

Production Function

the two operate in the same marketplace, one group may operate closer to best practice than the other. Education and health care are two sectors in which numerous public/private performance comparisons have been conducted, with the empirical evidence being mixed.

A third hypothesis asserts that regulation matters. Ill-designed regulatory frameworks inhibit best practice by diverting resources away from production toward compliance. Thoughtfully designed regulatory frameworks can enhance best practice by providing incentives for organizations to operate efficiently and to improve their efficiency through time. One regulatory context in which theoretical predictions have been quantified by empirical investigation concerns the impact of alternative forms of environmental control on organizational performance. In this context, however, the private cost of reduced efficiency must be balanced against the social benefits of environmental protection.

The theoretical concept of production and cost frontiers is universally accepted. However, the empirical implementation of best-practice frontiers and their use in the policy arena has attracted some criticism. One criticism asserts that the mathematical framework is necessarily incomplete and fails to incorporate the objectives of and constraints faced by the organization and its stakeholders. In 1976 George J. Stigler claimed that “waste is not a useful economic concept. Waste is error within the framework of modern economic analysis” (p. 216). A second criticism asserts that the empirical implementation fails to control adequately for variation in the operating environment, thereby confusing variation in operating efficiency with variation in factors beyond the control of management. This concern has inhibited the use of best-practice frontiers in evaluating the relative performance of educational institutions, health care providers, and regulated utilities.

SEE ALSO *Fixed Coefficients Production Function; Production; Production Function*

BIBLIOGRAPHY

- Aigner, Dennis J., C. A. Knox Lovell, and Peter Schmidt. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6 (1): 21–37.
- Charnes, Abraham, William W. Cooper, and Edwardo Rhodes. 1978. Measuring the Efficiency of Decision-Making Units. *European Journal of Operational Research* 2 (6): 429–444.
- Farrell, Michael James. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A, General* 120 (3): 253–281.
- Meeusen, Wim, and Julien van den Broeck. 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18 (2): 435–444.

Stigler, George J. 1976. The Xistence of X-Efficiency. *American Economic Review* 66 (1): 213–216.

C. A. Knox Lovell

PRODUCTION FUNCTION

The principal activity of a firm is to produce a good or provide a service, that is, to turn inputs into output. To represent this process, economists use an abstract model of production. The central concept in this model is the *production function*. A production function is a mathematical description of the various technical production possibilities faced by a firm. Algebraically, it is written as

$$q = f(x_1, x_2, \dots, x_n) \quad (1)$$

where q represents the flow of output produced and x_1, \dots, x_n are the flows of inputs, each measured in physical quantities—for example, the number of bushels of corn produced and the number of tractors and workers utilized. Often, production functions appear in textbooks written with two inputs as $q = f(k, l)$, where k denotes the amount of capital and l denotes the amount of labor. To simplify, we will use this production function in the remainder of the entry. Equation (1) is assumed to provide, for any conceivable set of inputs, the engineer's solution to the problem of how to best (most efficiently) combine different quantities of those inputs to get the output. Therefore, a production function can be understood as a constraint on the activities of producers that is imposed by the existing technology.

It is important to stress that, as noted above, equation (1) is essentially an *engineering* relationship. As such, it allows for no testing of economic hypotheses. Actual observed data are the results of economic decisions in which the production function is but one constraint. However, the key question from an *economic* point of view is how the levels of output and inputs are chosen by profit-maximizing firms. Thus, economists use production functions in conjunction with marginal productivity theory (see below) to provide explanations of factor prices and the levels of factor utilization. Whereas the engineering production function captures the maximum level of output that can be achieved if the given inputs are efficiently employed, the economic production function reflects the “best-practice” use of the available input and output combinations.

PROPERTIES OF THE PRODUCTION FUNCTION

The *marginal physical product* of an input is the additional output that can be produced by employing one more unit of that input while holding all other inputs constant.

Algebraically, $\frac{\partial q}{\partial k}$ is the marginal physical product of capital, and $\frac{\partial q}{\partial \ell}$ is the marginal physical product of labor.

It is assumed that both marginal products are positive, that is, $\frac{\partial q}{\partial k} > 0$ and $\frac{\partial q}{\partial \ell} > 0$ (a negative marginal product means that using more of the input in question results in less output being produced). It is also usually assumed that the production process exhibits diminishing marginal productivity. This means that successive additions of one factor while keeping the other one constant yields smaller and smaller increases of output, that is, $\frac{\partial^2 q}{\partial k^2} < 0$ and $\frac{\partial^2 q}{\partial \ell^2} < 0$.

Factor elasticity (ε) is the percentage change in output in response to an infinitesimal percentage change in a factor given that all other factors are held fixed, that is, $\varepsilon_\ell = \frac{\partial q}{\partial \ell} \frac{\ell}{q}$ and $\varepsilon_k = \frac{\partial q}{\partial k} \frac{k}{q}$.

The *marginal rate of technical substitution* (*MRTS*) shows the rate at which labor can be substituted for capital while holding output constant at the level q_0 , that is,

$MRTS(\ell \text{ for } k) = -\left. \frac{dk}{d\ell} \right|_{q=q_0}$. It can also be shown that

the *MRTS* equals the negative of the ratio of marginal productivities, that is, $MRTS(\ell \text{ for } k) = -\left. \frac{dk}{d\ell} \right|_{q=q_0} =$

$-\frac{\partial q}{\partial \ell} / \frac{\partial q}{\partial k}$. This expression indicates that the gain in output from increasing ℓ slightly is exactly balanced by the loss of output from suitably decreasing k (so as to keep output constant at the level q_0). It is important to note that for large ratios of k to ℓ , the *MRTS* is a large positive number, indicating that a large amount of capital can be given up if one more unit of labor becomes available. On the other hand, when a large amount of labor is already in use, the *MRTS* is low, indicating that only a small amount of capital can be exchanged for an additional unit of labor if output is to be held constant. This is the so-called property of *diminishing MRTS*, which states that progressively reducing the amount of one input while maintaining a constant output level will require progressively large increases of the other input. Diminishing *MRTS* requires both positive and diminishing marginal productivities and $\frac{(\partial q / \partial \ell)}{\partial k} > 0$, that is, that an increase in capital leads to a higher marginal productivity of labor.

If the production function is given by $q = f(k, \ell)$ and all inputs are multiplied by the same positive constant $t > 1$, the degree of *returns to scale* can be classified as follows: (1) constant if $f(tk, t\ell) = tf(k, \ell) = tq$; (2) increasing if $f(tk, t\ell) > tf(k, \ell) = tq$; and (3) decreasing if $f(tk, t\ell) < tf(k, \ell) = tq$.

The *MRTS* is a useful measure of substitutability of one factor for another, given output. However, it depends on the units in which both labor and capital are measured. An alternative measure, independent of the units of measurement, is the *elasticity of substitution* (σ). It measures the percentage change in the relative amount of the factors employed resulting from a given percentage change in the relative marginal products (that is, the *MRTS*). Assuming there are only two factors of production: $\sigma = \frac{d(k/\ell)}{(k/\ell)} / \frac{d(MRTS)}{MRTS}$. Given that k/ℓ and

MRTS move in the same direction, $\sigma > 0$. However, this is not true in the case of a production function with more than two inputs, in which case there are different definitions of the elasticity of substitution (see Chambers 1994, pp. 27–36). A high σ means that the *MRTS* does not change much relative to k/ℓ . In the extreme $\sigma = \infty$, the two inputs are said to be perfect substitutes. On the other hand, a low σ means that the *MRTS* will change by a substantial amount as k/ℓ varies. For example, if $\sigma = 0$, the inputs are used in fixed proportions and substitution is not possible.

MOST COMMON PRODUCTION FUNCTIONS

Fixed proportions (or Leontief): $q = \min(\alpha_1 k, \alpha_2 \ell)$, $\alpha_1, \alpha_2 > 0$. This production function is characterized by $\sigma = 0$, as the marginal products are not defined. In this case, capital and labor must be used in fixed proportions. This production function offers a good approximation to many real-world industrial processes.

Cobb-Douglas: $q = Ak^{\alpha_1} \ell^{\alpha_2}$. This is the most ubiquitous form in empirical analyses at the macroeconomic level. The marginal product of capital equals $\frac{\partial q}{\partial k} = \alpha_1(q/k) > 0$ (and $\frac{\partial^2 q}{\partial k^2} = \alpha_1(\alpha_1 - 1)(q/k^2) < 0$ provided $0 < \alpha_1 < 1$); and that of labor equals $\frac{\partial q}{\partial \ell} = \alpha_2(q/\ell) > 0$ (and $\frac{\partial^2 q}{\partial \ell^2} = \alpha_2(\alpha_2 - 1)(q/\ell^2) < 0$ provided $0 < \alpha_2 < 1$). This production function can exhibit any degree of returns to scale, as $A(tk)^{\alpha_1}(t\ell)^{\alpha_2} = At^{\alpha_1 + \alpha_2} k^{\alpha_1} \ell^{\alpha_2} = t^{\alpha_1 + \alpha_2} q$. In this production function, $\sigma = 1$, that is, the elasticity of substitution does not vary with the combination of factors used.

CES: $q = \gamma[\delta k^\rho + (1 - \delta)\ell^\rho]^{-\epsilon/\rho}$, where $\gamma > 0$ is an efficiency parameter; $0 \leq \delta \leq 1$ is a distribution parameter;

Production Function

$\rho \leq 1$ is the substitution parameter; and $\varepsilon \geq 0$ denotes the degree of returns to scale. For this production function, $f(tk, tl) = \varepsilon q$; and ρ is the substitution parameter, which equals $\rho = \frac{\sigma - 1}{\sigma}$. This implies that the elasticity of substitution is $\sigma = 1/(1 - \rho)$.

Translog: $\log q = \beta_0 + \beta_k \log k + \beta_l \log l + \beta_{kk} (\log k)^2 + \beta_{ll} (\log l)^2 + \beta_{lk} (\log k)(\log l)$. Mathematically, this is a second-order expansion, which is easy to implement empirically. In this case, estimates of marginal products and the elasticity of substitution are functions of the coefficients and the input levels. This form is flexible in that it imposes no assumptions on the elasticity of substitution. Other widely used flexible forms include the quadratic and the square root production functions (Beattie and Taylor 1985; Chambers 1994).

DUALITY

A major development since the late 1960s has been the dual formulation of production theory. This approach consists in recovering through the profit or cost functions the properties of the underlying production function. The cost function $TC(q, w, r)$ represents the minimum cost of producing output for any set of input costs (see Beattie and Taylor 1985, chap. 6, and Chambers 1994). The dual approach is very convenient in applied work because it deals directly with observed economic data generated by markets (that is, factor prices and output). For example, the Cobb-Douglas cost function dual of the production function $q = Ak^{\alpha_1} l^{\alpha_2}$ is given by the expression $TC = kq^{1/\nu} r^{\alpha_1/\nu} w^{\alpha_2/\nu}$, where k is a constant, r is the user cost of capital, w is the wage rate, and $\nu = \alpha_1 + \alpha_2$ is the degree of returns to scale.

TECHNICAL PROGRESS

Technical progress in economics refers to the impact of the adoption of new techniques on production (or cost). Analytically, the simplest way to represent technical progress is through a shift in the production function over time, that is, $q = f(k, l, t)$, where t is an index of the level of technology. In the analysis of time series data (one economic unit observed over time), time is used as a proxy for t . Technical progress is measured by how output changes as time elapses with the input bundle held constant. The rate of technical progress is defined as $T(k, l, t) = \frac{\partial \ln f(k, l, t)}{\partial t}$. The representation of technical progress this way, though convenient, is very unrealistic, for it assumes that technical progress does not require new inputs and, further, that the production function maintains the same form as time elapses (see Chambers 1994, chap. 6).

PRODUCTION FUNCTIONS IN APPLIED WORK

In this section, first, we provide some useful references for the reader interested in estimating production functions empirically; second, we provide examples of applications in microeconomics; and finally, examples of applications at the macroeconomic level.

Hands On: Estimating Production Functions Given the availability of computers and sophisticated software packages, estimation of production functions (that is, the use of statistical methods applied to real data in order to obtain values of the relevant parameters, such as the factor elasticities, the elasticity of substitution, or the degree of returns to scale) does not present serious problems from the technical and data points of view. Kenneth Wallis (1979, chap. 2) offers a classical and very accessible introduction to the estimation process. Ernst Berndt (1991, chap. 9) also offers a hands-on approach. R. L. Thomas (1993, chap. 11) offers a modern treatment with discussion of recent advances in time-series econometrics, such as unit roots and cointegration analyses. Estimation of engineering production functions requires the availability of data in physical terms for output and inputs. See Sören Wibe (1984) for a survey of estimation of engineering production functions. At aggregate levels (sectors or total economy), there are now many databases that contain series of output and inputs.

As noted above, economists use production functions in conjunction with marginal productivity theory to provide explanations of factor prices and the levels of factors utilization. Observed prices, output, and inputs are generated by a set of simultaneous relationships, and so it is inappropriate to estimate the production function as a single equation treating capital and labor as exogenous variables. In the simplest case, assuming perfect competition in product and factor markets, the prices of output, capital, and labor are exogenous. In the case of the Cobb-Douglas production function, the marginal productivity conditions are given by the equality of the factor prices to the marginal productivities, that is $\frac{\partial q}{\partial k} = \alpha_1(q/k) = r$ and $\frac{\partial q}{\partial l} = \alpha_2(q/l) = w$, where r denotes the price of capital (that is, the user cost of capital) and w denotes the price of labor (that is, the wage rate). The firm's optimal output and input levels result from jointly estimating these two equations together with the production function $q = Ak^{\alpha_1} l^{\alpha_2}$.

Examples of Applications in Microeconomics Efficient production decisions require that the marginal productivity of an input equals its market price. Therefore, comparing the marginal productivity with input prices permits

the analyst to test whether inputs are allocated efficiently. Pranab Bardhan's "Size, Productivity, and Returns to Scale" (1973) is a seminal paper in the literature. Bardhan estimated farm production functions to test the efficiency of land and labor allocations in rural areas of developing economies. Likewise, Hanan Jacoby (1993) illustrates how such tests can be refined and the difficulties in estimating marginal labor productivities.

The degree of returns to scale displayed is of great interest to market regulators. Mergers between rival companies reduce costs and potentially enhance welfare if they are in processes that are subject to increasing returns to scale. Similarly, the argument for granting exclusive franchises to operators providing certain public services (most notably public utilities) hinges upon the assumption that the activity displays increasing returns to scale over the relevant range of output (see Train 1991).

Elasticities of substitution are important for understanding how changes in the price of one input impact the demand for others. Ernst Berndt and David Wood (1975), in the aftermath of the first oil shock, showed that capital and energy were complements in U.S. manufacturing. This implies that fiscal incentives to stimulate investment would promote greater energy use.

Examples of Applications in Macroeconomics In broad terms, aggregate production functions are estimated empirically in macroeconomic work for the following purposes: (1) to obtain measures of the elasticity of substitution between the factors, and the factor-demand price elasticities; such measures are used for predicting the effects upon the distribution of the national income of changes in technology or factor supplies; (2) to apportion total growth into the accumulation of factors of production and technical change between two periods; (3) to test theories and quantify their predictions; and (4) to assess the likely effects of macroeconomic policies. Likewise, much work in international trade and labor economics uses the production function as one of their main pillars.

However, the most important application of production functions in macroeconomics is in the field of growth theory. Since the 1980s, the field of growth theory has mushroomed with the development of the so-called endogenous growth models (see, for example, Barro and Sala-i-Martin 1995; Aghion and Howitt 1998; and Valdés 1999). These models posit production functions with increasing returns to scale, an elasticity of capital of unity, externalities, or some combination of these. In order to assess the importance of these assumptions, economists have estimated aggregate production functions for entire economies, for the manufacturing sector, or for more narrowly defined industry aggregates with a view to testing if the real world is characterized by such phenomena. Paul

Romer (1987), Robert Hall (1990), Ricardo Caballero and Richard Lyons (1992), N. Gregory Mankiw et al. (1992), David Backus et al. (1992), Susanto Basu and John Fernald (1995, 1997), Craig Burnside (1996), and others have attempted to document the empirical importance of the phenomena of increasing returns and externalities hypothesized by the new growth models.

The other important application of production theory in the area of growth is referred to as *growth accounting*. Here the purpose is to decompose overall output growth into the contributions of factor accumulation and technological progress. In this case, the factor elasticities are not estimated econometrically but assumed to be equal to the factor shares (under profit maximization and competitive markets), which can be obtained from the national accounts. In fact, the purpose of growth accounting is to estimate residually the contribution of technical progress (the shift in the production function) to overall growth. In growth rates, the production function $q = f(k, l, t)$ can be written as $\dot{q} = \varepsilon_k \dot{k} + \varepsilon_l \dot{l} + T$ where ε_k and ε_l are the factor elasticities. The only unknown here is T , the rate of technical progress, which can be obtained as $T = \dot{q} - \varepsilon_k \dot{k} - \varepsilon_l \dot{l}$.

THE AGGREGATION PROBLEM

Macroeconomic work assumes that the economy produces a single homogenous output with "quantities" of homogenous inputs. In most applied work, this assumption is not even discussed, and applied economists use published (aggregate) data from the national accounts, for example. However, at the aggregate level, output and inputs are not physical quantities. The output of a nation (GDP) must be expressed in monetary units. Capital is not measured as a collection of machines, but as the result of a series of investments (thus also expressed in monetary units) added up through the perpetual inventory method and assuming some rate of depreciation. In other words, in aggregate work, economists actually estimate $V = F(J, L)$, and not $q = f(k, l)$, where V is real aggregate value added, L is total employment, and J is the deflated or constant-price value of the stock of capital. When these variables are used as arguments in a production function, they present a serious problem. This is that aggregate production functions cannot be, in general, derived theoretically. If one asks for the conditions under which a series of microproduction functions can be properly aggregated so as to yield an aggregate production function, such conditions are so stringent that it is difficult to believe that actual economies can satisfy them. In other words, for practical purposes, aggregate production functions do not exist (Felipe and Fisher 2003, 2006).

This led a number of economists to ask the following question: If aggregate production functions do not exist,

Production Function

what do applied economists find when they get aggregate data and estimate a regression? The answer is that all they do is approximate the accounting identity that relates definitionally the value of total output to the sum of the value of total inputs, that is, $V_t \equiv W_t + \Pi_t \equiv w_t L_t + r_t J_t$, where V is real value added, W is the total wage bill in real terms, Π denotes total profits (*operating surplus* in the NIPA [National Income and Product Accounts] terminology) also in real terms, w is the average real wage rate, L is employment, r is the average ex-post real profit rate, and J is the deflated or constant-price value of the stock of capital. The symbol \equiv indicates that the expression above is an accounting identity. In a series of papers, Jesus Felipe (2001), Felipe and F. Gerard Adams (2005), and Felipe and John S. L. McCombie (2001, 2003) have shown that this identity can be rewritten as $V = F(J, L)$. This argument explains why aggregate production functions, despite the fact that they do not exist, tend to “work” (that is, the fit is high when estimated econometrically and, at least at times, the factor elasticities approximate the factor shares), and why it deprives most applied work with aggregate production functions of any meaning and significance.

SEE ALSO *Cambridge Capital Controversy; Physical Capital; Production; Smith, Vernon L.*

BIBLIOGRAPHY

- Aghion, Philippe, and Peter Howitt. 1998. *Endogenous Growth Theory*. Cambridge, MA: MIT Press.
- Backus, David, Patrick J. Kehoe, and Timothy J. Kehoe. 1992. In Search of Scale Effects in Trade and Growth. *Journal of Economic Theory* 58 (2): 377–409.
- Bardhan, Pranab. 1973. Size, Productivity, and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture. *Journal of Political Economy* 81 (6): 1370–1386.
- Barro, Robert, and Xavier Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw-Hill. 2nd ed., 2004, Cambridge, MA: MIT Press.
- Basu Susanto, and John G. Fernald. 1995. Are Apparent Spillovers a Figment of Specification Error? *Journal of Monetary Economics* 36 (1): 165–188.
- Basu Susanto, John G. Fernald. 1997. Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy* 105 (2): 249–283.
- Beattie, Bruce R., and C. Robert Taylor. 1985. *The Economics of Production*. New York: Wiley.
- Berndt, Ernst. 1991. *The Practice of Econometrics: Classic and Contemporary*. Reading, MA: Addison-Wesley.
- Berndt, Ernst, and David O. Wood. 1975. Technology, Prices, and the Derived Demand for Energy. *Review of Economics and Statistics* 57 (3): 259–268.
- Burnside, Craig. 1996. Production Function Regressions, Returns to Scale, and Externalities. *Journal of Monetary Economics* 37 (2): 177–201.
- Caballero, Ricardo J., and Richard K. Lyons. 1992. External Effects in U.S. Procylical Productivity. *Journal of Monetary Economics* 29 (2): 209–225.
- Chambers, Robert G. 1988. *Applied Production Analysis: A Dual Approach*. Cambridge, U.K.: Cambridge University Press.
- Felipe, Jesus. 2001. Endogenous Growth, Increasing Returns, and Externalities: An Alternative Interpretation of the Evidence. *Metroeconomica* 52 (4): 391–427.
- Felipe, Jesus, and F. Gerard Adams. 2005. ‘A Theory of Production’: The Estimation of the Cobb-Douglas Function, a Retrospective View. *Eastern Economic Journal* 31 (3): 427–446.
- Felipe, Jesus, and Franklin M. Fisher. 2003. Aggregation in Production Functions: What Applied Economists Should Know. *Metroeconomica* 54 (2–3): 208–262.
- Felipe, Jesus, and Franklin M. Fisher. 2006. Aggregate Production Functions, Neoclassical Growth Models, and the Aggregation Problem. *Estudios de Economia Aplicada* 24 (1): 127–163.
- Felipe, Jesus, and J. S. L. McCombie. 2001. The CES Production Function, the Accounting Identity, and Occam’s Razor. *Applied Economics* 33 (10): 1221–1232.
- Felipe, Jesus, and J. S. L. McCombie. 2003. Some Methodological Problems with the Neoclassical Analysis of the East Asian Miracle. *Cambridge Journal of Economics* 54 (5): 695–721.
- Hall, Robert E. 1990. Invariance Properties of Solow’s Productivity Residual. In *Growth/Productivity/Employment: Essays to Celebrate Bob Solow’s Birthday*, ed. Peter Diamond, 71–112. Cambridge, MA: MIT Press.
- Jacoby, Hanan. 1993. Shadow Wages and Peasant Family Labor Supply: An Econometric Application to the Peruvian Sierra. *Review of Economic Studies* 60: 903–921.
- Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107 (2): 407–437.
- Romer, Paul. 1987. Crazy Explanations for the Productivity Slowdown. In *National Bureau of Economic Research Macroeconomics Annual 1987*, 163–202. Cambridge, MA: MIT Press.
- Thomas, R. L. 1993. *Introductory Econometrics*. 2nd ed. London and New York: Longman.
- Train, Kenneth. 1991. *Optimal Regulation: The Economic Theory of Natural Monopoly*. Cambridge, MA: MIT Press.
- Valdés, Benigno. 1999. *Economic Growth: Theory, Empirics, and Policy*. Northampton, MA: Edward Elgar.
- Wallis, Kenneth F. 1979. *Topics in Applied Econometrics*. 2nd ed. Minneapolis: University of Minnesota Press.
- Wibe, Sören. 1984. Engineering Production Functions: A Survey. *Economica* 51: 55–64.

Jesus Felipe
Aashish Mehta