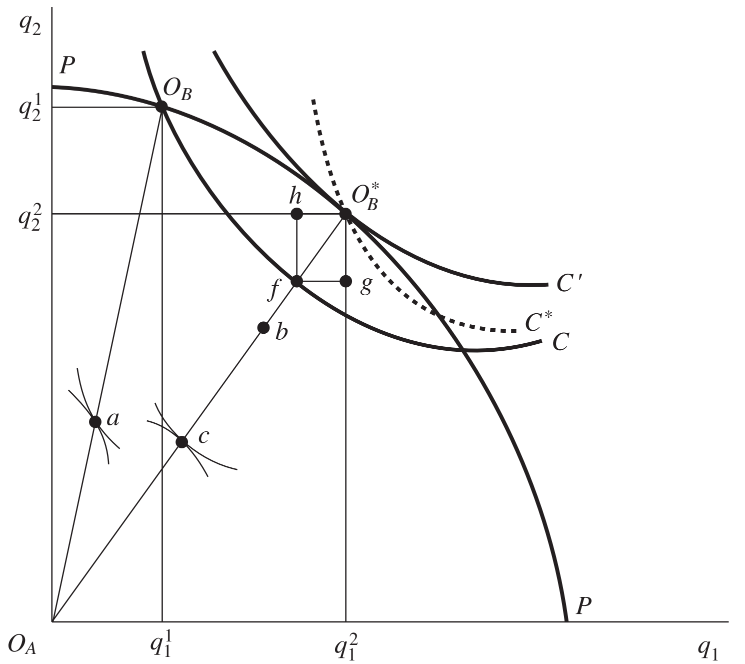


*Figure 3.2*

Pareto criterion, the compensation principle does not support a move away from a first-best state such as production at $O_B^*$ with distribution point *c* corresponding to Scitovsky indifference curve $C'$. Thus, the compensation criterion, like the Pareto criterion, cannot be used to rank two first-best states. A movement from one to the other would not be supported regardless of which is used as a starting point. The compensation criterion, on the other hand, gives a means of comparing all pairs of second-best states and for comparing all second-best states with all first-best states.

**The Reversal Paradox**

An important class of problems in applying the compensation principle falls under the general heading of the *reversal paradox* pointed out by Scitovsky (1941). For the case where gainers can potentially compensate losers, a conclusion that one position is better than another is not always warranted. *One must ask, also, whether the losers can bribe the gainers not to make the move.* The crux of the argument is presented in Figure 3.3. The production possibility curve is *PP*, and the two bundles to be compared are $O_B$ and $O_B^*$. Each of the bundles is distributed such that the corresponding Scitovsky indifference curves cross. In other words, both are second-best states because neither indifference curve is tangent to the production possibility curve. The curves $C_1$ and $C_2$ correspond to points *a* and *c* on the contract curves, respectively. Now, by the compensation principle, production at $O_B^*$ with distribution at point *c* is better than production at $O_B$ with distribution at point *a* because production at $O_B^*$ can generally be redistributed such that all are actually

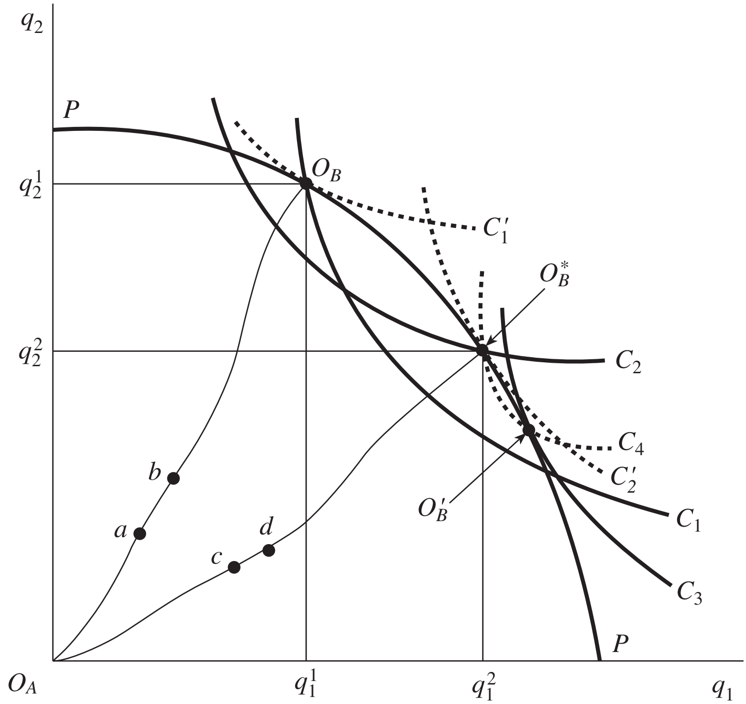


*Figure 3.3*

better off at point $d$ (where distribution at point $d$ corresponds to Scitovsky curve $C_2'$, which lies above curve $C_1$ and is associated with improved welfare for both individuals). However, by this criterion, $O_B^*$ is only potentially better off. Compensation is not actually paid. Because compensation is not actually paid, a reversal problem arises. That is, the new state with production at $O_B^*$ and distribution at point $c$ is a second-best state with Scitovsky curve $C_2$. Thus, according to Figure 3.3, there must be some distribution – say, at point $b$ – such that production at $O_B$ is preferred to production at $O_B^*$ by the Scitovsky criterion (where distribution at point $b$ corresponds to the Scitovsky curve $C_1'$, which is associated with improved welfare for both individuals as compared with $C_2$). Thus, each is preferred to the other.

This reversal occurs because in each case a given distribution of the first bundle is compared with all possible distributions of the alternative bundle. The reversal paradox suggests that all distributions of the initial bundle should also be considered. In other words, a *reversal test* (sometimes called the *Scitovsky reversal test*) is passed if one determines, first, that gainers can bribe losers to make a change and, second, that losers cannot bribe gainers not to make the change. Unless the reversal test is passed in addition to the Kaldor–Hicks compensation test, one cannot really say that one state is even potentially preferred to another.

Some additional points that must be borne in mind with respect to the Scitovsky reversal paradox are as follows:

1. *The reversal paradox occurs only in comparing two second-best bundles*. It does not arise if one of the bundles is a first-best or Pareto-efficient bundle. For example, in Figure 3.3, if production at $O_B^*$ with distribution corresponding to indifference curve $C_2$ is compared to production at $O_B'$ with distribution corresponding to indifference curve $C_3$, a reversal problem does not occur.
2. *The reversal paradox does not always occur in comparing two second-best bundles even though compensation is not actually paid*. For example, in Figure 3.3, production at $O_B^*$ with distribution corresponding to Scitovsky indifference curve $C_4$ does not lead to a paradox when compared to production at $O_B$ and distribution corresponding to Scitovsky curve $C_1$. The paradox occurs only when the relevant Scitovsky curves cross in the interior of the feasible production region. This problem may not occur when income distributions do not change substantially.

**Intransitive Rankings**

If the compensation criteria (both the direct Kaldor–Hicks and Scitovsky reversal tests) are employed to rank all possible states, a further problem can arise even if the reversal problem is not encountered. That is, compensation tests can lead to intransitive welfare rankings when more than two states are compared.[2] This problem arises when, for example, one must choose among, say, *states where all the alternative policies are of a second-best nature* (that is, there is no single policy in the policy set that leads to a bundle of goods distributed with the Scitovsky community indifference curve tangent to the production possibility curve). In Figure 3.4, given the production possibility curve *PP*, bundle $O_B^2$ is preferred to $O_B^1$, $O_B^3$ is preferred to $O_B^2$ and bundle $O_B^4$ is preferred to $O_B^3$, using the compensation test. However, $O_B^1$ is also preferred to $O_B^4$. Hence, the Kaldor–Hicks compensation test leads to welfare rankings that are intransitive. But note that some form of distortion exists for each bundle because the Scitovsky indifference curves are not tangent to the production possibility curve in any of the four cases. All the bundles are of a second-best nature.

Suppose, on the other hand, that one policy results in a bundle of goods that is economically efficient (with the Scitovsky indifference curve tangent to the production possibility curve). For example, consider bundles $O_B^1$, $O_B^2$, $O_B^3$ and $O_B^5$. Here, $O_B^5$ is clearly the optimum choice. There is no desire, once at $O_B^5$, to return to bundles $O_B^1$, $O_B^2$ or $O_B^3$. As a second example, suppose that the bundles to be compared are $O_B^1$, $O_B^2$, $O_B^3$ and $O_B^6$. Again, once at $O_B^6$, no potential gain is generated in returning to $O_B^1$, $O_B^2$ or $O_B^3$. Hence, no ambiguity is encountered in choosing a top-ranked policy if the policy set contains exactly one first-best state. Thus, as with the reversal problem discussed earlier, intransitivity occurs only when all the bundles being compared are generated from second-best policies.[3]

Consider, on the other hand, one further case where the possibilities consist of $O_B^1$, $O_B^5$ and $O_B^6$. In this case, the Kaldor–Hicks compensation test shows that $O_B^5$ is preferred to

2. The results in this section are due to Gorman (1955).
3. Partly in response to the problems associated with using the compensation principle as a basis for welfare comparison, Arrow (1951) developed the impossibility theorem, which proves that no reasonable rule exists for combining rankings of various states of society by individuals into a societal ranking. See the further discussion in Section 3.4.

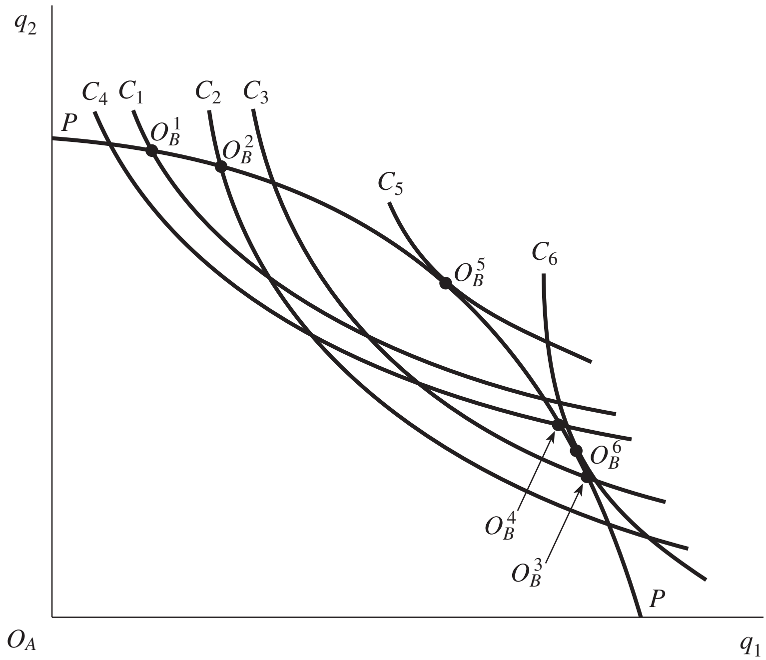


*Figure 3.4*

$O_B^1$. The possibility associated with $O_B^6$ is not preferred to $O_B^1$ even though $O_B^6$ is a first-best bundle. Among these three states, however, the rankings are not complete because, once at either $O_B^5$ or $O_B^6$, the compensation test does not suggest a move to either of the other states. In other words, the compensation test does not lead to a ranking of policy sets containing more than one first-best state.

## 3.2 UTILITY POSSIBILITY CURVES AND THE POTENTIAL WELFARE CRITERION

Another approach related to the choice of alternative income distributions and the reversal problem is based on the concept of utility possibility curves introduced by Samuelson (1947, 1956). To develop this approach, consider Figure 3.5 where the utilities of two individuals, $A$ and $B$, are represented. The utility of individual $B$ is measured on the vertical axis, while that for individual $A$ is measured along the horizontal axis. Three utility possibility curves are represented, each of which is derived by changing the distribution of a given bundle of goods along a contract curve. For example, $Q_2Q_2$ shows the maximum utility both individuals can receive from a fixed production at $O_B$ in Figure 3.3, $Q_1Q_1$ corresponds to a different bundle of goods, and so on.

To demonstrate the reversal paradox, consider $Q_2Q_2$ and $Q_1Q_1$. Points $a$ and $b$ represent particular distributions of the bundle from which $Q_2Q_2$ is derived. Similarly, points $c$ and $d$ represent particular distributions of the bundle from which $Q_1Q_1$ is derived.

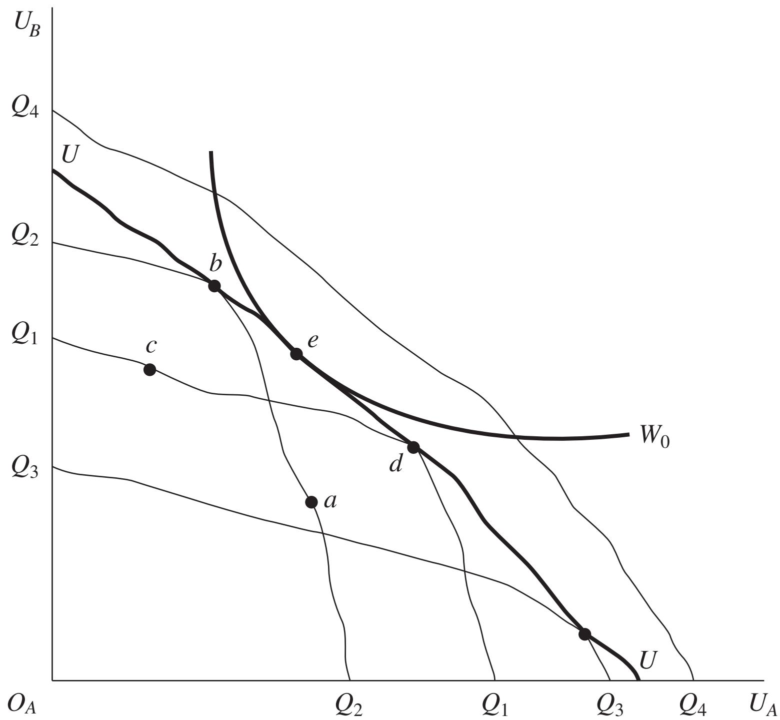


*Figure 3.5*

Suppose that the initial distribution is at point *c*. Then one can redistribute production when moving from $Q_1Q_1$ to $Q_2Q_2$ such that both individuals *A* and *B* would be better off at point *b* than at point *c*. Similarly, one could redistribute the other bundle so that both are better off at point *d* than at point *a*. The paradox arises because point *d* lies to the northeast of point *a*, while point *b* lies to the northeast of point *c*. Thus, one comparison implies a preference for the production bundle associated with $Q_1Q_1$, whereas the other comparison favors the production bundle associated with $Q_2Q_2$. This paradoxical situation would not arise if compensation were actually paid.

These results correspond directly with the analysis in commodity space in Figure 3.3. Points *a* and *c* in Figure 3.5 correspond to distributions that are second-best states. In other words, these points correspond to points *a* and *c* in Figure 3.3, which are also distributions giving rise to second-best states. Note that points *b* and *d* in Figure 3.5 and points *b* and *d* in Figure 3.3 correspond to first-best states.

If one considers all possible production bundles that can be obtained from a given production possibility frontier and all possible distributions of these bundles (which, in utility space, corresponds to considering utility possibility frontiers, such as $Q_3Q_3$, associated with all other possible production bundles), then the *grand utility possibility frontier UU* can be constructed as an envelope of the utility possibility frontiers. All points on this envelope curve correspond to first-best optima, that is, bundles distributed such that the Scitovsky curves are tangent to the production possibility curve.

Samuelson (1950) has argued that even if gainers can profitably bribe losers into accepting a movement and the losers cannot profitably bribe the gainers into rejecting it (that is, both the Kaldor–Hicks and Scitovsky criteria are satisfied), a potential gain in welfare is not necessarily attained. He argues that one has to consider all possible bundles and all possible distributions of these bundles before statements can be made about potential gains. The problem then amounts to selecting one among many first-best states for which there is no solution unless a social ranking of first-best utility possibilities can be determined. He proposes an alternative *potential welfare criterion*, which is demonstrated in Figure 3.5. Simply stated, *if there is some utility frontier such as $Q_4Q_4$ that lies entirely on or outside another utility frontier – $Q_2Q_2$, for example – owing perhaps to technological change, then any position on this new frontier is clearly at least potentially superior to any position on the old one.* Only if the new frontier lies entirely outside the other, however, are potential increases in real income necessarily obtained. Of course, this criterion can be used to compare either grand utility possibility frontiers before and after, say, a technological change or utility possibilities associated with given alternative production bundles.

In the absence of a rule for ranking alternative first-best utility possibilities, Samuelson (1950) argues that this is the only appropriate criterion to apply. In a strict sense, this argument is correct. But in a practical sense, this approach leads to few cases in which beneficial empirical evidence can be developed for policy-makers (see Section 8.3 for a discussion of the related empirical approach). On the other hand, the arguments in favor of this approach are based on an attempt to determine optimal policy without relying on policy-maker preferences or judgment. Such information is not necessary in a practical policy-making setting where political institutions exist for the express purpose of providing policy-makers to make such choices.

## 3.3 THE SOCIAL WELFARE FUNCTION

Because the potential welfare criterion often may not be satisfied even if utility possibility curves can be identified, economic inquiry has continued to search for a rule that can rank all states of society and thus determine which first-best state on the grand utility possibility frontier represents *the* social optimum. In theory, the social welfare function is such a concept. The *social welfare function* is simply a function – say, $W(U_A, U_B)$ – of the utility levels of all individuals such that a higher value of the function is preferred to a lower one. The assumption that the social welfare function is determined by the utilities of all individuals has been called the *fundamental ethical postulate* by Samuelson (1947) and is a cornerstone of democratic societies. Such a welfare function is called a *Bergsonian welfare function* after Abram Bergson (1938), who first used it.

The properties one would expect in such a social welfare function with respect to the utilities of individuals are much like those one would expect in an individual's utility function with respect to the quantities of commodities consumed. That is, one would expect that (1) an increase in the utility of any individual holding others constant increases social welfare (the Pareto principle); (2) if one individual is made worse off, then another individual must be made better off to retain the same level of social welfare; and (3) if some individual has a very high level of utility and another individual has a very low level of utility, then society is willing to give up some of the former individual's utility to obtain even a somewhat smaller increase in the latter individual's utility, with the intensity of this trade-off depending upon the degree of inequality.

The properties described above suggest the existence of welfare contours such as $W_0$ in Figure 3.5, which correspond conceptually to indifference curves for individual utility functions. By property (1), social welfare increases by moving to higher social welfare contours, either upward or to the right. By property (2), the social welfare contours have negative slope. By property (3), the welfare contours are convex to the origin.

Social welfare is maximized by moving to the highest attainable social welfare contour, which thus leads to tangency of the grand utility possibility frontier with the resulting social welfare contour such as at point *e* in Figure 3.5.[4] This tangency condition is sometimes called the *fourth optimality condition*. This condition, together with conditions in equations (2.1), (2.2) and (2.3), completely characterizes the social optimum.

## 3.4 LIMITATIONS OF THE SOCIAL WELFARE FUNCTION APPROACH

Although a social welfare function is a convenient and powerful concept in theory, its practical usefulness has been illusory. Many attempts have been made to specify a social welfare function sufficiently to facilitate empirical usefulness but none have been widely accepted. Apparently, little hope exists for determining a social welfare function on which general agreement can be reached. The major approaches that have been attempted include (1) the subjective approach, (2) the basic axiomatic approach and (3) the moral justice approach.

The subjective approach is represented by those who postulate a complete functional form for the social welfare function on subjective ethical grounds. Early students of the utilitarian school (for example, Bentham 1961, first published 1823) believed that changes in happiness should simply be added over individuals,

$$W = U^1 + U^2 + U^3 + \dots \tag{3.1}$$

4. Note that the slope of the welfare contour can be represented by

$$-\frac{\partial W/\partial U_A}{\partial W/\partial U_B} \equiv -\frac{W_{U_A}}{W_{U_B}}$$

if $W(U_A, U_B)$ is continuous and first derivatives exist. The slope of the utility possibility frontier is

$$-\frac{\partial U_B/\partial q_1}{\partial U_A/\partial q_1} = -\frac{\partial U_B/\partial q_2}{\partial U_A/\partial q_2}.$$

Thus, the tangency condition can be represented mathematically by

$$\frac{W_{U_A}}{W_{U_B}} = \frac{\partial U_B/\partial q_i}{\partial U_A/\partial q_i},\ i = 1, 2.$$

Cross-multiplying yields

$$\frac{\partial W}{\partial U_B}\left(\frac{\partial U_B}{\partial q_i}\right) = \frac{\partial W}{\partial U_A}\left(\frac{\partial U_A}{\partial q_i}\right),\ i = 1, 2,$$

which, simply stated, implies that the marginal social significance of consumption must be equated across individuals for each commodity.

where $U^i$ represents the utility of individual *i*. A positive net gain is then viewed as grounds for policy implementation. This implies that the welfare contours such as $W_0$ in Figure 3.5 should be straight lines with slope −1. Others argue that a functional form should be used that reflects positive benefits from increases in equality consistent with 'normal' distributional judgments (see, for example, Blackorby and Donaldson 1990). One social welfare function that reflects inequality aversion is the form,

$$W = \frac{1}{1-\rho}[(U^1)^{1-\rho} + (U^2)^{1-\rho} + (U^3)^{1-\rho} + \ldots]. \quad (3.2)$$

The problem here is that the term 'normal' is ambiguous. For example, agreement cannot be reached on the appropriate level of inequality aversion, for example, the appropriate value of $\rho$ in equation (3.2).[5]

The axiomatic approach, on the other hand, attempts to investigate the existence and form of the social welfare function mathematically based on a set of plausible underlying axioms about individual preferences and how they count to society. The most celebrated of these efforts is Arrow's (1951) *impossibility theorem*. This theorem addresses the question of whether a general rule exists that can rank social states based only on the way these states are ranked by individual members of society. Arrow showed that no such rule exists under the following plausible requirements:

1. *The domain of decisions is unrestricted.*
2. *The Pareto principle applies.*
3. *Dictatorship is ruled out.*
4. *Rankings are independent of irrelevant alternatives.*

An example of a rule that does not work is majority voting. Arrow's (1951) results suggest that social preferences are determined by a dictator (or a group that acts as a dictator), that the intensity of preferences of individuals rather than simple rankings matters (see Kemp and Ng 1977), or that one of the other axioms such as independence of irrelevant alternatives does not apply (see Sen 1970). Accordingly, Arrow's work has spawned a voluminous literature on possibility theorems by relaxing his axioms in various ways (see Sen 1982 or Fishburn 1973 for surveys). A major practical problem with this approach is that even under weaker axioms where voting works, the transactions costs of compiling votes or rankings of all individuals on each policy issue are prohibitive.

The moral justice approach argues that basic axiomatic examinations following Arrow fail because majority groups acting selfishly will prefer to eliminate consideration of minority interests. This failure can be addressed by admitting moral considerations such as *impartiality* and *economic justice*. Suppose that society consists of three individuals and a change is considered that takes \$1000 from one individual to give \$300 to each of the other two. If the three individuals were to vote selfishly knowing who the benefactors are, the majority would favor the change. On the other hand, if the voting were done not knowing who would pay and who would receive (a *veil of ignorance*), then the change

5. To demonstrate the different inequality aversions possible with the function in equation (3.2), note that it reduces to (3.1) when $\rho = 0$, it approaches (3.4) below when $\rho$ approaches infinity, and it approaches the multiplicative form, $W = U^1 U^2 U^3 \ldots$ when $\rho$ approaches 1. See Boadway and Bruce (1984, ch. 5).

would be unanimously rejected. Alternatively, moral concerns for equal treatment of individuals (*impartiality*) have led some to support value judgments whereby the social welfare function treats individuals symmetrically, for example, $W(U^A, U^B) = W(U^B, U^A)$. The contours of the social welfare function in Figure 3.5 are then symmetric about a 45° line from the origin. If all individuals have identical utility functions then the utility possibility frontier is also symmetric about the 45° line and optimality is achieved by perfect equality. On the other hand, if one individual receives proportionally more utility from consuming the same bundles of goods as another, then such a welfare function would, in effect, assign different weights to the consumption of the individuals.

Harsanyi (1953, 1955) gave the first formal treatment of moral considerations by distinguishing between an individual's personal preferences and moral preferences. In his work, moral preferences are the rankings of a rational individual given that the individual does not know which set of personal preferences he or she will have. Under a relatively weak set of assumptions, Harsanyi (1953, 1955) shows with this approach that the social welfare function is a weighted sum of individual utilities,[6]

$$W = \alpha_1 U^1 + \alpha_2 U^2 + \alpha_3 U^3 + \ldots . \tag{3.3}$$

Further imposing impartiality (symmetry), the welfare function in (3.3) reduces to the Benthamite welfare function in (3.1) with equal weights. This welfare function has been called the *just social welfare function* (see Mueller 1979).

Other moral considerations, however, tend to suggest a stronger concept of equality. Moral considerations in economic welfare issues are often called *rights* to economic justice. Various value judgments or ethical postulates representing these moral considerations include the right to consume what one produces, the right to subsistence, the right to economic liberty and the right to economic equality.[7] With these considerations, taking \$1000 from a very rich individual to give \$300 to each of two poor individuals may be preferred on moral grounds. The most celebrated work in this area is Rawls's (1971) *Theory of Justice*. This theory, which is really more of a philosophy than a theory, contends that policy should be evaluated by the welfare of the most miserable person in society. This implies a social welfare function of the form

$$W = \min\,(U^1, U^2, U^3, \ldots). \tag{3.4}$$

In a more general framework, Arrow (1973) and Harsanyi (1975) show that this choice would be supported by individuals' moral preferences only under infinite risk aversion about the vested interests and preferences to be assumed. With other levels of risk aversion, the welfare function in (3.2) is obtained. Arrow concludes that the possibilities of discovering a theory of justice are remote given the diversity of ethical beliefs in society.

Virtually all of these moral consideration approaches suggest a criterion of distribu-

6. The Harsanyi assumptions are that both personal and moral preferences satisfy the von Neumann–Morgenstern axioms of choice, that each individual has an equal probability of taking on any individual's personal preferences, and that two states are socially indifferent if they are indifferent for every individual. Thus, choices are made according to expected utility given uncertainty about individual preferences.
7. For a more detailed review of the theories of ethical income distribution and economic justice, see Boadway and Bruce (1984, ch. 6).

tional optimality that tends, in some sense, toward either equality or equal weighting. However, even these two simple alternatives represented by (3.1) and (3.4) differ drastically in their implications. With (3.4), the worst-off individual becomes a dictator while, with (3.1), individuals who have very small utilities (and marginal utilities) tend not to matter. Nevertheless, each of these functions can be supported by a plausible set of axioms.[8] Thus, while axiomatic developments have added to the sophistication of social welfare function specification efforts, the effect has been to shift the level of disagreement from the function itself to the axioms that support it. Agreement on the set of axioms appears to be no more possible than agreement on the form of the social welfare function.

In summary, efforts to reach a unique social welfare function have not gained widespread acceptance in spite of great effort by a host of social choice theorists and moral philosophers. Thus, no generally acceptable or objective way to make interpersonal comparisons of utility exists.[9] In spite of the lack of agreement on form, adoption of specific alternative social welfare functions is still advocated from time to time in the literature. Some observed policy choices that strictly redistribute income by, for example, taxing the rich to give to the poor, cannot be advocated or explained with other economic criteria used for policy evaluation. Even if a social welfare function is determined, however, a host of practical problems arise in any practical application. The social welfare function approach requires that individual utilities are cardinally measurable so that intensities of preferences can be compared. In contrast to this approach, Pareto and compensation criteria assume only that utility can be measured ordinally. Thus, much greater practical applicability is attained even though the associated social ordering is not sufficiently complete to identify a unique social optimum or resolve questions of income distribution.

In applied welfare economics, the notion of a social welfare function is useful conceptually but one should keep in mind that a welfare function cannot be specified for practical purposes. However, this does not mean that the study of welfare economics is impractical because the function cannot be specified. Even those who are critical of welfare economics for this reason must agree that economists can make a useful contribution by pointing out who loses and who gains, as well as the magnitude of losses and gains caused for various groups by particular policies.

To summarize the welfare function controversy, it suffices to quote a notable welfare economist, E.J. Mishan (1973, pp. 747–8):

> The social welfare function, even when it is more narrowly defined as a ranking of all conceivable combinations of individual welfare, remains but a pleasing and nebulous abstraction. It cannot be translated into practical guidance for economic policy. Even if there were no fundamental obstacles to its construction, or even if one could think up reasonable conditions under which a social welfare function could exist, there would remain the virtually impossible task of arranging for society to rank unambiguously all conceivable combinations of the individual welfares and moreover – in order to utilise this massive apparatus – to discover (without much cost)

8. For example, Maskin (1978) and Sen (1982) find that any social welfare function with unrestricted domain that satisfies independence of irrelevant alternatives, the Pareto principle, anonymity, separability of unconcerned individuals, and cardinality with interpersonal comparability must be of the form in (3.1). However, simply replacing cardinality with ordinality and adding a minimal equity assumption (the best-off individual's preferences can never be served when they conflict with all worse-off individuals' preferences) results in the Rawlsian social welfare function in (3.4).
9. For further discussion of the difficulties related to determination of a social welfare function, see Atkinson (1970) and Sen (1973).

the effect on the welfare of each person in society (in terms of utilities, goods, or money) of the introduction of alternative economic organisations. For only if we have such data can we rank all existing and future economic possibilities and pronounce some to be socially superior to others. Although one can always claim that 'useful insights' have emerged from the attempts to construct theoretical social welfare functions, the belief that they can ever by translated into useful economic advice is chimerical.

In contrast, the more pedestrian welfare criteria, although analyzed in abstract terms, can be translated into practical propositions. Modern societies do seek to rank projects or policies by some criterion of economic efficiency and to take account also of distributional consequences.

## 3.5 POTENTIAL VERSUS ACTUAL GAINS

Because the social welfare function is a concept upon which general agreement has not been reached and because the potential welfare criterion is one that renders many policy alternatives noncomparable, the compensation principle has emerged as the criterion that is empirically the most widely applicable. But this state of affairs underscores the controversy about whether compensation should actually be paid when adopting policy changes that satisfy the criterion. If possible, should the gainers from a new policy actually compensate the losers so that 'everyone' is actually made better off? Should a policy change be recommended only on the basis of 'potential' gains alone, given that, if the change is made, someone is actually made worse off? As an example, the United Automobile Workers (UAW) union has taken the stand that new technology that displaces workers should not be introduced unless the workers are fully compensated for their losses. This is a case where the *potential* gains criterion is not supported. But to the extent that the UAW represents displaced workers, objections from the losing groups are not surprising.

However, an economist can often analyze the distributional impacts of policy choices without getting into the issue of compensation. For example, suppose one did an analysis of the impact of removing quotas on the importation of steel into the USA. A proper analysis would show the separate effects on government revenues, producers, consumers and the like (possibly by disaggregated groups if, say, several groups of consumers are affected differently). Thus, the losers, the gainers, and the magnitudes of losses and gains would be identified. Such an analysis would be useful to government officials who are elected or appointed to decide, among other things, the issue of compensation. In fact, a welfare analysis that does not adequately indicate individual group effects may be misleading or useless to government officials who have the authority to make interpersonal comparisons. Thus, as emphasized in Chapter 1, studies on the impact of policy choices can be done using welfare economics without getting into the debate as to what 'ought to be'.

## 3.6 PRACTICAL APPLIED POLICY ANALYSIS: THE RELATIONSHIP OF GENERAL EQUILIBRIUM AND PARTIAL EQUILIBRIUM ANALYSIS

The practical applicability of the various criteria for policy evaluation depends on the potential for empirical implementation and on the intuitive understandability of policymakers. Both the social welfare function and potential welfare criteria suffer in both respects. First, consumer utility cannot be measured sufficiently for empirical implementation under general conditions and, second, units of measurement for utility and social welfare are abstract and not well understood by policy-makers. The Pareto and compensation criteria, however, can be implemented in terms of individuals' *willingness to pay* and *willingness to accept* the effects of policies and projects. As demonstrated throughout this book, these measures can be reported in monetary terms that are generally empirically feasible and well understood.

A second problem with the framework used thus far for practical policy analysis is that an abstract general equilibrium framework has been used to investigate possibilities for identifying potential social gains through application of compensation criteria. Such a general equilibrium framework is highly useful for understanding the nature of problems encountered in application of compensation criteria, but it is not very helpful for analyzing and quantifying the implications of specific policies or projects involving markets and prices for specific goods. Policy-makers are generally concerned with impacts on specific markets and specific types of agents in society. The remainder of this book focuses on measuring individual, market and group-specific welfare effects by first concentrating on partial equilibrium models.

To facilitate the transition from general equilibrium analysis to the analysis of specific markets and agents, consider Figure 3.6. Figure 3.6(a) illustrates a production possibility curve *PP*, a Scitovsky indifference curve *C*, and a first-best equilibrium at $(q_1^*,q_2^*)$, which attains a tangency of the production possibility curve and Scitovsky curve at prices $p_1^*$ and $p_2^*$ for goods $q_1$ and $q_2$, respectively. Figure 3.6(b) illustrates the supply and demand curves for $q_1$, which are derived from Figure 3.6(a) by varying the price $p_1$. The supply curve is found by plotting the absolute value of the slope of *PP* for each level of $q_1$. In other words, it is found by varying the price $p_1$ holding price $p_2$ fixed at $p_2^*$ and finding the corresponding tangency of the price line with slope $-p_1/p_2^*$ to the production possibility curve. This slope is the social marginal cost of $q_1$ in terms of $q_2$, that is, the value of $q_2$ that must be given up to gain an additional unit of $q_1$ at prices $p_1$ and $p_2^*$. As the amount of $q_1$ increases, social marginal cost increases. The demand curve in Figure 3.6(b) is the graph of the absolute value of the slope of the Scitovsky indifference curve *C* in Figure 3.6(a). It corresponds to varying the price $p_1$ holding price $p_2$ fixed at $p_2^*$ and finding the corresponding tangency of the price line with slope $-p_1/p_2^*$ to the Scitovsky curve. Thus, the Scitovsky curve has a social marginal willingness-to-pay (WTP) interpretation. That is, at each point on curve *C* the slope is the maximum amount of $q_2$ society is willing to give up to gain an additional unit of $q_1$ at prices $p_1$ and $p_2^*$. As society has more of $q_1$, the social marginal WTP, or social marginal benefit (MB), declines.

At the Pareto optimal level of $q_1$, denoted by $q_1^*$ in Figure 3.6, the marginal WTP is just equal to the marginal cost of $q_1$, so it is impossible to identify any potential economic social gains in moving from this point. In Figure 3.6(b) in particular, this result is noted by considering movements to the right and left of $q_1^*$. For example, for a movement to the right, say to $q_1^1$, the marginal cost is greater than marginal WTP and, hence, losses are associated with moving from $q_1^*$ to $q_1^1$. To the left, say at $q_1^2$, marginal WTP exceeds marginal cost so net social benefits are possible in moving from $q_1^2$ to $q_1^*$. Finally, note that under the assumptions of Section 2.5 the competitive mechanism results in a market equilibrium at product price $p_1^*$ and quantity $q_1^*$, which attains Pareto efficiency given price $p_2^*$ for good $q_2$.

Chapters 4 through 7 use the approach of Figure 3.6(b) assuming the prices in other

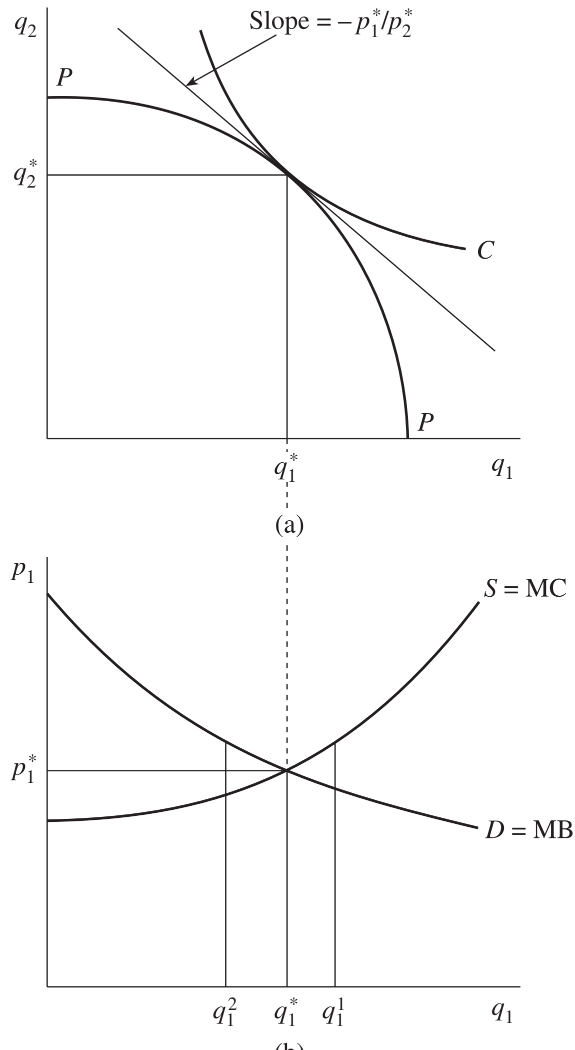


*Figure 3.6*

markets are fixed. Focusing on a single market while assuming that equilibrium in other markets is unaffected constitutes a partial equilibrium approach in contrast to the framework of Chapters 2 and 3, which uses a general equilibrium approach. A partial equilibrium approach is useful for illuminating how specific policies affect specific markets and groups of consumers and producers in specific markets. However, one must bear in mind that specific policies with specific effects in a given market can have additional general equilibrium implications in other markets. These are considered in Chapter 9.[10]

10. The results illustrated in the simple graphical model of Figure 3.6 are developed rigorously in the Appendix to Chapter 9 in the context of a market economy with many markets.

## 3.7 SUMMARY

This chapter has focused on the compensation principle and the social welfare function as devices to aid policy-makers in using resources optimally. The compensation principle states that state *B* is preferred to state *A* if, in making the move from state *A* to state *B*, the gainers can compensate the losers such that everyone could be made better off – that is, if the WTP of the gainers exceeds the WTA of the losers. The principle is based on potential rather than actual compensation. Some could actually be made worse off from a policy change, yet the change would be supported if the gainers *could* have compensated the losers so that everyone could have been better off. Because the principle is based on potential rather than actual gains, two problems arise: the reversal paradox and the intransitivity problem. However, even though the criterion is based on potential gains, these problems can arise only if no first-best bundle is considered.

The concept of a utility possibility curve was introduced, and a parallel was drawn between the utility possibilities approach to welfare economics versus that based on production possibility frontiers and Scitovsky indifference curves. The notion of a potential welfare criterion was introduced. If this criterion is adhered to by policy-makers, all possible bundles of goods have to be considered together with all possible distributions of these bundles. Such an approach is usually not empirically practical (although it is considered further in Section 8.3).

Because the compensation principle cannot rank first-best bundles, the concept of a welfare function was introduced. If such a function were available and agreed upon, the optimum organization could be obtained. But because agreement on such a function cannot be reached, the compensation principle is apparently the most widely applicable, yet also empirically practical, criterion. However, one of the problems with the principle is that it is based on potential rather than actual gains. Thus, in any policy context, the payment of compensation is a matter that must be decided by policy-makers endowed with the authority to determine income distributional issues.

# 4. Welfare measurement for the producer

Production is a crucial component of economic activity. Many firms produce the goods and services consumed by the general public. Some industries (for example, steel) are made up of only a few firms, whereas others are made up of many firms (for example, the US wheat-producing industry). Also, some firms are multinational in scope (that is, they manufacture and sell products in more than one country).

Increasing attention has been devoted to the question of how well producing firms and marketing firms perform. That is, do they price competitively or not? Does competitive pricing lead to economic efficiency? Do producers adequately consider the costs imposed on society by environmental pollution? Are producers affected by government policy (and, if so, by how much)? For example, what is the effect on the steel industry and on consumers of the use of import quotas?

This chapter focuses on the producer welfare measurement. Several different producer welfare measures are defined and compared. The chapter proceeds on a rather simple level, dealing with only a single competitive firm under the assumption of profit maximization. Clearly, firms may operate with more criteria in mind than only profit maximization. A vast literature exists in positive economics about firms that maximize sales, growth, market shares and the like. Although the approach developed in this chapter, and throughout the rest of this book, could be used in contexts in which firms have such objective criteria, the major emphasis is on profit maximization because it is the more popular and, perhaps, more relevant approach. The profit-maximization approach requires some generalization when price or production is random. For example, a more appropriate assumption is that producers maximize the expected utility from profits. But this distinction makes no difference when prices and production are nonrandom, as in the case of this chapter. The case with risk (random prices and production) where utility maximization may be important is introduced in Chapter 12 and considered in detail in Appendix Sections 12.A and 14.A.

Welfare analysis of an individual firm, even in a competitive industry, is often required. Examples are the analysis of effects of selective labor strikes against individual firms and the analysis of effects of antitrust litigation brought against a single processor or retailer by a single producer of a product. However, in most cases, empirical welfare analysis focuses on an entire industry. This requires aggregating producer responses and effects. This issue will be delayed until Chapter 8. The concepts derived in this chapter, however, form the basis for aggregation.

## 4.1 THE PROFIT-MAXIMIZING FIRM

The basic framework for the analysis of this chapter is the neoclassical model as presented in Figure 4.1. Under the assumption of profit maximization, the producer equates price and marginal cost (MC) assuming price exceeds the minimum of the average variable cost