# 2 Analytical Foundations

## 2.1 INTRODUCTION

The objective of this chapter is to provide an introduction to the analytical foundations of production economics. Our goal is not to provide a detailed exposition of the foundations; this is available in the references provided at the end of the book. Rather, our objective is to provide an analytical foundation that is sufficient to enable the reader to conduct econometric analyses of various types of productive efficiency, which is the ultimate objective of the book. By productive efficiency we mean the degree of success producers achieve in allocating the inputs at their disposal and the outputs they produce, in an effort to meet some objective. Thus in order to measure productive efficiency it is first necessary to specify producers' objectives and then to quantify their degrees of success. This book is primarily concerned with the development of econometric techniques for estimating their degrees of success.

At an elementary level, the objective of producers can be as simple as seeking to avoid waste, by obtaining maximum outputs from given inputs or by minimizing input use in the production of given outputs. In this case the notion of productive efficiency corresponds to what we call *technical* efficiency, and the waste avoidance objective of producers becomes one of attaining a high degree of technical efficiency. Chapter 3 is concerned with the development of econometric techniques for the estimation of technical efficiency.

At a higher level, the objective of producers might entail the production of given outputs at minimum cost, or the utilization of given inputs to maximize revenue, or the allocation of inputs and outputs to maximize profit. In these cases productive efficiency corresponds to what we call *economic* efficiency, and the objective of producers becomes one of attaining a high degree of economic (cost, revenue, or profit) efficiency. Chapters 4–6 are concerned with the development of econometric techniques for the estimation of cost and profit efficiency. We pay little attention to revenue efficiency, because the econometric techniques used to estimate cost efficiency can be modified easily to estimate revenue efficiency.

We begin this chapter by providing an analytical framework for describing the physical structure of production technology, in which multiple inputs are used to produce multiple outputs. This framework is based solely on information on the quantities of the inputs and the outputs. The structure of production technology is initially described in terms of feasible sets of inputs and outputs. Attention then moves to the boundaries of these sets, since the boundaries represent weakly efficient production activities. Eventually the structure of production technology is described in terms of distance functions, evocatively named since they provide measures of the distance of a production activity to the boundary of production possibilities. Distance functions are thus intimately related to the measurement of technical efficiency. Indeed since price information is not exploited, technical efficiency is the only type of efficiency that can be studied using distance functions.

The tools of duality theory are then used to obtain several related economic representations of the structure of production technology, using information on both the quantities and the prices of the inputs and the outputs, and assuming that producers attempt to solve an economic optimization problem. Economic representations of production technology include cost, revenue, and profit frontiers. These economic frontiers are then used as standards against which to measure cost, revenue, and profit efficiency.

Once these frontier representations of production technology have been introduced, productive efficiency is then defined in terms of distance to a particular frontier. Technical efficiency is defined in terms of distance to a production frontier, and economic efficiency is defined in terms of distance to an economic (cost, revenue, or profit) frontier. Whereas technical efficiency is a purely physical notion that can be measured without recourse to price information and without having to impose a behavioral objective on producers, cost, revenue, and profit efficiency are economic concepts whose measurement requires both price information and the imposition of an appropriate behavioral objective on producers.

The remainder of this chapter is organized as follows.

The subject of Section 2.2 is a description of the general structure of production technology with which multiple inputs are used to produce multiple outputs. This structure is described in Section 2.2.1 in terms of the graph of production technology and the corresponding input sets and output sets. In Section 2.2.2 we temporarily make the simplifying assumption that multiple inputs are used to produce a single output, which enables us to describe the structure of production technology in terms of a production frontier. In Section 2.2.3 we revert to the original assumption that multiple inputs are used to produce multiple outputs, and we describe the structure of production technology in terms of input distance functions and output distance functions. We also describe the relationship between output distance functions and the production frontier in the event that only a single output is produced. In Section 2.2.4 the structure of production technology is defined in terms of dual economic frontiers, namely cost, revenue, and profit frontiers. These economic frontiers apply to both the single-output case and the multiple-output case. In Section 2.2.5 we introduce variable cost frontiers and variable profit frontiers. These frontiers are relevant in the short run, or whenever a subset of inputs is fixed and therefore not freely adjustable by producers.

Section 2.3 is concerned with the measurement of technical efficiency. In Section 2.3.1 it is assumed that multiple inputs are used to produce a single output, and output-oriented technical efficiency is defined relative to a production frontier. In Section 2.3.2 it is assumed that multiple inputs are used to produce multiple outputs, and input distance functions and output distance functions are used to provide input-oriented and output-oriented definitions of technical efficiency.

Section 2.4 is concerned with the measurement and decomposition of economic efficiency. It is assumed that producers use multiple inputs to produce multiple outputs. In Section 2.4.1 cost efficiency is defined relative to a cost frontier. Cost efficiency is then decomposed into its two components, input-oriented technical efficiency and input allocative efficiency. The logic behind the decomposition is that both types of inefficiency (excessive input use and misallocation of inputs) are costly, and it is desirable to be able to identify the sources of cost inefficiency. In Section 2.4.2 revenue efficiency is defined relative to a revenue frontier. Revenue efficiency is then decomposed into output-oriented technical efficiency and output allocative efficiency. Once again, both types of inefficiency (output shortfall and an inappropriate output mix) are costly in terms of forgone revenue, and it is desirable to be able to identify the sources of revenue inefficiency. In Section 2.4.3 profit efficiency is defined relative to a profit frontier. Profit efficiency can be decomposed in two ways, depending on the orientation of its technical efficiency component. In Section 2.4.4 we introduce the notions of variable-cost efficiency and variable-profit efficiency for use in situations in which producers seek to minimize cost or to maximize profit in the presence of fixed inputs.

Section 2.5 provides a guide to the relevant literature.

## 2.2 PRODUCTION TECHNOLOGY

### 2.2.1 Representing Technology with Sets

We assume that producers use a nonnegative vector of inputs, denoted $x = (x_1, \dots, x_N) \in R^N_+$, to produce a nonnegative vector of outputs, denoted $y = (y_1, \dots, y_M) \in R^M_+$. Although the analytical foundations developed in this chapter readily accommodate zero values for some inputs and some outputs, much of the econometric analysis developed in subsequent chapters is based on logarithmic functional forms that do not easily accommodate nonpositive values of variables, and so in later chapters we will assume that $x \in R^N_{++}$ and $y \in R^M_{++}$. The initial task is to characterize the set of feasible production activities. A primitive characterization is provided in Definition 2.1 and illustrated in Figure 2.1.

> **Definition 2.1:** The *graph* of production technology, $GR = \{(y, x)\colon x \text{ can produce } y\}$, describes the set of feasible input–output vectors.

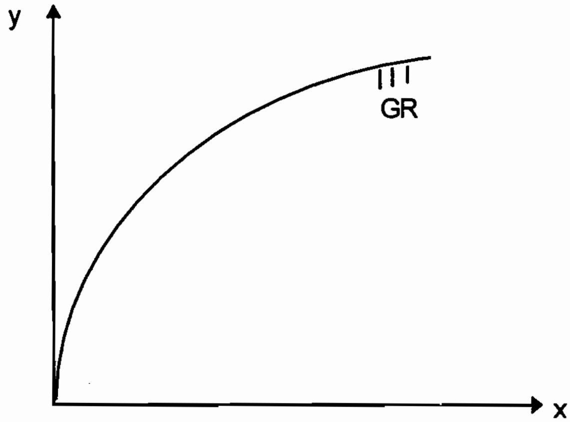


Figure 2.1 The Graph of Production Technology ($M = 1, N = 1$)

Figure 2.1 illustrates the graph of production technology in the single-input, single-output case. The graph, also known as the production possibilities set, is the set of input–output combinations bounded below by the $x$ axis and bounded above by the curve emanating from the origin. Soon the curve that provides the upper boundary of the graph will be given a name.

$GR$ is assumed to satisfy the following properties:

$G1$: $(0, x) \in GR$ and $(y, 0) \in GR \Rightarrow y = 0$.

$G2$: $GR$ is a closed set.

$G3$: $GR$ is bounded for each $x \in R_{+}^{N}$.

$G4$: $(y, x) \in GR \Rightarrow (y, \lambda x) \in GR$ for $\lambda \geq 1$.

$G5$: $(y, x) \in GR \Rightarrow (\lambda y, x) \in GR$ for $0 \leq \lambda \leq 1$.

Property $G1$ states that any nonnegative input vector can produce at least zero output and that there is no free lunch. $G2$ guarantees the existence of technically efficient input and output vectors. $G3$ guarantees that finite input cannot produce infinite output. $G4$ and $G5$ are weak monotonicity properties that guarantee the feasibility of radial expansions of feasible inputs and radial contractions of

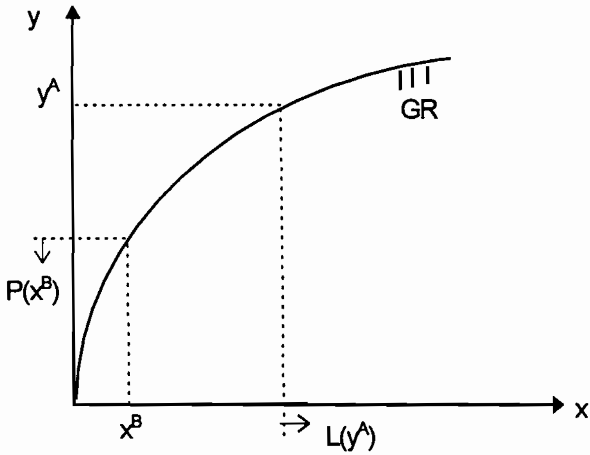


Figure 2.2 The Input Sets and Output Sets of Production Technology ($M = 1, N = 1$)

feasible outputs. These two properties are occasionally replaced with the single stronger monotonicity property

$$G6\text{: } (y, x) \in GR \Rightarrow (y', x') \in GR\ \forall (y', -x') \le (y, -x).$$

$G6$ guarantees the feasibility of any increase in feasible inputs, including but not limited to a radial increase, and also guarantees the feasibility of any reduction in feasible outputs, including but not limited to a radial contraction. $G4$ and $G5$ are often referred to as weak disposability properties, and $G6$ as a strong (or free) disposability property.

*GR* is not generally required to be a convex set. However on occasion this property is required, and so convexity is listed as a final property.

G7: *GR* is a convex set.

A second characterization of the set of feasible production activities is provided in Definition 2.2 and illustrated in Figures 2.2 and 2.3.

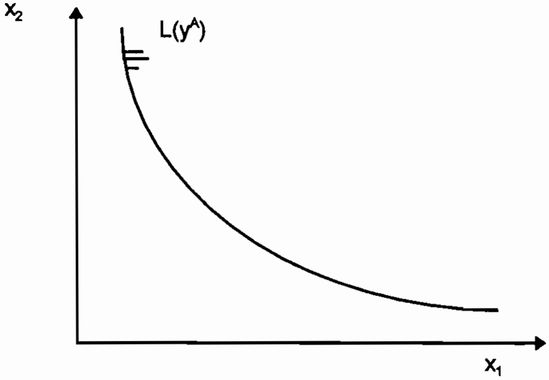


Figure 2.3 The Input Sets of Production Technology ($N = 2$)

**Definition 2.2:** The *input sets* of production technology, $L(y) = \{x: (y, x) \in GR\}$, describe the sets of input vectors that are feasible for each output vector $y \in R^M_+$.

In Figure 2.2 $L(y^A)$ is the set of inputs on the interval $[x^A, +\infty)$. In Figure 2.3 $L(y^A)$ is the region bounded below by the curve. Soon the curve will be given a name.

Since $GR$ is assumed to satisfy certain properties and since the input sets $L(y)$ are defined in terms of $GR$, it follows that the input sets $L(y)$ satisfy the following properties:

$L1$: $0 \notin L(y)$ for $y \geq 0$ and $L(0) = R^N_+$.

$L2$: The sets $L(y)$ are closed.

$L3$: $x$ is finite $\Rightarrow x \notin L(y)$ if $y$ is infinite.

$L4$: $x \in L(y) \Rightarrow \lambda x \in L(y)$ for $\lambda \geq 1$.

$L5$: $L(\lambda y) \subseteq L(y)$ for $\lambda \geq 1$.

If the weak monotonicity properties $G4$ and $G5$ are replaced with the strong monotonicity property $G6$, then the weak monotonicity properties $L4$ and $L5$ are replaced with the strong monotonicity property

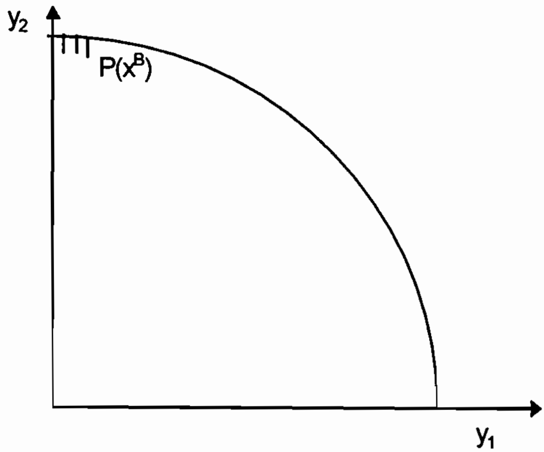


Figure 2.4 The Output Sets of Production Technology ($M = 2$)

$L6$: $x' \geq x \in L(y) \Rightarrow x' \in L(y)$ and $y' \geq y \Rightarrow L(y') \subseteq L(y)$,

which states that inputs and outputs are strongly, or freely, disposable. Finally, a convexity property is occasionally added to the list of properties satisfied by the input sets $L(y)$:

$L7$: $L(y)$ is a convex set for $y \in R^M_+$.

It should be noted that $G7$ is sufficient, but not necessary, for $L7$.

A third characterization of the set of feasible production activities is provided in Definition 2.3 and illustrated in Figures 2.2 and 2.4.

**Definition 2.3:** The *output sets* of production technology, $P(x) = \{y: (y, x) \in GR\}$, describe the sets of output vectors that are feasible for each input vector $x \in R^N_+$.

In Figure 2.2 $P(x^B)$ is the set of outputs on the interval $[0, y^B]$. In Figure 2.4 $P(x^B)$ is the region bounded above by the curve. Soon the curve will be given a name.

Since the output sets $P(x)$ are defined in terms of $GR$ and since $GR$ is assumed to satisfy certain properties, the output sets $P(x)$ satisfy properties corresponding to those satisfied by $GR$. These properties are:

$P1$: $P(0) = \{0\}$.

$P2$: $P(x)$ is a closed set.

$P3$: $P(x)$ is bounded for $x \in R^N_+$.

$P4$: $P(\lambda x) \supseteq P(x)$ for $\lambda \geq 1$.

$P5$: $y \in P(x) \Rightarrow \lambda y \in P(x)$ for $\lambda \in [0, 1]$.

If the weak monotonicity properties $G4$ and $G5$ are replaced with the strong monotonicity property $G6$, then the weak monotonicity properties $P4$ and $P5$ are strengthened to

$P6$: $x' \geq x \Rightarrow P(x') \supseteq P(x)$ and $y \leq y' \in P(x) \Rightarrow y \in P(x)$,

which states that $P(x)$ satisfies strong, or free, disposability of inputs and outputs. Finally, a convexity property

$P7$: $P(x)$ is a convex set for $x \in R^N_+$.

is occasionally required. As in the case of $G7$ and $L7$, $G7$ is sufficient, but not necessary, for $P7$.

We now turn our attention to the boundaries of the sets depicted in Figures 2.1–2.4.

**Definition 2.4:** The *input isoquants* Isoq $L(y) = \{x\colon x \in L(y), \lambda x \notin L(y), \lambda < 1\}$ describe the sets of input vectors capable of producing each output vector $y$ but which, when radially contracted, become incapable of producing output vector $y$.

**Definition 2.5:** The *input efficient subsets* Eff $L(y) = \{x\colon x \in L(y), x' \leq x \Rightarrow x' \notin L(y)\}$ describe the sets of input vectors capable of producing each output vector $y$ but which, when contracted in any dimension, become incapable of producing output vector $y$.

In Figure 2.5 Isoq $L(y)$ is the curve mentioned beneath Definition 2.2 that provides the lower boundary of the input set $L(y)$. Since it is a lower boundary representing one notion of minimal input use, it provides an appealing standard against which to measure the technical efficiency of input use. Eff $L(y)$ is the darkened portion of Isoq $L(y)$, and includes only the downward-sloping portion of Isoq $L(y)$. The upward-sloping portion of Isoq $L(y)$ not contained in Eff $L(y)$ is occasionally described as belonging to the *uneconomic region* of input space. Since Eff $L(y) \subseteq$ Isoq $L(y)$, Eff $L(y)$ provides a more

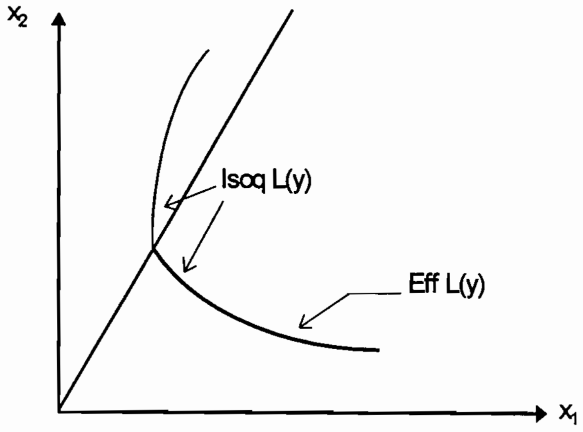


Figure 2.5 The Input Isoquant and the Input Efficient Subset ($N = 2$)

stringent standard against which to measure the technical efficiency of input use.

The weak monotonicity property $L4$ allows Eff $L(y) \subset$ Isoq $L(y)$. Unfortunately the strong monotonicity property given in the first part of $L6$ also allows Eff $L(y) \subset$ Isoq $L(y)$, as in the case of a fixed-proportions Leontief technology. An even stronger version of the first part of $L6$ is required for Eff $L(y) =$ Isoq $L(y)$. Some functional forms employed in the econometric analysis of efficiency, such as Cobb–Douglas, do have the property that Eff $L(y)$ = Isoq $L(y)$, making the distinction irrelevant. Others, such as translog, have the property that Eff $L(y) \subset$ Isoq $L(y)$, making the distinction potentially important. We return to this distinction in Section 2.3.1.

**Definition 2.6:** The *output isoquants* Isoq $P(x) = \{y\colon y \in P(x), \lambda y \notin P(x), \lambda > 1\}$ describe the sets of all output vectors that can be produced with each input vector $x$ but which, when radially expanded, cannot be produced with input vector $x$.

**Definition 2.7:** The *output efficient subsets* Eff $P(x) = \{y\colon y \in P(x), y' \geq y \Rightarrow y' \notin P(x)\}$ describe the sets of all output vectors that can be produced with each input vector $x$ but which, when expanded in any dimension, cannot be produced with input vector $x$.

In Figure 2.6 Isoq $P(x)$ is the curve mentioned beneath Definition 2.3 that provides the upper boundary of the output set $P(x)$. Since it

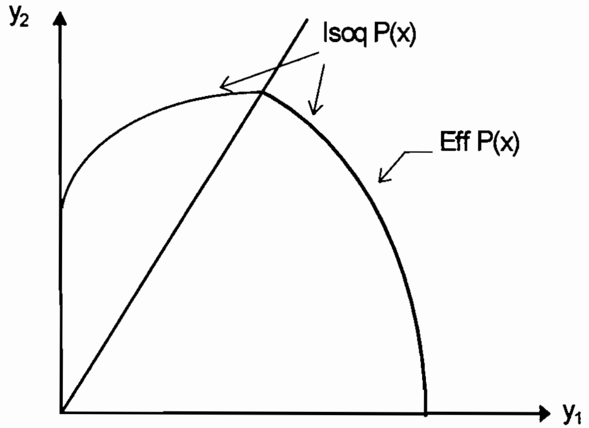


Figure 2.6 The Output Isoquant and the Output Efficient Subset ($M = 2$)

is an upper boundary characterizing one notion of maximum output producible with input vector $x$, it provides a standard against which to measure the technical efficiency of output production. Eff $P(x)$ is the darkened portion of Isoq $P(x)$, and includes only the downward-sloping portion of Isoq $P(x)$. By analogy with the concept of an uneconomic region in input space, that portion of Isoq $P(x)$ not included in Eff $P(x)$ might be characterized as belonging to the uneconomic region of output space. Since Eff $P(x) \subseteq$ Isoq $P(x)$, Eff $P(x)$ provides a more stringent standard against which to measure the technical efficiency of output production. If the output sets $P(x)$ satisfy the weak monotonicity property $P5$, then Eff $P(x) \subset$ Isoq $P(x)$ is allowed, and again the fixed-proportions Leontief technology shows that even the strong monotonicity property given in the second part of $P6$ allows for Eff $P(x) \subset$ Isoq $P(x)$. As in the case of Eff $L(y)$ and Isoq $L(y)$, the relationship between Eff $P(x)$ and Isoq $P(x)$ depends on the functional form used to characterize the structure of production technology. We return to this distinction in Section 2.3.1.

### 2.2.2 Production Frontiers

We now provide a functional characterization of the boundary of the graph of production technology. Since the boundary represents the maximum output that can be obtained from any given input vector (or, alternatively, the minimum input usage required to produce any given output vector), it represents another standard against which to measure the technical efficiency of production. We begin with a characterization of the boundary of the graph of production technology in the multiple-input, multiple-output case.

> **Definition 2.8:** A *joint production frontier* is a function $F(y, x) = 0$ having the properties $\text{Isoq}\, L(y) = \{x\colon F(y, x) = 0\}$ and $\text{Isoq}\, P(x) = \{y\colon F(y, x) = 0\}$.

A joint production frontier is also referred to as a production possibilities frontier, or a transformation frontier. It is rarely used in empirical analysis, and we will not discuss it further. The notion it characterizes, that of the boundary of the graph of production technology when multiple inputs are used to produce multiple outputs, is more easily characterized by means of input distance functions and output distance functions, to which we turn in Section 2.2.3.

A single-output specification of production activity is valid in two circumstances. It is obviously valid in the rare event that only a single output is produced. It is also valid in the more likely event that multiple outputs are produced, provided that the outputs can be aggregated into a single composite output $y = g(y_1, \ldots, y_M)$. In either case the joint production frontier introduced in Definition 2.8 collapses to a production frontier. Alternatively, Definitions 2.2 and 2.3 can be used to obtain a production frontier, introduced in Definition 2.9 and illustrated in Figure 2.7.

> **Definition 2.9:** A *production frontier* is a function $f(x) = \max\{y\colon y \in P(x)\} = \max\{y\colon x \in L(y)\}$.

Since the production frontier is defined in terms of the output sets $P(x)$ and the input sets $L(y)$, both of which satisfy certain properties, so does $f(x)$. These properties are

$f1$: $f(0) = 0$.

$f2$: $f$ is upper semicontinuous on $R_+^N$.

$f3$: $f(x) > 0 \Rightarrow f(\lambda x) \to +\infty$ as $\lambda \to +\infty$.

$f4$: $f(\lambda x) \geq f(x)$, $\lambda \geq 1$ for $x \in R_+^N$.

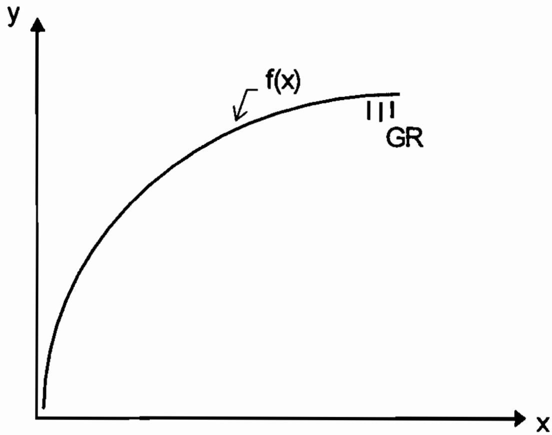


Figure 2.7 A Production Frontier

If the weak monotonicity property $L4$ is replaced by the strong monotonicity property given in the first part of $L6$, the weak monotonicity property $f4$ becomes

$f5$: $x' \geq x \Rightarrow f(x') \geq f(x)$.

If the convexity property $L7$ is imposed, then

$f6$: $f$ is quasiconcave on $R^N_+$.

In Figure 2.7 the production frontier $f(x)$ describes the maximum output that can be produced with any given input vector. Remembering that only a single output is being produced, it follows from Definitions 2.4, 2.5, and 2.9 that $L(y) = \{x: f(x) \geq y\}$, Isoq $L(y) = \{x: f(x) = y, f(\lambda x) < y, \lambda < 1\}$, and Eff $L(y) = \{x: f(x) = y, x' \leq x \Rightarrow f(x') < y\}$. That is, the input sets $L(y)$ consist of all input vectors capable of producing at least scalar output $y$. The input isoquants Isoq $L(y)$ consist of all input vectors capable of producing scalar output $y$ and which, when radially contracted, are incapable of producing scalar output $y$. The input efficient subsets Eff $L(y)$ consist of all input vectors capable of producing scalar output $y$ and which, when contracted in any dimension, are incapable of producing scalar output $y$.

The production frontier provides the upper boundary of production possibilities, and the input–output combination of each producer is located on or beneath the production frontier. The central problem in the measurement of technical efficiency is to measure the distance from the input–output combination of each producer to the production frontier. Two notions of distance are introduced in Section 2.2.3, and used extensively in Section 2.3 to measure technical efficiency.

### 2.2.3 Distance Functions

When multiple inputs are used to produce multiple outputs, Shephard's (1953, 1970) distance functions provide a functional characterization of the structure of production technology. Input distance functions characterize input sets, and output distance functions characterize output sets. Not only do distance functions characterize the structure of production technology, it turns out that they are intimately related to the measures of technical efficiency that will be introduced in Section 2.3. However the major role distance functions play is in duality theory. Just as under certain conditions a (single-output) production frontier is dual to a (single-output) cost frontier, also under certain conditions an input distance function is dual to a cost frontier and an output distance function is dual to a revenue frontier. Although the main role played by distance functions is in duality theory, they are not without empirical value. They can be estimated econometrically to provide measures of technical efficiency when producers use multiple inputs to produce multiple outputs. However they have rarely been used for this purpose.

The input distance function is introduced in Definition 2.10 and illustrated in Figures 2.8 and 2.9.

**Definition 2.10:** An *input distance function* is a function $D_I(y, x) = \max\{\lambda\colon x/\lambda \in L(y)\}$.

An input distance function adopts an input-conserving approach to the measurement of the distance from a producer to the boundary of production possibilities. It gives the maximum amount by which a producer's input vector can be radially contracted and still remain feasible for the output vector it produces. In Figure 2.8 the

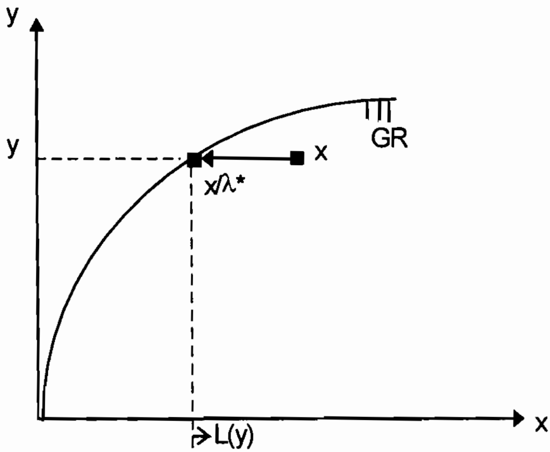


Figure 2.8 An Input Distance Function ($M = 1, N = 1$)

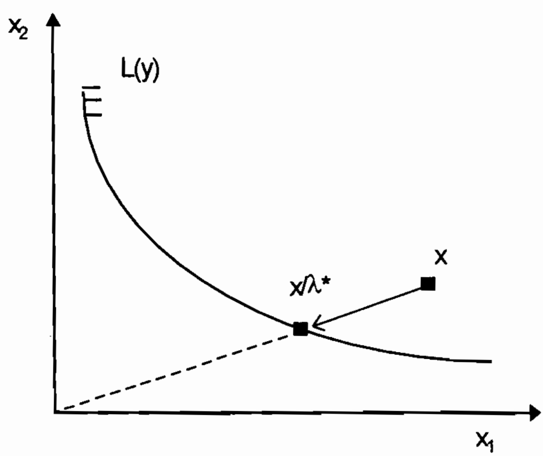


Figure 2.9 An Input Distance Function ($N= 2$)

scalar input $x$ is feasible for output $y$, but $y$ can be produced with smaller input $(x/\lambda^*)$, and so $D_I(y, x) = \lambda^* > 1$. In Figure 2.9 the input vector $x$ is feasible for output $y$, but $y$ can be produced with the radially contracted input vector $(x/\lambda^*)$, and so $D_I(y, x) = \lambda^* > 1$.

Since the input distance function $D_I(y, x)$ is defined in terms of the input sets $L(y)$, which satisfy certain properties, the input distance function satisfies a corresponding set of properties given by

$D_I1$: $D_I(0, x) = +\infty$ and $D_I(y, 0) = 0$.

$D_I2$: $D_I(y, x)$ is an upper-semicontinuous function.

$D_I3$: $D_I(y, \lambda x) = \lambda D_I(y, x)$ for $\lambda > 0$.

$D_I4$: $D_I(y, \lambda x) \geq D_I(y, x)$ for $\lambda \geq 1$.

$D_I5$: $D_I(\lambda y, x) \leq D_I(y, x)$ for $\lambda \geq 1$.

If the weak monotonicity properties $L4$ and $L5$ are replaced with the strong monotonicity property $L6$, then the weak monotonicity properties $D_I4$ and $D_I5$ are replaced with

$D_I6$: $D_I(y, x') \geq D_I(y, x)$ for $x' \geq x$ and $D_I(y', x) \leq D_I(y, x)$ for $y' \geq y$.

If the convexity property $L7$ holds, then

$D_I7$: $D_I(y, x)$ is a concave function in $x$.

It should be clear from Definition 2.10 and Figures 2.8 and 2.9 that $L(y) = \{x\colon D_I(y, x) \geq 1\}$ and that Isoq $L(y) = \{x\colon D_I(y, x) = 1\}$. Thus the input isoquant, which we have already mentioned as one possible standard against which to measure the technical efficiency of input use, corresponds to the set of input vectors having an input distance function value of unity. All other feasible input vectors have input distance function values greater than unity.

The output distance function is introduced in Definition 2.11 and illustrated in Figures 2.10 and 2.11.

**Definition 2.11:** An *output distance function* is a function $D_O(x, y) = \min\{\mu\colon y/\mu \in P(x)\}$.

An output distance function takes an output-expanding approach to the measurement of the distance from a producer to the boundary of production possibilities. It gives the minimum amount by which an output vector can be deflated and still remain producible with a given input vector. In Figure 2.10 scalar output $y$ can be produced with input $x$, but so can larger output $(y/\mu^*)$, and so $D_O(x, y) = \mu^* < 1$. In Figure 2.11 the output vector $y$ is producible with input $x$, but

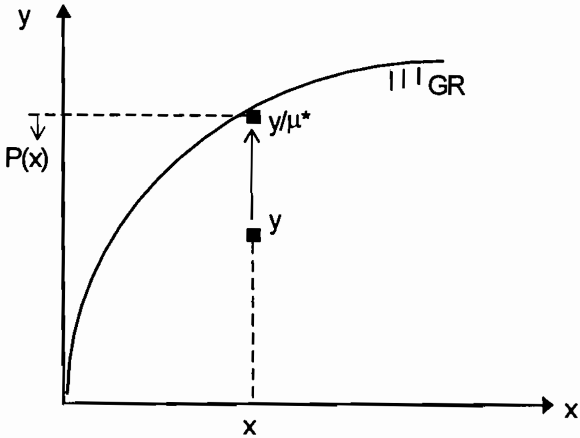


Figure 2.10 An Output Distance Function ($M = 1, N = 1$)

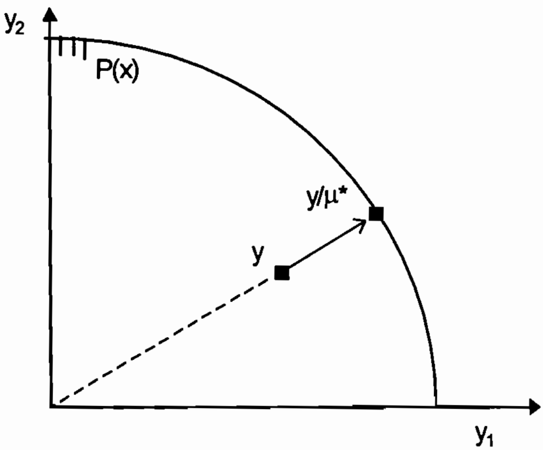


Figure 2.11 An Output Distance Function ($M = 2$)

so is the radially expanded output vector ($y/\mu^*$), and so $D_O(x, y) = \mu^* < 1$.

Since an output distance function $D_O(x, y)$ is defined in terms of the output sets $P(x)$, which satisfy certain properties, the corresponding output distance function satisfies the properties

$D_O1$: $D_O(x, 0) = 0$ and $D_O(0, y) = +\infty$.

$D_O2$: $D_O(x, y)$ is a lower-semicontinuous function.

$D_O3$: $D_O(x, \lambda y) = \lambda D_O(x, y)$ for $\lambda > 0$.

$D_O4$: $D_O(\lambda x, y) \le D_O(x, y)$ for $\lambda \ge 1$.

$D_O5$: $D_O(x, \lambda y) \le D_O(x, y)$ for $0 \le \lambda \le 1$.

If the weak monotonicity properties *P*4 and *P*5 are replaced with the strong monotonicity property *P*6, then $D_O4$ and $D_O5$ are replaced with the strong monotonicity property

$D_O6$: $D_O(x', y) \le D_O(x, y)$ for $x' \ge x$ and $D_O(x, y') \le D_O(x, y)$ for $y' \le y$.

Finally if the convexity property *P*7 holds, then

$D_O7$: $D_O(x, y)$ is a convex function in $y$.

It should be clear from Definition 2.11 and Figures 2.10 and 2.11 that $P(x) = \{y\colon D_O(x, y) \le 1\}$ and that Isoq $P(x) = \{y\colon D_O(x, y) = 1\}$. Thus the output isoquant, which we have mentioned as a possible standard against which to measure the technical efficiency of output production, corresponds to the set of output vectors having an output distance function value of unity. All other feasible output vectors have output distance function values less than unity.

Distance functions provide a characterization of the structure of production technology when multiple inputs are used to produce multiple outputs. However if only a single output is produced, an output distance function is related to the production frontier introduced in Definition 2.9. It should be apparent from Figure 2.10 that this relationship is given by $D_O(x, y) = y/f(x) \le 1$. No such relationship exists between an input distance function and a production frontier. However if a single input is used to produce multiple outputs, then $D_I(y, x) = x/g(y) \ge 1$, where $g(y)$ is an input requirements frontier, or an inverse production frontier.

### 2.2.4 Cost, Revenue, and Profit Frontiers

Thus far we have employed information on the quantities of inputs and outputs to describe the structure of production technology. We

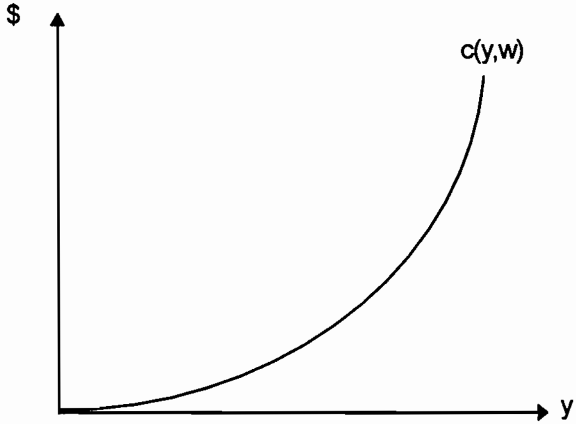


Figure 2.12 A Cost Frontier ($M = 1$)

now add information on the prices of the inputs and the outputs, together with a behavioral assumption, in order to provide additional characterizations of the structure of production technology. These characterizations are provided by the cost, revenue, and profit frontiers. While a joint production frontier describes the best that can be achieved technically, these three frontiers describe the best that can be achieved economically, and so they provide standards against which the economic performance of producers can be measured.

We begin by assuming that producers face a strictly positive vector of input prices given by $w = (w_1, \ldots, w_N) \in R^N_{++}$. We also assume that producers attempt to minimize the cost of producing the output vector $y$ they choose to produce, that cost being $w^T x = \Sigma_n w_n x_n$. Then either the input sets or the input distance function can be used to derive a cost frontier, introduced in Definition 2.12 and illustrated in Figure 2.12.

**Definition 2.12:** A *cost frontier* is a function $c(y, w) = \min_x\{w^T x: x \in L(y)\} = \min_x\{w^T x: D_I(y, x) \geq 1\}$.

If only a single output is produced, the second equality in Definition 2.12 becomes $c(y, w) = \min_x\{w^T x: y \leq f(x)\}$. In Figure 2.12 the cost frontier $c(y, w)$ shows the minimum expenditure required to produce any scalar output, given input prices. The expenditure of each producer must be on or above $c(y, w)$. Thus the cost frontier provides a standard against which to measure the performance of producers for whom the cost minimization assumption is deemed appropriate.

Since the input sets $L(y)$ and the input distance function $D_I(y, x)$ satisfy certain properties, so does the cost frontier $c(y, w)$, which is obtained from them. These properties are

$c1$: $c(0, w) = 0$ and $c(y, w) > 0$ for $y \geq 0$.

$c2$: $c(y, \lambda w) = \lambda c(y, w)$ for $\lambda > 0$.

$c3$: $c(y, w') \geq c(y, w)$ for $w' \geq w$.

$c4$: $c(y, w)$ is a concave function in $w$.

$c5$: $c(y, w)$ is a continuous function in $w$.

$c6$: $c(\lambda y, w) \leq c(y, w)$ for $0 \leq \lambda \leq 1$.

$c7$: $c(y, w)$ is lower semicontinuous in $y$.

If the weak monotonicity property $L5 \Leftrightarrow D_I5$ is replaced with the strong monotonicity property expressed in the second half of $L6 \Leftrightarrow D_I6$, then the weak monotonicity property $c6$ is replaced with

$c8$: $c(y', w) \leq c(y, w)$ for $0 \leq y' \leq y$.

Finally if $G7$ holds, then

$c9$: If $GR$ is convex, then $c(y, w)$ is a convex function in $y$.

If $c(y, w) = \min_x\{w^T x: D_I(y, x) \geq 1\}$ satisfies conditions $\{c1–c5, c7, c8\}$, then $c(y, w)$ is *dual* to $D_I(y, x)$ in the sense that $D_I(y, x) = \min_w\{w^T x: c(y, w) \geq 1\}$ satisfies properties $\{D_I1–D_I3, D_I6, D_I7\}$. In this case $D_I(y, x)$ and $c(y, w)$ provide equivalent representations of the structure of production technology, under the assumption of cost-minimizing behavior in the presence of exogenously determined input prices. Under these circumstances certain features of the structure of production technology, such as its returns to scale properties, can be inferred from the structure of the cost frontier. Thus, for example, constant returns to scale in production can be characterized as

$$L(\lambda y) = \lambda L(y) \Leftrightarrow D_I(\lambda y, x) = \lambda^{-1} D_I(y, x)$$
$$\Leftrightarrow c(\lambda y, w) = \lambda c(y, w), \qquad \lambda > 0.$$

Although our interest centers on situations in which not all producers succeed in minimizing the cost of producing their chosen output vector, the duality relationship linking a cost frontier with an input distance function remains critical for the measurement and decomposition of cost efficiency. This will become apparent in Section 2.4.1.

If the cost frontier $c(y, w)$ is differentiable with respect to input prices, then Shephard's (1953) lemma states that

$$x(y, w) = \nabla_w c(y, w)$$

Thus the vector of cost-minimizing input demand equations can be obtained as the input price gradient of the cost frontier. For a producer operating on the cost frontier, $w^T x = c(y, w)$ and $x = x(y, w)$. For a producer operating above the cost frontier, $w^T x > c(y, w)$ and $x \neq x(y, w)$. The properties of $x(y, w)$ are inherited from those of $c(y, w)$ from which they are derived. For example, if $c(y, w)$ is concave and twice continuously differentiable in $w$, then cost-minimizing input demand equations cannot slope upwards with respect to their own prices. The empirical implication of Shephard's lemma is that it allows $x(y, w)$, alone or in conjunction with $c(y, w)$, to be used to infer certain features of the structure of production technology. It will become apparent in later chapters that these cost-minimizing input demand equations play a central role in econometric analyses of cost efficiency.

Next we assume that producers face a strictly positive vector of output prices given by $p = (p_1, \ldots, p_M) \in R^M_{++}$ and that they seek to maximize the revenue $p^T y = \Sigma_m p_m y_m$ obtainable from the input vector $x$ at their disposal. Then either the output sets $P(x)$ or the output distance function $D_O(x, y)$ can be used to derive a revenue frontier, introduced in Definition 2.13 and illustrated in Figure 2.13.

**Definition 2.13:** A *revenue frontier* is a function $r(x, p) = \max_y \{p^T y: y \in P(x)\} = \max_y \{p^T y: D_O(x, y) \leq 1\}$.

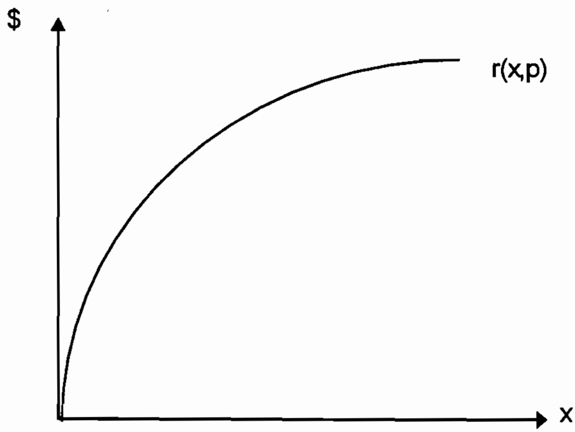


Figure 2.13 A Revenue Frontier ($N = 1$)

If only a single output is produced, the second equality in Definition 2.13 becomes $r(x, p) = \max_y\{py\colon y \le f(x)\} = pf(x)$, since $D_O(x, y) = y/f(x)$. In Figure 2.13 $r(x, p)$ shows the maximum revenue obtainable from any scalar input $x$, given output prices, and so observed revenue must be on or beneath $r(x, p)$. Thus the revenue frontier provides a standard against which to measure the performance of producers for whom the revenue maximization assumption is deemed appropriate.

Since the output sets $P(x)$ and the output distance function $D_O(x, y)$ satisfy certain properties, so does the revenue frontier $r(x, p)$, which is obtained from them. These properties are

$r1$: $r(0, p) = 0$ and $r(x, p) > 0$ for $x \geq 0$.

$r2$: $r(x, \lambda p) = \lambda r(x, p)$ for $\lambda > 0$.

$r3$: $r(x, p') \geq r(x, p)$ for $p' \geq p$.

$r4$: $r(x, p)$ is a convex function in $p$.

$r5$: $r(x, p)$ is a continuous function in $p$.

$r6$: $r(\lambda x, p) \geq r(x, p)$ for $\lambda \geq 1$.

$r7$: $r(x, p)$ is upper semicontinuous in $x$.

If the weak monotonicity property $P5 \Leftrightarrow D_O5$ is replaced with the strong monotonicity property expressed in the second half of $P6 \Leftrightarrow D_O6$, then the weak monotonicity property $r6$ is replaced with

$r8$: $r(x', p) \geq r(x, p)$ for $x' \geq x$.

Finally, if $G7$ holds, then

$r9$: If $GR$ is convex, then $r(x, p)$ is a concave function in $x$.

If $r(x, p) = \max_y\{p^T y\text{: } D_O(x, y) \leq 1\}$ satisfies conditions $\{r1\text{–}r5, r7, r8\}$, then $r(x, p)$ is *dual* to $D_O(x, y)$ in the sense that $D_O(x, y) = \max_p\{p^T y\text{: } r(x, p) \leq 1\}$ satisfies conditions $\{D_O1\text{–}D_O3, D_O6, D_O7\}$. In this case $D_O(x, y)$ and $r(x, p)$ provide equivalent representations of the structure of production technology, under the assumption of revenue-maximizing behavior in the presence of exogenously determined output prices. Just as in the case of duality between $D_I(y, x)$ and $c(y, w)$, duality between $D_O(x, y)$ and $r(x, p)$ enables certain features of the structure of production technology to be inferred from $r(x, p)$. In addition, this duality relationship plays a central role in the measurement and decomposition of revenue efficiency, which is the subject of Section 2.4.2.

If the revenue frontier $r(x, p)$ is differentiable with respect to output prices, then a derivative property similar to Shephard's lemma yields

$$y(x,p) = \nabla_p r(x,p),$$

so that the vector of revenue-maximizing output supply equations is obtained as the output price gradient of the revenue frontier. For a producer operating on the revenue frontier, $p^T y = r(x, p)$ and $y = y(x, p)$. For a producer operating beneath the revenue frontier, $p^T y < r(x, p)$ and $y \neq y(x, p)$. The properties of $y(x, p)$ are inherited from those of $r(x, p)$ from which they are derived. Thus, for example, if $r(x, p)$ is convex and twice continuously differentiable in $p$, then revenue-maximizing output supply equations cannot slope downwards with respect to their own prices. Although revenue frontiers are rarely employed in the econometric analysis of productive efficiency, these revenue-maximizing output supply equations would provide a suitable framework for analysis, just as the cost minimizing input demand

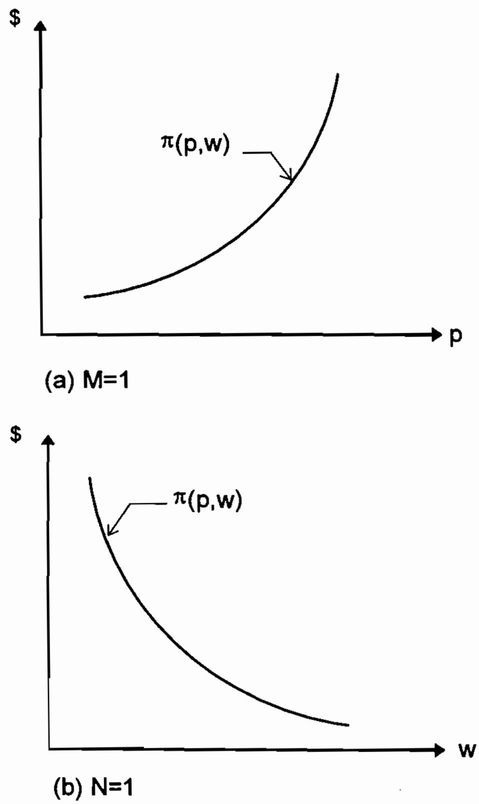


Figure 2.14 A Profit Frontier

equations do in the measurement and decomposition of cost efficiency.

We now assume that producers face strictly positive input prices $w \in R^N_{++}$ and strictly positive output prices $p \in R^M_{++}$, and attempt to maximize the profit $\{p^T y - w^T x\}$ they obtain from using $x \in R^N_+$ to produce $y \in R^M_+$. The graph of production technology can be used to obtain a profit frontier, which is introduced in Definition 2.14 and illustrated in Figure 2.14.

> **Definition 2.14:** A *profit frontier* is a function $\pi(p, w) = \max_{y,x}\{p^T y - w^T x: (y, x) \in GR\}$.

If successful cost-minimizing behavior is assumed, then $\pi(p, w) = \max_y\{p^Ty - c(y, w)\}$, since $w^Tx = c(y, w)$. Alternatively, if successful revenue-maximizing behavior is assumed, then $\pi(p, w) = \max_x\{r(x, p) - w^Tx\}$, since $p^Ty = r(x, p)$. In Figure 2.14 $\pi(p, w)$ shows the maximum profit obtainable from a given scalar output price in panel (a) and the maximum profit obtainable from a given scalar input price in panel (b). In each panel observed profit must be on or beneath $\pi(p, w)$, and so $\pi(p, w)$ provides a standard against which to measure the performance of producers for whom the profit maximization objective is deemed appropriate.

Since $GR$ satisfies certain properties, so does $\pi(p, w)$. In addition to the properties listed in Section 2.2.1, we assume that $GR$ exhibits strict decreasing returns to scale, since constant returns to scale would imply that either $\pi(p, w) = 0$ or $\pi(p, w) = +\infty$. With this in mind, the properties satisfied by $\pi(p, w)$ are

$\pi1$: $\pi(p', w) \geq \pi(p, w)$ for $p' \geq p$.

$\pi2$: $\pi(p, w') \leq \pi(p, w)$ for $w' \geq w$.

$\pi3$: $\pi(\lambda p, \lambda w) = \lambda\pi(p, w)$ for $\lambda > 0$.

$\pi4$: $\pi(p, w)$ is a convex function in $(p, w)$.

If $\pi(p, w) = \max_{y,x}\{p^Ty - w^Tx\colon (y, x) \in GR\}$ satisfies $\{\pi1\text{–}\pi4\}$, then $\pi(p, w)$ and $GR$ are *dual*, in the sense that $GR = \{(y, x)\colon p^Ty - w^Tx \leq \pi(p, w)\}$ satisfies $\{G1\text{–}G3, G6, G7\}$. In this case $\pi(p, w)$ and $GR$ provide equivalent representations of the structure of production technology, under the assumption of profit-maximizing behavior in the presence of exogenously determined output prices and input prices.

If $\pi(p, w)$ is differentiable, then Hotelling's (1932) lemma states that

$$y(p,w) = \nabla_p \pi(p,w),$$

$$-x(p,w) = \nabla_w \pi(p,w).$$

Thus the vectors of profit-maximizing output supply and input demand equations can be obtained from the profit frontier as the output price gradient and the negative of the input price gradient, respectively. For a producer operating on the profit frontier, $\{p^T y - w^T x\} = \pi(p, w)$ and $y = y(p, w), x = x(p, w)$. For a producer operating beneath the profit frontier, $\{p^T y - w^T x\} < \pi(p, w)$ and either $y \neq y(p, w)$ or $x \neq x(p, w)$ or both.

Particularly in the single-output case, it is frequently convenient to work with a normalized profit frontier. Since the profit frontier $\pi(p, w)$ is homogeneous of degree +1 in $(p, w)$, it is possible to divide maximum profit $\pi(p, w)$ by $p > 0$ to obtain

**Definition 2.15:** Let $M = 1$. A *normalized profit frontier* $\pi^*(w/p) = \pi(p, w)/p = \max_{y,x}\{y - (w/p)^T x: (y, x) \in GR\}$.

The normalized profit frontier $\pi^*(w/p)$ is nonincreasing, convex, and homogeneous of degree 0 in $(w; p)$. For a normalized profit frontier Hotelling's lemma generates

$$-x\left(\frac{w}{p}\right) = \nabla_{(w/p)}\pi^*\left(\frac{w}{p}\right),$$

$$y\left(\frac{w}{p}\right) = \pi^*\left(\frac{w}{p}\right) - \left(\frac{w}{p}\right)^T \nabla_{(w/p)}\pi^*\left(\frac{w}{p}\right).$$

A normalized profit frontier can also be defined in the multiple-output case, with the normalizing price being any positive output price or any positive input price.

### 2.2.5 Variable Cost Frontiers and Variable Profit Frontiers

It may not be possible for producers to minimize cost, or to maximize profit, if some inputs are fixed, perhaps by contractual arrangement. In this case the focus shifts from a cost frontier to a variable cost frontier, and from a profit frontier to a variable profit frontier. In this section we briefly discuss the properties of variable cost frontiers and variable profit frontiers. We leave the treatment of variable revenue frontiers to the reader.

We begin by assuming that producers use variable input vector $x \in R_+^N$, available at prices $w \in R_{++}^N$, and fixed input vector $z \in R_+^Q$, to produce output vector $y \in R_+^M$. We assume that producers seek to minimize the variable cost $w^T x$ required to produce $y$, given technology and $(w, z)$. A variable cost frontier can be defined as

**Definition 2.16:** A *variable cost frontier* is a function $vc(y, w, z) = \min_x\{w^T x: (y, x, z) \in GR\}$.

The variable cost frontier $vc(y, w, z)$ shows the minimum expenditure on variable inputs required to produce output vector $y$ when variable input prices are $w$ and fixed input quantities are $z$. Consequently $w^T x \geq vc(y, w, z)$, and $vc(y, w, z)$ becomes the standard against which to measure the performance of producers for whom the variable cost minimization assumption is appropriate.

The properties satisfied by $vc(y, w, z)$ are similar to those satisfied by the cost frontier $c(y, w)$. Thus $vc(y, w, z)$ is a nonnegative function that is homogeneous of degree +1 and concave in $w$ for given $(y, z)$. Under a weak monotonicity property akin to *G*6, $vc(y, w, z)$ is also nondecreasing in $w$ for given $(y, z)$, nondecreasing in $y$ for given $(w, z)$, and nonincreasing in $z$ for given $(y, w)$. Under a convexity property akin to *G*7, $vc(y, w, z)$ is a convex function in $(y, z)$ for given $w$. If $vc(y, w, z)$ is differentiable in the elements of $w$, then Shephard's lemma states that variable input demand equations are given by $x(y, w, z) = \nabla_w vc(y, w, z)$. Finally, if $vc(y, w, z)$ is differentiable in the elements of $z$, then a vector of shadow prices for the fixed inputs is given by $q^s = -\nabla_z vc(y, w, z)$.

We now assume that producers sell their outputs at prices $p \in R^M_{++}$ and attempt to maximize variable profit, the difference between total revenue $p^T y$ and variable cost $w^T x$, given technology and $(p, w, z)$. A variable profit frontier can be defined as

**Definition 2.17:** A *variable profit frontier* is a function $v\pi(p, w, z) = \max_{y,x}\{p^T y - w^T x: (y, x, z) \in GR\}$.

The variable profit frontier $v\pi(p, w, z)$ shows the maximum excess of total revenue over variable cost when output prices are $p$, variable input prices are $w$, and fixed input quantities are $z$. It follows that $(p^T y - w^T x) \leq v\pi(p, w, z)$, and $v\pi(p, w, z)$ becomes the standard against which to measure the performance of producers for whom the variable profit maximization assumption is appropriate.

Suppose that *GR* satisfies properties akin to {*GR*1–*GR*3, *GR*6, *GR*7}, extended from $(y, x)$ to $(y, x, z)$. Suppose also that *GR* is a cone, so that technology satisfies constant returns to scale in $(y, x, z)$. Then the variable profit frontier $v\pi(p, w, z)$ is a nonnegative function that is (i) convex and homogeneous of degree +1 in $(p, w)$ for given

$z$; (ii) nondecreasing in $p$ and nonincreasing in $w$ for given $z$; and (iii) nondecreasing, concave, and homogeneous of degree +1 in $z$ for given $(p, w)$. If $v\pi(p, w, z)$ is differentiable in the elements of $(p, w)$, then Hotelling's lemma states that output supply equations are given by $y(p, w, z) = \nabla_p v\pi(p, w, z)$ and variable input demand equations are given by $x(p, w, z) = -\nabla_w v\pi(p, w, z)$. Finally if $v\pi(p, w, z)$ is differentiable in the elements of $z$, $\nabla_z v\pi(p, w, z) = q^s$, $q^s$ being a vector of shadow prices of the fixed inputs. Note that this notion of shadow prices differs from the notion introduced within the context of a variable cost frontier because the objective of producers is different.

## 2.3 TECHNICAL EFFICIENCY

In Section 2.2 we introduced various types of frontiers. A production frontier exploits only input and output quantity data, while cost, revenue, and profit frontiers exploit input and/or output quantity data, together with input and/or output price data and a behavioral assumption as well. The next step is to introduce measures of distance to each of these frontiers, with distances providing measures of technical or economic efficiency. In this section we introduce a pair of measures of distance to a production frontier, these distances providing measures of technical efficiency. In Section 2.3.1 we define technical efficiency, and we introduce the two measures of technical efficiency. In Section 2.3.2 we discuss the two measures of technical efficiency under the assumption that producers produce a single output, and in Section 2.3.3 we discuss the two measures of technical efficiency under the assumption that producers produce multiple outputs.

### 2.3.1 Definitions and Measures of Technical Efficiency

Generally speaking, technical efficiency refers to the ability to minimize input use in the production of a given output vector, or the ability to obtain maximum output from a given input vector. A formal definition of technical efficiency, due to Koopmans (1951), is provided in Definition 2.18. Two special cases of Koopmans' definition, the first being input oriented and the second being output oriented, are provided in Definitions 2.19 and 2.20. Following these definitions of technical efficiency, two measures of technical efficiency, the first being input oriented and the second being output oriented, are provided in Definitions 2.21 and 2.22. These measures were first proposed by Debreu (1951) and Farrell (1957), and so they are often referred to jointly as the Debreu–Farrell measures of technical efficiency.

**Definition 2.18:** An output–input vector $(y, x) \in GR$ is *technically efficient* if, and only if, $(y', x') \notin GR$ for $(y', -x') \geq (y, -x)$.

**Definition 2.19:** An input vector $x \in L(y)$ is *technically efficient* if, and only if, $x' \notin L(y)$ for $x' \leq x$ or, equivalently, $x \in \text{Eff } L(y)$.

**Definition 2.20:** An output vector $y \in P(x)$ is *technically efficient* if, and only if, $y' \notin P(x)$ for $y' \geq y$ or, equivalently, $y \in \text{Eff } P(x)$.

Definition 2.18 calls a feasible output–input vector technically efficient if, and only if, no increase in *any* output or decrease in *any* input is feasible. Definition 2.19 holds the output vector fixed and calls a feasible input vector technically efficient if, and only if, no reduction in *any* input is feasible (recall Definition 2.5). Definition 2.20 holds the input vector fixed and calls a feasible output vector technically efficient if, and only if, no increase in *any* output is feasible (recall Definition 2.7). Thus technical efficiency is defined in terms of membership in an efficient subset. Definitions 2.19 and 2.20 of technical efficiency should be carefully compared with the following two measures of technical efficiency.

**Definition 2.21:** An *input-oriented measure of technical efficiency* is a function $TE_I(y, x) = \min\{\theta: \theta x \in L(y)\}$.

**Definition 2.22:** An *output-oriented measure of technical efficiency* is a function $TE_O(x, y) = [\max\{\phi: \phi y \in P(x)\}]^{-1}$.

Definitions 2.21 and 2.22 measure technical efficiency in terms of equiproportionate contraction of all inputs and equiproportionate expansion of all outputs, respectively. If no equiproportionate contraction of *all* inputs is feasible, that input vector is called technically efficient, whereas if no equiproportionate expansion of *all* outputs is feasible, that output vector is called technically efficient. Recalling Definitions 2.4 and 2.6, the Debreu–Farrell measures of technical efficiency associate technical efficiency with membership in isoquants. Since membership in isoquants is necessary, but not sufficient, for membership in efficient subsets, it follows that technical efficiency on the basis of the Debreu–Farrell measures is necessary, but not sufficient, for technical efficiency on the basis of Koopmans' definitions.

Throughout this book, and also throughout the vast majority of econometric work, technical efficiency is measured radially, using isoquants as standards, following Debreu and Farrell. It is natural to ask why technical efficiency is not measured relative to the more exacting standards provided by efficient subsets, following Koopmans. The answer is that radial measures have nice properties, and using efficient subsets as standards would require nonradial measures. The properties of the two Debreu–Farrell measures of technical efficiency are given in Proposition 2.1.

**Proposition 2.1:** The input-oriented measure of technical efficiency $TE_I(y, x)$ satisfies the properties:

(i) $TE_I(y, x) \leq 1$.

(ii) $TE_I(y, x) = 1 \Leftrightarrow x \in \text{Isoq } L(y)$.

(iii) $TE_I(y, x)$ is nonincreasing in $x$.

(iv) $TE_I(y, x)$ is homogeneous of degree $-1$ in $x$.

(v) $TE_I(y, x)$ is invariant with respect to the units in which $y$ and $x$ are measured.

The output-oriented measure of technical efficiency $TE_O(x, y)$ satisfies the properties:

(i) $TE_O(x, y) \leq 1$.

(ii) $TE_O(x, y) = 1 \Leftrightarrow y \in \text{Isoq } P(x)$.

(iii) $TE_O(x, y)$ is nondecreasing in $y$.

(iv) $TE_O(x, y)$ is homogeneous of degree $+1$ in $y$.

(v) $TE_O(x, y)$ is invariant with respect to the units in which $x$ and $y$ are measured.

All but one of these properties are desirable properties that any measure of technical efficiency should satisfy. Since both $TE_I(y, x)$ and $TE_O(x, y)$ are used to measure technical efficiency throughout the book, it is worthwhile to discuss these properties. The first property is a normalization property, which states that both $TE_I(y, x)$ and $TE_O(x, y)$ are bounded above by unity. [Elsewhere in the literature $TE_O(x, y)$ is frequently defined without the reciprocal operation; for such a definition $TE_O(x, y)$ is bounded below by unity.] The third property is a weak monotonicity property, which states that $TE_I(y, x)$ does not increase when usage of any input increases and that $TE_O(x, y)$ does not decrease when production of any output increases. The fourth property is a homogeneity property, which states that an equiproportionate change in all inputs results in an equivalent change in the opposite direction in $TE_I(y, x)$ and an equiproportionate change in all outputs results in an equivalent change in the same direction in $TE_O(x, y)$. The final property is an invariance property, which states that if the units in which any output or any input is measured are changed (say, from acres to hectares), efficiency scores are unaffected.

The second property is the only undesirable property. It states that $TE_I(y, x)$ and $TE_O(x, y)$ use the relaxed standards Isoq $L(y)$ and Isoq $P(x)$, rather than the more stringent standards Eff $L(y)$ and Eff $P(x)$, to measure technical efficiency. Property (ii) can be strengthened to $TE_I(y, x) = 1 \Leftrightarrow x \in \text{Eff } L(y)$ and $TE_O(x, y) = 1 \Leftrightarrow y \in \text{Eff } P(x)$, provided production technology satisfies a sufficiently strong monotonicity condition to guarantee that Eff $L(y)$ = Isoq $L(y)$ and Eff $P(x)$ = Isoq $P(x)$. This strong monotonicity condition is satisfied by a Cobb–Douglas production frontier, but it is not necessarily satisfied by a flexible production frontier such as translog, and the translog functional form is widely used in empirical efficiency analysis. Alternatively, technical efficiency can be defined relative to efficient subsets, but this would require replacing the radial Debreu–Farrell measures with nonradial measures that would not satisfy the homogeneity and invariance properties. In the econometric literature the tradeoff between radial efficiency measures satisfying the undesirable property (ii) and nonradial efficiency measures failing to satisfy the desirable properties (iv) and (v) has been resolved in favor of Debreu and Farrell.

In some cases it is appropriate to adopt an input-conserving approach to the measurement of technical efficiency, whereas in other cases it is appropriate to adopt an output-expanding approach. This makes it desirable to know the nature of the relationship between the input oriented measure and the output-oriented measure. The relationship between $TE_I(y, x)$ and $TE_O(x, y)$ is given by

**Proposition 2.2:** $TE_I(y, x) = TE_O(x, y)\ \forall (y, x) \in GR \Leftrightarrow L(\lambda y) = \lambda L(y) \Leftrightarrow P(\lambda x) = \lambda P(x)$.

Thus $TE_I(y, x)$ and $TE_O(x, y)$ assign the same technical efficiency score to a producer if, and only if, technology is homogeneous of degree +1; that is, technology is characterized by constant returns to scale. This condition is stringent, and is unlikely to be satisfied in empirical work. Consequently it is very likely that a ranking of producers will be sensitive to the orientation of the efficiency measurement. This makes it essential to choose the orientation with care.

### 2.3.2 Single-Output Production Frontiers and the Measurement of Technical Efficiency

We now consider the case in which producers use multiple inputs to produce a single output. In this case we can use Definition 2.9 to simplify the two Debreu–Farrell measures of technical efficiency. The single-output versions of these technical efficiency measures are given in Definitions 2.23 and 2.24, and illustrated in Figures 2.15–2.17.

**Definition 2.23:** If only a single output is produced, an *input-oriented measure of technical efficiency* is given by the function $TE_I(y, x) = \min\{\theta: y \leq f(\theta x)\}$.

**Definition 2.24:** If only a single output is produced, an *output-oriented measure of technical efficiency* is given by the function $TE_O(x, y) = [\max\{\phi: \phi y \leq f(x)\}]^{-1}$.

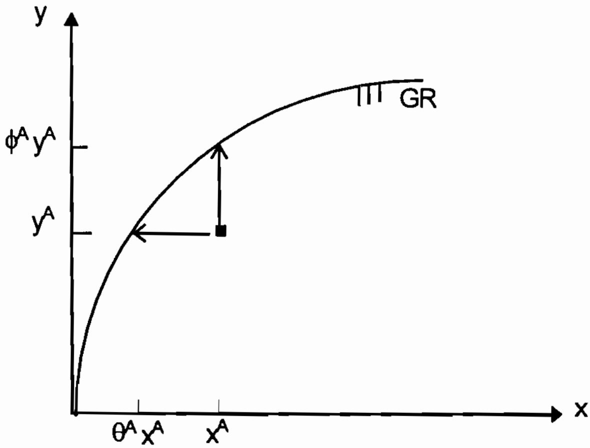


Figure 2.15 Input-Oriented and Output-Oriented Measures of Technical Efficiency ($M = 1, N = 1$)

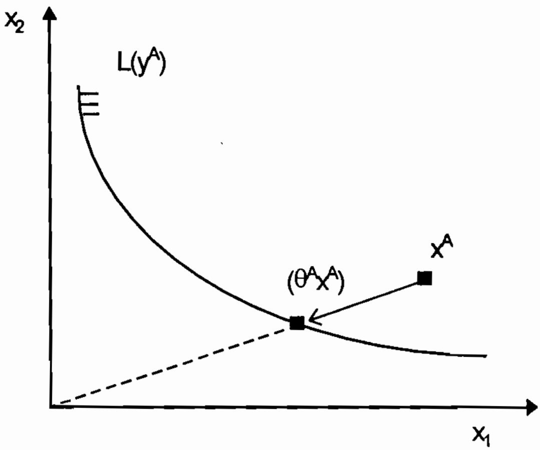


Figure 2.16 An Input-Oriented Measure of Technical Efficiency ($N = 2$)

Figure 2.15 uses the production frontier $f(x)$ to illustrate both measures of technical efficiency. A producer using $x^A$ to produce $y^A$ is technically inefficient, since it operates beneath $f(x)$. $TE_I(y^A, x^A)$ measures the maximum contraction of $x^A$ that enables continued

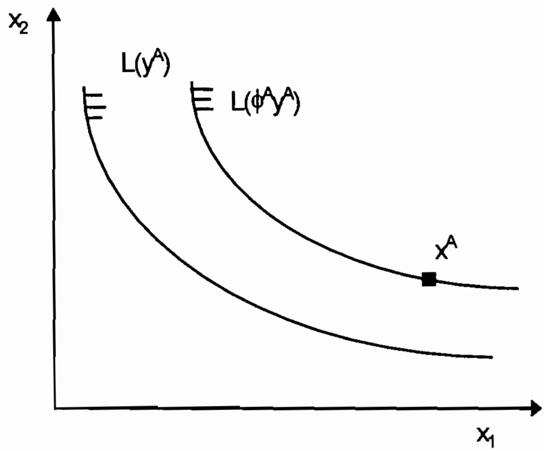


Figure 2.17 An Output-Oriented Measure of Technical Efficiency ($N = 2$)

production of $y^A$, and $TE_I(y^A, x^A) = \theta^A < 1$, since $y^A = f(\theta^A x^A)$. $TE_O(x^A, y^A)$ measures the reciprocal of the maximum expansion of $y^A$ that is feasible with $x^A$, and $TE_O(x^A, y^A) = (\phi^A)^{-1} < 1$, since $\phi^A y^A = f(x^A)$. Figure 2.16 uses the input set $L(y)$ and its isoquant Isoq $L(y)$ to illustrate the input-oriented measure of technical efficiency. $TE_I(y^A, x^A)$ measures the maximum radial contraction in $x^A$ that enables continued production of $y^A$, and $TE_I(x^A, y^A) = \theta^A < 1$, since $\theta^A x^A \in$ Isoq $L(y^A)$. Figure 2.17 uses the input set $L(y)$ and its isoquant Isoq $L(y)$ to illustrate the output-oriented measure of technical efficiency. The reciprocal of $TE_O(x^A, y^A)$ measures the maximum expansion of $y^A$ that is feasible with inputs $x^A$, and $TE_O(x^A, y^A) = (\phi^A)^{-1} < 1$, since $x^A \in$ Isoq $L(\phi^A y^A)$.

### 2.3.3 Multiple-Output Distance Functions and the Measurement of Technical Efficiency

We now assume that producers use multiple inputs to produce multiple outputs. The analytical framework is very similar to that of Section 2.3.2; the only difference is that the (single-output) production frontier is replaced with distance functions. Input distance functions are used to define an input-oriented measure of technical

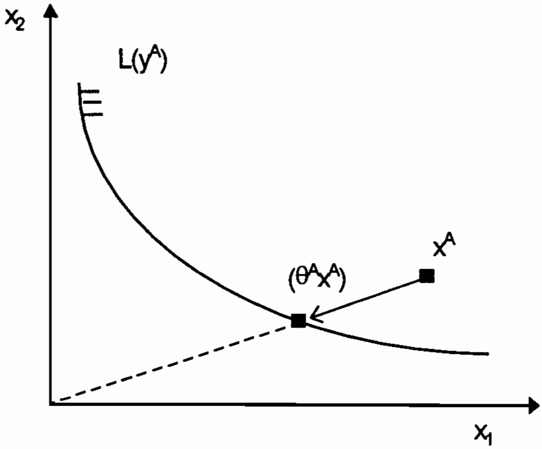


Figure 2.18 An Input-Oriented Measure of Technical Efficiency ($N = 2$)

efficiency, and output distance functions are used to define an output-oriented measure of technical efficiency. The following two definitions are straightforward extensions of Definitions 2.23 and 2.24 to the multiple-output case.

> **Definition 2.25:** If any number of outputs is produced, an *input-oriented measure of technical efficiency* is given by the function $TE_I(y, x) = \min\{\theta: D_I(y, \theta x) \geq 1\}$.

> **Definition 2.26:** If any number of outputs is produced, an *output-oriented measure of technical efficiency* is given by the function $TE_O(x, y) = [\max\{\phi: D_O(x, \phi y) \leq 1\}]^{-1}$.

In Figure 2.18 the input-oriented measure of the technical efficiency of producer $(x^A, y^A)$ is given by $TE_I(y^A, x^A) = \theta^A < 1$, since $\theta^A x^A \in \text{Isoq } L(y^A)$. In Figure 2.19 the output-oriented measure of the technical efficiency of producer $(x^A, y^A)$ is given by $TE_O(x^A, y^A) = (\phi^A)^{-1} < 1$, since $\phi^A y^A \in \text{Isoq } P(x^A)$.

The efficiency measures introduced in Definitions 2.25 and 2.26, and illustrated in Figures 2.18 and 2.19, look very much like the distance functions introduced in Definitions 2.10 and 2.11, and illustrated in Figures 2.9 and 2.11. This should not be surprising, since distance functions provide radial measures of the distance from an

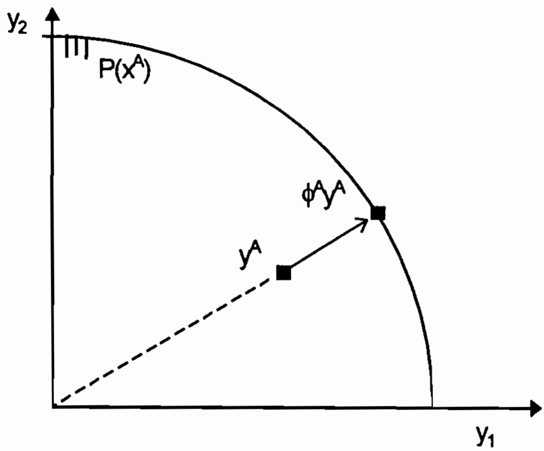


Figure 2.19 An Output-Oriented Measure of Technical Efficiency ($M = 2$)

output–input bundle to the boundary of production technology. The relationships between distance functions and radial efficiency measures are provided in

**Proposition 2.3:** $TE_I(y, x) = [D_I(y, x)]^{-1}$ and $TE_O(x, y) = D_O(x, y)$.

The input-oriented measure of technical efficiency $TE_I(y, x)$ is the reciprocal of the input distance function $D_I(y, x)$, and the output-oriented measure of technical efficiency $TE_O(x, y)$ coincides with the output distance function $D_O(x, y)$. Thus as we indicated in Section 2.2.3, distance functions are intimately related to the measurement of technical efficiency. This should be clear from a comparison of the properties of $TE_I(y, x)$ and $TE_O(x, y)$ given in Proposition 2.1 with the properties of $D_I(y, x)$ and $D_O(x, y)$ given in Section 2.2.3.

## 2.4 ECONOMIC EFFICIENCY

In Section 2.3 we introduced a pair of measures of technical efficiency, for both the single-output case and the multiple-output