APPLIED SYSTEMS ANALYSIS

Engineering Planning and Technology Management

Richard de Neufville
Massachusetts Institute of Technology

McGraw-Hill Publishing Company
New York St. Louis San Francisco Auckland Bogotá Caracas Hamburg
Lisbon London Madrid Mexico Milan Montreal New Delhi Oklahoma City
Paris San Juan São Paulo Singapore Sydney Tokyo Toronto

CHAPTER 2

PRODUCTION FUNCTIONS

2.1 CONCEPT

Production functions constitute the most basic models of applied systems analysis. They are attractive because they are both simple and powerful. They are simple in that they may consist of only one formula or computer routine; yet they are powerful in that this single expression can effectively summarize an enormous amount of detailed engineering data. Production functions thus serve as a fundamental means to discuss alternative designs of a system.

A *production function* represents the *technically efficient* transformation of *physical resources* into *products*. The words stressed are important to the definition and need to be carefully understood.

A production function is an engineering model first of all. It represents a relationship between physical quantities. Specifically, it describes how various resources—the inputs to the process—combine to form some products—the outputs of the transformation. A production function for air transportation, for example, would relate the several inputs to the process (aircraft of different types, fuel, pilots, ground equipment, etc.) to the various outputs (number of passengers, tons of cargo carried, etc.)

A production function represents *technically efficient* combinations of resources. It is technically efficient in that each point on the production function represents the maximum product that can be obtained from any given set

of resources. The production function therefore excludes any lesser amount of product that would come from a wasteful or technically inefficient use of these resources. The production function for air transportation thus expresses the maximum that can be carried for any number of pilots, aircraft, and so on—not the lesser amount that would arise if management were careless, pilots were lazy or the aircraft were flown in circles. By definition, a production function is the locus of all technically efficient combinations of resources.

Semantic caution: Ordinary language confuses two distinct concepts: *technical efficiency*, described here, and *economic efficiency*, which in brief is whether a design is best economically (see Chapter 4). The one does not imply the other. Technical efficiency is a necessary but not a sufficient condition for economic efficiency. The latter can clearly only be obtained if resources are not wasted. A design may be technically efficient but economically inefficient, however. For example, a design for a steel building might be technically efficient in that it could carry the most load for the least steel, and yet be economically inefficient compared to a concrete structure—due perhaps to the much higher cost of steel in that region.

The production function can represent very complex situations. It quite easily incorporates hundreds and even thousands of different resources, as for example in a linear programming problem (see Section 5.2). More commonly, production functions deal with only a handful of variables. This is because of our own desire to simplify the description of the design into something we can easily comprehend.

The principal limitation of the production function is that it can now, as a practical matter, only handle a limited number of products. Production functions are thus typically expressed in terms of a single product. This limitation is in part due to the lack of appropriate theory to describe processes that create multiple products, as Section 2.5 describes. It is also due to our limited ability to optimize efficiently over multiple products or objectives, which Chapter 8 describes.

In view of this limitation, most of the subsequent discussion focusses on production functions for single products. This does not mean that production functions can in practice only describe situations with one product; it does mean that they apply principally when—as is generally the case—different products can be aggregated into a single category. A production function thus would describe an automobile factory in terms of its production of "cars" rather than in terms of two-door cars, four-door cars with and without automatic transmission, and so on.

2.2 SINGLE OUTPUT

Single-output production functions are both the most common kind and the easiest to describe. Mathematically, they are of the form

$$Y = g(\mathbf{X}) = g(X_1, \ldots, X_i, \ldots, X_n)$$

where Y is the product and $\mathbf{X}$ is the vector of n different resources X_i.

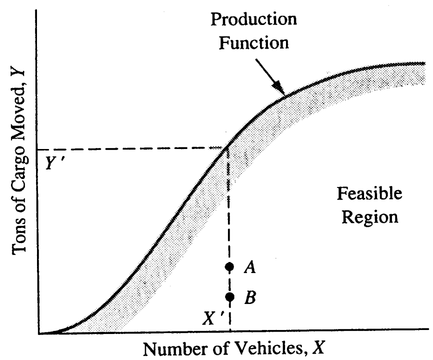

FIGURE 2.1
The production function is the upper bound to the feasible region of production.

By definition, the production function is technically efficient. This means that Y is the maximum product that can be obtained for any set of resources $\mathbf{X}$. To illustrate this, consider the hypothetical production function in Figure 2.1, describing the "tons of cargo moved" as a product in terms of the "number of vehicles" as the single resource. For any set of resources X', there are many levels of product that could be achieved, depending on how efficiently the system is organized. One might get as little as A or B. The maximum one can obtain, Y', is the point on the production function corresponding to X'. The entire production function itself is the locus of all points Y'. It is the solid line in Figure 2.1.

The production function is thus the bound to the feasible region of production. This *feasible region* consists of all the levels it is possible to produce with each X'. It thus includes levels A and B as well as Y'. When products and resources are shown as in Figure 2.1, the feasible region is below the production function.

A production function in terms of many resources is a surface. In three dimensions (one product, two resources) this is easy to visualize: it is a shell or dome of some sort. In a higher number of dimensions—in *hyperspace*, that is—the nature of the surface is more difficult to imagine. Yet it always has the same kinds of properties as in conventional space.

2.3 MATHEMATICAL REPRESENTATIONS

There are two basic ways of representing production functions: the deductive and the inductive. Each has quite different implications for the cost and accuracy of the analysis. Both are useful in practice.

The deductive approach uses a convenient functional form, whose parameters are subsequently estimated statistically to reflect reality as closely as possible. The advantage of this approach is that the choice of the right form leads to relatively simple calculations and to results that are easy to interpret. The disadvantage is that any chosen functional form implies a number of restrictions; it is a pattern that has only a limited number of shapes. The choice of a functional form thus will distort our view of reality to some degree and may hide important features of the production function.

The inductive approach synthesizes the production function from a detailed understanding of the physical processes involved. The mathematical representation may be in some explicit functional form, but it is more likely to be generated by extensive calculations that simulate the performance of different designs. This approach can have the merit of being much more accurate. It has the twin disadvantages of requiring far more detailed, laborious efforts, and of not having a functional form that can be easily interpreted.

Deductive models. The best-known and most widely cited deductive model of the production function is the Cobb-Douglas model:

$$Y = a_0 \prod X_i^{a_i} = a_0 X_1^{a_1} X_2^{a_2} \ldots X_n^{a_n}$$

[Note: the symbol Π represents the multiplication to all i terms. It is similar to $\sum$. Both will be subscripted only where several possible indices for terms might cause ambiguity.]

The Cobb-Douglas model dominated the literature and practice until the early 1970s, until its deficiencies became too obvious and practical alternatives became available through the development of high-speed computers. It has several attractive features:

- It can, by suitable choice of the parameters a_i, represent a broad range of the important characteristics of production functions (see Section 2.4).
- These parameters can be immediately interpreted in terms of the characteristics of the production functions, which is a great pedagogical and practical advantage.
- It is easy to estimate statistically. This is because the multiplicative equation can be transformed into a linear function by expressing variables in terms of their logarithms:

$$\log\ Y = a_0 + \sum a_i \log X_i$$

The parameters can then be estimated by statistical regression analysis designed to find the best fit of a line to data.

The weakness of the Cobb-Douglas function lies in its inability to represent real situations fully. The model inherently implies that certain features of the production process—as defined by the a_i as described in Section 2.4—remain the

Examples of Cobb-Douglas Model

A classic use of the Cobb-Douglas model was Nerlove's (1965) analysis of the production of electric power. He related output to the inputs of labor, capital, and fuel. By statistical regression he obtained

$$\text{output} = a_0(\text{labor})^{0.78}(\text{capital})^{0.00}(\text{fuel})^{0.61}$$

Further analysis underlines the difficulties created by the Cobb-Douglas requirement that its parameters be constant throughout the range of inputs and output. When Nerlove divided his analysis by level of production he obtained quite different results. For the lowest levels of production he obtained

$$\text{output} = a_0(\text{labor})^{1.45}(\text{capital})^{0.18}(\text{fuel})^{1.29}$$

and for the highest levels he got

$$\text{output} = a_o(\text{labor})^{0.84}(\text{capital})^{0.10}(\text{fuel})^{0.75}$$

These later results indicate that the first model simplified the real situation by, in effect, imposing single values of the a_i parameters (such as the 0.78 exponent for labor) when different values should apply for different sizes of plants.

same regardless of the level of production. These constancies may not exist (see box).

The model that is now widely recommended is the translog model:

$$\log Y = a_0 + \sum a_i \log X_i + \sum \sum a_{ij}(\log X_i)(\log X_j)$$

It is essentially the Cobb-Douglas model supplemented with interaction terms between resources.

The translog model overcomes the obvious disadvantages of the Cobb-Douglas formulation, while maintaining most of its advantages. The extra interactive terms make it possible to represent processes with complex features that the Cobb-Douglas model cannot cope with. Yet its linearity permits estimation of the parameters by standard techniques. The calculations are extensive, but proceed quickly on modern computers. The principal drawbacks of the translog model are that the characteristics of the production function cannot be read directly from the model, and it does require considerable effort and data to obtain.

Inductive models. Inductive models of the production function are based on a detailed understanding of the mechanisms that produce a particular product. Typically this perspective is gained by use of technical relationships. These models are thus widely known as engineering models of the production function.

The technical relationships can sometimes be expressed analytically, in terms of one or more explicit equations. These situations are convenient because they make it possible to compute the production function rapidly with the assistance of modest computers or even calculators. Unfortunately, such analytic situations are relatively rare; they mostly occur when the production process occurs in a fairly homogeneous, continuous force field. Moving cargo up a river is a classic example of this (see detailed discussion in Section 2.6).

Most frequently, the engineering production functions have to be simulated through quite extensive computer programs. These are typically expensive since they require enormous effort to identify the relevant data and create the program. The actual calculations may not be too difficult once the program is finished, but a substantial overall effort is required in any case. The simulations typically mimic the detailed design an engineer would have to undertake to establish the result or product of a particular design or combination of inputs. The computer program must thus include all the information required, specifically, extensive tables of discrete data (e.g., strength of specific sizes of steel plate) or non-analytic experimental results (e.g., engine performance). The production function is established by generating hundreds, if not thousands, of designs and identifying the envelope to the production possibilities. This outer limit is, by definition, the production function.

Examples of both an analytic and a simulation engineering production function appear in the accompanying box. Since the models cannot be expressed in a simple equation, they are shown graphically by means of contour lines of equal levels of productions. These contours, known technically as isoquants, are described in detail in the next section.

Inductive Models of Production Functions

A. Analytic Model: Cargo Transport in a River

The basic equation is that the effective push generated by the engines in the barges must equal the resistance of the water to the barges. Both elements can be expressed in terms of the design parameters of the barges: their horsepower, length, width, and depth. Hoffmeister and de Neufville (1973, 1976) solved the equation to obtain the product, ton-kilometers per hour, as a function of the horsepower and size of the barge (see Figure 2.2*a* and Section 2.6).

B. Simulation Model: Thermal Efficiency in Electricity

In another classic example, Cootner and Lof (1965, 1975) established the production function for thermal efficiency in steam generation of electricity in terms of boiler temperature and pressure. The isoquants, in Figure 2.2*b*, were obtained from engineering data and industry estimates. Note that the scales are logarithmic.

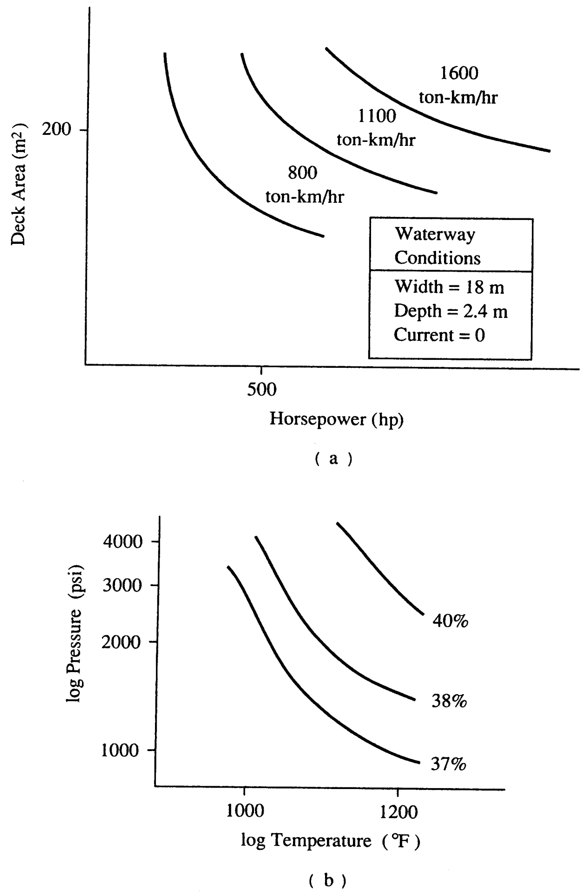

FIGURE 2.2
Examples of engineering production functions.

2.4 CHARACTERISTICS

The characteristics of the production function—its shape, slope, and smoothness—are important determinants of the kind of optimization techniques that can be usefully applied. It is therefore necessary to understand these characteristics both in general and in detail. Descriptions of production functions use a wide range of expressions that differ from those traditional in mathematics. They are based on concepts that have specific physical and economic interpretations, and are particularly useful for our purposes. Because they are different, they require particular attention. The following paragraphs develop these concepts.

Note carefully that few of the expressions used to describe production functions apply naturally, if at all, to situations where there are several products. This is a reflection of the fact that the profession has only recently begun to analyze systems in their full detail.

Isoquants. An isoquant is a locus on the production function of all equal levels of product. The term "isoquant" is, in fact, simply a word constructed to mean "equal quantity." Isoquants are easier to visualize when the production function is a shell in three-dimensional space: they are then simply cuts of the shell at specified levels of product. Figure 2.3 illustrates this for a hypothetical production function describing ton-kilometers per hour produced by barges built with different horsepower and deck area. As the figure suggests, the isoquant is like a contour on a mountain.

Isoquants are useful for describing production functions, particularly those that do not have convenient functional forms. They are thus almost always used to illustrate engineering production functions, as was done for the models illustrated in Figures 2.2.

The isoquants also illustrate an important phenomenon: any level of product can in general be obtained by many different combinations of resources, each technically efficient. Figure 2.2*a* shows that one can move a given amount of cargo either by moving a small barge quickly or a larger barge more slowly. Both these combinations, indeed all those on an isoquant, are technically efficient. There are thus no technical grounds to choose between the designs. The choice then rests with economic considerations, as Chapter 4 explains.

Isoquants for real situations often have a distinct shape: they tend to be asymptotic to the axes of the inputs, as indicated in Figure 2.2. This shape

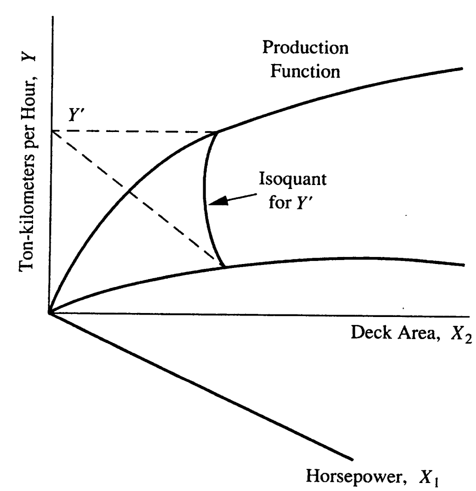

FIGURE 2.3
Example of an isoquant.

reflects the general necessity of having some of each input in the process in order to attain the production. Thus in the production function for river transport shown by Figure 2.2*a* we must have both capacity (represented by deck area > 0) and propulsion (horsepower > 0) in order to attain any output. In a few cases, inputs can substitute absolutely for each other. Either copper or aluminum can, for example, be used to make electric cable. In cases involving the possibility of absolute substitution, the isoquants can intercept the axes of the inputs. The general rule in practice, however, is that they do not intercept the axes but move away from them as they do in Figure 2.2.

Marginal products. A marginal product is the change in output due to a unit change in a specific input. Formally, for the production function $Y = g(\mathbf{X})$, the marginal product MP_i for input X_i is

$$MP_i = \frac{\partial Y}{\partial X_i}$$

This concept now only applies to single output production functions. No practical equivalent term has been developed for multiple-product models.

Empirically, marginal products typically follow a distinctive pattern, that of diminishing monotonically as X_i increases. This phenomenon results from the fact that the first additions of an input to a process generally do the most good; additional quantities gradually have less and less beneficial effect. Using again the example of moving cargo through a stream, greater amounts of horsepower in the engine do not increase speed proportionately since resistance grows exponentially with speed: the marginal product of horsepower thus diminishes steadily. Figure 2.4 illustrates this behavior.

This trend is known as the "Law of Diminishing Marginal Products." The term is inaccurate, unfortunately, and can be misleading. There is nothing formal or necessary about the pattern; exceptions to the "Law" exist. The counterexamples are usually found at relatively low levels of production, when a particular

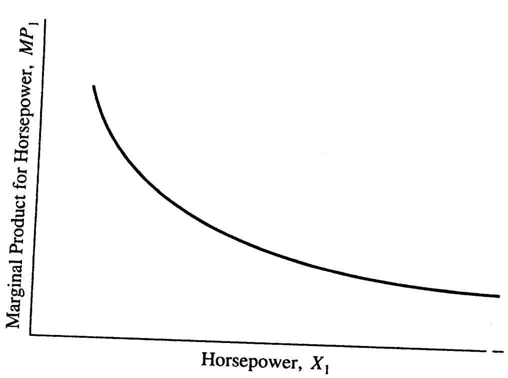

FIGURE 2.4
An example of diminishing marginal product.

input does not contribute significantly until a "critical mass" of it exists. The marginal product of uranium in terms of nuclear power is an obvious instance of this situation. The marginal product of some catalysts is another, and there are many others.

Marginal products can be easily calculated if the production function has a convenient form (see box). Otherwise, they may have to be calculated using small increments of X_i and the corresponding changes in product.

Marginal rates of substitution. A marginal rate of substitution is a special relationship between two inputs. It is the rate at which marginal increases in one input must substitute for marginal decreases in another input so that the total product remains constant. Figure 2.5 illustrates the definition in two dimensions. As the figure suggests, the marginal rate of substitution can be interpreted as the slope of the isoquant.

Any marginal rate of substitution can be expressed in terms of marginal products, and this is the form most useful for practice. Since by definition the changes in each input X_i and X_j lead to no net change in product, we can write:

$$\Delta X_i MP_i + \Delta X_j MP_j = 0$$

Marginal Products for a Cobb-Douglas Production Function

Consider a production function in the general Cobb-Douglas form:

$$Y = a_0 X_1^{a_1} \ldots X_i^{a_i} \ldots X_n^{a_n}$$

The marginal products are

$$MP_i = a_0 X_1^{a_1} \ldots a_i X_i^{a_i - 1} \ldots X_n^{a_n}$$

This may also, and more conveniently, be expressed by extracting the term a_i/X_i and writing $MP_i = (a_i/X_i)Y$.

The marginal product MP_i is diminishing with X_i whenever the coefficient $a_i < 1.0$. This is because the expression is multiplied by $X_i^{a_i - 1}$, which diminishes in X_i when $(a_i - 1)$ is negative.

A quick inspection of a Cobb-Douglas production function thus determines which marginal products are diminishing. Notice that some may diminish and others might not. Notice also that the Cobb-Douglas form implies that these relationships are constant throughout the function.

For Nerlove's basic model of the electric power industry

$$\text{output} = a_0(\text{labor})^{0.78}(\text{capital})^{0.00}(\text{fuel})^{0.61}$$

all marginal products are diminishing. This is also true for the higher level of production. But for the lowest level of production, this is not so (see Section 2.3). This model represents a typical result.

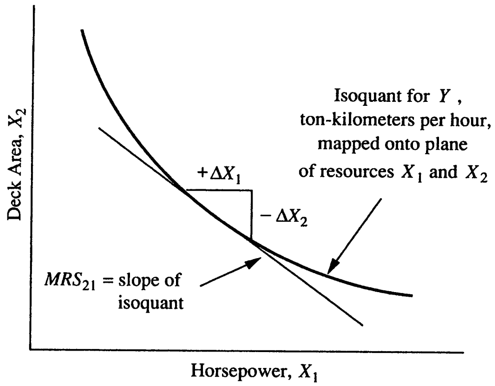

FIGURE 2.5
The marginal rate of substitution is the slope of the isoquant in two dimensions.

so that the marginal rate of substitution of X_i for X_j is

$$MRS_{ij} = \frac{\Delta X_i}{\Delta X_j} = -\left(\frac{MP_j}{MP_i}\right)$$

Two points are to be noted here: first, the marginal rate of substitution is the inverse ratio of the marginal products; second, the slope is negative as suggested by Figure 2.5 (see box). The marginal rate of substitution is useful in defining optimum combinations of resources, that is optimum designs (see Section 4.4).

Returns to scale. The returns to scale describe how fast output changes relative to the size of the production process. Formally they are defined as the ratio of the rate of change in output due to a proportional change in all inputs simultaneously,

---

Marginal Rates of Substitution for a Cobb-Douglas Production Function

For the model

$$Y = a_0 \prod X_i^{a_i}$$

the marginal rate of substitution is

$$MRS_{ij} = -\left(\frac{MP_j}{MP_i}\right) = -\left(\frac{a_j}{a_i}\right)\left(\frac{X_i}{X_j}\right)$$

The model implies that this ratio is constant for any given ratio of X_i and X_j in the design. This may not actually be the case. In Nerlove's model for electric power, the marginal rates of substitution were different at low and high levels of production.

---

and of that proportional change. The mathematical expression thus refers to an initial level of production, $Y' = g(\mathbf{X})$, a constant indicating the increase in scale or use of all inputs, s, and a consequent second level of production, $Y'' = g(s\mathbf{X})$. The returns to scale of the production function are thus

$$RTS = \left(\frac{Y''/Y'}{s}\right) = \frac{\{g(s\mathbf{X})/g(\mathbf{X})\}}{s}$$

Both returns to scale and marginal products refer to rates of change in the production function. But their concepts differ substantially. Most importantly, returns to scale characterize the variation in the production function along rays drawn through the origins of the inputs, rather than in specified directions parallel to the axes of the inputs. Figure 2.6 depicts this distinction (see box). Marginal products also are explicitly expressed in units of product, whereas returns to scale refer to a nondimensional rate.

In practice, the relevant question about the returns to scale is whether they are greater, equal to, or less than one. That is, whether the returns to scale are increasing, constant, or decreasing. The answer has important implications for the strategy to be followed in creating large systems.

The most interesting situation is that of increasing returns to scale, $RTS > 1.0$. This means that bigger production processes, bigger plants specifically, are inherently more productive than smaller ones. For a production function with increasing returns to scale, doubling all inputs would, for example, more than double the output. More would thus be produced by using available resources in one larger plant than in two smaller ones half the size. Increasing returns to scale thus indicate that, purely from the point of view of the production technology, it is advantageous to consolidate production in large plants rather than disperse it in a few small ones.

These implications of increasing returns to scale for design strategy generally hold when the economic factors are incorporated into the analysis. They may not, however, for either of two significant reasons. Firstly, the costs of the inputs

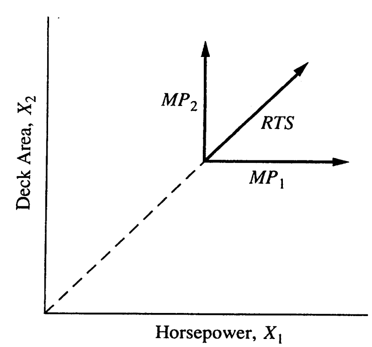

FIGURE 2.6
Directions in which the rate of change in output is measured for marginal products and returns to scale.

Returns to Scale for a Cobb-Douglas Production Function

For the model

$$Y' = a_0 \prod X_i^{a_i}$$

a scale increase by a factor of s of each input X_i increases the product to

$$Y'' = s(\exp \sum a_i) Y'$$

[Note s(exp) should be read as "s to the power of."] The returns to scale are

$$RTS = \frac{s(\exp \sum a_i)}{s}$$

For the Cobb-Douglas production function, the returns to scale are thus increasing if $\sum a_i > 1.0$, constant if $\sum a_i = 1.0$, and decreasing if $\sum a_i < 1.0$. In the electric power industry analyzed by Nerlove, the returns to scale were increasing.

The Cobb-Douglas model implies that returns to scale are the same throughout the process. This may not be so; Nerlove found that $\sum a_i = 2.92$ for lower levels of production and 1.69 for the higher levels.

---

may also increase with scale, thereby counterbalancing the physical advantages of the increasing returns to scale. Secondly, the concept of returns to scale is not properly defined for multiple output processes, since the "product" is not unique. Attempts to apply the concept of returns to scale to such situations can thus give misleading results. Section 4.5 describes these problems in detail.

Increasing returns to scale are systematically inherent in a broad range of technologies. Many of these have the common characteristics that their products are a function of volume whereas the inputs vary with surface. One example is shipping, in which the amount carried is a volume and the effort required is principally a function of the wetted surface of the ship; hence the motivation for supertankers and other very large ships. Other examples are pipelines, the production of electricity using boilers, and especially the chemical industries whose plants are typically combinations of pipes and containers.

Constant returns to scale are typically associated with processes in which higher levels of output are achieved by replicating some basic unit of production, rather than by increasing the size of this unit. The trucking industry thus normally shows constant returns to scale: the size of its vehicles is limited by both traffic regulations and the designs of roads and bridges, so that to move more one generally has to use more vehicles rather than to increase their size.

Decreasing returns to scale are also possible, although apparently rare in practice. They can exist when the production is so large that it is difficult to coordinate and control. Such might be the case of a large, nationwide industry.

To the extent that the causes of the decreasing returns to scale are administrative and organizational, however, they may be solved. This might be done by breaking up the large production process into many smaller units, each with no less than constant returns to scale. If it were possible to operate with these smaller units, then the overall production function would have constant returns to scale, and the single large unit of production would simply be technically inefficient (that is, at some interior point of the feasible region of production—as point A in Figure 2.1).

Convexity of the feasible region. The overall shape of the production possibilities is of interest to the analyst because it indicates which methods can be used to optimize the system, and whether this process will be difficult or not. Specifically, for example, this shape crucially determines whether we can optimize a system using linear programming (Chapter 5) or dynamic programming (Chapter 7). The essential issue with regard to the shape is whether the feasible region is convex.

A *convex feasible region* is one that has no reentrant boundaries, no edges that intrude into or dent the space. Figure 2.7 illustrates the situation. It is to be stressed that the convexity of the feasible region does not simply depend on the shape of its boundary. Indeed, the same boundary can be associated with both convex and nonconvex feasible regions. Figure 2.8 illustrates this case, with feasible regions for production (left) and for cost (right). The feasible region for costs is above the boundary, since it is always possible to pay more. Cost functions are discussed in detail in Section 4.4.

Formally, a feasible region is convex if every straight line between any two points in the region lies entirely in the region. Mathematically, if **A** and **B** are vectors to two points A and B on the function that bounds the region, the vector **T**:

$$\mathbf{T} = w\mathbf{A} + (1 - w)\mathbf{B} \qquad 0 \leq w \leq 1$$

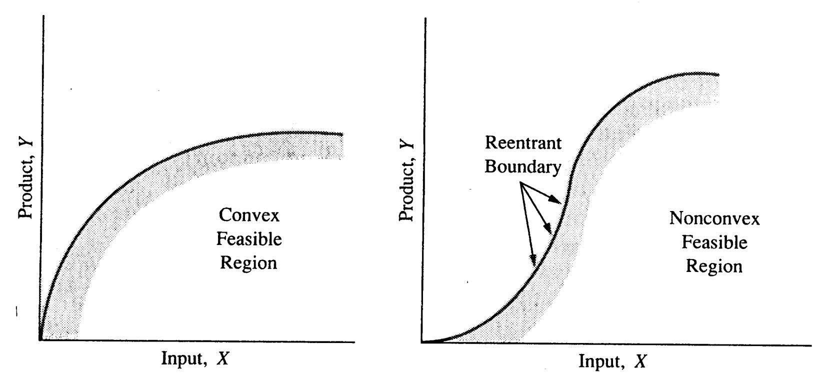

FIGURE 2.7
Convex and nonconvex feasible regions for production functions.

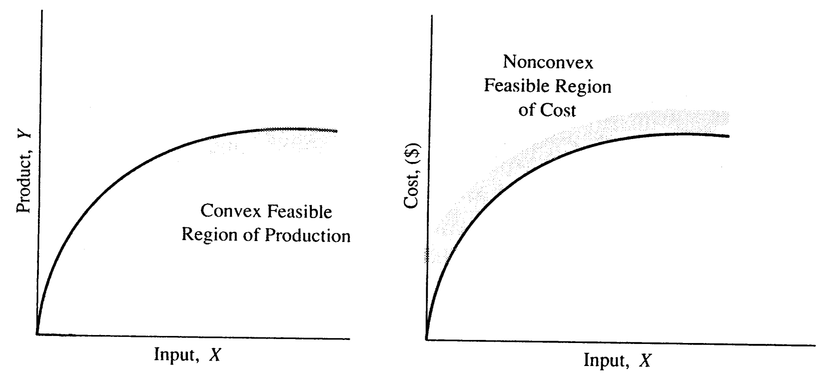

FIGURE 2.8
Convexity of the feasible region depends on the shape of the space.

represents every point on straight line between A and B, and all **T** must describe points inside the region for it to be convex. This property defines a test to distinguish between convex and nonconvex regions. Figure 2.9 indicates the situation.

Convexity of a feasible region is highly desirable because it facilitates optimization. This property guarantees that the function being optimized will be unimodal, that there will be only one optimum set, either a point or a surface. Where there is convexity, the optimization process does not have to be concerned about local optima that could be suboptimal globally: the optimum found will be the global optimum. This result leads to great efficiency in computation: it

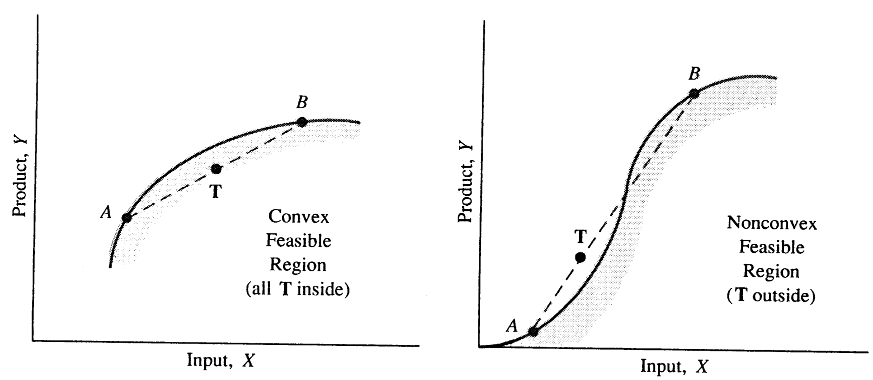

FIGURE 2.9
The test for convexity of the feasible region is that all T be part of the region.

Compatibility of Diminishing Marginal Products and Increasing Returns to Scale

Consider Nerlove's basic model of the electric power industry

$$\text{output} = a_0(\text{labor})^{0.78}(\text{capital})^{0.00}(\text{fuel})^{0.61}$$

All marginal products are diminishing, $a_i < 1.0$. However, the returns to scale are increasing, $\sum a_i = 1.39 > 1.0$.

This result leads to great efficiency in computation: it means that the optimization does not have to search the entire feasible region to obtain the optimum, it can simply start anywhere and follow any single path of gradual improvements, which necessarily lead to the optimum (see Chapter 4).

Production functions bound convex feasible regions if both the marginal products and the returns to scale are monotonically decreasing or constant with increasing inputs. The role of returns to scale is to be emphasized, as beginners frequently assume that decreasing marginal products imply that the production function flattens out in all directions, as a dome does. This is not correct. It is possible to have both decreasing marginal products and increasing returns to scale (see box). The production function associated with this situation is saddle-shaped, and the feasible region is not convex.

2.5 MULTIPLE OUTPUTS

It is often necessary to deal with the fact that a production process results in not just a single but multiple outputs. The difficulty in this regard is that the tools for dealing with such situations are not well developed. This section thus merely presents some of the issues, laying the basis for further discussion of analytic methods in later chapters.

Examples of multiple output processes abound. For example: refining crude petroleum leads to gasoline, fuel oil, and tars; metallurgical processing of zinç also produces cadmium, a close chemical relative; et cetera. Sometimes the several products are fairly subtle. The transportation of passengers by some scheduled service, say an airline, produces the trips or the movement of people, and a frequency or convenience of the service. In general then, the output of a production process is a vector of products:

$$\mathbf{Y} = (Y_1, \ldots, Y_j, \ldots, Y_m)$$

The description of a multiple output process is complicated by the fact that it is often difficult to distinguish between inputs and outputs. Consider the development of an airport in a region. One of the consequences of the operation of this facility, in addition to the air transport, is noise. As a result of the process, noise is in some sense a product. On the other hand it could equally be argued that one of the inputs to the production of air transport is a loss of quiet. Because of these kinds of ambiguities analysts now tend to avoid the explicit distinction between inputs and outputs.

The general mathematical form of a production function for a process with multiple outputs is thus

$$g(\mathbf{Y}, \mathbf{X}) = 0$$

This formulation simply describes the relationships between characteristics of the process, and avoids the need to apply the label of input or output to any of the factors involved.

A great difficulty in working with a multiple output production function is that the classical concepts used to describe production functions are unavailable. Indeed, since there is no single product, since in fact it may not be quite clear what is a product, the concept of marginal product does not apply. Likewise, it is difficult to define returns to scale since the multiple outputs may not all change at the same rate as the scale of inputs varies.

In practice, the standard way of analyzing production processes with multiple outputs has been to focus on a single major product, and to treat the other consequences as by-products of secondary importance. More recently, a variety of multiobjective methods of analysis have become available, and these are discussed in Chapter 8. Neither of these approaches is fully satisfactory so far, and much research is being done on the topic.

2.6 APPLICATIONS

To optimize the design of river transportation, engineers at the Massachusetts Institute of Technology developed a production function for the system of towboats and barges. This case indicates how such functions can be derived, and illustrates the concepts of marginal products and returns to scale. The use of this production function for optimization is shown as an application in Section 4.6.

Inland water transportation, on both rivers and lakes, is used worldwide. It is often the most heavily traveled form of communication in developing regions where roads are sparse or nonexistent. It is thus frequently critical to define the right system for its particular context.

A system of inland water transportation typically consists of towboats and flat-bottomed barges. Barges can be self-propelled but this design is inefficient since it keeps the expensive propulsive unit idle when the barge is being loaded and unloaded; it is only really justified when the cargo must cross extensive open water and each barge requires power for safety.

The physics of combinations of barges and towboats have been extensively documented by analysis and experiments in tow-tanks. These studies have led to empirical equations from which it is possible to predict the performance of any system in a specific stream—as defined by its width, depth, and current. The purely technical aspects of the system can be considered well-known. But

these do not themselves provide any useful guidance on how to select the best power plant for a raft of barges, let alone the best combination of both for a given situation. To get this guidance, one can build on the technical details to construct the production function.

The essence of the technical equations is that the effective push of the towboats equals the resistance of the flotilla as it moves through the water. The empirical equations describing these terms are very messy to write (see Hoffmeister and de Neufville, 1973 for details) but can be solved quickly by computer for any combination of the design parameters.

The two key design parameters for the system are the size of the barges, as represented by the deck area, and the horsepower of the towboats. These can be considered the two inputs to the system. The output is then the weight of cargo involved in any period, expressed in ton-kilometers/hour for example. The production function is then the locus of optimal outputs for each combination of horsepower and deck area.

As the production function consists of an infinite number of points, it is difficult to deal with in detail. Designers need to define overall views of the production function as expressed by the marginal products and returns to scale. When each point on the production function is generated individually by a computer program from given inputs, these views are easy to generate.

The marginal products are simply defined by varying one parameter when the others are held constant. For our case, for example, Figure 2.10 gives the output for a given deck area and varying horsepower; this is, in effect, a

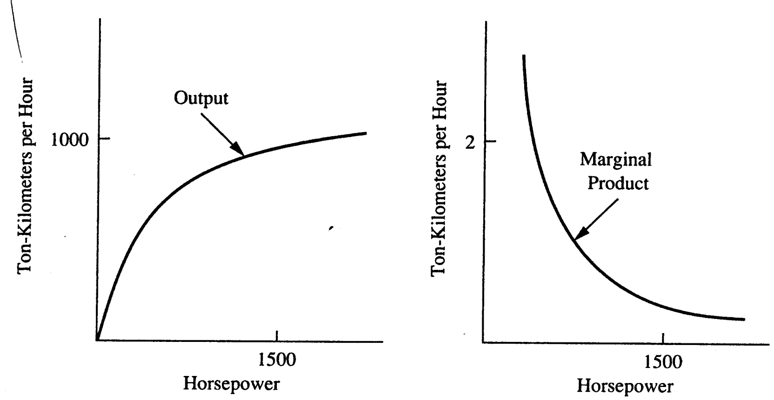

FIGURE 2.10
Total output of River Transport (left) and marginal product (right) for horsepower for a specified level of the other design parameters (Deck Area = 200 m^2), and specified water conditions (stream width = 30 m, stream depth = 6 m, current = 0).

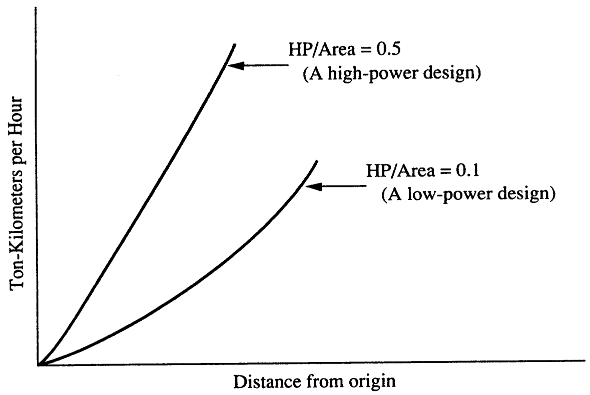

FIGURE 2.11
Increasing returns to scale for River Transport for designs with different Horsepower/Deck Area ratios, for specified water conditions (stream width = 18 m; stream depth = 2.4 m; current = 0).

"vertical slice" through the production function viewed as a "dome" of output rising from different combinations of inputs. From this curve the marginal product is derived by either calculating the slope or taking the derivative. In this case they are uniformly decreasing, as should be expected.

The returns to scale can be explored simply by plotting the change in output obtained by scaling up uniformly all parameters of a particular design. If these plots bend upward, as they do in Figure 2.11, then larger designs give greater output per unit input and there are economies of scale. The results here are again in the expected direction of economies of scale for ships, but are limited because the constraints of the river limit the benefits of larger size of the flotilla of barges. (Experienced mariners will note that uniformly scaling up all aspects of a design is not the best way to proceed in a real situation—the application in Section 4.6 makes this point in detail).

From the nature of marginal products and returns to scale the analyst can determine whether the feasible region of the production functions is convex or not. In this case it is nonconvex because of the increasing returns to scales. This fact limits the kinds of optimization techniques that can be used on the production function (specifically, linear programming of Chapter 5 is excluded).

REFERENCES

Cootner, P., and Lof, G., (1965). *Water Demand for Steam Electric Generation*, Johns Hopkins, Baltimore, MD. Also Chapter 7 in de Neufville and Marks (1974).

de Neufville, R., and Marks, D. H., (1974). *Systems Planning and Design: Case Studies in Modeling, Optimization and Evaluation*, Prentice-Hall, Englewood Cliffs, NJ.

Hoffmeister, J., and de Neufville, R., (1973, 1974). Chapter 6 in de Neufville and Marks (1974), also "Optimizing the Supply of Inland Water Transport," *ASCE Water Resources Journal*, 99, August 1973, p. 293–308.

Nerlove, M., (1965). *Estimation and Identification of Cobb-Douglas Production Functions*, Rand McNally, Chicago, IL. Also Chapter 9 in de Neufville and Marks (1974).

Pindyck, R. S., and Rubinfeld, D. L., (1989). "Production," Ch. 6, and "The Cost of Production," Ch. 7, in *Microeconomics*, Macmillan, New York.

Sage, A. P., (1983). "Production and the Theory of the Firm," Ch. 2, *Economic Systems Analysis*, North Holland, New York.

PROBLEMS

2.1. *Feasible Regions*
State what types of returns to scale and marginal products each of the following functions exhibit: (assume $X, Y \geq 0$)

(*a*)	$23X^{0.3}Y^{0.4}$	(*b*)	$12X^{1.6}Y^{0.5}$
(*c*)	$2X^{0.4}Y^{0.8}$	(*d*)	$5X + 3XY - 4Y - 2Y^2$
(*e*)	$X + 2Y^4 - Y^2Z^3/3$	(*f*)	$7X^{0.3}Y^{0.5}$
(*g*)	$16X^{1.1}Y^{0.5}$	(*h*)	$X - 4Y^2 + 2XY$
(*i*)	$10X^{0.1}Y^{0.3}$	(*j*)	$5X^{0.3}Y^{0.4}Z^{0.4}$
(*k*)	$1.1X^{1.3}Y^{0.3}$	(*l*)	$2Xe^Y$
(*m*)	$X + Y^2 + 1$	(*n*)	$4X + Y$

Which of the above functions define convex feasible regions over their entire range?

What can you conclude in your own words about the relationship between returns to scale and marginal products?

2.2. *Thought Problems*

(*a*) A chemical plant uses heat to convert raw materials (A and B) into a product. The reaction must take place at a certain high pressure, which is maintained by a spherical reaction chamber. What sorts of marginal products would you expect for each of the four resources (A, B, heat, and chamber size)? Note that for a sphere,

$$\text{volume} = (4/3)r^3, \qquad \text{surface area} = 4r^2$$

and (since pressure is independent of chamber size) the thickness of the chamber wall is also independent of size. What can you say about returns to scale in this process?

(*b*) Two major inputs in Midwestern Plains farming are land and labor. What would you expect about the marginal returns on the inputs? (For example, Soviet farms frequently employ an order of magnitude more people per acre than would a typical Minnesota farmer. How would the marginal benefit of employing one more person on the Soviet farm compare to that of one more person on the Minnesota farm?) Looking at historical trends, what are the returns to scale like in Plains farming? Would you expect the same results in an area characterized by fertile valleys and rocky hillsides? Why?

(*c*) Transoceanic shipping companies can obtain ships of differing weights and horsepower to achieve an output of tons-mi/yr carried. From what you know about trends in ship size over the last 20 years (tankers are one well-publicized example), would you expect that there are economies of scale, diseconomies of scale, or constant returns to scale? How would you expect productivity to change as ship weight is increased for a given horsepower engine? As horsepower is increased, for a given ship weight? What does this imply about the marginal returns for weight and horsepower?

(*d*) In American trucking, all operators have essentially the same kind of trucks, with size and axle load limited by regulation. Employees tend to receive the same wages, depending on whether they are in unions or not. What sort of returns to scale would you expect in the size of trucking firms? Why? What sort of marginal returns would you expect for the number of employees and the number of trucks used by the company?

(*e*) An advertising campaign is being organized for a new product. TV minutes and magazine inches will be purchased and the advertiser wants as many people as possible exposed to the product's name. What marginal returns and returns to scale would you expect?

2.3. *Translations*

For each part, describe the production function in equation form and draw it, indicating key points.

(*a*) A trade journal shows that the watermelon output per acre is roughly the square root of the number of seeds planted, for outputs between 50 and 300 watermelons per acre. The cost of seeds is \$2/1000 seeds for the first 10,000 seeds and \$1.50/1000 seeds after that. Watermelons sell for about \$5 each.

(*b*) For small buildings, construction of a housing unit (Y) requires the use as primary materials of either 5 tons of cement for a concrete house, or of 3 tons of structural steel for a metal frame house. When, however, more than 10 units are built together, the construction code requires additional facilities, such as parking lots. The output per unit material is then reduced by one-third. The unit prices of cement and steel in place are \$1000 and \$2000, respectively. The value of a housing unit is \$40,000.

2.4. *Production Function I*

(*a*) Describe the marginal products, marginal rates of substitution, and the returns to scale for the production function:

$$Z = 10X^{0.1}Y^{0.3}$$

(*b*) Is the feasible region convex? Explain.

2.5. *Production Function II*

Same as 2.4 (*a*), for: $Z = 10X^{0.2}Y^{0.4}$

2.6. *Production Function III*

As 2.4 (*a*), for: $Z = 2\ \log_e X + 4\ \log_e Y$

2.7. *Production Function IV*

As 2.4 (*a*), for: $Z = 0.3X^{0.8}Y^{0.6}$

2.8. *Vi-Tall Cereal*

Tab Booleigh, chief cook at Health Foods Inc., makes Vi-Tall Cereal from rye and bran. Her recipe is as follows: 1 lb of Vi-Tall requires either 1 1/2 lbs of rye or 1 lb of bran. For batches of 10 lbs or more, spoilage reduces the output per unit of grain by 1/3. The prices per pound of rye and bran are \$2 and \$4 respectively. A pound of Vi-Tall sells for \$7.50.

(*a*) Does the production function define a convex, feasible region?

(*b*) Plot the isoquants for 4 and 12 lbs.

(*c*) What is the *MRS* of rye for bran?

(*d*) What are the *MP*s with respect to rye and bran?

2.9. *Plywood Factory*

Charles B. Ord, known to his friends as Chip, makes plywood from hard and soft wood using either of three processes with the following characteristics:

	Board Feet of Wood		
Process	**Hard**	**Soft**	**Ply**
1	1000	500	1333
2	1000	1850	2500
3	1000	3200	4000

(*a*) What is the *MP* for soft wood when Chip has 1000 bd. ft of hard and 3200 of soft wood?

(*b*) How much hard and soft wood would Chip use if the factory manager assigned 500 bd.ft of hard wood each to Process 1 and Process 2?

(*c*) Establish and draw the isoquant for 10,000 bd. ft of plywood.

(*d*) What is the *MRS* between hard and soft wood at (1000, 1850)?

2.10. *Timothy Burr*

Tim operates a logging mill that chews up pine and balsam. Having analyzed its records, Tim finds that the marginal product of pine is inversely proportional to the square root, and that the marginal product of balsam varies with the inverse 1/4 power. What is the marginal rate of substitution? What can you say about the returns to scale of the operation?