as $C(w_1, w_2, \ldots, w_n, y)$, defined as the minimum level of cost for any set of input wages $w_1, w_2, \ldots, w_n$, at a given level of output y, so that

$$C(w_1, w_2, \ldots, w_n, y) \text{ solves } \min_{x_1, x_2, \ldots, x_n} C \equiv \sum_{j=1}^{n} w_j x_j \quad \text{such that } y = f(x_1, x_2, \ldots, x_n) \qquad *(8.2.57)$$

For the cost function, the optimal factor inputs satisfy

$$x_j = \frac{\partial C}{\partial w_j} \qquad *(8.2.58)$$

where, since the cost function is concave in w_j,

$$\frac{\partial x_j}{\partial w_j} = \frac{\partial^2 C}{\partial w_j^2} < 0, \quad \text{all } j \qquad *(8.2.59)$$

as in (8.2.37). The cost curve of (8.2.14) then corresponds to the cost function with given input wages

$$C(y) = C(\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_n, y). \qquad *(8.2.60)$$

Of these various functions, the ones most frequently estimated using econometric techniques are the production function (8.2.1), the cost curve (8.2.14), factor demand functions (8.2.28), and the cost function (8.2.57).

8.3 Estimation of production functions

A basic problem in applied econometrics is that of estimating the production function, representing the technological relationship between output and factor inputs.[10] In most empirical applications the production function gives output y as a function of only two homogeneous inputs–labor L and capital K:

$$y = f(L, K). \qquad *(8.3.1)$$

Data for the estimation include cross-section or time-series data on some or all three variables and related variables, such as prices and wages. Output is typically measured as value added per year, deflated for price changes in time-series studies. It can also, however, be measured as physical units of output per year or gross value of output per year. The inputs should, in theory, be measured in terms of *services* of the input per unit of time, but such data are generally not available, so they are instead typically measured by the amount of the input utilized or available in the production process. Labor input is typically measured as manhours employed per year, but it is also sometimes measured as number of employees. Capital input is typically measured by the net capital stock (net of depreciation), but it is also sometimes measured by the gross capital stock and by certain direct measures (e.g., number of tractors in use for agriculture). Among the other inputs that could be included in the production function are raw materials, fuel, and land. Furthermore, labor and capital can be disaggregated, e.g., into skilled and unskilled labor and, for capital, plant and equipment.

[10] For surveys of production functions and their estimation see Walters (1963, 1968), Frisch (1965), Hildebrand and Liu (1965), Nerlove (1967), Solow (1967), and Ferguson (1969).

Of these variables the one that creates the most problems is the capital input. While data on output and labor are generally available, data on capital are either not available or of questionable validity. Enormously complex problems of measurement arise with respect to capital as an input to the production process. First, capital generally represents an aggregation of very diverse components, including various types of machines, plant, inventories, etc. Even machines of the same type may cause aggregation problems if they are of different vintages, with different technical characteristics, particularly different levels of productivity or efficiency. Second, some capital is rented but most is owned. For the capital stock that is owned, however, it is necessary to impute rental values to take account of capital services. Such an imputation depends, in part, on depreciation of capital. Depreciation figures are generally unrealistic, however, since they entail both tax avoidance by the firm and the creation by the tax authorities of incentives to invest via accelerated depreciation. Third, there is the problem of capacity utilization. Only capital that is actually utilized should be treated as an input, so measured capital should be adjusted for capacity utilization. Accurate data on capacity utilization are, however, difficult or impossible to obtain.[11] Other problems could be cited as well, but all these suggest that, if at all possible, the use of an explicit measure of the capital stock should be avoided, since it is virtually impossible to find data adequately representing capital stock.

To estimate the production function requires the further development of its properties, leading to the specification of an explicit functional form. In particular, it is generally assumed that the production function satisfies the properties

$$f(0, K) = f(L, 0) = 0 \qquad (8.3.2)$$

$$\frac{\partial f}{\partial L} \geqslant 0, \quad \frac{\partial f}{\partial K} \geqslant 0 \qquad (8.3.3)$$

$$\frac{\partial^2 f}{\partial L^2} \leqslant 0, \quad \frac{\partial^2 f}{\partial K^2} \leqslant 0, \quad \frac{\partial^2 f}{\partial L^2}\frac{\partial^2 f}{\partial K^2} - \left(\frac{\partial^2 f}{\partial L\, \partial K}\right)^2 \geqslant 0. \qquad (8.3.4)$$

[11] One approach to capacity utilization is to assume that the percentage of capital utilized is the same as the percentage of labor utilized, and therefore to reduce total capital available by the (labor) unemployment rate. There are various problems with this approach, however. For example, to the extent that capital is owned, the cost of using unemployed capital is less than that of using unemployed labor, suggesting that labor unemployment might exceed capital unemployment.

Here (8.3.2) indicates that both factor inputs are indispensable in the production of output, (8.3.3) states that both marginal products are nonnegative, and (8.3.4) states that the Hessian matrix of second-order partial derivatives of the production function is negative semidefinite, ensuring the proper curvature of the isoquants.

The production function (8.3.1) can, in certain cases, exhibit certain *returns-to-scale* phenomena at particular points. Thus at the point (L, K) the production function exhibits local

$$\left\{\begin{matrix}\textit{constant}\\ \textit{increasing}\\ \textit{decreasing}\end{matrix}\right\} \textit{returns to scale} \quad \text{if } f(\lambda L, \lambda K) \left\{\begin{matrix} = \\ > \\ < \end{matrix}\right\} \lambda f(L, K), \quad \text{all } \lambda > 1. \tag{8.3.5}$$

The constant-returns-to-scale case, that in which the production function exhibits (global) constant returns to scale for all positive λ, is that in which it is positive homogeneous of degree one (sometimes called "linearly homogeneous"), satisfying

$$f(\lambda L, \lambda K) = \lambda f(L, K), \quad \text{all } \lambda > 0, \quad \text{all } (L, K). \qquad {}^*(8.3.6)$$

In this case, at any levels of the inputs, scaling both inputs by the same multiplicative factor scales output by the same multiplicative factor. Then Euler's theorem on homogeneous functions implies that

$$\frac{\partial f}{\partial L} L + \frac{\partial f}{\partial K} K = f(L, K). \tag{8.3.7}$$

This condition implies, from (8.2.6), assuming perfect competition, that

$$wL + rK = pf(L, K). \qquad {}^*(8.3.8)$$

Here the left-hand side is total income, the sum of labor income and capital income, w and r being the wages rates of labor and capital, respectively. The right-hand side is the value of output, given as output price times the level of output. Condition (8.3.8) thus states that, assuming profit maximization and perfect competition, a constant-returns-to-scale production function implies that total income equals total output. This result is sometimes called the "adding-up theorem." More generally, the production function is positive homogeneous of degree h if

$$f(\lambda L, \lambda K) = \lambda^h f(L, K), \quad \text{all } \lambda > 0, \quad \text{all } (L, K) \tag{8.3.9}$$

the case $h = 1$ being that of constant returns to scale. If the production function is homogeneous of degree h and $h > 1$, then it exhibits (global) increasing returns to scale, while if $h < 1$ it exhibits (global) decreasing returns to scale.[12]

[12] Of course, production functions need not be homogeneous of any degree. A local measure of returns to scale is given by the *elasticity of production* at the point (L, K):

$$\epsilon(L, K) = \frac{L}{y}\frac{dy}{dL} = \frac{K}{y}\frac{dy}{dK} \quad \text{where} \quad \frac{dL}{L} = \frac{dK}{K}$$

and thus is defined for an equal proportional change in each of two inputs. See Problem 8-G.

The production function is said to be *homothetic* if it can be expressed as

$$y = F[g(L, K)] \tag{8.3.10}$$

where F is a monotonic increasing function of a single variable and g is a function that is homogeneous of degree one in L and K. The case of homogeneity of degree one of the production function, as represented by (8.3.6), is thus a special case of homotheticity. Homotheticity ensures that all isoquants, as in Figure 8.1, are "radial blowups" of a given isoquant, since the isoquants passing through a given ray from the origin all have the same slope.

Another important property of production functions, in addition to that of returns to scale, is that of the *substitutability of inputs* for one another. A local measure of such substitutability is the *elasticity of substitution* σ, defined as the ratio of the proportionate change in the ratio of factor inputs (called "factor proportions") to the proportionate change in the ratio of marginal products (the marginal rate of technical substitution at given levels of inputs):[13]

$$\sigma = \frac{d \ln (K/L)}{d \ln (\mathrm{MP}_L/\mathrm{MP}_K)} = \frac{d \ln (K/L)}{d \ln (\mathrm{MRTS}_{LK})} \qquad {}^*(8.3.11)$$

In this definition the numerator involves the ratio of capital to labor, while the denominator involves the ratio of the marginal product of labor to that of capital, ensuring that σ is nonnegative.

Assuming perfect competition and profit maximization, the ratio of the marginal products is the ratio of the factor prices, as in (8.2.7). Thus σ can, under these assumptions, be written

$$\sigma = \frac{d \ln (K/L)}{d \ln (w/r)} = \frac{d(K/L)/(K/L)}{d(w/r)/(w/r)} = \frac{(w/r)}{(K/L)}\frac{d(K/L)}{d(w/r)} \qquad {}^*(8.3.12)$$

The elasticity of substitution is thus a measure of how rapidly factor proportions change for a change in relative factor prices. It is therefore a measure of the curvature of the isoquants. Figure 8.3 illustrates σ by showing isoquants for each of two production functions. In this case isoquant 1 exhibits greater elasticity of substitution than isoquant 2, since the same change in relative factor prices elicits for 1 a greater change in factor proportions, shown geometrically as the change in the slope of the ray from the origin to the tangency between isocost and isoquant.

[13] See Allen (1938)

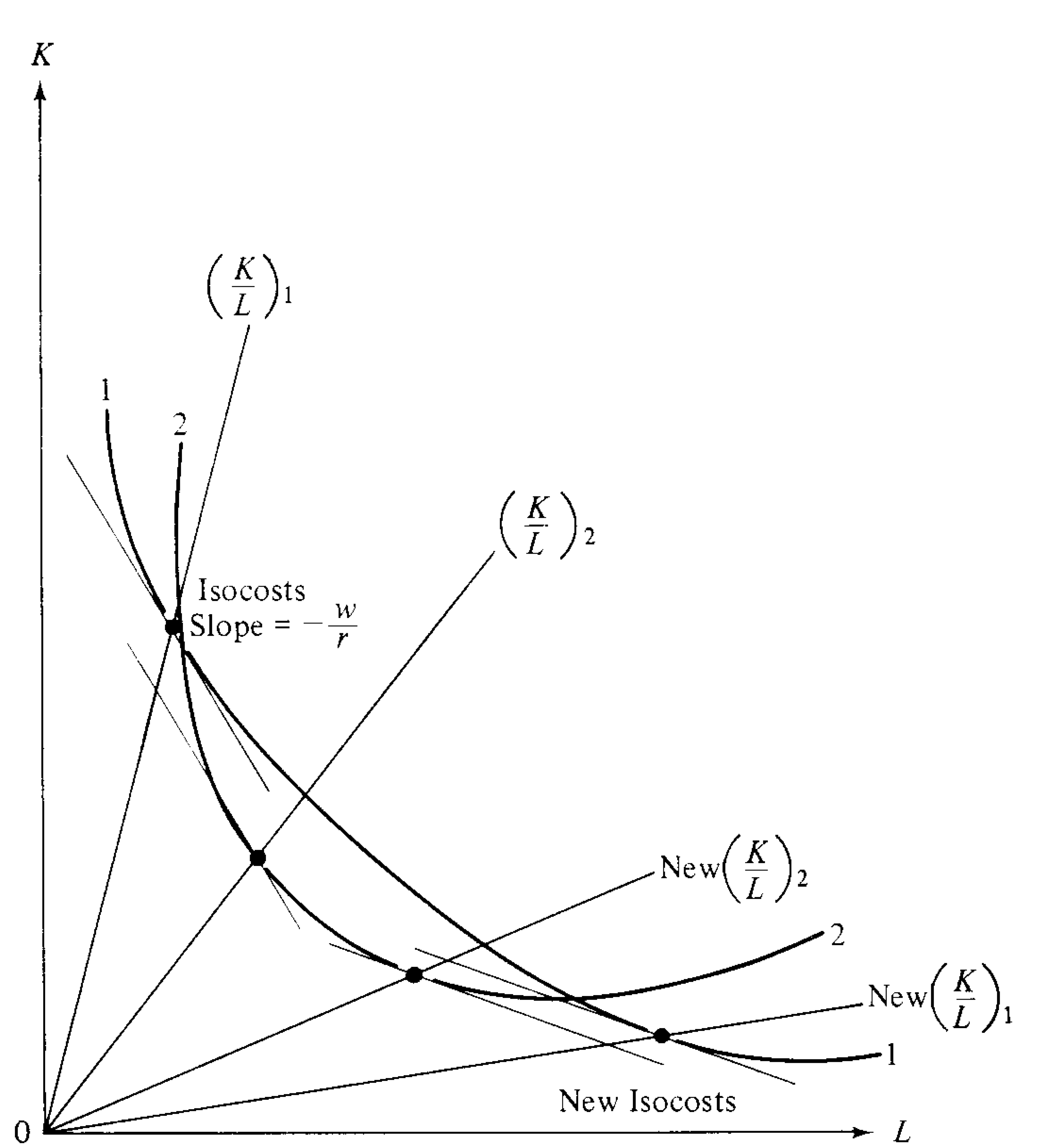

Figure 8.3 Elasticity of Substitution of Isoquant 1 > Isoquant 2

One of the most widely used production functions for empirical estimation is the *Cobb-Douglas production function*, of the form[14]

$$y = AL^\alpha K^\beta \qquad *(8.3.13)$$

where A, α, and β are fixed positive parameters. This specification is identical to that of the last chapter for constant elasticity demand functions. In this case the exponents are the elasticities of output with respect to each input:

$$\alpha = \frac{L}{y}\frac{\partial y}{\partial L}, \quad \beta = \frac{K}{y}\frac{\partial y}{\partial K}, \qquad 0 < \alpha < 1, \ 0 < \beta < 1, \ \alpha + \beta \leqslant 1. \qquad (8.3.14)$$

The constancy of these elasticities is a characteristic of the Cobb-Douglas production function, and the inequalities in (8.3.14) ensure that conditions (8.3.2)–(8.3.4) are satisfied. The sum of the elasticities is the degree of homogeneity of the function, since

$$f(\lambda L, \lambda K) = A(\lambda L)^\alpha(\lambda K)^\beta = \lambda^{\alpha+\beta}AL^\alpha K^\beta = \lambda^{\alpha+\beta}f(L, K). \qquad (8.3.15)$$

[14] See Douglas (1948) and Nerlove (1965).

The Cobb-Douglas function is linear in the logarithms of the variables. Considering cross-section studies, the Cobb-Douglas function for the ith firm, after taking logarithms and adding a stochastic disturbance term u_i to account for variations in the technical or productive capabilities of the ith firms, is[15]

$$\ln y_i = a + \alpha \ln L_i + \beta \ln K_i + u_i \qquad (a = \ln A). \qquad *(8.3.16)$$

It is assumed here that the parameters α and β (and also the prices) are the same for all firms, differences among firms being summarized by the u_i. One way of estimating the parameters a, α, and β is to estimate this equation directly, given data on output y_i, labor input L_i, and capital input K_i. Since such data are often not available, especially data on capital, the function has generally been estimated indirectly. Even if these data were available, however, a direct estimation of (8.3.16) would be a somewhat questionable procedure, since the explanatory variables $\ln L_i$ and $\ln K_i$ are endogenous variables, jointly determined with $\ln y_i$, and are not independent of the stochastic disturbance term, leading to a problem of simultaneous-equations estimation, specifically an endogenous explanatory variable. They also tend not to be independent of one another, leading to a possible problem of multicollinearity. Furthermore, the variance of the stochastic disturbance term need not be constant, leading to a problem of heteroskedasticity.

The classical approach to estimating the Cobb-Douglas production function is to assume perfect competition and profit maximization, so conditions (8.2.6) are applicable. These conditions require that marginal productivity equal the real wage:

$$\frac{\partial y_i}{\partial L_i} = \alpha\frac{y_i}{L_i} = \frac{w}{p}, \qquad \frac{\partial y_i}{\partial K_i} = \beta\frac{y_i}{K_i} = \frac{r}{p}. \qquad (8.3.17)$$

These conditions can be written

$$\alpha = \frac{wL_i}{py_i}, \qquad \beta = \frac{rK_i}{py_i}. \qquad *(8.3.18)$$

Here the common denominator is py_i, the value of output. The numerator wL_i is payments to labor, and the other numerator, rK_i, is payments to capital. Thus, these conditions require that labor's share of total income be the parameter α, while the share of capital be the parameter β. Since the total value of output equals total income (the sum of labor income and capital income),

$$py_i = wL_i + rK_i \qquad (8.3.19)$$

[15] An additive stochastic disturbance term here means that in the original formulation the stochastic disturbance is multiplicative, (8.3.13) taking the form

$$y_i = AL_i^\alpha K_i^\beta e^{u_i}$$

The multiplicative nature of this stochastic disturbance term is justified mainly by convenience.

conditions (8.3.18) and (8.3.19) require that

$$\alpha + \beta = 1. \tag{8.3.20}$$

This condition is precisely the condition that the Cobb-Douglas function exhibit constant returns to scale.

Assuming constant returns to scale, equation (8.3.16) implies that

$$\ln y_i = a + \alpha \ln L_i + (1 - \alpha) \ln K_i + u_i \tag{8.3.21}$$

which, further, implies that

$$\ln\left(\frac{y_i}{L_i}\right) = a + (1 - \alpha)\ln\left(\frac{K_i}{L_i}\right) + u_i. \qquad *(8.3.22)$$

This equation is the production function in *intensive form*, relating output per worker to the capital-labor ratio. Estimating this equation yields an estimate of $1 - \alpha$, the elasticity of output with respect to capital, where α is the elasticity with respect to labor. Using this equation rather than (8.3.16) also reduces the problems of multicollinearity and heteroskedasticity; the use of ratios to reduce the problem of heteroskedasticity having been discussed in Section 6.3.

An alternative method of estimation, assuming constant returns to scale, perfect competition, and profit maximization, is based on the share of labor income in output. From (8.3.17) and constant returns to scale

$$\alpha = \frac{wL_i}{py_i} = s_L, \qquad \beta = 1 - \alpha \qquad *(8.3.23)$$

where s_L is the share of labor in national income. Thus the shares yield direct estimates of both α and β under these assumptions.[16] This method requires no data on capital inputs, either in total [as in (8.3.16)] or relative to labor [as in (8.3.22)], but it does depend on the assumption of constant returns to scale and hence cannot be used to test hypotheses about returns to scale.

Assuming constant returns to scale, perfect competition, and profit maximization, the marginal-productivity equation (8.3.17) implies a log linear relation between output per worker and the real wage:

$$\ln \frac{y_i}{L_i} = \ln \frac{w}{p} - \ln \alpha. \tag{8.3.24}$$

Adding a stochastic disturbance term to this relation, to account for errors made by firms in choosing inputs so as to maximize profits, leads to a regression equation. The estimated intercept then provides an estimate of the (negative of the logarithm of the) elasticity α.

[16] With cross-section or time-series data the shares can be estimated as the *geometric* means of shares calculated for each production unit or at each time period. See Problem 8-K.

There are, then, at least four different methods of estimating the parameters of the production function, involving alternative assumptions and econometric problems.[17] The first is that of estimating *the production function itself* in log linear form, (8.3.16). This method requires no further assumptions, e.g., as to returns to scale, but it typically leads to econometric problems of simultaneity (endogenous explanatory variable), multicollinearity, and heteroskedasticity. The second method is that of estimating the *intensive production function* in log linear form, (8.3.22). This method reduces the problems of multicollinearity and heteroskedasticity, but it does require the assumption of constant returns to scale and hence cannot be used to test for increasing or decreasing returns to scale. It also has an endogenous explanatory variable. The third and fourth methods, those of *factor shares*, (8.3.23), and of the *marginal productivity relation*, (8.3.24), respectively, eliminate the simultaneity, multicollinearity, and heteroskedasticity problems, but require the assumptions of constant returns to scale, perfect competition, and profit maximization. None of these methods dominates the others. Each is appropriate in particular situations, depending upon what can be assumed and what is to be investigated.[18] The resulting para-

[17] See Walters (1963) and Nerlove (1965). A fifth method is discussed in the next footnote.

[18] A fifth method is to estimate the simultaneous system consisting of the production function and the first order conditions for profit maximization

$$y_i = AL_i^{\alpha} K_i^{\beta} e^{u_i}$$

$$\frac{\partial y_i}{\partial L_i} = \frac{\alpha y_i}{L_i} = \frac{w}{p} e^{v_i}$$

$$\frac{\partial y_i}{\partial K_i} = \beta \frac{y_i}{K_i} = \frac{r}{p} e^{w_i}.$$

Here u_i is a technical disturbance term, affecting the efficiency of the production process, and v_i and w_i are economic disturbance terms, affecting the attainment of the two profit-maximization conditions. Taking logarithms gives the linear system

$$\ln y_i = a + \alpha \ln L_i + \beta \ln K_i + u_i$$

$$\ln y_i = -\ln \alpha + \ln L_i + \ln \frac{w}{p} + v_i$$

$$\ln y_i = -\ln \beta + \ln K_i + \ln \frac{r}{p} + w_i$$

which is the structural form for a system in which $\ln y_i$, $\ln L_i$, and $\ln K_i$ are the endogenous variables and $\ln w/p$ and $\ln r/p$ are the exogenous variables (assuming perfect competition). See Marschak and Andrews (1944), Nerlove (1965), Hildebrand and Liu (1965), Zellner, Kmenta, and Dreze (1966), Griliches and Ringstad (1971), and Problem 8-I. The first method of estimation entails estimating only the first equation of this system. Estimating the complete system is generally superior to estimating only the first equation from both economic and econometric standpoints. From an economic standpoint estimating the complete system expresses the assumption that the data reflect both the behavior of the decision maker (the firm) and the technology, while the first equation reflects only the technology. From an econometric standpoint the estimator of only the first equation involves simultaneous-equations bias, so the estimators will be biased and inconsistent, as discussed in Chapter 11.

meter estimates will generally be different, and there is little evidence to suggest which estimates come closest to true values.

Table 8.1 presents some estimates of the Cobb-Douglas production function for the macroeconomy of a nation or state using time-series data and the technique of least squares, as applied to (8.3.16). The discussion of the previous section referred, however, to a single firm. Estimates of production relationships for macroeconomies, such as those of Table 8.1, are based upon the further assumption that the macroeconomic entity acts as if it were representative of the underlying microeconomic entities.[19] The index i then ranges over time.

The four alternative estimates for the United States and the two alternative estimates for New Zealand in Table 8.1 are based on alternative ways of measuring inputs and output. Douglas concluded, based on the results reported in this table and other results (some based on cross-section rather than time-series data), that production exhibits approximately constant returns to scale. He also concluded that the factors of production receive approximately the share they would receive under competitive conditions, given as the elasticity of output with respect to the factor. Later authors have questioned these conclusions, however. One criticism was based on the multicollinearity in the data used. Another was based on the condition that the total value of output equal total income (8.3.19), which creates a bias of the estimated production function toward these results.[20] To show this bias, using index numbers in (8.3.16), it follows that (ignoring the stochastic disturbance term)

$$\ln \frac{y_i}{\bar{y}_i} = \alpha \ln \frac{L_i}{\bar{L}_i} + \beta \ln \frac{K_i}{\bar{K}_i} \tag{8.3.25}$$

where $\bar{y}_i$, $\bar{K}_i$, and $\bar{L}_i$ are base-year quantities of output, capital, and labor, respectively, for the ith firm.[21] But if y_i, K_i, and L_i do not vary appreciably from the base quantities, the ratios are close to unity, so

$$\ln \frac{y_i}{\bar{y}_i} \approx \frac{y_i}{\bar{y}_i} - 1, \ \ln \frac{L_i}{\bar{L}_i} \approx \frac{L_i}{\bar{L}_i} - 1, \ \ln \frac{K_i}{\bar{K}_i} \approx \frac{K_i}{\bar{K}_i} - 1. \tag{8.3.26}$$

Thus (8.3.25) implies that

$$\frac{y_i}{\bar{y}_i} \approx \alpha \frac{L_i}{\bar{L}_i} + \beta \frac{K_i}{\bar{K}_i} + (1 - \alpha - \beta) \tag{8.3.27}$$

[19] Formally, under certain aggregation conditions, it may be possible to aggregate microeconomic production functions into macroeconomic production functions. The aggregation conditions here are comparable to those for a household, as discussed in Section 7.7. Several new issues arise here, however, with regard to aggregation. One is that of *reswitching*, where different ratios of inputs are used at different ratios of input prices. Others are *efficiency* and *technical change*, which are both affected by and affect aggregation of micro production functions into macro production functions. Also some exogeneity assumptions change (e.g., factor prices). On the general problems of aggregation see Walters (1963), Green (1964), and Problem 8-K. For a study of efficiency and aggregation see Houthakker (1955-6), who derived a macro Cobb-Douglas production function on the basis of micro fixed-coefficients (input-output) production functions [introduced in (8.3.33)], assuming a specific probability distribution (the Pareto distribution) of firms over possible values of the input coefficients. Generalizations and related approaches appear in Johansen (1972) and Sato (1975).

[20] See Cramer (1969). The bias toward constant returns to scale is an example of the practical problem stemming from the aggregation problem, as previously discussed in Section 7.7. Aggregate output is calculated from the total value of payments to factors of production, so using this value for output to test for returns to scale is questionable. Similarly, if aggregate capital data are constructed by subtracting the value of labor input from the value of output and deflating then a test of returns to scale is also questionable.

[21] The intercept drops out of the equation, since, if

$$y_i = AL_i^{\alpha}K_i^{\beta}, \quad \bar{y}_i = A\bar{L}_i^{\alpha}\bar{K}_i^{\beta}$$

then, taking ratios,

$$\frac{y_i}{\bar{y}_i} = \left(\frac{L_i}{\bar{L}_i}\right)^{\alpha} \left(\frac{K_i}{\bar{K}_i}\right)^{\beta}.$$

Taking logarithms of this equation yields (8.3.25).

Table 8.1. Estimates of the Cobb-Douglas Production Function

Country, Time Period	*Labor Elasticity* α	*Capital Elasticity* β	*Returns to Scale* $\alpha + \beta$	*Average Labor Share* s_L
United States I 1899–1922	0.81 (0.15)	0.23 (0.06)	1.04	0.61
United States II 1899–1922	0.78 (0.14)	0.15 (0.08)	0.93	0.61
United States III 1899–1922	0.73 (0.12)	0.25 (0.05)	0.98	0.61
United States IV 1899–1922	0.63 (0.15)	0.30 (0.05)	0.93	0.61
New Zealand I 1915–1916 and 1918–1935	0.42 (0.11)	0.49 (0.03)	0.91	0.52
New Zealand II 1923–1940	0.54 (0.02)			0.54
New South Wales, Australia 1901–1927	0.78 (0.12)	0.20 (0.08)	0.98	
Victoria, Australia 1902–1929	0.84 (0.34)	0.23 (0.17)	1.07	

Source Douglas (1948).

NOTE Numbers in parentheses are standard errors.

so that

$$py_i = \left(\alpha p \frac{\bar{y}_i}{\bar{L}_i}\right) L_i + \left(\beta p \frac{\bar{y}_i}{\bar{K}_i}\right) K_i + (1 - \alpha - \beta)p\bar{y}_i. \quad (8.3.28)$$

Comparing this equation to (8.3.19), however, it follows that

$$\alpha p \frac{\bar{y}_i}{\bar{L}_i} \approx w, \quad \beta p \frac{\bar{y}_i}{\bar{K}_i} \approx r, \quad (1 - \alpha - \beta)p \approx 0. \quad (8.3.29)$$

These results imply that

$$\alpha + \beta \approx 1 \quad (8.3.30)$$

which means returns to scale are approximately constant, and

$$\frac{w\bar{L}_i}{p\bar{y}_i} \approx \alpha, \quad \frac{r\bar{K}_i}{p\bar{y}_i} \approx \beta \quad (8.3.31)$$

which means that factor shares are approximately the elasticities, α and β, the shares received under competitive conditions. Thus, assuming only small variations in output and inputs, the form of the production function and the equality of the values of output and income imply that the production function exhibits approximately constant returns to scale and that factor shares are approximately the elasticities.

A second example of the Cobb-Douglas production function is the estimation by Kimbell and Lorant of a production function for physicians' services.[22] The data were obtained from an American Medical Association survey of physician activities in 1970. Altogether there were 844 observations on physicians in both solo and group practices. The estimated function is

$$\begin{aligned} \ln (py) = 2.826 &+ \underset{(0.052)}{0.255} \ln h + \underset{(0.037)}{0.708} \ln d + \underset{(0.030)}{0.302} \ln a \\ &+ \underset{(0.042)}{0.074} \ln r, \quad R^2 = 0.906. \end{aligned} \quad (8.3.32)$$

Here py is gross revenue from medical practice, a measure of output for the heterogeneous services provided by physicians; h is the average number of hours worked by (full-time) physicians in the practice; d is the number of (full-time equivalent) physicians in the practice; a is the number of (full-time equivalent) allied health personnel (e.g. nurses) employed by the practice; and r is the number of rooms used in the practice, a measure of capital input. According to these results, the elasticity of gross revenue with respect to physicians' hours is 0.255, so a 10% increase in hours would increase gross revenue by about 2.6%. The

[22] See Kimbell and Lorant (1974)

elasticity for aides implies that increasing the number of aides by one-third would increase gross revenue by about 10%. The sum of the elasticities is 1.084, which is significantly greater than unity at the 0.01 confidence level, indicating increasing returns to scale for physicians' services.

Another form of the production function is the *input-output production function*,[23]

$$y = \min\left(\frac{L}{a}, \frac{K}{b}\right), \quad a, b > 0. \quad *(8.3.33)$$

Here the isoquants are right-angled (L-shaped), as shown in Figure 8.4, and the production function permits no substitution between the inputs. The condition of profit maximization, given positive factor wages, is

$$\frac{L}{a} = \frac{K}{b} \quad (8.3.34)$$

that is, operation at the vertex of the isoquants. Then

$$a = \frac{L}{y}, \quad b = \frac{K}{y} \quad (8.3.35)$$

so the parameters a and b are, respectively, the input of labor per unit of output and the input of capital per unit of output—the fixed proportions of inputs to output. The equations in (8.3.35) are typically used to estimate the parameters a and b, which are called "technical coefficients." The estimation is typically based on a single observation, so regression techniques are not used. The estimated production function is used in input-output studies concerned with the interrelationships among productive sectors that arise from the fact that the inputs of any one sector consist of portions of the outputs of other sectors.[24]

One of the most widely used production functions in empirical work is the *constant elasticity of substitution* (CES) production function, of the form[25]

$$y = A[\delta L^{-\beta} + (1 - \delta)K^{-\beta}]^{-1/\beta}. \quad *(8.3.36)$$

[23] See Leontief (1951, 1966) and Chenery and Clark (1959).

[24] Let x_{ij} be the input of commodity i, as produced by sector i, that is used in the production of commodity j by sector j. If x_j is the output of sector j, then the technical coefficients comparable to (8.3.35) are given as

$$a_{ij} = \frac{x_{ij}}{x_j}, \quad i, j = 1, 2, \ldots, n.$$

See Intriligator (1971).

[25] See Arrow, Chenery, Minhas, and Solow (1961), Brown and de Cani (1963), and Minhas (1963). Note that β here plays an entirely different role from the β in the Cobb-Douglas production function.

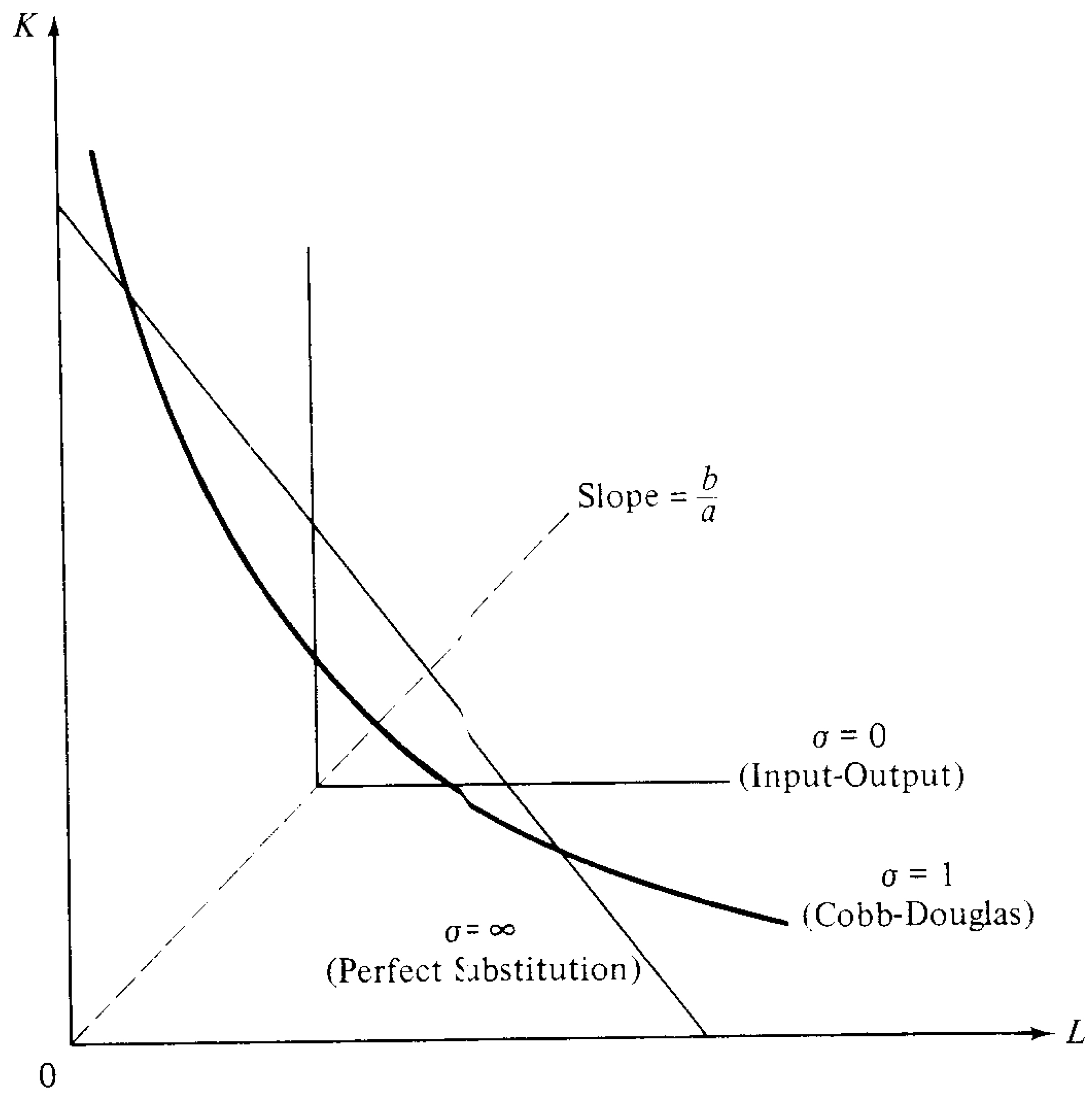

Figure 8.4 Isoquants of the CES Production Function Corresponding to Different Values of the Elasticity of Substitution, σ

The parameters defining this production function are

A: scale parameter, $A > 0$

δ: distribution parameter, $0 < \delta < 1$

β: substitution parameter, $\beta \geq -1$.

The name of the function is based upon the concept of the elasticity of substitution, σ, defined in (8.3.11). In general, the elasticity of substitution σ varies with K and L. Assuming σ is constant, however, and solving the resulting differential equation yields, in the constant-returns-to-scale case, precisely the CES function, where

$$\sigma = \frac{1}{1 + \beta} \qquad *(8.3.37)$$

justifying the interpretation of β as the substitution parameter. As defined in (8.3.11) σ must be nonnegative, so

$$\beta \geq -1. \qquad (8.3.38)$$

At the extreme value of $\beta = -1$ the CES function reduces to the linear function

$$y = A[\delta L + (1 - \delta)K] \quad \text{if} \quad \beta = -1, \text{ i.e., } \sigma = \infty. \qquad *(8.3.39)$$

The isoquants for this case are linear, the slope of each being $-\delta/(1-\delta)$. In this case of perfect substitution $\sigma = \infty$, meaning that certain slight changes in w/r would lead to discontinuous changes in K/L, e.g., from one boundary point to another. At the other extreme value for β, namely in the limit as β approaches ∞, from (8.3.37), σ approaches zero and, in this case, in the limit of the CES as $\beta \to \infty$, it approaches the input-output production function, as in (8.3.33):

$$y = \min\left(\frac{L}{a}, \frac{K}{b}\right) \quad \text{if} \quad \beta \to \infty, \text{ i.e., } \sigma \to 0. \qquad (8.3.40)$$

In the limit as β approaches zero, σ approaches unity; this is the case of the Cobb-Douglas production function, where, taking the limit as $\beta \to 0$, the CES approaches (8.3.13)

$$y = A_0 L^{\delta} K^{1-\delta} \quad \text{if} \quad \beta \to 0, \text{ i.e., } \sigma \to 1. \qquad (8.3.41)$$

Thus the CES is a family of production functions that includes, as special cases, the Cobb-Douglas, input-output, and linear production functions. The isoquants of these various cases are shown in Figure 8.4, and estimation of σ gives information on the curvature of the isoquants. It might be noted that the isoquants of the CES production function intersect the axes if $\sigma > 1$, and they are asymptotic to horizontal and vertical lines if $\sigma < 1$.[26]

The CES can be estimated by using the conditions of profit maximization (8.2.6). The marginal product of labor can be written

$$\frac{\partial y}{\partial L} = A'\left(\frac{y}{L}\right)^{1+\beta} \qquad (8.3.42)$$

where A' is a constant, so setting the marginal product equal to the real wage yields

$$A'\left(\frac{y}{L}\right)^{1+\beta} = \frac{w}{p}. \qquad (8.3.43)$$

Solving for output per manhour (labor productivity) y/L,

$$\frac{y}{L} = A''\left(\frac{w}{p}\right)^{1/1+\beta} \qquad (8.3.44)$$

so, taking logs, and using (8.3.37):

[26] See Problem 8-N. As noted there, the case $\sigma > 1$ is excluded if both factors are essential in that output is zero if either factor input is zero—conditions (8.3.2). The case $\sigma < 1$ is consistent with factors being essential in this sense.

$$\ln \frac{y}{L} = a + \frac{1}{1+\beta} \ln \frac{w}{p} = a + \sigma \ln \frac{w}{p}, \qquad a = \ln A''. \tag{8.3.45}$$

This equation relates output per worker to the real wage, where a and σ are constants, σ being the coefficient of ln (w/p). The special case of the Cobb-Douglas for which $\sigma = 1$ was presented earlier in (8.3.24). Equation (8.3.45), with an additive stochastic disturbance term on the right-hand side, can be estimated using least-squares regression. Alternatively the equation can be solved for the real wage and the resulting equation,

$$\ln \frac{w}{p} = a' + (1 + \beta) \ln \frac{y}{L} \tag{8.3.46}$$

in which the dependent and explanatory (exogenous) variables have switched roles, can be estimated to obtain $1/(1 + \beta)$ as an estimate of σ. Such an estimation could, for example, utilize cross-section data on output, y, labor, L, and the real wage, w/p, assuming the real wage is exogenous and all entities in the cross section use the same underlying production function. This was the approach used by Arrow, Chenery, Minhas, and Solow, who estimated σ in (8.3.45) using cross-section data for specific industries from 19 different countries over the period 1950–1956. They found that their estimates of σ tended to cluster below unity, with 10 out of 24 industries having an estimated σ statistically different from (and below) unity. Their approach was extended by Fuchs who, using the same data, distinguished developed from less developed countries in the sample of 19 countries. He showed, using analysis of covariance, that the developed and less developed countries exhibit different intercept a in (8.3.45), but the same σ, and he reestimated σ, using a dummy variable to reflect the different intercept in the developed countries.[27] His results for σ are presented in Table 8.2, where industries are arranged in order of increasing estimated σ. The estimates tend to cluster about unity, ranging from a low of 0.658 for clay products to a high of 1.324 for grain and mill products. Only one of the estimates is statistically significantly different from unity. This one exception is glass, for which the estimated σ is significantly above unity, indicating a greater ease of substitution between capital and labor than that indicated by the Cobb-Douglas function. Since this is the only such case out of 24 industries, the Fuchs study provides justification for continued use of the Cobb-Douglas function. Various other studies also find that the estimated elasticity of substitution does not differ significantly from unity, justifying use of the Cobb-Douglas production function.[28]

The CES production function can be extended to the case of nonconstant returns to scale, but homogeneous case, for which the function can be written

$$y = A[\delta L^{-\beta} + (1 - \delta)K^{-\beta}]^{-h/\beta} \tag*{*(8.3.47)}$$

where h is the degree of homogeneity of the function. This case reduces to (8.3.36) if $h = 1$, the constant-returns-to-scale case. The general function was estimated by Dhrymes using cross-section data on U.S. states.[29] Some of his results are shown in Table 8.3. From his results for h, most industries operate at

[27] See Fuchs (1963).

[28] See Griliches (1967), Zarembka (1970), and Griliches and Ringstad (1971). Griliches (1967) found only one industry (paper) out of 17 in which use of the Cobb-Douglas production function was not justified. It might be noted, however, that Nerlove (1967), in surveying over 40 papers, found conflicting estimates of the elasticity of substitution, with values ranging from 0.068 to 1.16. He concluded that the estimates are sensitive to the period under consideration and the concepts employed. In a later survey, Mayor (1969) found that studies using cross-section data obtain estimates of the elasticity of substitution close to unity while those using time-series obtain estimates considerably less than unity, clustering around one-half. Johansen (1972) attributes this difference between cross-section and time-series studies to the "putty-clay" nature of technology, according to which substitution possibilities are reduced once investment has occurred and capital is in place. Johansen suggests that the firm decides factor proportions before investment in new plant and equipment occurs and that, after this investment has occurred, subsequent decisions involve only the scale of operation. Cross-section estimates may reveal *ex-ante* substitution possibilities, before capital is in place, and so exhibit relatively high elasticities of substitution. Time-series estimates, by contrast, tend to reveal *ex-post* substitution possibilities, after capital is in place, and so exhibit relatively low elasticities of substitution. For further discussion of the putty-clay model, which distinguishes ex-ante and ex-post substitution possibilities, such as substitution possibilities ex-ante but fixed coefficients ex-post see Johansen (1959, 1972) and Bliss (1968).

[29] See Dhrymes (1965) and Kmenta (1967) and Zarembka (1970). See also Brown and de Cani (1963), where the derivation and estimation of the CES production function allowed for $h \neq 1$.

Table 8.2. Estimates of σ, the Elasticity of Substitution for 19 Countries, 1950–1956

Industry	*Estimated Elasticity of Substitution* σ	*Industry*	*Estimated Elasticity of Substitution* σ
Clay products	0.658 (0.197)	Furniture	1.043 (0.090)
Iron and steel	0.756 (0.112)	Bakery products	1.056 (0.105)
Sugar	0.898 (0.183)	Fats and oils	1.058 (0.181)
Dairy products	0.902 (0.080)	Misc. chemicals	1.060 (0.088)
Pulp and paper	0.912 (0.175)	Ceramics	1.078 (0.125)
Nonferrous metals	0.935 (0.197)	Lumber and wood	1.083 (0.141)
Knitting mills	0.948 (0.083)	Fruit and vegetable canning	1.086 (0.098)
Leather finishing	0.975 (0.100)	Basic chemicals	1.113 (0.104)
Textile spinning	0.976 (0.104)	Tobacco	1.215 (0.208)
Metal products	1.006 (0.166)	Glass	1.269 (0.096)
Printing and publishing	1.021 (0.085)	Cement	1.308 (0.217)
Electrical machinery	1.026 (0.214)	Grain and mill products	1.324 (0.167)

Source Fuchs (1963).

Table 8.3. Estimates of the CES Production Function for the United States

Industry	Elasticity of Substitution σ		Degree of Homogeneity h	
Machinery, except electrical	0.050	extremes	1.029	
Rubber products	1.984	extremes	1.092	
Textile mill products	0.936		closest to unity 0.997	extremes
Lumber and wood products	1.109		1.218	extremes
Furniture and fixtures	1.001	closest to unity	1.017	
Chemicals	0.506		1.042	
Food	0.469		1.044	

Source Dhrymes (1965). Results have been rounded.

or above constant returns to scale ($h = 1$), with textile mill products exhibiting the lowest degree of homogeneity. From his results for σ, most consumer goods (e.g., textile mill products, furniture) are produced with an elasticity of substitution of approximately unity, i.e., close to the Cobb-Douglas production function. Most producer goods (e.g., machinery, chemicals), however, are produced with an elasticity of substitution significantly below unity, approaching in some cases the input-output production function. However, other studies have arrived at radically different results for certain industries. The study by Ferguson, for example, of U.S. manufacturing industries, using time-series data from the U.S. Census for 18 industries, 1949–1961, found an estimate of σ for non-electrical machinery of 1.041 (0.04), in contrast to the Dhrymes value of 0.050, and for chemicals of 1.248 (0.072), in contrast to the Dhrymes value of 0.506. Some of the other industries yielded somewhat comparable estimates, however--for example, for textile mill products [1.104 (0.44) vs. 0.936], lumber and wood [0.905 (0.067) vs. 1.109], furniture and fixtures [1.123 (0.045) vs. 1.001], and food [0.241 (0.20 vs. 0.469].[30]

It has already been noted that the Cobb-Douglas production function is a special case of the CES production function, corresponding to an elasticity of substitution of unity. Conversely the CES production function can be viewed as a generalization of the Cobb-Douglas production function to the case of a non-unitary, but constant, elasticity of substitution. For example, expanding $\ln y$ in a Taylor's series approximation of the CES around $\beta = 0$ yields[31]

[30] See Ferguson (1965) and Nerlove (1967).

[31] See Kmenta (1967). This approximation can be used to estimate the parameters of the CES production function. Using this approach, Kmenta estimated σ as 0.672 and h as 1.179. Neither of these estimates, however, was significantly different from unity, so a Cobb-Douglas production function with constant returns to scale is not ruled out by his findings. It should be noted, however, that the estimated σ is not invariant to a change in units of measurement.

$$\ln y \approx a + h\delta \ln L + h(1 - \delta) \ln K - \frac{\beta h \delta(1 - \delta)}{2} (\ln L - \ln K)^2 \qquad (8.3.48)$$

The first several terms on the right are those of the Cobb-Douglas production function, and the last term accounts for $\sigma \neq 1$. This approximation is better the closer the elasticity of substitution is to unity, and it reduces to the Cobb-Douglas case if $\beta = 0$.

While the CES production function represents one generalization of the Cobb-Douglas production function, the Cobb-Douglas has also been generalized in several other ways. One such way is the *transcendental production function*, of the form[32]

$$y = AL^{\alpha}K^{\beta}e^{\alpha' L + \beta' K} \qquad A > 0, \qquad \alpha', \beta' \leqslant 0. \qquad (8.3.49)$$

This case reduces to the Cobb-Douglas if α' and β' vanish. Taking logarithms

$$\ln y = a + \alpha \ln L + \beta \ln K + \alpha' L + \beta' K \qquad (8.3.50)$$

so $\ln y$ is a linear function of the inputs L and K, as well as the logarithms of the inputs $\ln L$ and $\ln K$. For this function it is possible for marginal products to rise before eventually falling. This function also permits variable elasticity of production and variable elasticity of substitution over the range of inputs.

A second approach to generalizing the Cobb-Douglas production function is the *Zellner-Revankar production function*, of the form[33]

$$ye^{cy} = AL^{\alpha}K^{\beta}, \qquad c \geqslant 0. \qquad (8.3.51)$$

This case reduces to the Cobb-Douglas form if $c = 0$. Taking logarithms,

$$\ln y + cy = a + \alpha \ln L + \beta \ln K. \qquad (8.3.52)$$

This case is essentially the obverse of the transcendental case. In the transcendental case inputs and logarithms of inputs enter on the right-hand side, while in this case output and the logarithm of output enter on the left-hand side.

A third approach to generalizing the Cobb-Douglas production function is the *Nerlove-Ringstad production function*, of the form[34]

$$y^{1 + c \ln y} = AL^{\alpha}K^{\beta}, \qquad c \geqslant 0. \qquad (8.3.53)$$

This case reduces to the Cobb-Douglas form if $c = 0$. Taking logarithms,

$$(1 + c \ln y) \ln y = a + \alpha \ln L + \beta \ln K \qquad (8.3.54)$$

so $\ln y$ and $(\ln y)^2$ appear on the left-hand side.

[32] See Halter, Carter, and Hocking (1957). Note that α' and β' are not invariant to a change in units of measurement. For a discussion of other functional forms not necessarily related to the Cobb-Douglas production function see Heady and Dillon (1961).

[33] See Zellner and Revankar (1969).

[34] See Nerlove (1963) and Ringstad (1967).

A fourth approach to generalizing the Cobb-Douglas production function is the *translog production function*, of the form[35]

$$\ln y = a + \alpha \ln L + \beta \ln K + \gamma \ln L \ln K + \delta(\ln L)^2 + \epsilon(\ln K)^2. \quad *(8.3.55)$$

This function, which is quadratic in the logarithms of the variables, reduces to the Cobb-Douglas case if the parameters γ, δ, and ϵ all vanish; otherwise it exhibits nonunitary elasticity of substitution. In general this function is quite flexible in approximating arbitrary production technologies in terms of substitution possibilities. It provides a local approximation to any production frontier.[36]

The last several production functions are extensions of the Cobb-Douglas production function. The CES production function has also been generalized in different ways. One such generalization is the two-level production function.[37] For this function factors are combined according to the CES at one level to form "higher-level" factors, which are combined again according to the CES to produce output. An example is the production function

$$y = A\left\{[\delta_1 x_1^{-\beta_1} + (1 - \delta_1)x_2^{-\beta_1}]^{-\beta/\beta_1} + [\delta_2 x_3^{-\beta_2} + (1 - \delta_2)x_4^{-\beta_2}]^{-\beta/\beta_2}\right\}^{-1/\beta} \quad (8.3.56)$$

Here x_1 and x_2 are combined into a "higher-level" factor, where the elasticity of substitution is $(1 + \beta_1)^{-1}$, while x_3 and x_4 are combined with an elasticity of substitution of $(1 + \beta_2)^{-1}$. The "higher-level" inputs are then combined with an elasticity of substitution of $(1 + \beta)^{-1}$. Another generalization of the CES is the *VES production function*, i.e., the variable-elasticity-of-substitution production function.[38] For this function the elasticity of substitution varies with the factor proportions (the ratio of the inputs). Such a relationship can be estimated by regressing the log of output per worker on both the real wage [as in (8.3.45)] and the capital-labor ratio.[39]

[35] "Translog" is short for "transcendental logarithmic". See Christensen, Jorgenson, and Lau (1973) and Griliches and Ringstad (1971). More generally, for n inputs, the translog function is

$$\ln y = a + \sum_{i=1}^{n} \alpha_i \ln x_i + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln x_i \ln x_j$$

where x_i is the ith input and $\gamma_{ij} = \gamma_{ji}$. Note that this function is also not invariant to a change of units.

[36] It can also be applied to other frontiers, e.g., to demand functions or to price frontiers.

[37] See Sato (1967).

[38] See Sato and Hoffman (1968); see also Lu and Fletcher (1968), Revankar (1971), and Lovell (1973).

[39] See Hildebrand and Liu (1965). In most industries the coefficient of the capital-labor ratio is significant.

8.4 Estimation of cost curves and cost functions

Cost curves, based on economic theory, were developed in Section 8.2, equations (8.2.14)-(8.2.20), and illustrated in Figure 8.2. A variety of cost curves, including total, average, and marginal cost curves, have been estimated empirically for particular industries.[40]

A simple example of a total cost curve that satisfies the curvature postulated in Figure 8.2 is the *cubic cost curve*,

$$C = a_0 + a_1 y + a_2 y^2 + a_3 y^3 \quad *(8.4.1)$$

where a_0, a_1, a_2, and a_3 are given parameters. The average cost associated with the cubic cost curve is

$$\text{AC} = \frac{a_0}{y} + a_1 + a_2 y + a_3 y^2 \quad (8.4.2)$$

and marginal cost is given as

$$\text{MC} = a_1 + 2a_2 y + 3a_3 y^2. \quad (8.4.3)$$

For U-shaped average and marginal cost curves, as illustrated in Figure 8.2, the parameters must satisfy the restrictions

$$a_0 \geqslant 0,\ a_1 > 0,\ a_2 < 0,\ a_3 > 0,\ a_2^2 < 3a_3 a_1 \quad (8.4.4)$$

where a_0 is the fixed cost, the cost at zero output.

Empirical studies of cost curves typically estimate a long-run cost curve using cross-sectional data on firms in the industry, specifically data on total costs, output, and other relevant variables. Assuming that the same technology applies to all firms, that observed outputs are close to planned outputs, and that firms are seeking to minimize costs at each planned output level, it follows that the cost curve estimated from a scatter diagram of cost-output points represents an estimate of the long-run cost curve. The specific curve estimated is usually an average cost curve; and taking ratios as called for in such a curve reduces problems of heteroskedasticity.[41] In the long-run case, a_0 in the cubic cost curve (8.4.1), which is fixed cost, vanishes, and so the average cost curve in this case is

$$\text{AC} = a_1 + a_2 y + a_3 y^2. \quad (8.4.5)$$

[40] For surveys of cost curves see Johnston (1960) and Walters (1963, 1968). For cost functions see Shephard (1970).

[41] Note that if u is an additive stochastic disturbance term for the total cost curve and Var (u) $= Ky^2$, where K is a positive constant, then Var $(u/y) = K$; with this assumption the additive stochastic disturbance term in the average cost curve exhibits constant variance for all levels of output y.