Chapter 3

Costs, Learning Curves, and Scale Economies: From Simple to Multiple Regression

"Understanding of the underlying causes of the experience curve is still imperfect. The effect itself is beyond question. It is so universal that its absence is almost a warning of mismanagement or misunderstanding. Yet the basic mechanism that produces the experience curve effect is still to be adequately explained. . . . Our entire concept of competition, anti-trust, and non-monopolistic free enterprise is based on a fallacy if the experience curve effect is true."

BOSTON CONSULTING GROUP (1973, p. 2)

"The learning curve creates entry barriers and protection from competition by conferring cost advantages on early entrants and those who achieve large market shares."

A. MICHAEL SPENCE (1981, p. 68)

In this chapter we focus on econometric issues that are encountered in estimating the cost effects of scale economies and learning curves. This focus will help us to understand differences between simple and multiple regression, to assess the effects of incorrectly omitting variables from a regression equation, and to implement alternative ways of testing hypotheses.

The existence of scale economies, as well as the cost reductions due to learning curve effects, have important implications for market structure and economic welfare, since such phenomena can create barriers to entry, thereby protecting early entrants from effective market competition. On the other hand, acting in the interests of their shareholders, effective managers must focus a great deal of attention on these and other factors that might reduce their production costs.

According to one line of thinking, potential *economies of scale* can be exploited, resulting in average or unit cost decreases as the level of output increases per time period. If such economies of scale are available, then it may be rational for profit-maximizing or cost-minimizing managers to accelerate investment plans, reduce prices, and thereby achieve higher levels of production at lower unit cost than if scale economies were absent. Scale economies also have important implications for the competitive structure of an industry. But from what do economies of scale derive?

Economies of scale might arise because of the large fixed capital expenditures often required before any production can take place. Such reasoning has been used in explaining, for example, increasing returns to scale in the generation of electricity. Other sources of scale economies include physical-technical relationships and laws of nature. For example, in the case of boilers or pipelines, costs are typically closely related to the surface area of the boiler or pipeline, yet output (throughput) depends primarily on the potential volume. From basic geometry it is known that the area of a sphere or cylinder of constant proportions increases by the two-thirds power of its volume; this implies that the surface area increases less rapidly than the potential volume, suggesting that for both boilers and pipelines, one might expect economies of scale to be available. In the engineering literature this is known as the "two-thirds" or "six-tenths" rule, and it implies a rule of thumb that costs should increase by only about 60–67% as output or potential volume doubles.[1]

While these particular engineering examples are striking, one would not expect the two-thirds rule to apply to all types of production activities. Therefore obtaining estimates of scale economies is a very important issue; as we shall see, such a task can be carried out by using econometric methods.

Another important determinant of production costs is called the *learning curve* effect, a concept that is closely related to progress functions and experience curves. It is not clear precisely when the learning curve was first discovered.[2] A particularly well-known historical example occurred when it was noticed that, for certain emergency shipbuilding yards that were involved in the construction of the Liberty vessel during World War II, unit costs tended to fall over time as production experience accumulated, even when the annual level of operations remained unchanged; similar cost-reduction trends were noted for aircraft.[3] More generally, in numerous assembly line operations in which tasks are performed in a repetitive manner, it has frequently been found that workers tend to learn from their experiences, thereby reducing the time and labor costs required to perform prescribed tasks.[4] Even in operations that are less repetitive, such as the construction of coal-burning or nuclear power plants, learning curve effects have been found to occur.[5]

Today the notion that unit costs and unit prices tend to decline systematically in real terms as cumulative output increases is widespread and important in both the private and public sectors. In strategic management, for example, the existence of such experience or learning curve effects can provide a rationale for a pricing and marketing strategy in which producers initially price low (perhaps even below current marginal cost) in order to expand sales and gain market penetration rapidly, thereby quickly accumulating experience and exploiting the cost-reducing effects of such learning.[6] In modern industrial economics and marketing, the effects of learning curves on optimal pricing policies, make-or-buy decisions, market structure, and consumers' welfare are currently being modeled and analyzed, using increasingly sophisticated dynamic optimization techniques.[7] Finally, in the public policy sphere it has been argued by some that because of the existence of learning curves, it may make sense for governments to provide limited temporary protection to domestic manufacturers from foreign competitors.[8]

The above examples clearly demonstrate that learning or experience curves are very important in the formulation of strategy and policy. But how are they estimated, how might they be used in forecasting costs, and how are they related to scale economies?[9]

Chapter 3 is organized as follows. We begin by summarizing the economic theory of cost and production, define returns to scale, and characterize the effects of learning as shifts in production or cost functions. Next we derive estimating equations based on the Cobb-Douglas production function, discuss several measurement and econometric issues, and then present an overview of empirical research findings on scale economies and learning curves. Finally, we present a number of exercises based on classic data sets that involve you in the estimation of scale economies in the generation of electricity, as well as in the estimation and interpretation of learning curve effects in the manufacture of polyethylene and titanium dioxide. These exercises involve simple and multiple regressions, single and joint hypothesis tests, and the interpretation of R^2 in simple and multiple regressions.

3.1 THE UNDERLYING ECONOMIC THEORY OF COST AND PRODUCTION

The relationship among inputs and output is summarized by a production function. More specifically, denote flows of n input services as x_i, $i = 1, \ldots, n$, and the flow of output as y; the production function f indicates

the maximum possible output y given any combination of inputs x_i, $i = 1, \ldots, n$,

$$y = f(x_1, x_2, \ldots, x_n) \tag{3.1}$$

Essentially, therefore, the production function is an engineering relationship reflecting technology and the laws of nature.

While laws of nature do not change over time, our understanding of technology and nature has improved over the years, as has our ability to exploit technological possibilities. At any point in time, therefore, the various possible ways in which output y can be produced with differing combinations of inputs x_i can be thought of as a book containing alternative "blueprints." Since the number of pages in the book of such blueprints has grown with advances in knowledge, empirical analysts of production and cost relationships often add a variable into Eq. (3.1) to reflect such improvements in the state of technical knowledge. In this spirit, therefore, we insert a state of technical knowledge variable, called A, into Eq. (3.1) as follows:

$$y = f(x_1, x_2, \ldots, x_n; A) \tag{3.2}$$

One important characteristic of Eq. (3.2) is the notion of *returns to scale*. Suppose that all inputs are increased proportionately, say, by the factor μ, while A is held constant. If output y then increases by a factor greater than, equal to, or less than μ, returns to scale are said to be increasing, constant, or decreasing, respectively. As an example, if all inputs are increased by 100% and output increases by 115%, 100%, or 85%, then returns to scale are increasing (equal to about 1.15), constant (1.00), or decreasing (about 0.85), respectively. Further, *economies of scale* are typically computed as returns to scale minus 1; in the above example, economies of scale are positive (0.15), zero, or negative (−0.15), respectively.

While the production function (3.2) is essentially an engineering notion, economic content can be obtained by making assumptions concerning the economic behavior of firms. The two most common assumptions are those of profit maximization and cost minimization. Since the latter hypothesis is less stringent, we now adopt it.

The usual set of assumptions surrounding cost minimization behavior by firms includes the presumption that the level of output y produced by the firm is predetermined (it is not a contemporaneous endogenous variable chosen by the firm), that the prices of the n inputs, $p_1, \ldots, p_n$, are fixed and exogenous, and that the firm chooses its bundle of input quantities so as to minimize the total costs of producing y. This implies that, dual to the production function (3.2), there exists a *cost function* relating the minimum possible total cost $C \equiv \Sigma_i p_i x_i$ of producing a given level of output to the prices of the n inputs, the level of output y, and the state of technical knowledge A. The dual cost function can therefore be written as

$$C = g(p_1, p_2, \ldots, p_n, y; A) \tag{3.3}$$

Now define average or unit cost c as $c \equiv C/y$. If returns to scale are increasing, then doubling all inputs more than doubles output, and average cost falls; similarly, if returns to scale are decreasing, then doubling all inputs results in less than a doubling of output, and average cost increases; finally, if returns to scale are constant, then doubling all inputs results in an equiproportional doubling of output, and average cost is unaffected.

The relationship between returns to scale, output level, and average costs can be seen graphically by observing the long-run average cost curves in Fig. 3.1 For both average cost curves drawn in Fig. 3.1, at levels of output to the left of y_0, average cost is falling as output increases; therefore in this range of outputs, returns to scale are increasing. To the right of y_0, average cost is increasing with output; therefore in this range of outputs, returns to scale are decreasing. Finally, at the point y_0, average cost is at a minimum, and returns to scale are constant.

The above discussion on returns to scale focused on relationships among inputs and output, given that the state of technical knowledge A was unchanged. But how might improvements in the state of technical knowledge—increases in A—affect production and costs? It is convenient to think of such improvements as shifting the production function (or production possibility frontier) outward, since given the same combination of inputs, the maximum possible output increases with improvements in technical knowledge. This implies that the total and average costs of producing output decline with improvements in technical knowledge, that is, that the average cost curve shifts downward with increases in A.

One example of special interest here is the case of learning. To the extent that cumulative experience with the production of output results in improvements in technical knowledge, one might think of A as being affected by cumulative experience, or learning. In such cases the effects of learning are to change A in Eq. (3.3) from, say, A_0 to A_1 and thus to shift the cost curve in

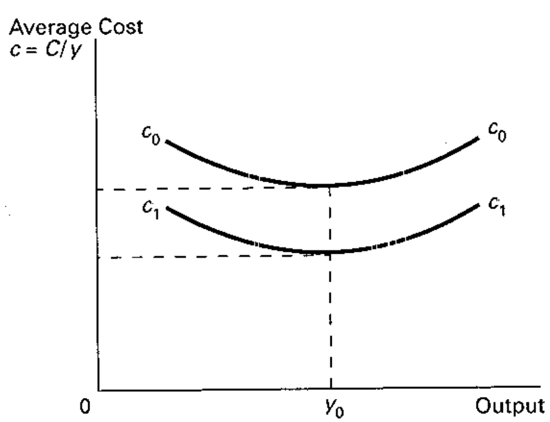

FIGURE 3.1

Fig. 3.1 downward from the curve $c_0 - c_0$ to the lower curve $c_1 - c_1$. Incidentally, it is worth noting that changes in returns to scale correspond with movements *along* the average cost curves in Fig. 3.1, whereas changes in the state of technical knowledge induce *shifts* in these curves.

In terms of the calculus, one can write the cost minimization problem facing the firm as a constrained optimization problem:

$$\min_{x_i} \mathcal{L} = \sum_{i=1}^{n} p_i x_i + \lambda[y - f(x_1, \ldots, x_n; A)] \tag{3.4}$$

where λ is the Lagrange multiplier.[10] The first-order conditions for minimization of costs are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= p_1 - \lambda f_1(x_1, x_2, \ldots x_n; A) = 0 \\ &\vdots \\ \frac{\partial \mathcal{L}}{\partial x_n} &= p_n - \lambda f_n(x_1, x_2, \ldots, x_n; A) = 0 \end{aligned} \tag{3.5}$$

and

$$\frac{\partial \mathcal{L}}{\partial \lambda} = y - f(x_1, x_2, \ldots, x_n; A) = 0 \tag{3.6}$$

where f_i is the first partial derivative (here, marginal product) of the production function (3.2) with respect to the ith input quantity, $i = 1, \ldots, n$.

Assuming that the second-order conditions for minimization are satisfied, we can solve for the cost function g in Eq. (3.3) dual to the production function f in Eq. (3.2) in a sequential manner, first by taking ratios of first-order conditions (3.5) to eliminate λ (reflecting cost minimization):

$$\begin{aligned} \frac{p_1}{p_n} &= \frac{f_1(x_1, x_2, \ldots, x_n; A)}{f_n(x_1, x_2, \ldots, x_n; A)} \\ &\vdots \\ \frac{p_{n-1}}{p_n} &= \frac{f_{n-1}(x_1, x_2, \ldots, x_n; A)}{f_n(x_1, x_2, \ldots, x_n; A)} \end{aligned}$$

then solving for each x_i by repeated substitution into Eq. (3.6)—call these $\hat{x}_i$—and finally by substituting these optimal $\hat{x}_i$ into the expression for total costs, $C \equiv \Sigma_i p_i \hat{x}_i$.

It might also be noted here that in many cases it is impossible to solve analytically for the cost function g dual to a particular production function f, particularly when the production function becomes mathematically complex. In the last two decades, however, there have been important advances in the economic theory of duality that now facilitate theoretical and empirical implementations of cost functions dual to rather general production functions. While such duality theory issues are very important for empirical research on costs and production (and are discussed in further detail in Chapter 9), they are beyond the scope of this chapter.[11]

3.2 BRIEF OVERVIEW OF LEARNING CURVE LITERATURE

The essential features underlying learning curves—curves relating unit real cost or real price to cumulative production—were briefly summarized in the introduction to this chapter. There it was emphasized that, given exposure to repetitive tasks, workers are likely to learn from cumulative experience how such tasks can be performed more quickly and efficiently. Not only are assembly line workers able to exploit their experiences, but so too are plant managers and other company officials. For example, plant engineers and managers can call upon their experience to improve the operations management of the plant, rearranging its layout, modifying job assignments, and reducing material wastes. Further, machines that are involved in the productive process are also affected by experience, not in the sense that they learn, but rather that they may undergo a number of technical improvements and in some cases can be replaced altogether by machines embodying newer technologies.[12]

The literature on the cost effects of learning has occasionally differentiated learning curve effects from experience curve effects: the former was confined to the learning and increased effectiveness of workers, while the latter incorporated the complete effects of experience, from workers' training to technical improvements to better management. Be they learning or experience curve effects, however, their impacts on costs are typically computed by developing a measure of cumulative previous production. For the moment we interpret cumulative previous production as a measure incorporating the effects of learning or experience.[13]

Consider, for example, the historical record of the Ford Motor Company in producing its Model T Ford over the 1909–1923 time period. Ford decided to make relatively standard models with few customized options.[14] Because of the experiences gained by management and labor, dramatic cost reductions occurred as cumulative production increased; these cost declines were then passed on to consumers in the form of price reductions.[15]

To see what happened at Ford, observe Fig. 3.2, which indicates the average list price in 1958 dollars of the Model T as cumulative production increased.[16] The vertical axis displays the natural logarithm of prices in thousands of 1958 dollars, while the horizontal axis displays the natural logarithm of cumulative production. A simple regression line with a negative slope of about −0.25 "fits" these data very nicely and suggests that price fell substantially with increases in cumulative production. As has been noted by a number of analysts, slopes of −0.20 to −0.30 on diagrams analogous to Fig. 3.2 are very common for a number of manufactured products.[17]

The learning curve has been formulated in a variety of ways. The simplest and most common form of the learning curve is as follows:[18]

$$c_t = c_1 n_t^{\alpha_c} e^{u_t} \qquad (3.7)$$

where

$c_t \equiv$ average or unit real costs of production in time period t (nominal average costs adjusted for inflation using a GNP-type deflator)

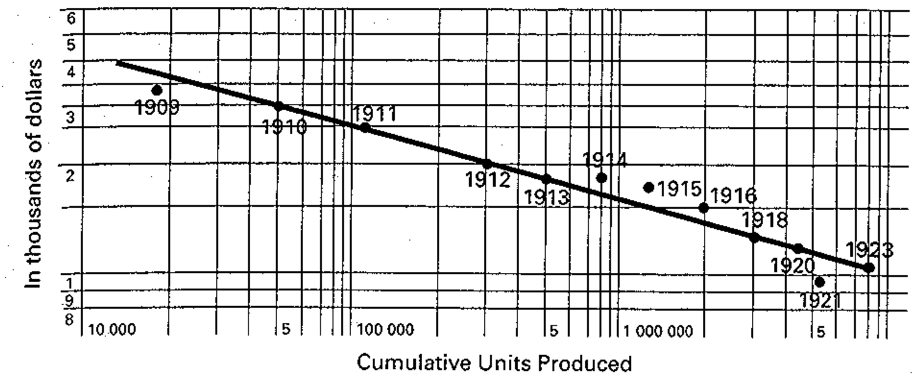

FIGURE 3.2
(*Source* W.J. Abernathy and K. Wayne, "Limits of the Learning Curve," *Harvard Business Review*, September/October, 1974.) Reprinted by permission of *Harvard Business Review*.

$c_1 \equiv$ real unit or average costs in the initial production time period
$n_t \equiv$ cumulative number of units of output produced up to (but not including) time period t
$\alpha_c \equiv$ elasticity of unit costs with respect to cumulative volume (typically negative)
$u_t \equiv$ a stochastic disturbance term reflecting the effects of inherent randomness in cost-production processes

Typically, it is assumed that the u_t disturbance term is independently and identically normally distributed with mean zero and constant covariance matrix (more on this later).

Equation (3.7) can be written in logarithmic form as

$$\ln c_t = \ln c_1 + \alpha_c \cdot \ln n_t + u_t \qquad (3.8)$$

The learning curve elasticity parameter α_c can then be estimated by least squares, provided that appropriate data on unit costs and production are available.[19]

The mathematics of logarithmic transformations implies that each time cumulative experience doubles, costs (prices) will decline to d% of its previous level, where

$$d = 2^{\alpha_c} \qquad (3.9)$$

Therefore if cost (price) declines to 80% of its previous level as cumulative production doubles, then the experience or learning curve is said to have an 80% slope. Simple calculations imply that the numerical relationship between d and α_c is approximately as follows:

α_c:	−0.50	−0.33	−0.25	−0.16
d:	0.71	0.80	0.84	0.89

An example is found in Fig. 3.2, where the slope of the simple regression line on Model T Fords is about −0.25; on the left-hand side of this figure it is noted that the experience curve has an 85% slope. From the above example an α_c estimate of about −0.25 implies a d estimate of about 0.84.

While other formulations of the learning curve relationship have been employed, it is the double logarithmic relationship (3.8) that has been most widely estimated in the empirical learning curve literature. In the next section of this chapter we examine what economic theory and what mathematical relationships could be called upon to justify employing an equation such as the popular learning curve (3.8).

3.3 DERIVATION OF COST FUNCTION BASED ON COBB-DOUGLAS PRODUCTION FUNCTION

In Section 3.1 we overviewed the economic theory underlying cost and production functions and focused on the notions of returns to scale and advances in technical knowledge. Then in the Section 3.2 we briefly outlined the learning curve literature. In this and the following section we synthesize these two literatures and move toward empirical implementation. First we employ economic theory and derive a Cobb-Douglas cost function equation, apparently related to the learning curve equation (3.8), whose parameters can be estimated by using linear multiple regression techniques. Then in Section 3.4 we explicitly integrate the learning curve equation (3.8) with the Cobb-Douglas cost function.

It is worth emphasizing that our focus here is the rather simple one of obtaining and interpreting an estimable learning curve equation that is consistent with the theory of cost and production. Students who are interested in pursuing more complex mathematical issues further may wish to consult a microeconomic theory textbook, such as those referenced in footnote 11.

To obtain a cost function whose parameters have clear interpretations, it is useful to specify a particular functional form for the production function f in Eq. (3.2). A very convenient, albeit somewhat restrictive, functional form is known as the Cobb-Douglas production function. Similar mathematical representations were used by the economist Knut Wicksell as early as 1896, but this form became particularly well known following an important published study by Charles Cobb and Paul Douglas [1928].[20]

In several of the empirical exercises in this chapter you will deal with the estimation of returns to scale in the generation of electricity based on the three inputs of capital, labor, and fuel. We therefore begin by writing a slightly more general version of the constant returns to scale, two-input functional form than that originally presented by Cobb and Douglas in their 1928 paper. In particular, we write the three-input Cobb-Douglas production function as

$$y = A \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot x_3^{\alpha_3} \tag{3.10}$$

where α_1, α_2, and α_3 are unknown parameters to be estimated. If each of the x's is multiplied by the proportionality factor μ, then substitution into Eq. (3.10) implies that the resulting output equals $y \cdot \mu^r$, where $r \equiv \alpha_1 + \alpha_2 + \alpha_3$. This means that returns to scale for the Cobb-Douglas function, hereafter denoted as r, equal the sum of the exponents in Eq. (3.10), that is,

$$\text{Returns to scale} \equiv r = \alpha_1 + \alpha_2 + \alpha_3 \tag{3.11}$$

Further, economies of scale are computed as $r - 1$. If, for example, r equalled 1.15, 1.00, or 0.85, then returns to scale would be increasing, constant, or decreasing, respectively, and economies of scale would be positive (0.15), zero, or negative (−0.15), respectively.

To derive the cost function dual to the Cobb-Douglas production function, substitute Eq. (3.10) for $f(x_1, x_2, \ldots, x_n)$ into Eq. (3.4) and rewrite the corresponding first-order conditions in Eqs. (3.5) and (3.6). After rearranging, the first-order conditions in Eqs. (3.5) turn out to be, for this Cobb-Douglas case,

$$x_i = \lambda \cdot y \cdot \alpha_i/p_i \qquad i = 1, 2, 3 \tag{3.12}$$

Now take ratios of these first-order conditions in Eq. (3.12) to eliminate λ and y; for example, take ratios x_2/x_1 and x_3/x_1. Given these ratios, solve for x_2 and x_3, substitute into Eq. (3.6), and rearrange to solve for, say, x_1 as a function of y, A, α_i, and p_i, $i = 1, 2, 3$. This yields the derived demand for x_1 as a function of the level of output y, the parameters of the Cobb-Douglas production function, and the prices of the three inputs. Follow analogous procedures to obtain derived demands for x_2 and x_3. Then, using these analytical expressions for x_1, x_2, and x_3, obtain the functional form for the cost function g dual to the Cobb-Douglas production function (3.10) as

$$C = p_1x_1 + p_2x_2 + p_3x_3 \tag{3.13}$$

which, after simple but tedious algebraic computations, turns out to be

$$C = k \cdot y^{1/r} \cdot p_1^{\alpha_1/r} \cdot p_2^{\alpha_2/r} \cdot p_3^{\alpha_3/r} \tag{3.14}$$

where

$$k = r \cdot [A \cdot \alpha_1^{\alpha_1} \cdot \alpha_2^{\alpha_2} \cdot \alpha_3^{\alpha_3}]^{-1/r} \tag{3.15}$$

and r is the returns-to-scale parameter defined in Eq. (3.11).

While the Cobb-Douglas cost function (3.14) and (3.15) appears formidably nonlinear, a linear specification that is amenable to conventional linear multiple regression techniques can be obtained by taking natural logarithms. This logarithmic transformation yields the much simpler equation

$$\ln C = \ln k + (1/r) \cdot \ln y + (\alpha_1/r) \cdot \ln p_1 + (\alpha_2/r) \cdot \ln p_2 + (\alpha_3/r) \cdot \ln p_3 \tag{3.16}$$

where ln denotes the natural logarithm. Notice in particular that if one estimated Eq. (3.16) using conventional linear multiple regression techniques (ignoring for the moment the absence of a stochastic disturbance term), the reciprocal of the estimated coefficient on the ln y variable would provide a direct estimate of returns to scale. However, and this will turn out to be very important, since r is a constant parameter, returns to scale with this Cobb-Douglas cost function cannot vary with the level of output; this implies that the average cost curve in Fig 3 1 corresponding to the Cobb-Douglas cost function would always slope downward if $r > 1$, would be a horizontal straight line if $r = 1$, and would always slope upward if $r < 1$

One other additional constraint can be imposed before estimating Eq. (3.16). Specifically, according to economic theory, in order that cost functions are well behaved, it is necessary that they be homogeneous of degree 1 in input prices, no matter the extent of scale economies.

The underlying intuition for homogeneity of degree 1 in input prices is as follows: For a given level of output, if all input prices double, one would expect that total costs should also double. (Note that this constraint involving prices and costs, holding output fixed, is different from the returns-to-scale relationship whereby $r \equiv \alpha_1 + \alpha_2 + \alpha_3$ See any microeconomic theory text for further details.) This homogeneity of degree 1 in input prices (also called linear homogeneity) implies the constraint that the coefficients on the input price variables in Eq. (3.16) sum to unity, that is, that

$$\alpha_1/r + \alpha_2/r + \alpha_3/r = (\alpha_1 + \alpha_2 + \alpha_3)/r = 1 \quad (3.17)$$

This restriction can be imposed by solving Eq (3.17) for, say, α_3/r:

$$\alpha_3/r = 1 - \alpha_1/r - \alpha_2/r \quad (3.18)$$

substituting Eq (3.18) into Eq. (3.16), collecting terms, and then rearranging This results in

$$\ln C - \ln p_3 = \ln k + (1/r) \cdot \ln y + (\alpha_1/r) \cdot (\ln p_1 - \ln p_3) + (\alpha_2/r) \cdot (\ln p_2 - \ln p_3) \quad (3.19)$$

A convenient way of rewriting Eq (3.19) is as follows:

$$\ln C^* = \beta_0 + \beta_y \cdot \ln y + \beta_1 \cdot \ln p_1^* + \beta_2 \cdot \ln p_2^* \quad (3.20)$$

where

$$\begin{aligned} \ln C^* &\equiv \ln C - \ln p_3 \\ \ln p_1^* &\equiv \ln p_1 - \ln p_3 \\ \ln p_2^* &\equiv \ln p_2 - \ln p_3 \\ \beta_0 &\equiv \ln k \\ \beta_y &\equiv (1/r) \\ \beta_1 &\equiv \alpha_1/r \\ \beta_2 &\equiv \alpha_2/r \end{aligned} \quad (3.21)$$

Equation (3.20) highlights the linearity and empirical simplicity of the Cobb-Douglas cost function equation

An important issue now emerges: If one estimated the parameters in Eq. (3.20) using linear multiple regression techniques, how could one employ them to recover estimates of the underlying Cobb-Douglas production function parameters and returns to scale? As was noted earlier and implied in the fifth line of Eqs. (3.21), $r \equiv 1/\beta_y$; similarly, by using the last two lines of Eqs (3.21),

$$\alpha_1 = \beta_1 \cdot r = \beta_1/\beta_y \quad \text{and} \quad \alpha_2 = \beta_2 \cdot r = \beta_2/\beta_y \quad (3.22)$$

Finally, employing Eq (3.18) and the bottom two lines of Eqs. (3.21) the implicit parameter α_3 can be recovered from the parameters directly estimated in Eq. (3.20) as

$$\alpha_3 = (1 - \beta_1 - \beta_2)/\beta_y \quad (3.23)$$

Hence by estimating parameters of the dual cost function equation (3.20), one can recover estimates of all other Cobb-Douglas production function parameters

In the exercises at the end of this chapter, you will become involved with estimation of parameters in Eq. (3.20) and will obtain estimates of returns to scale in the generation of electricity for various electric utility companies in the United States

3.4 INTEGRATING THE LEARNING CURVE WITH THE COBB-DOUGLAS COST FUNCTION

We have now discussed two apparently rather different strands of literature, one dealing with learning and experience curves and the other with Cobb-Douglas production and dual cost functions In this section we integrate these discussions.

It is useful to recall that the most common equation used to estimate parameters of the learning curve is that of Eq. (3.8), namely,

$$\ln c_t = \ln c_1 + \alpha_c \cdot \ln n_t + u_t \quad (3.8)$$

while the Cobb-Douglas cost function equation (with a stochastic disturbance term now included) was derived in Eq. (3.16) as

$$\ln C = \ln k + (1/r) \cdot \ln y + (\alpha_1/r) \cdot \ln p_1 + (\alpha_2/r) \cdot \ln p_2 + (\alpha_3/r) \cdot \ln p_3 + u_t \quad (3.16)$$

To integrate the learning curve and Cobb-Douglas discussions, in essence we ask what restrictions or changes must be made to Eq. (3.16) so that it reduces to the simple learning curve equation (3.8).

The first important difference to note between Eqs. (3.16) and (3.8) is that the variable reflecting the effects of learning and experience, measured in Eq. (3.8) by cumulative previous production and denoted n_t, is entirely absent from the Cobb-Douglas cost function equation (3.16) However, recall that the constant term in Eq. (3.16) is simply ln k, where k was defined in Eq. (3.15) to depend on A and where in turn A reflected advances in the state of knowledge. Surely advances in knowledge are closely related to learning curve effects, and therefore it should be relatively straightforward to integrate A with n_t.

In this spirit, define the state of knowledge in time period t as cumulative production up to time period t, all raised to the power α_c, where α_c is the experience curve elasticity parameter to be estimated, that is,

$$A_t \equiv n_t^{-\alpha_c} \qquad (3.24)$$

Substituting Eq. (3.24) into Eq. (3.14), adding time subscripts to the cost, output, and price variables, and then taking logarithms yield a modified version of the Cobb-Douglas cost function equation (3.16) with experience effects included:

$$\ln C_t = \ln k' + (\alpha_c/r) \cdot \ln n_t + (1/r) \cdot \ln y_t + (\alpha_1/r) \cdot \ln p_{1t} + (\alpha_2/r) \cdot \ln p_{2t} + (\alpha_3/r) \cdot \ln p_{3t} \qquad (3.25)$$

where k' is the same as k in Eq. (3.15) except that the effects of A are removed, that is,

$$k' = r \cdot [\alpha_1^{\alpha_1} \cdot \alpha_2^{\alpha_2} \cdot \alpha_3^{\alpha_3}]^{-1/r} \qquad (3.26)$$

While the Cobb-Douglas cost function equation (3.25) now has learning or experience curve effects included, it also contains a number of variables that are not found in the learning curve equation (3.8), such as the three input prices and the output variables.

With respect to the three input prices, one could simply make the grandiose assumption that over whatever time period the data spanned, the relative input prices were constant. In such a case, prices could simply be ignored, and their effect would be included in a composite intercept term. However, while such an assumption is convenient, it is heroic indeed.

One could instead make a different assumption that effectively removes the price variables as regressors from the Cobb-Douglas cost function equation. In particular, one could assume that the effects of the three input prices in Eq. (3.25) could be captured by using an appropriate deflator from the national income and product accounts, such as the deflator for gross national product (GNP). Denote such a GNP deflator as GNPD, set GNPD equal to the Cobb-Douglas function of prices in Eq. (3.25),

$$\ln \text{GNPD}_t \equiv (\alpha_1/r) \cdot \ln p_1 + (\alpha_2/r) \cdot \ln p_2 + (\alpha_3/r) \cdot \ln p_3 \qquad (3.27)$$

and then define total costs in constant dollars (denoted C_t') as total costs in current dollars (C_t) divided by GNPD$_t$, that is,

$$C_t' \equiv C_t/\text{GNPD}_t \quad \text{or} \quad \ln C_t' = \ln C_t - \ln \text{GNPD}_t \qquad (3.28)$$

If we solve Eq. (3.28) for ln C_t and then substitute Eq. (3.27) into Eq. (3.25), the Cobb-Douglas cost function equation becomes

$$\begin{aligned}\ln C_t &= \ln C_t' + \ln \text{GNPD}_t \\ &= \ln C_t' + (\alpha_1/r) \cdot \ln p_{1t} + (\alpha_2/r) \cdot \ln p_{2t} + (\alpha_3/r) \cdot \ln p_{3t} \\ &= \ln k' + (\alpha_c/r) \cdot \ln n_t + (1/r) \cdot \ln y_t + (\alpha_1/r) \cdot \ln p_{1t} \\ &\quad + (\alpha_2/r) \cdot \ln p_{2t} + (\alpha_3/r) \cdot \ln p_{3t} + u_t\end{aligned} \qquad (3.29)$$

Notice that since the terms involving price variables appear on both sides of the second equal sign in Eq. (3.29), the price terms can be subtracted from both sides, thereby generating a modified Cobb-Douglas cost equation that looks very much like the learning curve equation (3.8), specifically,

$$\ln C_t' = \ln k' + (\alpha_c/r) \cdot \ln n_t + (1/r) \cdot \ln y_t + u_t \qquad (3.30)$$

It is important to note that this similarity in the Cobb-Douglas and learning curve equations depends critically on the assumption that the effects of input prices on costs of production can be measured by using the GNP deflator, that is, that Eq. (3.27) is valid. Of course, that involves a rather strong assumption and is unlikely to be appropriate whenever the shares of the inputs in total costs of production for a particular company or firm differ significantly from the weights used by national income accountants in constructing the national GNP deflator. This aspect of the learning curve equation should therefore be viewed with considerable skepticism. With this strong caveat firmly in mind, however, let us now proceed with the final integration of the learning curve and the Cobb-Douglas cost function equations.

The major remaining differences between Eqs. (3.8) and (3.30) appear to be the fact that the dependent variable in the learning curve equation (3.8) is unit or average real cost, while in the modified Cobb-Douglas equation (3.30) the dependent variable is total real cost; further, an output variable appears as a right-hand side variable in the Cobb-Douglas equation (reflecting the effects of returns to scale), while such an output variable is omitted in the learning curve equation.

Since total and average costs are related by the identity $c_t \equiv C_t'/y_t$ (average real costs equal total real costs divided by the level of output in time period t), it follows that ln c_t = ln C_t' − ln y_t. Subtracting ln y_t from both sides of Eq. (3.30) gives

$$\begin{aligned}\ln C_t' - \ln y_t &= \ln c_t \\ &= \ln k' + (\alpha_c/r) \cdot \ln n_t + ((1-r)/r) \cdot \ln y_t + u_t\end{aligned} \qquad (3.31)$$

Notice that the dependent variables in Eqs. (3.8) and (3.31) are now identical and that the only remaining difference in these two equations is that ln y_t appears as a right-hand variable in the Cobb-Douglas cost function equation (3.31) but is not present in the learning curve equation (3.8).

How should one interpret this current output variable in the modified Cobb-Douglas cost equation (3.31)? If returns to scale were increasing (if $r > 1$), then the value of the parameter on the ln y_t variable (equal to $(1 - r)/r$) should be negative, indicating that, given the effects of experience, unit costs would fall with increases in current output. Similarly, if returns to scale were decreasing ($r < 1$), then the value of this parameter would be positive, implying that, given the effects of experience, unit costs would rise with increases in current output. Finally, if returns to scale were constant ($r = 1$), then the value of the $(1 - r)/r$ parameter would be zero, and the $\ln y_t$ term would drop out of the estimating equation (3.31).

This suggests the following. To obtain an estimating equation based on the Cobb-Douglas cost function but identical to that commonly used in the experience curve literature, one must also make an assumption concerning returns to scale. Specifically, if one makes the additional assumption that returns to scale are constant, that is, that $r = 1$, then $\ln y_t$ disappears completely from Eq. (3.31), leaving

$$\ln c_t = \ln k' + \alpha_c \cdot \ln n_t + u_t \tag{3.32}$$

an equation that is identical to the learning curve equation (3.8).[21]

The above discussion can be summarized as follows. To synthesize the estimating equations based on Cobb-Douglas cost functions with those on the effects of learning curves, one must make two assumptions: First, it must be assumed that the effects of input price changes on costs of production can be accurately measured by using a GNP deflator; this assumption is stated explicitly in Eq. (3.27). Second, one must also assume that returns to scale are constant, that is, that $r = 1$.

While each of these assumptions is subject to serious question, the second assumption involving returns to scale is of particular interest to us here. In particular, note that if returns to scale are not constant, then instead of estimating Eq. (3.32), one should estimate Eq. (3.31):

$$\ln c_t = \ln k' + (\alpha_c/r) \cdot \ln n_t + (1 - r)/r \cdot \ln y_t + u_t$$

which can be conveniently rewritten as

$$\ln c_t = \beta_0 + \beta_1 \cdot \ln n_t + \beta_2 \cdot \ln y_t + u_t \tag{3.33}$$

where

$$\beta_0 \equiv \ln k', \qquad \beta_1 \equiv \alpha_c/r, \qquad \text{and} \qquad \beta_2 \equiv (1 - r)/r \tag{3.34}$$

Because the learning curve equation (3.32) is based on the assumption of constant returns-to scale, in empirical work it may be of interest to test whether such a restriction is in fact supported by the data. Such a test is relatively straightforward, since by Eq. (3.34), $\beta_2 \equiv (1 - r)/r$. Specifically, one could simply estimate Eq. (3.33) by least squares regression methods and then test the null hypothesis of constant returns to scale ($r = 1$) against the alternative hypothesis of nonconstant returns to scale ($r \neq 1$) by employing a t-test and assessing whether, at a predetermined reasonable level of significance, the estimate of β_2 is statistically different from zero. If this hypothesis is rejected, then a key assumption underlying the most commonly used learning curve equation is invalid, and an equation such as Eq. (3.33) should be employed rather than Eq. (3.32). On the other hand, if the null hypothesis that $\beta_2 = 0$ is not rejected, then the constant returns to scale assumption underlying the learning curve equation is empirically validated.

Note also that if one estimates Eq. (3.33) by linear multiple regression techniques, one can employ Eq. (3.34) to recover indirectly an estimate of the learning curve elasticity parameter α_c and the returns-to-scale parameter r based on the direct estimates of β_1 and β_2. Specifically,

$$r = 1/(1 + \beta_2) \qquad \text{and} \qquad \alpha_c = \beta_1 \cdot r = \beta_1/(1 + \beta_2) \tag{3.35}$$

Unfortunately, since the estimates of the indirectly estimated parameters r and α_c involve nonlinear transformations of the directly estimated β_1 and β_2 parameters, one cannot in general directly employ the estimated standard errors of β_1 and β_2 to compute confidence intervals for r and α_c.[22]

3.5 ECONOMETRIC ISSUES

Having derived and synthesized estimating equations based on the learning curve and the Cobb-Douglas function, we now turn attention to important econometric issues, beginning with measurement issues and an examination of the bias effects of incorrectly omitting an explanatory variable.

3.5.1 Measurement Issues

A number of important measurement issues arise in implementing empirically the learning curve model. Here we briefly mention two important issues.[23] First, according to the learning curve framework, deflated unit costs should be the dependent variable. Since unit costs are typically calculated in current year dollars, to compute real (inflation adjusted) costs, one must deflate the cost data using some type of deflator. In many studies a GNP or CPI deflator is employed. Montgomery and Day [1985, p. 215] argue that an overall deflator such as the GNP deflator is preferable to an industry-specific output deflator, since if the latter is used, the learning curve effect may be "washed out" owing to the price deflator already capturing a substantial portion of the learning-induced productivity gains in that particular industry. Note also that the unit cost series should include all costs—labor, capital, energy, materials, and so on. Some learning curve studies focus only on labor costs and therefore are not comparable with studies using more comprehensive input cost measures unless labor and other input costs change at the same rate over time.

Second, since cost data are often difficult to obtain and, even if available, may be proprietary, a substantial number of empirical learning curve studies have substituted the deflated unit cost variable with a real price variable and then regressed the logarithm of real price on a constant term and the logarithm of cumulative production. This procedure introduces a major complication and makes separate identification of the learning curve effect very difficult if not impossible.

To see this, consider two firms with identical learning curve experiences. Assume one firm adopted a penetration pricing strategy in which the initial price was very low, thereby increasing demand, market share, and production, while the other firm adopted a "cream-skimming" pricing strategy in which the initial price was very high and was lowered only gradually. If one regressed price on cumulative production for these two firms, one would obtain very different estimates of the learning curve elasticity, even though the potential learning curve effects were identical. Had unit cost data been employed instead, however, this discrepancy would not have occurred. This demonstrates that it is preferable to employ unit cost rather than price data to estimate learning curve relationships, since unlike the case with unit cost, use of price data confounds the effects of pricing strategy with learning curves.

3.5.2 Omitted Variable Bias

Recall from Eq. (3.33) that in the Cobb-Douglas cost function equation with nonconstant returns to scale, the equation to be estimated has as right-hand variables a constant term, the logarithm of cumulative previous production $\ln n_t$, the logarithm of current output $\ln y_t$, and a disturbance term u_t:[24]

$$\ln c_t = \beta_0 + \beta_1 \cdot \ln n_t + \beta_2 \cdot \ln y_t + u_t \tag{3.33}$$

On the other hand, the equation most commonly estimated in the learning curve literature is Eq. (3.8), which can be rewritten as

$$\ln c_t = \alpha_0 + \alpha_1 \cdot \ln n_t + \omega_t \tag{3.36}$$

where $\alpha_0 \equiv \ln c_1$, $\alpha_1 \equiv \alpha_c$, and ω_t is a random disturbance term.

Suppose that one incorrectly estimated Eq. (3.36) when the parameters of Eq. (3.33) should have been estimated instead. Will the least squares estimate of α_1 be larger, equal to, or smaller than the least squares estimate of β_1? What bias results from incorrectly omitting $\ln y_t$ from the learning curve equation (3.36), and on what will the magnitude of the bias depend? These issues all concern the *omitted variable bias* and are very important. In our context, omitted variable bias issues involve (1) examining whether the learning curve elasticity is underestimated or overestimated when returns to scale are nonconstant, that is, when $\ln y_t$ is incorrectly omitted from Eq. (3.36), and (2) determining on what the magnitude of the underestimation or overestimation depends.

To analyze these omitted variable bias issues, it will be convenient to introduce a new equation, often called an auxiliary regression equation, in which the omitted right-hand variable is related to the included right-hand variable and a stochastic disturbance term ϵ_t,

$$\ln y_t = \delta_0 + \delta_1 \cdot \ln n_t + \epsilon_t \tag{3.37}$$

Now label the least-squares estimates of the α's, β's, and δ's in Eqs. (3.33), (3.36), and (3.37) as a's, b's, and d's, respectively. The omitted variable bias questions can then be stated simply as follows: What is the relationship between a_1 (the least squares estimate of α_1 in Eq. (3.36)) and b_1 (the least squares estimate of β_1 in Eq. (3.33))? On what does the difference between a_1 and b_1 depend?

Since the mathematics underlying this question are presented in a number of econometric theory textbooks,[25] here we simply note that the relationship between a_1 and b_1 can be shown to be as follows:

$$a_1 = b_1 + d_1 b_2 \quad \text{or} \quad a_1 - b_1 = d_1 b_2 \tag{3.38}$$

The *bias* from omitting $\ln y_t$ from Eq. (3.36) is simply equal to $a_1 - b_1$, which from Eq. (3.38) can be seen to equal $d_1 b_2$. This omitted variable bias will therefore be zero only if at least one of the following two conditions is satisfied:

1. $d_1 = 0$, that is, the logarithms of current output and cumulative previous production in Eq. (3.37) are uncorrelated;
2. $b_2 = 0$, that is, current average cost in Eq. (3.33) does not depend on current production. Alternatively, by Eq. (3.35), returns to scale are constant.

If neither of these two conditions is satisfied, then an omitted variable bias will result, implying that if one estimated the learning curve elasticity parameter using Eq. (3.36) and incorrectly omitting $\ln y_t$, one would obtain a biased estimate of the true learning curve parameter. But will the bias be positive or negative, that is, will the learning curve elasticity be underestimated or overestimated?

To calculate the sign of the bias, recall from Eq. (3.38) that

$$a_1 - b_1 = d_1 \cdot b_2$$

Therefore the sign of the bias depends on the sign of the product of d_1 and b_2. Since cumulative production is the sum over time of current production, n_t and y_t are positively correlated, and d_1 will be positive. On the other hand, whether b_2 is positive, zero, or negative depends on whether returns to scale are decreasing, constant, or increasing, that is, whether $r < 1$, $r = 1$, or $r > 1$.[26] With d_1 positive this therefore implies that

- $a_1 - b_1 > 0$ if $r < 1$ (there are decreasing returns to scale),
- $a_1 - b_1 = 0$ if $r = 1$ (there are constant returns to scale), and
- $a_1 - b_1 < 0$ if $r > 1$ (there are increasing returns to scale).

As was noted in the introduction, in many cases, one might expect returns to scale to be increasing, so it is plausible to expect that in many cases the last of the above three cases will occur. Further, since a_1 and b_1 are typically negative, $a_1 - b_1 < 0$ corresponds with b_1 being smaller in absolute value than a_1. In such a case, estimation of the simple learning curve equation (3.36) yields a larger estimate of the learning curve elasticity (in absolute value) than if one includes the current output variable as in Eq. (3.33); hence in this case the learning curve elasticity is overestimated in absolute value.

The interpretation of this result is as follows: By incorrectly omitting the current output variable, one attributes to the learning curve elasticity what in fact is due in part to the effects of returns to scale. In the learning curve context, omitted variable bias could be very important.

Fortunately, in the exercises at the end of this chapter you will have an opportunity to verify these numerical relationships among least squares estimates and to assess the magnitude of the omitted variable bias.

3.5.3 Individual and Joint Hypothesis Tests

A second set of econometric issues that merits attention here involves hypothesis testing. The computer output of typical regression program software includes t-statistics on each of the estimated coefficients, as well as an equation F-statistic. We now examine hypothesis testing in the context of the integrated Cobb-Douglas production–learning curve model, where it is assumed that the disturbance term appended to Eq. (3.33) is independently and identically normally distributed with mean zero and constant variance.

As noted earlier, to test the null hypothesis of constant returns to scale against the alternative hypothesis of nonconstant returns, all one need do is estimate Eq. (3.33) by least squares and then perform a t-test on β_2; the null hypothesis is $\beta_2 = 0$, and the alternative hypothesis is $\beta_2 \neq 0$.

Similarly, one could test the null hypothesis that the learning curve effect is zero. Having estimated Eq. (3.33) by least squares, this hypothesis test could also be performed by using a t-test; the null hypothesis is $\beta_1 = 0$, while the alternative hypothesis is $\beta_1 \neq 0$.

Suppose, however, that one wanted to test the above two hypotheses simultaneously. In this case the joint null hypothesis that both the learning curve elasticity is zero and returns to scale are constant would correspond with the joint null hypothesis that $\beta_1 = \beta_2 = 0$; the alternative hypothesis would of course be that $\beta_1 \neq 0, \beta_2 \neq 0$. Note that under the joint null hypothesis, Eq. (3.33) reduces to an equation in which real unit cost is simply regressed on a constant term; no other regressors remain.

To perform such a joint hypothesis test, however, one should employ an F-statistic rather than individual t-statistics. The reason is that the individual t-tests are not, in general, independent; by contrast, the calculation underlying the F-statistic properly accounts for the dependence among the individual hypothesis tests. In this particular case the dependence between the individual t-tests depends primarily on the covariance between the least squares estimates b_1 and b_2.

Because the F-test statistic properly accounts for dependence between the two individual t-tests, an inference based on the F-statistic might not necessarily agree with that based on the two t-tests. Specifically, any of the following six cases is possible:

1. reject the joint null on the basis of the F-statistic, but do not reject each separate null on the basis of the individual t-tests;
2. reject the joint null on the basis of the F-statistic, reject one individual hypothesis on the basis of a t-test, and do not reject other individual hypothesis on the basis of a t-test;
3. reject the joint null on the basis of the F-statistic, and reject each separate null on the basis of individual t-tests;
4. do not reject the joint null on the basis of the F-statistic, and do not reject each separate null on the basis of individual t-tests;
5. do not reject the joint null on the basis of the F-statistic, reject one individual hypothesis on the basis of a t-test, and do not reject other individual hypothesis on the basis of a t-test;
6. do not reject the joint null on the basis of the F-statistic, but reject each separate null on the basis of individual t-tests.

Although they are possible, in practice, cases 1 and 6 occur only very rarely. As we shall see with the learning curve exercises in which, on the basis of individual t-tests, the estimated learning curve elasticity is statistically significant but returns to scale are only marginally different from unity, cases 2 and 3 are relatively common, particularly with time series data.

3.5.4 Brief Overview of Empirical Findings on Returns to Scale and Learning Curves

We now briefly summarize the principal empirical findings on returns to scale and on learning curve elasticities. We begin with learning curves.

Over the years, hundreds of studies, many of them involving the Boston Consulting Group, have focused on estimating parameters of the learning or experience curves. Obviously, such a large number of studies cannot be summarized here. Rather, what we do is to comment on a recent survey by Pankaj Ghemawat [1985], published in the *Harvard Business Review*, in which 97 academic studies from the learning and experience curve literature are reviewed. The results of this summary are presented in Table 3.1.

In Ghemawat's survey, the 97 studies are classified according to the size of the learning or experience curve slope. Recall from our earlier discussion that the relationship between the slope of the learning curve d and the size of the learning curve elasticity α_c is, from Eq. (3.9),

$$d = 2^{\alpha_c}$$

Hence for each experience curve slope range presented in Ghemawat's study, we first employ the above relationship and calculate the corresponding range for the learning curve elasticity.

Table 3.1 Summary of Variation in Learning Curve Elasticity Estimates by Product

Learning Curve Elasticity (in Absolute Value)	Learning Curve Slope	Number of Products
0.63–0.74	0.60–0.64	3
0.52–0.62	0.65–0.69	3
0.42–0.51	0.70–0.74	10
0.33–0.41	0.75–0.79	23
0.25–0.32	0.80–0.84	30
0.16–0.24	0.85–0.89	26
0.08–0.15	0.90–0.94	6
0.01–0.07	0.95–0.99	1

Average learning curve slope: 0.85
Total number of products surveyed: 97

Source: Data from Pankaj Ghemawat "Building Strategy on the Experience Curve" *Harvard Business Review* March/April 1985 Exhibit II p. 146

As can be seen in Table 3.1, the largest number of products have learning curve elasticities in the range −0.25 to −0.32, with corresponding learning curve slopes of 0.80 to 0.84; this range covers approximately 30% of the products and studies examined. The vast majority of products (79 of the 97 examined) fall in the learning curve elasticity range of about −0.16 to −0.41, which corresponds with learning curve slopes ranging from 0.75 to 0.89. Hence while there is considerable variation among products in the size of the learning curve elasticities and slopes, it is quite unusual to obtain learning curve elasticity estimates smaller than 0.15 or larger than 0.41 (in absolute value).

Ghemawat notes in his survey that, in general, manufacturing activities are associated with larger learning curve elasticities than those from the raw materials purchasing, marketing, sales, or distribution companies. Moreover, Ghemawat finds a pattern to the learning curve elasticities, stating (p. 144) that "manufacturing costs decline particularly steeply in industries with standardized product ranges and complex, labor-intensive production processes such as the airframe assembly or machine tool businesses."

It might also be noted here that, like other experience and learning curve analysts, Ghemawat is rather cautious about interpreting these experience curves; in particular, he does not rule out the possibility that their magnitude may also reflect the presence of nonconstant returns to scale.[27] Recall that the advantage of the regression equation such as Eq. (3.33) is precisely that it permits separate estimation and identification of learning curve and returns-to-scale effects.

We now move on to a brief discussion of empirical findings on returns to scale. The first important point is that measures of returns to scale can be expected to vary as one moves from the plant level to the company and then

to the industry as a whole. In particular, even if there are increasing returns to scale at the plant level, companies can build additional plants such that doubling the number of plants doubles the output at any given level of cost. This may therefore show up at the company level as constant returns to scale. Similar arguments suggest that returns to scale may be constant at the industry level but could be increasing at the company level.

Since the number of industries is very large and the number of firms or companies is even larger, there is no compact way in which the voluminous empirical literature on returns to scale can be summarized. Given this, we focus here on but one particular industry in which the presence of alleged increasing returns to scale has brought about considerable government regulation, namely, the electric utility industry.

A number of studies have attempted to measure returns to scale in the generation of electricity over the last three decades; these studies have been summarized and critiqued by Thomas G. Cowing and V. Kerry Smith [1978]. The classic econometric study of returns to scale in the electric utility industry is by Marc Nerlove [1963] and is based on the Cobb-Douglas cost function examined earlier in this chapter. An attractive feature of Nerlove's study is that it includes a data appendix with all data listed. In the hands-on exercises at the end of this chapter you will have the opportunity to work with Nerlove's original data, as well as data of more recent vintage, to estimate returns to scale and attempt to replicate classic findings.

Cowing and Smith summarize a number of econometric studies, including those based on production functions, input demand models, cost functions, and profit functions. Evidence of increasing returns to scale is widespread, particularly in the era preceding the 1970s in the United States. Nerlove, for example, used 1955 data and obtained returns to scale estimates of 0.97 to a rather large 1.91; not surprisingly, he also found that returns to scale tended to fall as the size of the electric utility company increased. For the very largest utilities, returns to scale were approximately constant. Returns to scale estimates ranging from 1.1 to 1.2 are common and typical in the econometric studies surveyed by Cowing and Smith.

A particularly interesting study is that by Laurits R. Christensen and William H. Greene [1976], who used Nerlove's original 1955 data base and then updated it to 1970. Further, Christensen and Greene employed a translog cost function, a mathematical formulation that is considerably more general than the Cobb-Douglas form considered in this chapter. One attractive feature of the translog form is that it allows returns to scale to vary, depending on input prices and on the level of output.[28] Christensen and Greene [1976, p. 655] summarized their principal findings based on the translog function as follows:

> We find that in 1955 there were significant scale economies available to nearly all firms. By 1970, however, the bulk of U.S. electricity generation was by firms operating in the essentially flat area of the average cost curve. We conclude that a small number of extremely large firms are not required

MARC NERLOVE

Estimating Returns to Scale

Marc Nerlove's classic article on estimating returns to scale in the supply of electricity was not written in the traditional research ambience of academia or the private sector. Rather, it was begun while Nerlove was serving in the U.S. Army from 1957 to 1959. Nerlove undertook research on economic factors affecting the supply of electricity, since the adequacy of national energy supplies was clearly an issue of strategic importance. Although most of the research work was done while he was in the Army, Nerlove completed his article at the University of Minnesota in 1960. A major innovation in Nerlove's model of electricity supply was that it was firmly based on the theory of production; in fact, it is the first empirical application using the dual cost function approach. As is discussed in this chapter, a principal econometric finding in Nerlove's study was that while scale economies were substantial, they tended to decline with firm size.

Marc Nerlove was born in Chicago in 1933. At the age of 15 he enrolled as an undergraduate at the University of Chicago, majoring in mathematics. Graduating with honors at age 18, he became interested in economics and decided to pursue graduate studies at the Johns Hopkins University. His 1954 M.A. thesis focused on the demand for meat in the United States, and his 1956 Ph.D. dissertation dealt with estimating supply elasticities for corn, cotton, and wheat. With his Ph.D. degree in hand, Nerlove took a position in 1956 as an analytical statistician for the Agricultural Marketing Service of the U.S. Department of Agriculture. In 1959, after spending two years in the U.S. Army, he initiated an academic career at the University of Minnesota. Since then, Nerlove has held academic appointments at Stanford, Yale, Chicago, Northwestern, and, since 1982, the University of Pennsylvania, where he is now University Professor of Economics and Research Fellow at the International Food Policy Research Institute.

Nerlove's research has ranged widely, from the analysis of nonstatic expectations, distributed lags, the pooling of cross-sectional and time series data, time series approaches, and univariate and multivariate logit models to applications in agriculture, macroeconometric modeling, population economics, advertising, income distribution, and inventory behavior. He has been elected a Fellow of the Econometric Society, served as its President in 1981, is a Fellow of the American Statistical Association and the American Academy of Arts and Sciences, and is an elected Member of the National Academy of Sciences. Nerlove's thesis and articles have won publication awards from the American Farm Economics Association, and his 1958 Agricultural Handbook No. 141 from the U.S. Department of Agriculture, *Distributed Lags and Demand Analysis*, has been cited for an enduring publication commendation.

for efficient production and that policies designed to promote competition in electric power generation cannot be faulted in terms of sacrificing economies of scale.

In summary, the econometric literature on estimated returns to scale in the electric utility industry in the United States appears to suggest that substantial economies of scale have been available, that such scale economies may have been largely exploited by the early 1970s, and that today the bulk of electricity generation comes from firms generating electricity at the bottom of their average cost curves.

This completes our discussion of costs, scale economies, and learning curves. It is now time to move on and engage you in hands-on applications.

3.6 HANDS ON WITH COSTS, LEARNING CURVES, AND SCALE ECONOMIES

The purpose of the following exercises is to enable you to gain an empirical understanding of cost curves, scale economies, and learning curves, using a variety of simple and multiple regression techniques. Exercises 1, 2, and 3 are particularly useful. In Exercise 1 you examine typical learning curve data on costs and cumulative output and use the simple regression procedure to estimate a learning curve elasticity. Then in Exercise 2 you assess the possible omitted variable bias on learning curve elasticity estimates resulting from incorrectly assuming that returns to scale are constant. In this exercise you employ both simple and multiple regression procedures, and you perform a number of single and joint hypothesis tests. To help you better understand relationships among simple and multiple regressions, in Exercise 3 you compare the R^2 goodness-of-fit measure from various regressions and find that in one very strange data set (appropriately called WEIRD) you obtain a striking and memorable result.

In Exercise 4 you examine and attempt to replicate the returns-to-scale estimates in the classic study of the U.S. electric utility industry by Marc Nerlove. You also construct confidence intervals and diagnose the least squares residuals. Then in Exercise 5 you employ information from the residual diagnostics and, following Nerlove, estimate returns to scale using a number of alternative equation specifications. You also test for the empirical validity of the constant returns to scale assumption. In Exercise 6 you compare returns-to-scale estimates from Nerlove's 1955 data with the Christensen-Greene data for 1970, and you evaluate the Christensen-Greene finding that by 1970 the bulk of U.S. electricity generation was being produced by firms operating on relatively flat portions of their average cost curves.

In the next two exercises you focus on stochastic specification and information gained from examining least squares residuals. Specifically, in Exercise 7 you examine the effects of first-order autocorrelation on estimates of the learning curve elasticity and its statistical significance. In Exercise 8 you employ the Durbin-Watson test statistic (usually applied in a time series context) with Nerlove's cross-sectional data to help you in uncovering model misspecification.

The issue of whether returns to scale are identical for firms of varying size can also be examined by estimating subsets of the entire data set separately and then testing whether parameters are equal across subsets. In Exercise 9 you perform such a procedure, called the Chow test, using both the Nerlove 1955 data and the Christensen-Greene 1970 data sets for electric utilities.

Finally, in Exercise 10 you employ an estimated learning curve model to forecast unit costs in the future, after further learning will have occurred. You will also estimate the forecast error variance and will uncover some disconcerting implications of using small data sets to make such forecasts.

••••••••••

In the data diskette provided, you will find a subdirectory called CHAP3.DAT, having five data files. In these data files a number of data series are provided, including data from the Nerlove returns-to-scale study in the electricity industry on 145 electric utility companies in 1955 (named NERLOV), as well as updated data from Laurits R. Christensen and William H. Greene for 1970 on 99 electric utility companies (called UPDATE). Data files on unit costs and production in the manufacture of polyethylene (POLY) and titanium dioxide (TIO2) are also found in the CHAP3.DAT subdirectory. Finally, there is a rather strange data set in this file called WEIRD.

Note: Remember that before you can use the data diskette in this exercise, you must properly format the data files. For further information, refer back to Chapter 1, Section 1.3. MAKE SURE THAT ALL DATA FILES TO BE USED BELOW ARE PROPERLY FORMATTED.

At numerous times in the following exercises you will be asked to test an hypothesis using a "reasonable level of significance." Since what is "reasonable" is to some extent discretionary, you are asked to follow the convention of stating precisely what percent significance level you use to test hypotheses; in that way, others with different preferences can still draw their own inferences. Alternatively, you may wish to state the level of significance at which the parameter estimate is significantly different from zero or at which the null hypothesis is rejected.

EXERCISES

EXERCISE 1: Estimating Parameters of a Learning Curve

The purpose of this exercise is to have you estimate and interpret parameters of a learning curve. In this exercise you have a choice of using one of two data files, both of which are in your data diskette subdirectory CHAP3.DAT. One file, called POLY, contains annual data on deflated unit costs (UCOSTP), current production (PRODP), and cumulative production to year $t - 1$ (CUMP) for a typical manufacturer of polyethylene over the 13-year period from 1960 to 1972. The second file, called TIO2, contains annual data for the DuPont Corporation in its production of titanium dioxide from 1955 to 1970. The variables in this data file include *undeflated* unit costs (UCOSTT), current production (PRODT), cumulative production to year $t - 1$ (CUMT), and a GNP-type deflator (DEFL).[29] The YEAR variable is called YEARP in the POLY file and YEART in the TIO2 file. Notice that both files contain data on *costs* and production, which are typically more difficult to obtain than data on *prices* and production.

Choose one of these two data files, and then perform steps (a) and (b). *Important note.* If you are using the TIO2 data file, you will first need to deflate the UCOSTT data by dividing UCOSTT by DEFL.

(a) Take the logarithm of cumulative production up to time period $t - 1$, and name this variable LNCP. Do the same for unit cost (LNUC) and current output (LNY). Plot unit cost against cumulative production, and then plot LNUC against LNCP. Comment on the two plots and what they might imply for the mathematical form of the learning curve relationship.

(b) Using least squares, estimate the parameters of the simple learning curve model (3.8) in which LNUC is regressed on a constant and on LNCP. Interpret your estimate of the learning curve elasticity. Is this estimate reasonable? What is the corresponding slope of the learning curve? Using a reasonable level of significance, construct a confidence interval for your estimated learning curve elasticity, and then test the null hypothesis that the learning curve elasticity is zero against the alternative hypothesis that it is not equal to zero.

EXERCISE 2: Testing the Simple Learning Curve Specification

In this exercise you test the simple learning curve specification as a special case of the more general Cobb-Douglas cost function. You also consider the effects of incorrectly omitting a right-hand variable

As in Exercise 1, in this exercise you have a choice of using one of two data files, both of which are in your data diskette subdirectory CHAP3.DAT. One file, called POLY, contains annual data on deflated unit costs (UCOSTP), current production (PRODP), and cumulative production to year $t-1$ (CUMP) for a typical manufacturer of polyethylene over the 13-year period from 1960 to 1972. The second file, called TIO2, contains annual data for the DuPont Corporation in its production of titanium dioxide from 1955 to 1970 on undeflated unit costs (UCOSTT), current production (PRODT), cumulative production to year $t-1$ (CUMT), and a GNP-type deflator (DEFL). The YEAR variable is called YEARP in the POLY file and YEART in the TIO2 file.

Choose one of these two data files, and then perform steps (a) through (d). *Important note.* If you are using the TIO2 data file, you will first need to deflate the UCOSTT data by dividing UCOSTT by DEFL.

(a) Just as in Exercise 1, take logarithms of cumulative production up to time period $t-1$ (call it LNCP), unit cost (LNUC), and current output (LNY). Using ordinary least squares, now estimate the parameters of a *multiple* regression model based on Eq. (3.33) in which LNUC is regressed on an intercept, LNCP, and LNY. Using a reasonable level of significance, test the null hypothesis that because technology is characterized by constant returns to scale, the coefficient on LNY in this multiple regression equals zero What is the point estimate of returns to scale with this model and data?

(b) Suppose that you instead estimated the simple learning curve specification (3.36) with LNY omitted as a right-hand variable. To evaluate the bias consequences on the learning curve elasticity estimate of incorrectly omitting LNY, run the simple regression (3.36) as well as the auxiliary regression (3.37). Verify numerically that the analytical relationship between the two estimates of the learning curve elasticity is in fact expressed by Eq (3.38). In this particular case, why is the bias large or small?

(c) On the basis of your results from part (b), compare the R^2 from the simple regression equation (3.36) with that from the multiple regression equation (3.33). Why is the latter R^2 larger? Is this always true? Why?

(d) Using a reasonable level of significance, test the joint null hypothesis that both slope coefficients in the regression equation (3.33) of part (a) are simultaneously equal to zero Perform this test using the analysis-of-variance approach based on changes in the sum of squared residuals and on the basis of the R^2 approach Compare the inference based on the two individual t-tests with that based on the joint F-test Are these results mutually consistent? Why or why not?

EXERCISE 3: R^2 in Simple and Multiple Regressions: A Surprise

The purpose of this exercise is to provide you with a dramatic example of possible relationships among R^2 measures in simple and multiple regressions. The common experience of practitioners is that the R^2 from a multiple regression of Y on X1 and X2 is less than the sum of the R^2 from the two simple regressions, Y on X1 and Y on X2. As you will see in this exercise, however, this need not always be the case.

In your data subdirectory CHAP3.DAT is a data file called WEIRD. It contains 15 observations on three variables: Y, X1, and X2.[30]

(a) Print the data series on Y, X1, and X2, and compute the simple correlations among these three variables. Does anything look particularly strange? Why or why not?

(b) Comment on why, especially with time series data, R^2 from a multiple regression of Y on a constant and K regressors is typically less than the sum of the R^2 from K simple regressions of Y on a constant and X_k, $k = 1, \ldots, K$.

(c) Now do an ordinary least squares regression of Y on a constant and X1. Then do an ordinary least squares regression of Y on a constant and X2. Are there any striking results on t-statistics in these two regressions? Compute the sum of the R^2 measures from these two regressions

(d) Next do a multiple regression, using ordinary least squares, of Y on a constant, X1 and X2. Are there any striking results on t-statistics in this regression? What is the R^2 from this multiple regression? Compare this R^2 to the sum of the R^2 from the two regressions from part (c). Why is this result rather surprising? To what do you attribute it?

EXERCISE 4: Replicating Nerlove's Classic Results on Scale Economies

The purpose of this exercise is to engage you in attempting to replicate some of the principal returns-to-scale results reported by Nerlove in his classic 1955 article The equation estimated by Nerlove is Eq. (3.20). In the data file NERLOV, data are provided on total costs (COSTS) in millions of dollars, output (KWH) in billions of kilowatt hours, and prices of labor (PL), fuels (PF), and capital (PK) for 145 electric utility companies in 1955. There are 145 observations, and the observations are ordered in size, observation 1 being the smallest company and observation 145 the largest.

(a) Using the data transformation facilities of your computer software, generate the variables required to estimate parameters of Eq. (3.20) by least squares. In particular, for each of the 145 companies, form the variables LNCP3 (≡ ln (COSTS/PF)), LNP13 (≡ ln (PL/PF)), LNP23 (≡ ln (PK/PF)), and LNKWH (≡ ln (KWH)). Print the entire data series for LNKWH, and verify that the observations are ordered by size of output, that is, that the first observation is the smallest output company, whereas the last observation has the largest output.[31]

(b) Given the data for all 145 firms from (a), estimate Eq. (3.20) by least squares, where your constructed variable LNCP3 ≡ ln C^*, LNKWH ≡ ln y, LNP13 ≡ ln p_1^*, and LNP23 ≡ ln p_2^*. Nerlove [1963, p. 176] reports parameter estimates for β_y, β_1, and β_2 as 0.721, 0.562, and −0.003, respectively, with standard errors of 0.175, 0.198, and 0.192, respectively, and an R^2 of 0.931. Can you replicate Nerlove's results? (*Note* You will not be able to replicate Nerlove's results precisely. One reason for this is that he used common rather than natural logarithms; however, this should affect only the estimated intercept term. According to Nerlove, the data set published with his article is apparently an earlier one that includes errors, while a revised data set was used in the estimation. This final data set has never been found.)

(c) Using the estimates you obtained in part (b) and a reasonable level of significance, construct a confidence interval for β_y. Is the null hypothesis that $\beta_y = 1$ rejected? What does this imply concerning a test of the null hypothesis that returns to scale are constant? Using line 5 of Eqs. (3.21), compute the point estimate of returns to scale based on your estimate of β_y. Are estimated returns to scale increasing, constant, or decreasing? Are economies of scale positive, zero, or negative?

(d) According to Eq. (3.12), demands for each factor of production will be positive only if α_i is positive, $i = 1, 2, 3$. What is the implied estimate of α_2 from part (c)? Is it significantly different from zero? Why do you think Nerlove was unsatisfied with this estimate of α_2?

(e) Compute and plot the residuals from estimated regression equation (3.20). Nerlove noticed that if the residuals were plotted against the log of output, the pattern was U-shaped, residuals at small levels of output being positive, those at medium levels of output being negative, and those at larger levels of output again becoming positive. Do you find the same U-shaped pattern? How might this pattern of residuals be interpreted? Finally, what is the sample correlation of residuals with LNKWH across the entire sample? Why is this the case?

EXERCISE 5: Assessing Alternative Returns-to-Scale Specifications

Because of the pattern of residuals noted by Nerlove (see Exercise 4, particularly part (e)), Nerlove hypothesized that estimated returns to scale varied with the level of output. In this exercise you evaluate Nerlove's conjecture and assess alternative specifications that relax the assumptions implicit in Eq. (3.20). To facilitate grouping the data in this exercise, in the data file NERLOV in the CHAP3 subdirectory is a variable named ORDER; the first 29 values of this variable are numbered 101 to 129, the second set of 29 values are 201 to 229, and so forth, the final 29 values of the variable ORDER taking on the values 501 to 529.

(a) Following Nerlove, divide the sample of 145 firms into five subsamples, each having 29 firms. Recall that since the data are ordered by level of output, the first 29 observations will have the smallest output levels, whereas the last 29 observations will have the largest output levels. Then using least squares regression techniques, estimate parameters of Eq. (3.20) for each of these subsamples. Nerlove [1963, p. 176] reports the following results (estimated standard errors are in parentheses):

	Parameter Estimated			
	β_y	β_1	β_2	R^2
Subsample I	0.398	0.641	−0.093	0.512
	(0.079)	(0.691)	(0.669)	
Subsample II	0.668	0.105	0.364	0.635
	(0.116)	(0.275)	(0.277)	
Subsample III	0.931	0.408	0.249	0.571
	(0.198)	(0.199)	(0.189)	
Subsample IV	0.915	0.472	0.133	0.871
	(0.108)	(0.174)	(0.157)	
Subsample V	1.045	0.604	−0.295	0.920
	(0.065)	(0.197)	(0.175)	

How well can you replicate Nerlove's reported results? To what might you attribute any discrepancies? (*Note* See the brief discussion at the end of part (b) of Exercise 4.)

(b) On the basis of your parameter estimates of β_y in part (a), compute the point estimates of returns to scale in each of the five subsamples. What is the general pattern of estimated scale economies as the level of output increases? How might this pattern be interpreted? Does this suggest an alternative specification?

(c) Now construct data variables such that a regression equation (3.20) will be estimated, except that while each of the five subsamples has common estimated "slope" coefficients for β_1 and β_2, each of the five subsamples has a different intercept term and a different estimate of β_y. Given the results in part (b), why might such a specification be plausible? Estimate this expanded model, and assess your success in replicating Nerlove, who reported the five subsample estimates of β_y as being 0.394 (0.055), 0.651 (0.189), 0.877 (0.376), 0.908 (0.354), and 1.062 (0.169), respectively, where numbers in parentheses are standard errors. The common estimates of β_1 and β_2 reported by Nerlove are 0.435 (0.207) and 0.100 (0.196), respectively. Nerlove's reported R^2 was 0.950.

(d) For each of the five subsample estimates of β_y in part (c), compute the implied estimate of returns to scale. What is the general pattern of estimated scale economies as the level of output increases?

(e) How would you compare estimates in part (c) versus those in part (a)? In particular, since part (a) estimates constitute a special case of part (c), using an F-test and a reasonable level of significance, formulate and test the restrictions implicit in part (a) against the alternative hypothesis in part (c). Is the null hypothesis rejected or not rejected? Comment on your results.

(f) To exploit the fact that estimated returns to scale seemed to decline with the level of output in a nonlinear fashion, Nerlove formulated and estimated a slight generalization of Eq. (3.20) in which the variable $(\ln y)^2$ was added as a regressor; call the corresponding coefficient β_{yy}. Using the full sample of 145 observations, estimate the equation

$$\ln C^* = \beta_0 + \beta_y \cdot \ln y + \beta_{yy} \cdot (\ln y)^2 + \beta_1 \cdot \ln p_1^* + \beta_2 \cdot \ln p_2^*$$

by least squares. How well can you replicate Nerlove's reported results, which he reported as 0.151 (0.062), 0.117 (0.012), 0.498 (0.161), and 0.062 (0.151) for β_y, β_{yy}, β_1, and β_2, respectively, and an R^2 of 0.952? Now, using a reasonable level of significance, test the joint null hypothesis that returns to scale are constant, that is, that $\beta_y = 1$, $\beta_{yy} = 0$, against the null hypothesis that returns to scale are nonconstant, that is, that $\beta_y \neq 1$, $\beta_{yy} \neq 0$. How does inference based on the joint F-test compare with that based on the individual t-tests? Finally, since returns to scale in the above expanded model vary with the level of output and can be shown to equal $r = 1/(\beta_y + 2 \cdot \beta_{yy} \cdot \ln y)$, compute the implied range of returns-to-scale estimates using the median value of LNY in each of the five subsamples.

EXERCISE 6: Comparing Returns-to-Scale Estimates from 1955 with Updated 1970 Data

Nerlove's returns to scale results were based on 1955 data for 145 electric utility companies in the United States. These data have been updated to 1970 and have been used for estimation by Christensen and Greene [1976]. In this exercise you compare returns-to-scale estimates based on the 1955 and the 1970 data and then evaluate the Christensen-Greene finding that by 1970 the bulk of electricity generation in the United States came from firms operating very near the bottom of their average cost curves.

The 1970 data are presented in the CHAP3.DAT subdirectory data file called UPDATE. The 1970 data sample is smaller, consisting of 99 observations, and like the data in NERLOV, the observations are ordered by size of firm, as measured by kilowatt hour output. The variables in the UPDATE data file include the original Christensen-Greene observation number (OBSNO), total costs in millions of 1970 dollars (COST70), millions of kilowatt hours of output (KWH70), the price of labor (PL70), the rental price index for capital (PK70), and the price index for fuels (PF70). (Notice that the numbers "70" have been added to the COST, KWH, PL, PK, and PF variables to distinguish these 1970 updated data from the Nerlove 1955 data.)[32]

(a) Using the 1970 updated data for 99 firms, construct the appropriate variables needed to estimate Eq. (3.20) by least squares. In particular, for each of the 99 observations, generate the following variables: LNC70 ≡ ln (COST70/PF70), LNY70 ≡ ln (KWH70), LNP170 ≡ ln (PL70/PF70), and LNP270 ≡ ln (PK70/PF70), where your just-constructed LNC70 is the same as $\ln C^*$ in Eqs. (3.20) and (3.21), LNY70 is $\ln y$, LNP170 is $\ln p_1^*$, and LNP270 is $\ln p_2^*$. Compute the sample mean for KWH70, and compare it to the sample mean for KWH in Nerlove's 1955 data set, found in the data file NERLOV. On average, are firms generating larger amounts of electricity in 1970 than in 1955? What might you therefore expect in terms of returns-to-scale estimates for 1970 as compared to those for 1955? Why?

(b) Now estimate by least squares the parameters of the equation

$$\text{LNC70} = \beta_0 + \beta_y \cdot \text{LNY70} + \beta_1 \cdot \text{LNP170} + \beta_2 \cdot \text{LNP270} + \epsilon$$

and then construct a confidence interval for β_y, using a reasonable level of significance. Is the null hypothesis of constant returns to scale ($\beta_y = 1$) rejected? Using line 5 of Eqs. (3.21), calculate the implied estimate of returns to scale. Compare this result, based on the 1970 data, with that reported by Nerlove for his 1955 data (see Exercise 4, part (b), for a list of Nerlove's results). Are you surprised by these results? Why or why not?

(c) A slightly generalized version of Eq. (3.20) involves adding $(\text{LNY70})^2$ as a regressor. Such an equation,

$$\text{LNC70} = \beta_0 + \beta_y \cdot \text{LNY70} + \beta_{yy} \cdot (\text{LNY70})^2 + \beta_1 \cdot \text{LNP170} + \beta_2 \cdot \text{LNP270} + \epsilon$$

cannot be derived from the Cobb-Douglas production function (3.10) but has the advantage of permitting returns to scale to vary with the level of output. In particular, in the above equation, returns to scale can be shown to equal $r = 1/(\beta_y + 2 \cdot \beta_{yy} \cdot \ln y)$. Note also that the equation in part (b) of this exercise (analogous to Eq. (3.20) in the text) is a special case of the expanded equation here, being valid if and only if $\beta_{yy} = 0$. Using the 1970 data, estimate by least squares the parameters of the above expanded equation. Then, based on a reasonable level of significance, test the null hypothesis that returns to scale do not vary with the level of output, that is, test the null hypothesis that $\beta_{yy} = 0$ against the alternative hypothesis that $\beta_{yy} \neq 0$. Next test the joint null hypothesis that returns to scale are constant, that is, that $\beta_{yy} = 0$ and $\beta_y = 1$, against the alternative hypothesis that $\beta_{yy} \neq 0$, $\beta_y \neq 1$. Interpret these two different test results. Are they mutually consistent?

(d) Next, calculate the implied range of returns to scale by splitting the 1970 sample into five groups, ordered by size, where the first four groups consist of 20 firms each and the last group has only 19 firms. Estimate by least squares the parameters of the equation in part (b) separately for each of the five groups. For each group, compare the returns-to-scale estimates based on 1970 data with those reported by Nerlove and based on 1955 data, namely, 2.92, 2.24, 1.97, 1.84, and 1.69.

(e) Finally, how might one best evaluate the Christensen-Greene finding that by 1970 the bulk of U.S. electricity generation was being produced by firms operating "very close" to the bottom of their average cost curves? Do you agree or disagree with Christensen and Greene? Why?

EXERCISE 7: Autocorrelation in the Learning Curve Model

The purpose of this exercise is to assess whether disturbances in the simple and expanded learning curve models are independently distributed or appear to follow a first-order autoregressive pattern.

(a) Using the data described in Exercises 1 and 2 and least squares regression methods, choose either the polyethylene or the titanium dioxide data and estimate the parameters of the simple learning curve model (3.36) as well as the generalized learning curve model (3.33). On the basis of the Durbin-Watson test statistic and using a reasonable level of significance, test the null hypothesis of no autocorrelation against the alternative hypothesis of a first-order autoregressive disturbance structure.

(b) Using either the Hildreth-Lu or the Cochrane-Orcutt estimation procedures, estimate both the simple and generalized learning curve models under the assumption that disturbances follow a first-order autoregressive scheme. Compare the estimated learning curve elasticities, as well as their statistical significance, with and without allowing for first-order autocorrelation. Are there any surprises? Why or why not?

EXERCISE 8: Misspecification in the Nerlove Returns-to-Scale Model

The purpose of this exercise is to acquaint you with one somewhat unorthodox procedure for examining misspecification in the cross-sectional data set on electric utility companies in 1955. Even though the data are cross-sectional, the Durbin-Watson test statistic will be seen to provide some useful (although not particularly powerful) evidence concerning model misspecification.

(a) Using the 1955 data for 145 electric utility firms described in Exercise 4 (see especially part (a) of that exercise), estimate by least squares the parameters of the regression equation (3.20).

(b) A particular feature of this data set is that it is ordered by size of output, that is, the first observation is that from the firm with the smallest output in 1955, while the last observation is that from the firm with the largest output. If average cost curves are U-shaped, then returns to scale will vary with the level of output. The specification of Eq. (3.20), however, assumes that the degree of returns to scale is the same regardless of the level of output. If in fact returns to scale varied with the level of output but a regression equation like Eq. (3.20) were estimated by least squares where the data were ordered by size of output, what might the pattern of residuals look like? In what sense is this lack of independence among residuals in this cross-sectional context similar to first-order autocorrelation in the time series context?

(c) Given this different interpretation of first-order autocorrelation, use a reasonable level of significance and the Durbin-Watson test statistic from part (a) to test the null hypothesis that disturbances are independent. Interpret the alternative hypothesis. Would it make sense to reestimate this model by using the Hildreth-Lu or Cochrane-Orcutt procedures, or does the above analysis suggest instead that the apparent presence of autocorrelation here indicates a fundamental problem with specification that cannot simply be "fixed" by using generalized least squares? Defend your position carefully. (The results in the first parts of Exercise 5 might be of help.)

EXERCISE 9: Testing for the Equality of Coefficients in the 1955 and 1970 Returns-to-Scale Models

The purpose of this exercise is to give you experience in performing a number of tests on the equality of the returns-to-scale estimates based on various subsamples of the 1955 and 1970 electric utility companies. The data necessary for this study are described at the beginning of Exercises 4 and 6. The hypothesis tests involve use of the Chow test procedure.[33]

(a) The 145 companies in the 1955 data base are ordered by size of output. Disaggregate this sample into five groups with 29 observations each and estimate Eq. (3.20) for each group and for the entire 145-company sample. Using the Chow test procedure, test the null hypothesis that parameters in each of these five subgroups are equal. (*Note* Take special care in computing degrees of freedom.)

(b) The 99 companies in the 1970 data base are also ordered by size of output. Disaggregate this 1970 sample into five groups with 20 observations in each of the first four groups and 19 observations in the last group. Then estimate Eq. (3.20) for each group and for the entire 99-company sample. Using the Chow test procedure, test the null hypothesis that parameters in each of these five subgroups in 1970 are equal.

(c) Now pool the 1955 and 1970 data, estimate Eq. (3.20) for the pooled sample, and then test the null hypothesis that parameters in each of the ten subgroups (five in 1955 and five in 1970) are equal; also test the null hypothesis that parameters in 1970 (assumed to be the same for all companies in that year) equal those from 1955 (assumed to be the same for all companies in that year).

(d) Finally, using least squares procedures and a reasonable level of significance, estimate the generalized Cobb-Douglas model with $(\ln y)^2$ added as a regressor, that is, estimate the parameters of

$$\begin{aligned} \text{LNC70} = \beta_0 + \beta_y \cdot \text{LNY70} + \beta_{yy} \cdot (\text{LNY70})^2 \\ + \beta_1 \cdot \text{LNP170} + \beta_2 \cdot \text{LNP270} + \epsilon \end{aligned}$$

first for the entire 145-company sample in 1955, then for the entire 99-company sample in 1970, and finally for the pooled 1955–1970 sample. Using the Chow test procedure, test the null hypothesis of parameter equality in 1955 and 1970 for this generalized Cobb-Douglas model.

(e) Comment on how one might interpret the entire set of test results in parts (a) through (d), in particular, on how they might be of help in choosing a preferred specification.

EXERCISE 10: Forecasting Unit Cost after Further Learning Occurs

One of the principal reasons why analysts are interested in the effects of learning curves is that, if estimated reliably, the parameters can be used to forecast unit costs after further learning occurs. In fact, as was noted earlier in this chapter, if unit costs can reliably be forecasted to fall at a certain rate as cumulative production increases, in some cases it may be optimal to reduce current prices in order to increase current demand and production and thereby accelerate learning. The purpose of this exercise is to forecast unit costs after cumulative production doubles and to obtain an estimate of the forecast error variance.

Recall from Eq. (3.36) that the bivariate regression model underlying the simple form of the learning curve is $\ln c_t = \alpha_0 + \alpha_1 \cdot \ln n_t$; for notational simplicity, rewrite this equation as $z_t = \alpha_0 + \alpha_1 \cdot x_t$, where $z_t \equiv \ln c_t$ and $x_t \equiv \ln n_t$. When an identically and independently normally distributed random disturbance term is appended to this equation, and given a known future value of x_t denoted x_τ, as is shown in numerous econometric theory textbooks,[34] the estimated forecast error variance s_f^2 equals

$$\begin{aligned} s_{f,\tau}^2 &= \text{VAR}(a_0) + 2 \cdot x_\tau \cdot \text{COV}(a_0, a_1) + x_\tau^2 \cdot \text{VAR}(a_1) + s^2 \\ &= s^2 \cdot \left[1 + \frac{1}{T} + \frac{(x_\tau - \bar{x})^2}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right] \end{aligned}$$

where a_0 and a_1 are least squares estimates of α_0 and α_1, T is the number of observations used in the sample to estimate the parameters, $\bar{x}$ is the sample mean of x_t (here, $\ln n_t$) over the estimation sample, x_τ is the known value of x ($\ln n$) at time τ in the future, and s^2 is the sum of squared residuals in the estimation divided by $T - 2$.

(a) Using the data on unit costs and cumulative production of either the polyethylene or the titanium dioxide data discussed in Exercise 1, obtain least squares estimates of α_0 and α_1, their variances and covariances, and s^2.

(b) Given the above formula and the estimates in part (a), calculate the forecasted value of unit costs, as well as the estimated forecast error variance at a level of cumulative production equal to twice that at the last historical observation. *Note:* To do this, double the value of n_t at the last historical observation, and then form x_τ as the natural logarithm of this doubled n_τ.

(c) Given the forecasted value of unit costs in part (b) brought about by additional learning, as well as the estimated forecast error variance, use a reasonable level of significance and construct a confidence interval for forecasted unit costs at $x = x_\tau$.

(d) After examining the above equation for the estimated forecast error variance, comment on the effects of increasing the sample size on the magnitude of the estimated forecast error variance. Since the number of observations in data sets underlying typical learning curve studies is rather small (often fewer than 20 observations), how reliable would you expect forecasts of unit costs to be? How would reliability be affected by the difference between x_τ and the sample mean of x? With these points in mind, comment on the range of the confidence interval estimated in part (c).

CHAPTER NOTES

1. More extended discussions on the six-tenths rule, scale economies at the product-specific and plant level, and a list of important references are found in F. Michael Scherer [1980], Chapter 4.
2. In the industrial psychology literature, learning functions go back at least as far as 1930; see L. L. Thurstone [1930] and, for an attempt to provide a statistical framework in an experimental setting, William K. Estes [1950]. For historical overviews and references in the management context, see Louis E. Yelle [1979] and John M. Dutton, Annie Thomas, and John E. Butler [1983].

3 For shipbuilding, see Allan D Searle [1945], and for aircraft, see Kenneth A Middleton [1945]; also see Leonard Rapping [1965].

4. Classic discussions of the learning curve in production economics include Frank J. Andress [1954], Armen Alchian [1963], Werner Hirsch [1952], and Wilfred B. Hirschmann [1964].

5 For nuclear power, see Martin L. Zimmerman [1982], and for coal-burning plants, see Paul L Joskow and Nancy L. Rose [1985].

6. This pricing strategy is often identified with the Boston Consulting Group, which made it a central focus of its corporate consulting practice See Boston Consulting Group [1973, 1974, 1982], A Michael Spence [1981], and Arnoldo Hax and Nicolas J. Majluf [1982] for further discussion

7 For an overview, see Jean Tirole [1989], Chapters 6–10. Also see Robert J Dolan and Abel P. Jeuland [1981], Drew Fudenberg and Jean Tirole [1983], and Saman Majd and Robert S. Pindyck [1989] On make-or-buy decisions, see James J. Anton and Dennis A Yao [1987]

8 See, for example, Partha Dasgupta and Joseph Stiglitz [1985] and Elhanan Helpman and Paul R. Krugman [1985].

9 The relationship between scale economies and learning curve effects has often been unclear in the strategic management literature; see William Alberts [1989] for a critical review For an empirical study that carefully distinguishes these notions and relates them to pricing behavior and market structure in the chemical processing industries, see Marvin B. Lieberman [1984]

10 Notice that in this case, λ also equals the marginal cost of producing output, that is, $\partial \mathcal{L}/\partial y = \lambda$.

11 Discussions of duality theory can be found in numerous advanced microeconomic theory textbooks. See, for example, Walter Nicholson [1985], Chapter 2, or R. Robert Russell and Maurice Wilkinson [1979], especially Chapters 5 and 9 A more complete discussion is found in Hal R Varian [1984], especially Chapters 1 and 4

12 The interpretation of learning and experience curves is not without controversy. For a recent critique that includes questions of the direction of causality, see William W. Alberts [1989]

13. For further discussions of strategic implications of learning and experience curves, see Arnoldo Hax and Nicolas Majluf [1982; 1984, Chapter 6], George S. Day and David B. Montgomery [1983], Boston Consulting Group [1982], Michael E Porter [1980], and William W. Alberts [1989].

14 Ford did this in part to take advantage of the cost-reducing effects of learning by experience. However, learning was not the only reason; volume buying of input parts also reduced input prices and therefore costs of production.

15. The measurement of learning curve effects is complicated when one uses price rather than unit cost data; see Section 3 5 1 for further discussion.

16. Figure 3.2 is taken from William J Abernathy and Kenneth Wayne [1974].

17 See William Abernathy and Kenneth Wayne [1974], as well as David B. Montgomery and George S. Day [1985]; also see Thomas H. Naylor, John M Vernon, and Kenneth L. Wertz [1983], especially Chapter 12

18. More complex formulations are considered in, among others, A. Ronald Gallant [1968], N. Keith Womer and J Wayne Patterson [1983], N. Keith Womer [1984], and John McDonald [1987].

19. Learning curve effects are typically assumed to affect costs. The extent to which such cost reductions are passed on to consumers in the form of price decreases depends, of course, on market structure and other strategic pricing decisions. See Section 3 5.1 for further discussion

20 For an historical overview of the Cobb-Douglas form and its estimation, see Paul A Samuelson [1979].

21 Note, however, that the interpretation of the intercept term in the two equations still differs In Eq (3.32), ln k' reflects the effects of the α parameters as shown in Eq. (3 26), while in Eq (3.8) the intercept represents c_1—unit costs in the initial production time period

22 Estimation and statistical inference issues arising in models that are nonlinear in the parameters are beyond the scope of this chapter but are discussed in many graduate econometric theory textbooks. See, for example, Thomas B. Fomby, R. Carter Hill, and Stanley R Johnson [1984], Appendix, pp. 603–616, and George G. Judge et al. [1985], Chapter 6

23 These and other measurement issues are discussed in further detail in George S Day and David B. Montgomery [1983] and in Montgomery and Day [1985].

24. The additive disturbance term in the Cobb-Douglas cost function equation (3.33) can be derived from a multiplicative random disturbance term appended to the original Cobb-Douglas production function (3 10) Specifically, if the multiplicative disturbance term on Eq. (3.10) is denoted as u and the additive disturbance term on the Cobb-Douglas cost function equation (3 36) is v, then it can be shown that $v = (-1/r) \cdot \ln u$

25. See, for example, Arthur S. Goldberger [1968], Chapter 3, in which the omitted variable bias algebra is worked out for both the 2- and the k-regressor case, for $k > 2$. Other early derivations are found in Zvi Griliches [1957] and Henri Theil [1957].

26. See the discussion in the paragraph underneath Eq. (3.31)

27 For additional discussion on the interpretation of learning or experience curves, see William W. Alberts [1989] The effect of possible first-order autocorrelation of disturbances on the estimation of learning curve elasticities is considered by David Montgomery and George Day [1985].

28. The translog form is discussed and implemented empirically in Chapter 9

29. The polyethylene data have been in the public domain for some time and are usually attributed to some early studies by the Boston Consulting Group, a consulting firm that has specialized in analyzing the effects of learning on optimal pricing and production strategies. The DuPont titanium dioxide data are taken from the Federal Trade Commision Docket No 9108, U.S Government Printing Office, Washington, D.C. They are described in further detail by Pankaj Ghemawat [1986].

30 These data are taken from David Hamilton [1987]

31. The careful student will note that it is not clear on what basis company size ranking is determined. Although Nerlove suggests that it is on the basis of company output (KWH), in fact there are several minor discrepancies in this ranking; see, for example, observations whose ORDER numbers are 119–120, 324–325, and 408–409

32 Further, although not used in this exercise, the UPDATE data file also contains 1970 cost share data for labor, capital, and fuel, denoted SL, SK, and SF, respectively

33. The Chow test procedure is described in most standard econometrics textbooks, often under the name "tests for parameter equality" The original article is by Gregory C. Chow [1960]

34. See, for example, Robert S. Pindyck and Daniel L. Rubinfeld [1981], Chapter 8, especially Section 8.1, pp. 206–211.

CHAPTER REFERENCES

Abernathy, William J. and Kenneth Wayne [1974], "Limits of the Learning Curve," *Harvard Business Review*, 52:5, September/October, 109–119.

Alberts, William W. [1989], "The Experience Curve Doctrine Reconsidered," *Journal of Marketing*, 53:3, July, 36–49.

Alchian, Armen [1963], "Reliability of Progress Curves in Airframe Production," *Econometrica*, 31:4, October, 679–693.

Andress, Frank J. [1954], "The Learning Curve as a Production Tool," *Harvard Business Review*, 32:1, January/February, 87–97.

Anton, James J. and Dennis A. Yao [1987], "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement," *Rand Journal of Economics*, 18:1, 57–76.

Boston Consulting Group [1973], "The Experience Curve—Reviewed, II: History," *Perspectives*, No. 125.

Boston Consulting Group [1974], "The Experience Curve—Reviewed, III: Why Does It Work?" *Perspectives*, No. 128.

Boston Consulting Group [1982], *Perspectives on Experience*, Boston: Boston Consulting Group, Inc.

Chow, Gregory C. [1960], "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28:3, July, 591–605.

Christensen, Laurits R. and William H. Greene [1976], "Economies of Scale in U.S. Electric Power Generation," *Journal of Political Economy*, 84:4, Part 1, August, 655–676.

Cobb, Charles and Paul H. Douglas [1928], "A Theory of Production," *American Economic Review*, 18:1, Supplement, March, 139–165.

Cowing, Thomas G. and V. Kerry Smith [1978], "The Estimation of a Production Technology: A Survey of Econometric Analyses of Steam-Electric Generation," *Land Economics*, 54:2, May, 157–170.

Dasgupta, Partha and Joseph Stiglitz [1985], "Learning-by-Doing, Market Structure and Industrial and Trade Policies," London: Centre for Economic Policy Research, Discussion Paper No. 80, October.

Day, George S. and David B. Montgomery [1983], "Diagnosing the Experience Curve," *Journal of Marketing*, 47:2, Spring, 44–58.

Dolan, Robert J. and Abel P. Jeuland [1981], "Experience Curves and Dynamic Demand Models: Implications for Optimal Pricing Strategies," *Journal of Marketing*, 45:1, Winter, 52–73.

Dutton, John M., Annie Thomas, and John E. Butler [1983], "The History of Progress Functions as a Managerial Technology," *Business History Review*, 58:2, Summer, 204–233.

Estes, William K. [1950], "Towards a Statistical Theory of Learning," *Psychological Review*, 57:2, March, 94–107.

Fomby, Thomas B., R. Carter Hill, and Stanley R. Johnson [1984], *Advanced Econometric Methods*, New York: Springer-Verlag.

Fudenberg, Drew and Jean Tirole [1983], "Learning-by-Doing and Market Performance," *Bell Journal of Economics*, 14:2, Autumn, 522–530.

Gallant, A. Ronald [1968], "A Note on the Measurement of Cost/Quantity Relationships in the Aircraft Industry," *Journal of the American Statistical Association*, 63:324, December, 1247–1252.

Ghemawat, Pankaj [1985], "Building Strategy on the Experience Curve," *Harvard Business Review*, 63:2, March/April, 143–149.

Ghemawat, Pankaj [1986], "DuPont in Titanium Dioxide," Cambridge, Mass: Harvard Graduate School of Business, Case Study 9-385-140, 1984; revised June 1986.

Goldberger, Arthur S. [1968], *Topics in Regression Analysis*, New York: Macmillan.

Griliches, Zvi [1957], "Specification Bias in Estimates of Production Functions," *Journal of Farm Economics*, 39:1, March, 8–20.

Hamilton, David [1987], "Sometimes $R^2 > r^2_{yx1} + r^2_{yx2}$: Correlated Variables Are Not Always Redundant," *The American Statistician*, 41:2, May, 129–132.

Hax, Arnoldo and Nicolas J. Majluf [1982], "Competitive Cost Dynamics: The Experience Curve," *Interfaces*, 12:5, October, 50–61.

Hax, Arnoldo and Nicolas Majluf [1984], *Strategic Management: An Integrative Perspective*, Englewood Cliffs, N.J.: Prentice-Hall.

Helpman, Elhanan and Paul R. Krugman [1985], *Market Structure and Foreign Trade*, Cambridge, Mass.: MIT Press.

Hirsch, Werner [1952], "Manufacturing Progress Functions," *Review of Economics and Statistics*, 34:2, May, 143–155.

Hirschmann, Wilfred B. [1964], "Profit from the Learning Curve," *Harvard Business Review*, 42:1, January/February, 125–139.

Joskow, Paul L., and Nancy L. Rose [1985], "The Effects of Technological Change, Experience, and Environmental Regulation on the Construction Cost of Coal-Burning Generating Units," *Rand Journal of Economics*, 16:1, Spring, 1–27.

Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee [1985], *The Theory and Practice of Econometrics*, Second Edition, New York: John Wiley and Sons.

Lieberman, Marvin B. [1984], "The Learning Curve and Pricing in the Chemical Processing Industries," *Rand Journal of Economics*, 15:2, Summer, 213–228.

Majd, Saman and Robert S. Pindyck [1989], "The Learning Curve and Optimal Production under Uncertainty," *Rand Journal of Economics*, 20:3, Autumn, 331–343.

McDonald, John [1987], "A New Model for Learning Curves, DARM," *Journal of Business and Economic Statistics*, 5:3, July, 329–335.

Middleton, Kenneth A. [1945], "Wartime Productivity Changes in the Airframe Industry," *Monthly Labor Review*, 62:2, August, 215–225.

Montgomery, David B. and George S. Day [1985], "Experience Curves: Evidence, Empirical Issues, and Applications," Chapter 3.6 in Howard Thomas and David Gardner, eds., *Strategic Marketing and Management*, New York: John Wiley and Sons, pp. 213–238.

Naylor, Thomas H., John M. Vernon, and Kenneth L. Wertz [1983], *Managerial Economics: Corporate Economics and Strategy*, New York: McGraw-Hill.

Nerlove, Marc [1963], "Returns to Scale in Electricity Supply," Chapter 7 in Carl F. Christ, ed., *Measurement in Economics: Studies in Honor of Yehuda Grunfeld*, Stanford, Calif.: Stanford University Press, pp. 167–198.

Nicholson, Walter [1985], *Microeconomic Theory*, Third Edition, Chicago: The Dryden Press.

Pindyck, Robert S and Daniel L. Rubinfeld [1981], *Econometric Models and Economic Forecasts*, Second Edition, New York: McGraw-Hill

Porter, Michael E. [1980], *Competitive Strategy*, New York: The Free Press, Macmillan

Rapping, Leonard [1965], "Learning and World War II Production Functions," *Review of Economics and Statistics*, 47:1, February, 81–86

Russell, R Robert and Maurice Wilkinson [1979], *Microeconomics A Synthesis of Modern and Neoclassical Theory*, New York: John Wiley and Sons

Samuelson, Paul A. [1979], "Paul Douglas' Measurement of Production Functions and Marginal Productivities," *Journal of Political Economy*, 87:5, Part 1, October, 923–939.

Scherer, F Michael [1980], *Industrial Market Structure and Economic Performance*, Second Edition, Chicago: Rand McNally.

Searle, Allan D. [1945], "Productivity Changes in Selected Wartime Shipbuilding Programs," *Monthly Labor Review*, 61:6, December, 1132–1147

Spence, A. Michael [1981], "The Learning Curve and Competition," *Bell Journal of Economics*, 12:1, Spring, 49–70.

Theil, Henri [1957], "Specification Errors and the Estimation of Economic Relationships," *Review of the International Statistical Institute*, 25:1, 41–51.

Thurstone, L L. [1930], "The Learning Function," *Journal of General Psychology*, 3, 469–493.

Tirole, Jean [1989], *The Theory of Industrial Organization*, Cambridge, Mass.: MIT Press.

Varian, Hal R. [1984], *Microeconomic Analysis*, Second Edition, New York: W W. Norton

Womer, N. Keith [1984], "Estimating Learning Curves from Aggregated Monthly Data," *Management Science*, 30:8, August, 982–992.

Womer, N Keith and J Wayne Patterson [1983], "Estimation and Testing of Learning Curves," *Journal of Business and Economic Statistics*, 1:4, October, 265–272.

Yelle, Louis E. [1979], "The Learning Curve: Historical Review and Comprehensive Survey," *Decision Sciences*, 10:2, April, 302–328.

Zimmerman, Martin L. [1982], "Learning Effects and the Commercialization of New Energy Technologies: The Case of Nuclear Power," *Bell Journal of Economics*, 13:2, Autumn, 297–310.

FURTHER READINGS

Diewert, W. Erwin [1974], "Applications of Duality Theory," in Michael Intriligator and David Kendrick, eds , *Frontiers of Quantitative Economics*, Vol 2, Amsterdam: North-Holland. A classic development of duality theory and its applications

Johnston, J. [1960], *Statistical Cost Analysis*, New York: McGraw-Hill An early study of cost and production economics

Nerlove, Marc [1965], *Estimation and Identification of Cobb-Douglas Production Functions*, Chicago: Rand McNally A classic and detailed econometric study using the Cobb-Douglas function.

Rosenberg, Nathan [1982], *Inside the Black Box*, Cambridge, England: Cambridge University Press. See especially Chapter 6, "Learning by Using."

Walters, Alan A. [1963], "Production and Cost Functions: An Econometric Survey," *Econometrica*, Vol. 31, January/April, pp. 1–66 A survey of econometric findings through 1962 based on the Cobb-Douglas production function