and

$$(\beta_1 + \gamma_1) + (\beta_2 + \delta_1)t_2^* = (\beta_1 + \gamma_1 + \gamma_2) + (\beta_2 + \delta_1 + \delta_2)t_2^*.$$

These are linear restrictions on the coefficients. Collecting terms, the first one is

$$\gamma_1 + \delta_1 t_1^* = 0 \text{ or } \gamma_1 = -\delta_1 t_1^*.$$

Doing likewise for the second and inserting these in (8-3), we obtain

$$\text{income} = \beta_1 + \beta_2 \text{ age} + \delta_1 d_1(\text{age} - t_1^*) + \delta_2 d_2(\text{age} - t_2^*) + \epsilon.$$

Constrained least squares estimates are obtainable by multiple regression, using a constant and the variables

$$x_1 = \text{age},$$
$$x_2 = \text{age} - 18 \quad \text{if age} \geq 18 \text{ and 0 otherwise},$$

and

$$x_3 = \text{age} - 22 \quad \text{if age} \geq 22 \text{ and 0 otherwise}.$$

We can test the hypothesis that the slope of the function is constant with the joint test of $\delta_1 = 0$ and $\delta_2 = 0$. Whether the individual test, $\delta_1 = 0$ or $\delta_2 = 0$, is meaningful will depend on the context. In our example, the first of these is a test that graduating from high school (actually, reaching 18) does not carry with it any increase in the slope of the earnings function until age 22 is reached.

8.3. Nonlinearity in the Variables

It is useful at this point to write the linear regression model in a very general form: Let $\mathbf{z} = z_1, z_2, \ldots, z_L$ be a set of L independent variables; let $f_1, f_2, \ldots, f_K$ be K linearly independent functions of $\mathbf{z}$; let $g(y)$ be an observable function of y; and retain the usual assumptions about the disturbance. The linear regression model is

$$\begin{aligned} g(y) &= \beta_1 f_1(\mathbf{z}) + \beta_2 f_2(\mathbf{z}) + \cdots + \beta_K f_K(\mathbf{z}) + \epsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \epsilon \qquad \textbf{(8-4)} \\ &= \mathbf{x}'\boldsymbol{\beta} + \epsilon. \end{aligned}$$

By using logarithms, exponentials, reciprocals, transcendental functions, polynomials, products, ratios, and so on, this "linear" model can be tailored to any number of situations.

8.3.1. FUNCTIONAL FORMS

A commonly used form of regression model is the log-linear model

$$y = \alpha \prod_k X_k^{\beta_k} e^{\epsilon}$$

or

$$\begin{aligned} \ln y &= \ln \alpha + \sum_k \beta_k \ln X_k + \epsilon \\ &= \beta_1 + \sum_k \beta_k x_k + \epsilon. \end{aligned}$$

In this model, the coefficients are elasticities:

$$\left(\frac{\partial y}{\partial x_k}\right)\left(\frac{x_k}{y}\right) = \frac{\partial \ln y}{\partial \ln x_k} = \beta_k. \tag{8-5}$$

This formulation is particularly useful in studies of demand and production.

A hybrid of the linear and log-linear models is the semilog equation

$$y = e^{\beta_1 + \beta_2 x + \epsilon}$$

or

$$\ln y = \beta_1 + \beta_2 x + \epsilon.$$

A common use of semilog formulation is in exponential growth curves. If x is "time" t, then

$$\begin{aligned}\frac{d \ln y}{dt} &= \beta_2 \\ &= \text{average rate of growth of } y.\end{aligned}$$

Macroeconomic models are often formulated with autonomous time trends. For example, aggregate models of productivity will usually include a trend variable, as in

$$\ln\left(\frac{Q}{L}\right)_t = \beta_1 + \beta_2 \ln\left(\frac{K}{L}\right)_t + \delta t + \epsilon_t,$$

which provides an estimate of the "autonomous growth in productivity," usually attributed to technical change. In this equation, δ is the rate of growth of average product not attributable to increases in the use of capital.

Another useful formulation of the regression model is one with **interaction terms.** For example, a model relating braking distance D to speed S and road wetness W might be

$$D = \beta_1 + \beta_2 S + \beta_3 W + \beta_4 SW + \epsilon.$$

In this model,

$$\begin{aligned}\frac{\partial E[D]}{\partial S} &= \beta_2 + \beta_4 W \\ \frac{\partial E[D]}{\partial W} &= \beta_3 + \beta_4 S,\end{aligned}$$

which implies that the marginal effect of higher speed on braking distance is increased when the road is wetter (assuming that β_4 is positive). If it is desired to form confidence intervals or test hypotheses about these marginal effects, then the necessary standard error is computed as

$$\text{Var}\left(\frac{\partial \hat{E}[D]}{\partial S}\right) = \text{Var}[\hat{\beta}_2] + W^2 \text{Var}[\hat{\beta}_4] + 2W\,\text{Cov}[\hat{\beta}_2, \hat{\beta}_4],$$

and likewise for $\partial E[D]/\partial W$. A value must be inserted for W. The sample mean is a natural choice, but for some purposes, a specific value, such as an extreme value of W in this example, might be preferred.

A form of the linear model

$$y = \alpha + \beta g(x) + \epsilon$$

that has appeared in a number of studies is the **Box–Cox transformation,**[8]

$$g^{(\lambda)}(x) = \frac{x^{\lambda} - 1}{\lambda}. \qquad \textbf{(8-6)}$$

The linear model results if λ equals 1, whereas a log-linear or semilog model (depending on how y is measured) results if λ equals 0.[9] Other values of λ produce many different functional forms. For example, if λ equals -1, then the equation will involve the reciprocal of x. The Box–Cox model is a useful formulation that embodies many of the models we have considered as special cases. Of course, in itself, this usefulness is only a minor virtue; if λ is known, then we merely insert the known value and obtain a model linear in the transformed variables. But, except for the special cases of λ equal to $-1, 0$, or 1, it is hard to conceive of situations in which a particular value would be specified a priori. By treating λ as an additional unknown parameter in the equation, we obtain a tremendous amount of flexibility.[10] The cost of doing so is that the model then becomes nonlinear in its parameters. We shall study the model in detail in Chapter 10.

Despite their very complex functional forms, these models are **intrinsically linear** because they can be placed directly in the form of (8-4). As discussed in Chapter 6, the distinguishing feature of the linear model is not the relationship among the variables as such but the way the parameters enter the equation.

8.3.2. IDENTIFYING NONLINEARITY

If the functional form is not known a priori, then there are a few approaches that might help at least to identify any nonlinearity and provide some information about it from the sample. For example, if the suspected nonlinearity is with respect to a single regressor in the equation, then fitting a quadratic or cubic polynomial rather than a linear function may capture some of the nonlinearity. By choosing several ranges for the regressor in question and allowing the slope of the function to be different in each range, a piecewise linear approximation to the nonlinear function can be fit.

EXAMPLE 8.3 Nonlinear Cost Function

In a celebrated study of the U.S. electric power industry, Nerlove (1963) analyzed the production costs of 145 American electric generating companies. At the outset of the study, he fit a Cobb–Douglas cost function of the form

$$\log C = \beta_1 + \beta_q \log Q + \sum_k \beta_k \log P_k + \epsilon,^{11} \qquad \textbf{(8-7)}$$

[8]See Box and Cox (1964) and Zarembka (1974). A survey of the properties of this model appears in Spitzer (1982b). Some further results on estimation appear in Spitzer (1982a).

[9]For λ equal to zero, L'Hôpital's rule can be used to obtain the result of the transformation: $g^{(0)}(x) = [\lim_{\lambda\to 0} d(x^{\lambda} - 1)/d\lambda]/[\lim_{\lambda\to 0} d\lambda/d\lambda] = \ln x$.

[10]In principle, y could be modified by the Box–Cox transformation as well. Some important issues of specification will arise, however (that is, is the model linear or log-linear? What is the allowable range of the dependent variable?). We defer until Chapter 10 a detailed treatment of the Box–Cox transformation applied to the dependent variable.

[11]Readers who attempt to replicate the original study should note that Nerlove used common (base 10) logs in his calculations, not natural logs. This change creates some numerical differences.