6 Functional Form and Structural Change

6.1 Introduction

This chapter will complete our ~~formal development~~ analysis of the linear regression model. We begin by examining different aspects of the functional form of the regression model. Many different types of functions are *linear* by the definition in Section 2.3.1. By using different transformations of the dependent and independent variables, binary variables, and different arrangements of functions of variables, a wide variety of models can be constructed that are all estimable by linear least squares. Section 6.2 considers using binary variables to accommodate nonlinearities in the model. Section 6.3 broadens the class of models that are linear in the parameters. By using logarithms, quadratic terms, and interaction terms (products of variables), the regression model can accommodate a wide variety of functional forms in the data.

Section 6.4 examines the issue of specifying and testing for discrete change in the underlying model that generates the data, under the heading of **structural change.** In a time series context, this relates to abrupt changes in the economic environment, such as major events in financial (e.g., the world financial crisis of 2007-2009) or commodity markets (such as the several upheavals in the oil market). In a cross section, we can modify the regression model to account for discrete differences across groups such as different preference structures or market experiences of men and women.

6-2

6 FUNCTIONAL FORM AND STRUCTURAL CHANGE

6.1 INTRODUCTION

In this chapter, we are concerned with the functional form of the regression model. Many different types of functions are "linear" by the definition considered in Section 2.3.1. By using different transformations of the dependent and independent variables, and dummy variables and different arrangements of functions of variables, a wide variety of models can be constructed that are all estimable by linear least squares. Section 6.2 considers using binary variables to accommodate nonlinearities in the model. Section 6.3 broadens the class of models that are linear in the parameters. Sections 6.4 and 6.5 then examine the issue of specifying and testing for change in the underlying model that generates the data, under the heading of **structural change.**

6.2 USING BINARY VARIABLES

One of the most useful devices in regression analysis is the **binary,** or **dummy variable.** A dummy variable takes the value one for some observations to indicate the presence of an effect or membership in a group and zero for the remaining observations. Binary variables are a convenient means of building discrete shifts of the function into a regression model.

6.2.1 BINARY VARIABLES IN REGRESSION

Dummy variables are usually used in regression equations that also contain other quantitative variables. In the earnings equation in Example 4.3, we included a variable *Kids* to indicate whether there were children in the household, under the assumption that for many married women, this fact is a significant consideration in labor supply behavior. The results shown in Example 6.1 appear to be consistent with this hypothesis.

Example 6.1 Dummy Variable in an Earnings Equation

Table 6.1 following reproduces the estimated earnings equation in Example 4.3. The variable Kids is a dummy variable, which equals one if there are children under 18 in the household and zero otherwise. Since this is a **semilog equation,** the value of −0.35 for the coefficient is an extremely large effect, one which suggests that all other things equal, the earnings of women with children are nearly a third less than those without. This is a large difference, but one that would certainly merit closer scrutiny. Whether this effect results from different labor market effects that influence wages and not hours, or the reverse, remains to be seen. Second, having chosen a nonrandomly selected sample of those with only positive earnings to begin with, it is unclear whether the sampling mechanism has, itself, induced a bias in this coefficient.

AU: KT "semilog equation" not in chap. list

TABLE 6.1 Estimated Earnings Equation

$\ln earnings = \beta_1 + \beta_2\, age + \beta_3\, age^2 + \beta_4\, education + \beta_5\, kids + \varepsilon$

Sum of squared residuals:	599.4582
Standard error of the regression:	1.19044
R^2 based on 428 observations	0.040995

Variable	*Coefficient*	*Standard Error*	*t Ratio*
Constant	3.24009	1.7674	1.833
Age	0.20056	0.08386	2.392
Age^2	−0.0023147	0.00098688	−2.345
Education	0.067472	0.025248	2.672
Kids	−0.35119	0.14753	−2.380

In recent applications, researchers in many fields have studied the effects of **treatment** on some kind of **response.** Examples include the effect of college on lifetime income, sex differences in labor supply behavior as in Example 6.1 in salary structures in industries, and in pre- versus postregime shifts in macroeconomic models, to name but a few. These examples can all be formulated in regression models involving a single dummy variable:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \delta d_i + \varepsilon_i$$

One of the important issues in policy analysis concerns measurement of such treatment effects when the dummy variable results from an individual participation decision. For example, in studies of the effect of job training programs on post-training earnings, the "treatment dummy" might be measuring the latent motivation and initiative of the participants rather than the effect of the program itself. We will revisit this subject in Section 24.5.

It is common for researchers to include a dummy variable in a regression to account for something that applies only to a single observation. For example, in time-series analyses, an occasional study includes a dummy variable that is one only in a single unusual year, such as the year of a major strike or a major policy event. (See, for example, the application to the German money demand function in Section 22.3.5.) It is easy to show (we consider this in the exercises) the very useful implication of this:

A dummy variable that takes the value one only for one observation has the effect of deleting that observation from computation of the least squares slopes and variance estimator (but not R-squared).

6.2.2 SEVERAL CATEGORIES

When there are several categories, a set of binary variables is necessary. Correcting for seasonal factors in macroeconomic data is a common application. We could write a consumption function for quarterly data as

$$C_t = \beta_1 + \beta_2 x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \varepsilon_t,$$

Dummy variables are particularly useful in loglinear regressions. In a model of the form

$$\ln y = \beta_1 + \beta_2 x + \beta_3 d + \varepsilon,$$

the coefficient on the dummy variable, d, indicates a multiplicative shift of the function. The percentage change in $E[y|x,d]$ associated with the change in d is

$$\begin{aligned}\%\left(\Delta E[y \mid x,d]/\Delta d\right) &= 100\%\left\{\frac{E[y \mid x,d=1]-E[y \mid x,d=0]}{E[y \mid x,d=0]}\right\} \\ &= 100\%\left\{\frac{\exp(\beta_1+\beta_2 x+\beta_3)E[\exp(\varepsilon)]-\exp(\beta_1+\beta_2 x)E[\exp(\varepsilon)]}{\exp(\beta_1+\beta_2 x)E[\exp(\varepsilon)]}\right\} \\ &= 100\%\left[\exp(\beta_3)-1\right].\end{aligned}$$

Example 6.2 Value of a Signature

Example 4.12 we explored the relationship between (log of) sale price and surface area for 430 sales of Monet paintings. Regression results from the example are included in Table 6.2 below. The results suggest a strong relationship between area and price – the coefficient is 1.33372 indicating a highly elastic relationship and the t ratio of 14.70 suggests the relationship is highly significant. A variable (effect) that is clearly left out of the model is the effect of the artist's signature on the sale price. Of the 430 sales in the sample, 77 are for unsigned paintings. The results at the right of Table 6.2 include a dummy variable for whether the painting is signed or not. The results show an extremely strong effect. The regression results imply that

$$E[\text{Price}|\ln\text{Area},\text{Aspect},\text{Signature}) = \exp[-9.64+1.35\ln\text{Area} - .08\text{AspectRatio} + 1.23\ \text{Signature} + .993^2/2].$$

(See Section 4.6.) Computing this result for a painting of the same area and aspect ratio, we find the model predicts that the signature effect would be

$$100\% \times (\Delta E[\text{Price}]/\text{Price}) = 100\%[\exp(1.23) - 1] = 242\%.$$

The effect of a signature on an otherwise similar painting is to more than double the price. The estimated standard error for the signature coefficient is 0.1253. Using the delta method, we obtain an estimated standard error for $[\exp(b_3)-1]$ of the square root of $[\exp(b_3)]^2 \times .1253^2$, which is 0.4285. For the percentage difference of 242%, we have an estimated standard error of 42.85%.

Superficially, it is possible that the size effect we observed earlier could be explained by the presence of the signature. If the artist tended on average to sign only the larger paintings, then we would have an explanation for the counterintuitive effect of size. (This would be an example of the effect of multicollinearity of a sort.) For a regression with a continuous variable and a dummy variable, we can easily confirm or refute this proposition. The average size for the 77 sales of unsigned paintings is 1228.69 square inches. The average size of the other 353 is 940.812 square inches. There does seem to be a substantial systematic difference between signed and unsigned paintings, but it goes in the other direction. We are left with significant findings of both a size and a signature effect in the auction prices of Monet paintings. *Aspect Ratio*, however, appears still to be inconsequential.

There is one remaining aspect of this sample for us to explore. These 430 sales involved only 387 different paintings, Several sales involved repeat sales of the same painting. The assumption that observations are independent draws is violated, at least for some of them. We will examine this form of "clustering" in Chapter 11 in our treatment of panel data.

TABLE 6.2 Estimated Equations for Log Price

$$\ln price = \beta_1 + \beta_2\, lnArea + \beta_3\, aspect\ ratio + \beta_4\, signature + \varepsilon$$

Mean of log Price	.33274	
Number of observations	430	

Sum of squared residuals	519.17235	420.16787
Standard error	1.10266	0.99313
R-squared	0.33620	0.46279
Adjusted R-squared	0.33309	0.45900

		Standard			*Standard*	
Variable	***Coefficient***	***Error***	***t***	***Coefficient***	***Error***	***t***
Constant	-8.42653	0.61183	-13.77	-9.64028	.56422	-17.09
Ln area	1.33372	0.09072	14.70	1.34935	.08172	16.51
Aspect ratio	-.16537	0.12753	-1.30	-0.07857	.11519	-0.68
Signature	0.00000	0.00000	0.00	1.25541	.12530	10.02

6-6

TABLE 6.1 Estimated Earnings Equation

$ln\ earnings = \beta_1 + \beta_2\, age + \beta_3\, age^2 + \beta_4\, education + \beta_5\, kids + \varepsilon$

Sum of squared residuals:	599.4582
Standard error of the regression:	1.19044
R^2 based on 428 observations	0.040995

Variable	*Coefficient*	*Standard Error*	*t Ratio*
Constant	3.24009	1.7674	1.833
Age	0.20056	0.08386	2.392
Age^2	−0.0023147	0.00098688	−2.345
Education	0.067472	0.025248	2.672
Kids	−0.35119	0.14753	−2.380

In recent applications, researchers in many fields have studied the effects of **treatment** on some kind of **response.** Examples include the effect of college on lifetime income, sex differences in labor supply behavior as in Example 6.1 in salary structures in industries, and in pre- versus postregime shifts in macroeconomic models, to name but a few. These examples can all be formulated in regression models involving a single dummy variable:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \delta d_i + \varepsilon_i$$

One of the important issues in policy analysis concerns measurement of such treatment effects when the dummy variable results from an individual participation decision. For example, in studies of the effect of job training programs on post-training earnings, the "treatment dummy" might be measuring the latent motivation and initiative of the participants rather than the effect of the program itself. We will revisit this subject in Section 24.5.

It is common for researchers to include a dummy variable in a regression to account for something that applies only to a single observation. For example, in time-series analyses, an occasional study includes a dummy variable that is one only in a single unusual year, such as the year of a major strike or a major policy event. (See, for example, the application to the German money demand function in Section 22.3.5.) It is easy to show (we consider this in the exercises) the very useful implication of this:

23

A dummy variable that takes the value one only for one observation has the effect of deleting that observation from computation of the least squares slopes and variance estimator (but not R-squared).

6.2.2 SEVERAL CATEGORIES

When there are several categories, a set of binary variables is necessary. Correcting for seasonal factors in macroeconomic data is a common application. We could write a consumption function for quarterly data as

$$C_t = \beta_1 + \beta_2 x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \varepsilon_t,$$

where x_t is disposable income. Note that only three of the four quarterly dummy variables are included in the model. If the fourth were included, then the four dummy variables would sum to one at every observation, which would reproduce the constant term—a case of perfect multicollinearity. This is known as the **dummy variable trap.** Thus, to avoid the dummy variable trap, we drop the dummy variable for the fourth quarter. (Depending on the application, it might be preferable to have four separate dummy variables and drop the overall constant.)[1] Any of the four quarters (or 12 months) can be used as the base period.

The preceding is a means of *deseasonalizing* the data. Consider the alternative formulation:

$$C_t = \beta x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \delta_4 D_{t4} + \varepsilon_t. \tag{6-1}$$

Section 3.3

Using the results from ~~Chapter 3~~ on partitioned regression, we know that the preceding multiple regression is equivalent to first regressing C and x on the four dummy variables and then using the residuals from these regressions in the subsequent regression of deseasonalized consumption on deseasonalized income. Clearly, deseasonalizing in this fashion prior to computing the simple regression of consumption on income produces the same coefficient on income (and the same vector of residuals) as including the set of dummy variables in the regression.

insert next page msp b-8

6.2.3 SEVERAL GROUPINGS

The case in which several sets of dummy variables are needed is much the same as those we have already considered, with one important exception. Consider a model of statewide per capita expenditure on education y as a function of statewide per capita income x. Suppose that we have observations on all $n = 50$ states for $T = 10$ years. A regression model that allows the expected expenditure to change over time as well as across states would be

$$y_{it} = \alpha + \beta x_{it} + \delta_i + \theta_t + \varepsilon_{it}. \tag{6-2}$$

As before, it is necessary to drop one of the variables in each set of dummy variables to avoid the dummy variable trap. For our example, if a total of 50 state dummies and 10 time dummies is retained, a problem of "perfect multicollinearity" remains; the sums of the 50 state dummies and the 10 time dummies are the same, that is, 1. One of the variables in each of the sets (or the overall constant term and one of the variables in one of the sets) must be omitted.

Example 6.2 Analysis of Covariance

The data in Appendix Table F6.1 were used in a study of efficiency in production of airline services in Greene (1997b). The airline industry has been a favorite subject of study [e.g., Schmidt and Sickles (1984); Sickles, Good, and Johnson (1986)], partly because of interest in this rapidly changing market in a period of deregulation and partly because of an abundance of large, high-quality data sets collected by the (no longer existent) Civil Aeronautics Board. The original data set consisted of 25 firms observed yearly for 15 years (1970 to 1984), a "balanced panel." Several of the firms merged during this period and several others experienced strikes, which reduced the number of complete observations substantially. Omitting these and others because of missing data on some of the variables left a group of 10 full

[1] See Suits (1984) and Greene and Seaks (1991).

1 See Suits (1984) and Greene and Seaks (1991).

Insert on msp, 6-7 where indicated

Example 6.3 Genre Effects on Movie Box Office Receipts

Table 4.8 in Example 4.12 presents the results of the regression of log of box office receipts for 62 2009 movies on a number of variables including a set of dummy variables for genre: *Action*, *Comedy*, *Animated* or *Horror*. The left out category is "any of the remaining 9 genres" in the standard set of 13 that is usually used in models such as this one. The four coefficients are -.869, -.016, -.833, +.375, respectively. This suggests that, save for horror movies, these genres typically fare substantially worse at the box office than other types of movies. We note, the use of *b* directly to estimate the percentage change for the category, as we did in example 6.1 when we interpreted the coefficient of -.35 on *Kids* as indicative of a 35% change in income, is an approximation that works well when *b* is close to zero, but deteriorates as it gets far from zero. Thus, the value of -.869 above does not translate to an 87% difference between *Action* movies and other movies. Using the formula we used in Example 6.2, we find an estimated difference closer to [exp(-.869)-1] or about 58%.

where x_t is disposable income. Note that only three of the four quarterly dummy variables are included in the model. If the fourth were included, then the four dummy variables would sum to one at every observation, which would reproduce the constant term—a case of perfect multicollinearity. This is known as the **dummy variable trap.** Thus, to avoid the dummy variable trap, we drop the dummy variable for the fourth quarter. (Depending on the application, it might be preferable to have four separate dummy variables and drop the overall constant.)[1] Any of the four quarters (or 12 months) can be used as the base period.

The preceding is a means of *deseasonalizing* the data. Consider the alternative formulation:

$$C_t = \beta x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \delta_4 D_{t4} + \varepsilon_t. \tag{6-1}$$

Using the results from Chapter 3 on partitioned regression, we know that the preceding multiple regression is equivalent to first regressing C and x on the four dummy variables and then using the residuals from these regressions in the subsequent regression of deseasonalized consumption on deseasonalized income. Clearly, deseasonalizing in this fashion prior to computing the simple regression of consumption on income produces the same coefficient on income (and the same vector of residuals) as including the set of dummy variables in the regression.

insert next page →

6.2.3 SEVERAL GROUPINGS

The case in which several sets of dummy variables are needed is much the same as those we have already considered, with one important exception. Consider a model of statewide per capita expenditure on education y as a function of statewide per capita income x. Suppose that we have observations on all $n = 50$ states for $T = 10$ years. A regression model that allows the expected expenditure to change over time as well as across states would be

$$y_{it} = \alpha + \beta x_{it} + \delta_i + \theta_t + \varepsilon_{it}. \tag{6-2}$$

As before, it is necessary to drop one of the variables in each set of dummy variables to avoid the dummy variable trap. For our example, if a total of 50 state dummies and 10 time dummies is retained, a problem of "perfect multicollinearity" remains; the sums of the 50 state dummies and the 10 time dummies are the same, that is, 1. One of the variables in each of the sets (or the overall constant term and one of the variables in one of the sets) must be omitted.

Example 6.2 Analysis of Covariance

The data in Appendix Table F6.1 were used in a study of efficiency in production of airline services in Greene (1997b). The airline industry has been a favorite subject of study [e.g., Schmidt and Sickles (1984); Sickles, Good, and Johnson (1986)], partly because of interest in this rapidly changing market in a period of deregulation and partly because of an abundance of large, high-quality data sets collected by the (no longer existent) Civil Aeronautics Board. The original data set consisted of 25 firms observed yearly for 15 years (1970 to 1984), a "balanced panel." Several of the firms merged during this period and several others experienced strikes, which reduced the number of complete observations substantially. Omitting these and others because of missing data on some of the variables left a group of 10 full

2007a

[1] See Suits (1984) and Greene and Seaks (1991).

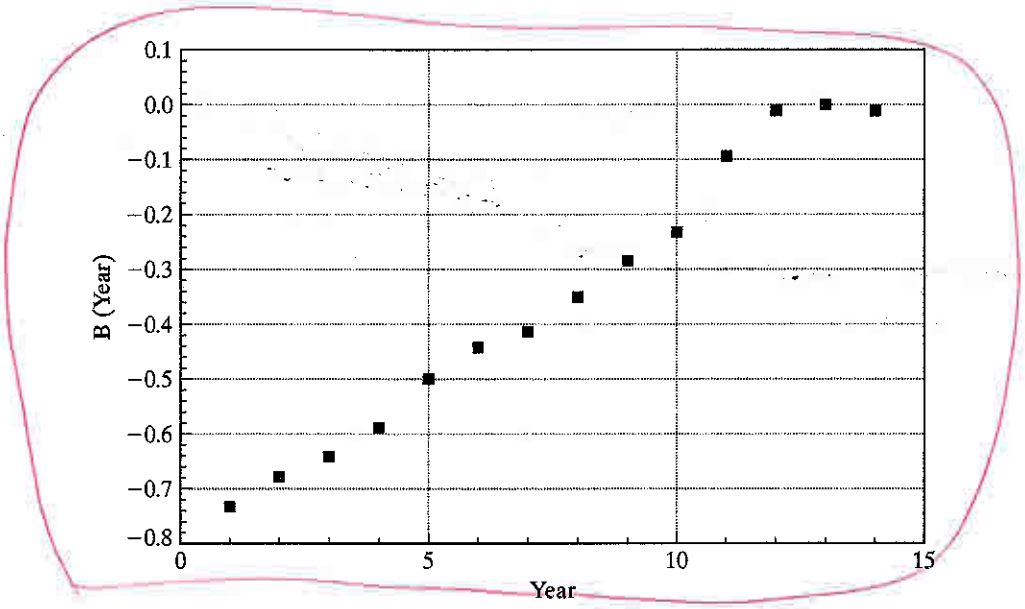

FIGURE 6.1 Estimated Year Dummy Variable Coefficients.

observations, from which we have selected six for the examples to follow. We will fit a cost equation of the form

$$\ln C_{i,t} = \beta_1 + \beta_2 \ln Q_{i,t} + \beta_3 \ln^2 Q_{i,t} + \beta_4 \ln P_{fuel\ i,t} + \beta_5\, Loadfactor_{i,t} + \sum_{t=1}^{14} \theta_t D_{i,t} + \sum_{i=1}^{5} \delta_i F_{i,t} + \varepsilon_{i,t}.$$

The dummy variables are $D_{i,t}$ which is the year variable and $F_{i,t}$ which is the firm variable. We have dropped the last one in each group. The estimated model for the full specification is

$$\ln C_{i,t} = 13.56 + 0.8866 \ln Q_{i,t} + 0.01261 \ln^2 Q_{i,t} + 0.1281 \ln P_{fi,t} - 0.8855\, LF_{i,t} + \text{time effects} + \text{firm effects}.$$

FIG 6.1

The year effects display a revealing pattern, as shown in Figure 6.1. This was a period of rapidly rising fuel prices, so the cost effects are to be expected. Since one year dummy variable is dropped, the effect shown is relative to this base year (1984).

TB 6.3

We are interested in whether the firm effects, the time effects, both, or neither are statistically significant. Table 6.2 presents the sums of squares from the four regressions. The F statistic for the hypothesis that there are no firm-specific effects is 65.94, which is highly significant. The statistic for the time effects is only 2.61, which is larger than the critical value

6.3

3

TABLE 6.2 F tests for Firm and Year Effects

Model	*Sum of Squares*	*Restrictions*	*F*	*Deg.Fr.*
Full model	0.17257	0	—	
Time effects only	1.03470	5	65.94	[5, 66]
Firm effects only	0.26815	14	2.61	[14, 66]
No effects	1.27492	19	22.19	[19, 66]

of 1.84, but perhaps less so than Figure 6.1 might have suggested. In the absence of the year-specific dummy variables, the year-specific effects are probably largely absorbed by the price of fuel.

6.2.4 THRESHOLD EFFECTS AND CATEGORICAL VARIABLES

In most applications, we use dummy variables to account for purely qualitative factors, such as membership in a group, or to represent a particular time period. There are cases, however, in which the dummy variable(s) represents levels of some underlying factor that might have been measured directly if this were possible. For example, education is a case in which we typically observe certain thresholds rather than, say, years of education. Suppose, for example, that our interest is in a regression of the form

$$income = \beta_1 + \beta_2\, age + effect\ of\ education + \varepsilon.$$

The data on education might consist of the highest level of education attained, such as high school (HS), undergraduate (B), master's (M), or Ph.D. (P). An obviously unsatisfactory way to proceed is to use a variable E that is 0 for the first group, 1 for the second, 2 for the third, and 3 for the fourth. That is, $income = \beta_1 + \beta_2\, age + \beta_3 E + \varepsilon$. The difficulty with this approach is that it assumes that the increment in income at each threshold is the same; β_3 is the difference between income with a Ph.D. and a master's and between a master's and a bachelor's degree. This is unlikely and unduly restricts the regression. A more flexible model would use three (or four) binary variables, one for each level of education. Thus, we would write

$$income = \beta_1 + \beta_2\, age + \delta_B\, B + \delta_M\, M + \delta_P\, P + \varepsilon.$$

The correspondence between the coefficients and income for a given age is

$$\begin{aligned}
&\text{High school:} && E[income \mid age, HS] = \beta_1 + \beta_2\, age,\\
&\text{Bachelor's:} && E[income \mid age, B] = \beta_1 + \beta_2\, age + \delta_B,\\
&\text{Master's:} && E[income \mid age, M] = \beta_1 + \beta_2\, age + \delta_M,\\
&\text{Ph.D.:} && E[income \mid age, P] = \beta_1 + \beta_2\, age + \delta_P.
\end{aligned}$$

The differences between, say, δ_P and δ_M and between δ_M and δ_B are of interest. Obviously, these are simple to compute. An alternative way to formulate the equation that reveals these differences directly is to redefine the dummy variables to be 1 if the individual has the degree, rather than whether the degree is the highest degree obtained. Thus, for someone with a Ph.D., all three binary variables are 1, and so on. By defining the variables in this fashion, the regression is now

$$\begin{aligned}
&\text{High school:} && E[income \mid age, HS] = \beta_1 + \beta_2\, age,\\
&\text{Bachelor's:} && E[income \mid age, B] = \beta_1 + \beta_2\, age + \delta_B,\\
&\text{Master's:} && E[income \mid age, M] = \beta_1 + \beta_2\, age + \delta_B + \delta_M,\\
&\text{Ph.D.:} && E[income \mid age, P] = \beta_1 + \beta_2\, age + \delta_B + \delta_M + \delta_P.
\end{aligned}$$

Instead of the difference between a Ph.D. and the base case, in this model δ_P is the marginal value of the Ph.D. How equations with dummy variables are formulated is a matter of convenience. All the results can be obtained from a basic equation.

6.2.5 Treatment Effects and Difference in Differences Regression

Researchers in many fields have studied the effect of a **treatment** on some kind of **response.** Examples include the effect of going to college on lifetime income [Dale and Krueger (2002)], the effect of cash transfers on child health [Gertler (2004)], the effect of participation in job training programs on income [LaLonde (1986)] and pre- versus postregime shifts in macroeconomic models [Mankiw (2006)], to name but a few. These examples can be formulated in regression models involving a single dummy variable:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \delta D_i + \varepsilon_i,$$

where the shift parameter, δ, measures the impact of the treatment or the policy change (conditioned on **x**) on the sampled individuals. In the simplest case of a comparison of one group to another,

$$y_i = \beta_1 + \beta_2 D_i + \varepsilon_i,$$

we will have $b_1 = (\bar{y} \mid D_i = 0)$, that is, the average outcome of those who did not experience the intervention, and $b_2 = (\bar{y} \mid D_i = 1) - (\bar{y} \mid D_i = 0)$, the difference in the means of the two groups. In the Dale and Krueger (2002) study, the model compared the incomes of students who attended elite colleges to those who did not. When the analysis is of an intervention that occurs over time, such as Krueger's (1999) analysis the Tennessee STAR experiment in which school performance measures were observed before and after a policy dictated change in class sizes, the treatment dummy variable will be a period indicator, $D_t = 0$ in period 1 and 1 in period 2. The effect in β_2 measures the change in the outcome variable, e.g., school performance, pre- to post-intervention; $b_2 = \bar{y}_1 - \bar{y}_0$.

The assumption that the treatment group does not change from period 1 to period 2 weakens this comparison. A strategy for strengthening the result is to include in the sample a group of **control observations** that do not receive the treatment. The change in the outcome for the **treatment group** can then be compared to the change for the **control group** under the presumption that the difference is due to the intervention. An intriguing application of this strategy is often used in clinical trials for health interventions to accommodate the **placebo effect**. The placebo "effect" is a controversial, but apparently tangible outcome in some clinical trials in which subjects "respond" to the treatment even when the treatment is a decoy intervention, such as a sugar or starch pill in a drug trial. [See Hróbjartsson and Peter C. Götzsche, 2001]. A broad template for assessment of the results of such a clinical trial is as follows: The subjects who receive the placebo are the controls. The outcome variable – level of cholestrol for example – is measured at the baseline for both groups. The treatment group receives the drug; the control group receives the placebo, and the outcome variable is measured post treatment. The impact is measured by the difference in differences,

$$E = [(\bar{y}_{exit} \mid treatment) - (\bar{y}_{baseline} \mid treatment)] - [(\bar{y}_{exit} \mid placebo) - (\bar{y}_{baseline} \mid placebo)].$$

The presumption is that the difference in differences measurement is robust to the placebo effect *if it exists*. If there is no placebo effect, the result is even stronger (assuming there is a result).

AU: OK to spell out "e.g." in text?

AU: Four KTs in paragraph not in chap. list

AU: Are subs in equation OK set italics?

An increasingly common social science application of treatment effect models with dummy variables is in the evaluation of the effects of discrete changes in policy.[2] A pioneering application is the study of the Manpower Development and Training Act (MDTA) by Ashenfelter and Card (1985). The simplest form of the model is one with a pre- and post treatment observation on a group, where the outcome variable is y, with

$$y_{it} = \beta_1 + \beta_2 T_t + \beta_3 D_i + \beta_4 T_t \times D_i + \varepsilon, \; t = 1,2. \qquad (6-3)$$

In this model, T_t is a dummy variable that is zero in the pre- treatment period and one after the treatment and D_i equals one for those individuals who received the "treatment." The change in the outcome variable for the "treated" individuals will be

$$(y_{i2}|D_i = 1) - (y_{i1}|D_i = 1) = (\beta_1 + \beta_2 + \beta_3 + \beta_4) - (\beta_1 + \beta_3) = \beta_2 + \beta_4.$$

For the controls, this is

$$(y_{i2}|D_i = 0) - (y_{i1}|D_i = 0) = (\beta_1 + \beta_2) - (\beta_1) = \beta_2.$$

The **difference in differences** is

$$[(y_{i2}|D_i = 1) - (y_{i1}|D_i = 1)] - [(y_{i2}|D_i = 0) - (y_{i1}|D_i = 0)] = \beta_4.$$

In the multiple regression of y_{it} on a constant, T, D and TD, the least squares estimate of β_4 will equal the difference in the changes in the means,

$$\begin{aligned} b_4 = {} & (\bar{y} \mid D = 1, Period\ 2) - (\bar{y} \mid D = 1, Period\ 1) \\ & - (\bar{y} \mid D = 0, Period\ 2) - (\bar{y} \mid D = 0, Period\ 1) \\ = {} & \Delta\bar{y}|treatment - \Delta\bar{y}|control. \end{aligned}$$

The regression is called a difference in differences estimator in reference to this result.

[2] Surveys of literatures on treatment effects, including use of D-i-D estimators are provided by Imbens and Wooldridge (2009) and Millimet, Smith and Vytlacil (2008).

When the treatment is the result of a policy change or event that occurs completely outside the context of the study, the analysis is often termed a **natural experiment**. Card's (1990) study of a major immigration into Miami in 1979 discussed in Example 6.5 is an ~~example~~ application.

Example 6.5 A Natural Experiment: The Mariel Boatlift

A sharp change in policy can constitute a natural experiment. An example studied by Card (1990) is the Mariel boatlift from Cuba to Miami (May-September, 1980) which increased the Miami labor force by 7%. The author examined the impact of this abrupt change in labor market conditions on wages and employment for non-immigrants. The model compared Miami to a similar city, Los Angeles. Let i denote an individual and D denote the "treatment," which for an individual would be equivalent to "lived in a city that experienced the immigration." For an individual in either Miami or Los Angeles, the outcome variable is

(Y_i) = 1 if they are unemployed and 0 if they are employed.

Let c denote the city and let t denote the period, before (1979) or after (1981) the immigration. Then, the unemployment rate in city c at time t is $E[y_{i,0}|c,t]$ if there is no immigration and it is $E[y_{i,1}|c,t]$ if there is the immigration. These rates are assumed to be constants. Then,

$$E[y_{i,0}|c,t] = \beta_t + \gamma_c \quad \text{without the immigration,}$$
$$E[y_{i,1}|c,t] = \beta_t + \gamma_c + \delta \quad \text{with the immigration.}$$

The effect of the immigration on the unemployment rate is measured by δ. The natural experiment is that the immigration occurs in Miami and not in Los Angeles, but is not a result of any action by the people in either city. Then,

$$E[y_i|M,79] = \beta_{79} + \gamma_M \quad \text{and} \quad E[y_i|M,81] = \beta_{81} + \gamma_M + \delta \quad \text{for Miami,}$$
$$E[y_i|L,79] = \beta_{79} + \gamma_L \quad \text{and} \quad E[y_i|L,81] = \beta_{81} + \gamma_L \quad \text{for Los Angeles.}$$

It is assumed that unemployment growth in the two cities would be the same if there were no immigration. If neither city experienced the immigration, the change in the unemployment rate would be

$$E[y_{i,0}|M,81] - E[y_{i,0}|M,79] = \beta_{81} - \beta_{79} \quad \text{for Miami,}$$
$$E[y_{i,0}|L,81] - E[y_{i,0}|L,79] = \beta_{81} - \beta_{79} \quad \text{for Los Angeles.}$$

If both cities were exposed to migration,

$$E[y_{i,1}|M,81] - E[y_{i,1}|M,79] = \beta_{81} - \beta_{79} + \delta \quad \text{for Miami}$$
$$E[y_{i,1}|L,81] - E[y_{i,1}|L,79] = \beta_{81} - \beta_{79} + \delta \quad \text{for Los Angeles.}$$

Only Miami experienced the migration (the "treatment"). The difference in differences that quantifies the result of the experiment is

$$\{E[y_{i,1}|M,81] - E[y_{i,1}|M,79]\} - \{E[y_{i,0}|L,81] - E[y_{i,0}|L,79]\} = \delta.$$

The author examined changes in employment rates and wages in the two cities over several years after the boatlift. The effects were surprisingly modest given the scale of the experiment in Miami.

One of the important issues in policy analysis concerns measurement of such treatment effects when the dummy variable results from an individual participation decision. In the clinical trial example given earlier, the control observations (it is assumed) do not know they they are in the control group. The treatment assignment is exogenous to the experiment. In contrast, in Keueger and Dale's study, the assignment to the treatment group, attended the elite college, is completely voluntary and determined by the individual. A crucial aspect of the analysis in this case is to accommodate the almost certain outcome that the "treatment dummy" might be measuring the latent motivation and initiative of the participants rather than the effect of the program itself. That is the main appeal of the natural experiment approach – it more closely (possibly exactly) replicates the exogenous treatment assignment of a clinical trial.[3] We will examine some of these cases in Chapters 8 and 18.

[3] See Angrist and Krueger (2001) and Angrist and Pischke (2010) for discussions of this approach.

where $t_1^* = 18$ and $t_2^* = 22$. To combine all three equations, we use

$$income = \beta_1 + \beta_2\, age + \gamma_1 d_1 + \delta_1 d_1\, age + \gamma_2 d_2 + \delta_2 d_2\, age + \varepsilon. \quad \textbf{(6-3)}$$

This relationship is the dashed function in Figure 6.2. The slopes in the three segments are β_2, $\beta_2 + \delta_1$, and $\beta_2 + \delta_1 + \delta_2$. To make the function **piecewise continuous**, we require that the segments join at the knots—that is,

$$\beta_1 + \beta_2 t_1^* = (\beta_1 + \gamma_1) + (\beta_2 + \delta_1) t_1^*$$

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

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

Doing likewise for the second and inserting these in (6-3), we obtain

$$income = \beta_1 + \beta_2\, age + \delta_1 d_1\, (age - t_1^*) + \delta_2 d_2\, (age - t_2^*) + \varepsilon.$$

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

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

and

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

We can test the hypothesis that the slope of the function is constant with the joint test of the two restrictions $\delta_1 = 0$ and $\delta_2 = 0$.

6.3 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

$$\begin{aligned} g(y) &= \beta_1 f_1(\mathbf{z}) + \beta_2 f_2(\mathbf{z}) + \cdots + \beta_K f_K(\mathbf{z}) + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon \\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon. \end{aligned} \quad \textbf{(6-4)}$$

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

6.3.1 FUNCTIONAL FORMS

A commonly used form of regression model is the **loglinear model,**

$$\ln y = \ln \alpha + \sum_k \beta_k \ln X_k + \varepsilon = \beta_1 + \sum_k \beta_k x_k + \varepsilon.$$

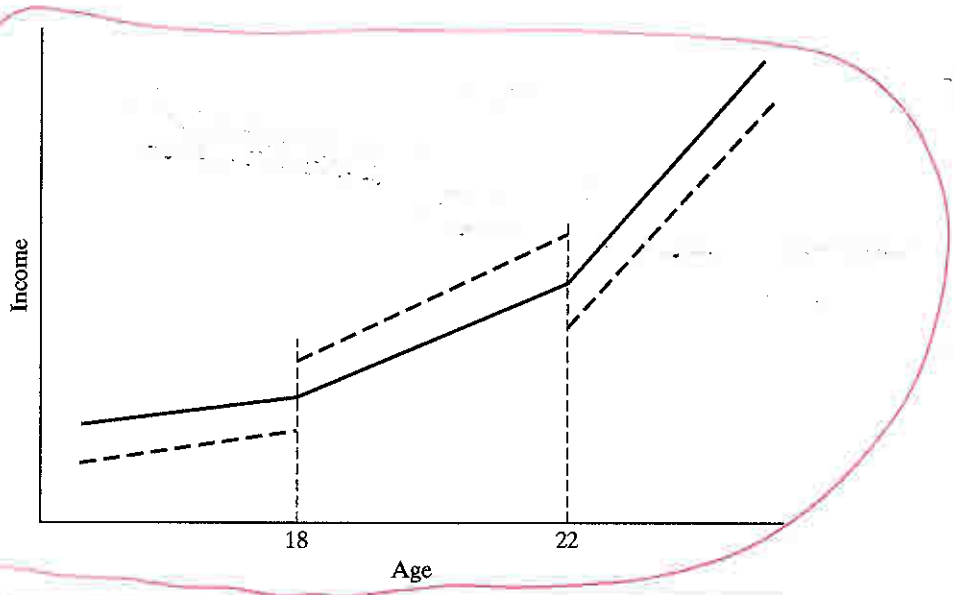

FIGURE 6.2 Spline Function.

6.3.1 Piecewise Linear REGRESSION

If one is examining income data for a large cross section of individuals of varying ages in a population, then certain patterns with regard to some age thresholds will be clearly evident. In particular, throughout the range of values of age, income will be rising, but the slope might change at some distinct milestones, for example, at age 18, when the typical individual graduates from high school, and at age 22, when he or she graduates from college. The **time profile** of income for the typical individual in this population might appear as in Figure 6.2. Based on the discussion in the preceding paragraph, we could fit such a regression model just by dividing the sample into three subsamples. However, this would neglect the continuity of the proposed function. The result would appear more like the dotted figure than the continuous function we had in mind. Restricted regression and what is known as a **spline** function can be used to achieve the desired effect.[1]

The function we wish to estimate is

$$E[income \mid age] = \begin{cases} \alpha^0 + \beta^0\, age & \text{if } age < 18, \\ \alpha^1 + \beta^1\, age & \text{if } age \geq 18 \text{ and } age < 22, \\ \alpha^2 + \beta^2\, age & \text{if } age \geq 22. \end{cases}$$

The threshold values, 18 and 22, are called **knots.** Let

$$d_1 = 1 \quad \text{if } age \geq t_1^*,$$
$$d_2 = 1 \quad \text{if } age \geq t_2^*,$$

[1] An important reference on this subject is Poirier (1974). An often-cited application appears in Garber and Poirier (1974).

where $t_1^* = 18$ and $t_2^* = 22$. To combine all three equations, we use

$$income = \beta_1 + \beta_2\, age + \gamma_1 d_1 + \delta_1 d_1\, age + \gamma_2 d_2 + \delta_2 d_2\, age + \varepsilon.$$

This relationship is the dashed function in Figure 6.2. The slopes in the three segments are β_2, $\beta_2 + \delta_1$, and $\beta_2 + \delta_1 + \delta_2$. To make the function **piecewise continuous,** we require that the segments join at the knots—that is,

$$\beta_1 + \beta_2 t_1^* = (\beta_1 + \gamma_1) + (\beta_2 + \delta_1) t_1^*$$

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

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

Doing likewise for the second and inserting these in (6-3), we obtain

$$income = \beta_1 + \beta_2\, age + \delta_1 d_1\, (age - t_1^*) + \delta_2 d_2\, (age - t_2^*) + \varepsilon.$$

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

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

and

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

We can test the hypothesis that the slope of the function is constant with the joint test of the two restrictions $\delta_1 = 0$ and $\delta_2 = 0$.

6.3 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

$$\begin{aligned} g(y) &= \beta_1 f_1(\mathbf{z}) + \beta_2 f_2(\mathbf{z}) + \cdots + \beta_K f_K(\mathbf{z}) + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon \\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon. \end{aligned} \tag{6-4}$$

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

6.3.1 FUNCTIONAL FORMS

A commonly used form of regression model is the **loglinear model,**

$$\ln y = \ln \alpha + \sum_k \beta_k \ln X_k + \varepsilon = \beta_1 + \sum_k \beta_k x_k + \varepsilon.$$

where $t_1^* = 18$ and $t_2^* = 22$. To combine all three equations, we use

$$income = \beta_1 + \beta_2\, age + \gamma_1 d_1 + \delta_1 d_1\, age + \gamma_2 d_2 + \delta_2 d_2\, age + \varepsilon. \quad \textbf{(6-3)}$$

This relationship is the dashed function in Figure 6.2. The slopes in the three segments are β_2, $\beta_2 + \delta_1$, and $\beta_2 + \delta_1 + \delta_2$. To make the function **piecewise continuous**, we require that the segments join at the knots—that is,

$$\beta_1 + \beta_2 t_1^* = (\beta_1 + \gamma_1) + (\beta_2 + \delta_1) t_1^*$$

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

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

Doing likewise for the second and inserting these in (6-3), we obtain

$$income = \beta_1 + \beta_2\, age + \delta_1 d_1\, (age - t_1^*) + \delta_2 d_2\, (age - t_2^*) + \varepsilon.$$

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

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

and

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

We can test the hypothesis that the slope of the function is constant with the joint test of the two restrictions $\delta_1 = 0$ and $\delta_2 = 0$.

6.3 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

$$\begin{aligned} g(y) &= \beta_1 f_1(\mathbf{z}) + \beta_2 f_2(\mathbf{z}) + \cdots + \beta_K f_K(\mathbf{z}) + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon \\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon. \end{aligned} \quad \textbf{(6-4)}$$

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

6.3.1 FUNCTIONAL FORMS

A commonly used form of regression model is the **loglinear model,**

$$\ln y = \ln \alpha + \sum_k \beta_k \ln X_k + \varepsilon = \beta_1 + \sum_k \beta_k x_k + \varepsilon.$$

In this model, the coefficients are elasticities:

$$\left(\frac{\partial y}{\partial x_k}\right)\left(\frac{x_k}{y}\right) = \frac{\partial \ln y}{\partial \ln x_k} = \beta_k. \tag{6-5}$$

In the loglinear equation, measured changes are in proportional or percentage terms; β_k measures the percentage change in y associated with a 1 percent change in x_k. This removes the units of measurement of the variables from consideration in using the regression model. An alternative approach sometimes taken is to measure the variables and associated changes in standard deviation units. If the data are "standardized" before estimation using $x_{ik}^* = (x_{ik} - \bar{x}_k)/s_k$ and likewise for y, then the least squares regression coefficients measure changes in standard deviation units rather than natural units or percentage terms. (Note that the constant term disappears from this regression.) It is not necessary actually to transform the data to produce these results; multiplying each least squares coefficient b_k in the original regression by s_k/s_y produces the same result.

A hybrid of the linear and loglinear models is the semilog equation

$$\ln y = \beta_1 + \beta_2 x + \varepsilon. \tag{6-6}$$

We used this form in the investment equation in Section 5.2,

$$\ln I_t = \beta_1 + \beta_2 (i_t - \Delta p_t) + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t,$$

where the log of investment is modeled in the levels of the real interest rate, the price level, and a time trend. In a semilog equation with a time trend such as this one, $d \ln I/dt = \beta_5$ is the average rate of growth of I. The estimated value of -0.00566 in Table 6.1 suggests that over the full estimation period, after accounting for all other factors, the average rate of growth of investment was -0.566 percent per year.

The coefficients in the semilog model are partial- or semi-elasticities; in (6-6), β_2 is $\partial \ln y/\partial x$. This is a natural form for models with dummy variables such as the earnings equation in Example 5.1. The coefficient on *Kids* of -0.35 suggests that all else equal, earnings are approximately 35 percent less when there are children in the household.

The quadratic earnings equation in Example 6.1 shows another use of nonlinearities in the variables. Using the results in Example 6.1, we find that for a woman with 12 years of schooling and children in the household, the age-earnings profile appears as in Figure 6.3. This figure suggests an important question in this framework. It is tempting to conclude that Figure 6.3 shows the earnings trajectory of a person at different ages, but that is not what the data provide. The model is based on a cross section, and what it displays is the earnings of different people of different ages. How this profile relates to the expected earnings path of one individual is a different, and complicated question.

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

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

In this model,

$$\frac{\partial E[D \mid S, W]}{\partial S} = \beta_2 + \beta_4 W,$$

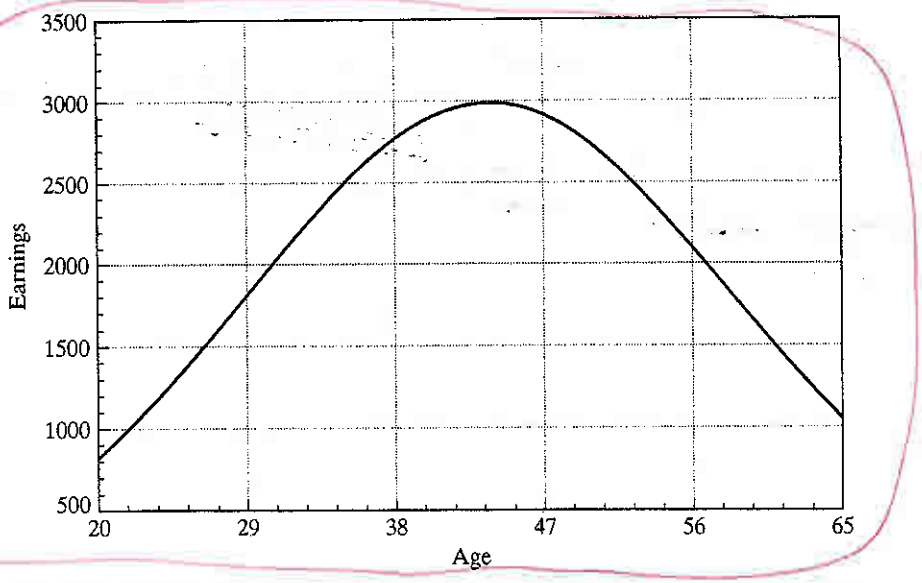

FIGURE 6.3 Age-Earnings Profile.

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 from

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

and similarly for $\partial E[D \mid S, W]/\partial W$. 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.

6.3.2 IDENTIFYING NONLINEARITY

If the functional form is not known a priori, then there are a few approaches that may 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 6.3 Functional Form for a Nonlinear Cost Function

In a celebrated study of economies of scale in the U.S. electric power industry, Nerlove (1963) analyzed the production costs of 145 American electricity generating companies. This study produced several innovations in microeconometrics. It was among the first major applications of statistical cost analysis. The theoretical development in Nerlove's study was the first to

6.3.4 IDENTIFYING NONLINEARITY

If the functional form is not known a priori, then there are a few approaches that may 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 6.6 Functional Form for a Nonlinear Cost Function

In a celebrated study of economies of scale in the U.S. electric power industry, Nerlove (1963) analyzed the production costs of 145 American electricity generating companies. This study produced several innovations in microeconometrics. It was among the first major applications of statistical cost analysis. The theoretical development in Nerlove's study was the first to show how the fundamental theory of duality between production and cost functions could be used to frame an econometric model. Finally, Nerlove employed several useful techniques to sharpen his basic model.

The focus of the paper was economies of scale, typically modeled as a characteristic of the production function. He chose a Cobb–Douglas function to model output as a function of capital, K, labor, L, and fuel, F:

$$Q = \alpha_0 K^{\alpha K} L^{\alpha L} F^{\alpha F} e^{\varepsilon i},$$

where Q is output and ε_i embodies the unmeasured differences across firms. The economies of scale parameter is $r = \alpha_K + \alpha_L + \alpha_F$. The value 1 indicates constant returns to scale. In this study, Nerlove investigated the widely accepted assumption that producers in this industry enjoyed substantial economies of scale. The production model is loglinear, so assuming that other conditions of the classical regression model are met, the four parameters could be estimated by least squares. However, he argued that the three factors could not be treated as exogenous variables. For a firm that optimizes by choosing its factors of production, the demand for fuel would be $F^* = F^*(Q, P_K, P_L, P_F)$ and likewise for labor and capital, so certainly the assumptions of the classical model are violated.

In the regulatory framework in place at the time, state commissions set rates and firms met the demand forthcoming at the regulated prices. Thus, it was argued that output (as well as the factor prices) could be viewed as exogenous to the firm and, based on an argument by Zellner, Kmenta, and Dreze (1966), Nerlove argued that at equilibrium, the *deviation* of costs from the long run optimum would be independent of output. (This has a testable implication which we will explore in Chapter 8.) Thus, the firm's objective was cost minimization subject to the constraint of the production function. This can be formulated as a Lagrangean problem,

$$\text{Min}_{K,L,F}\, P_K K + P_L L + P_F F + \lambda(Q - \alpha_0 K^{\alpha K} L^{\alpha L} F^{\alpha F}).$$

The solution to this minimization problem is the three factor demands and the multiplier (which measures marginal cost). Inserted back into total costs, this produces an (intrinsically linear) loglinear cost function,

$$P_K K + P_L L + P_F F = C(Q, P_K, P_L, P_F) = r A Q^{1/r} P_K^{\alpha K/r} P_L^{\alpha L/r} P_F^{\alpha F/r} e^{\varepsilon i/r},$$

or

$$\ln C = \beta_1 + \beta_q \ln Q + \beta_K \ln P_K + \beta_L \ln P_L + \beta_F \ln P_F + u_i, \qquad \textbf{(6-7)}$$

where $\beta_q = 1/(\alpha_K + \alpha_L + \alpha_F)$ is now the parameter of interest and $\beta_j = \alpha_j / r$, $j = K, L, F$. Thus, the duality between production and cost functions has been used to derive the estimating equation from first principles.

A complication remains. The cost parameters must sum to one; $\beta_K + \beta_L + \beta_F = 1$, so estimation must be done subject to this constraint. This restriction can be imposed by regressing $\ln(C/P_F)$ on a constant, $\ln Q$, $\ln(P_K/P_F)$, and $\ln(P_L/P_F)$. This first set of results appears at the top of Table 6.3.

Initial estimates of the parameters of the cost function are shown in the top row of Table 6.3. The hypothesis of constant returns to scale can be firmly rejected. The t ratio is $(0.721-1)/0.0174 = -16.03$, so we conclude that this estimate is significantly less than 1 or, by implication, r is significantly greater than 1. Note that the coefficient on the capital price is negative. In theory, this should equal α_K/r, which (unless the marginal product of capital is negative) should be positive. Nerlove attributed this to measurement error in the capital price variable. This seems plausible, but it carries with it the implication that the other coefficients are mismeasured as well. [Christensen and Greene's (1976) estimator of this model with these data produced a positive estimate. See Section 10.4.2.]

The striking pattern of the residuals shown in Figure 6.4 and some thought about the implied form of the production function suggested that something was missing from the model. In theory, the estimated model implies a continually declining average cost curve, which in turn implies persistent economies of scale at all levels of output. This conflicts with the textbook notion of a U-shaped average cost curve and appears implausible for the data. Note the three clusters of residuals in the figure. Two approaches were used to ~~analyze~~ extend the model.

Readers who attempt to replicate Nerlove's study should note that he used common (base 10) logs in his calculations, not natural logs. A practical tip: to convert a natural log to a common log, divide the former by $\log_e 10 = 2.302585093$. Also, however, although the first 145 rows of the data in Appendix Table F6.2 are accurately transcribed from the original study, the only regression listed in Table 6.3 that can be reproduced with these data is the first one. The results for Groups 1 – 5 in the table have been recomputed here and do not match Nerlove's results. Likewise, the results in Table 6.4 have been recomputed and do not match the original study. ~~The fact that the first full sample regression can be reproduced while none of the others can remains a puzzle.~~

In the context of the econometric model, the restriction has a testable implication by the definition in Chapter 5. But, the underlying economics require this restriction—it was used in deriving the cost function. Thus, it is unclear what is implied by a test of the restriction. Presumably, if the hypothesis of the restriction is rejected, the analysis should stop at that point, since without the restriction, the cost function is not a valid representation of the production function. We will encounter this conundrum again in another form in Chapter 10. Fortunately, in this instance, the hypothesis is not rejected. (It is in the application in Chapter 10.)

A Durbin–Watson test of correlation among the residuals (see Section 20.7) revealed to the author a substantial autocorrelation. Although normally used with time series data, the Durbin–Watson statistic and a test for "autocorrelation" can be a useful tool for determining the appropriate functional form in a cross-sectional model. To use this approach, it is necessary to sort the observations based on a variable of interest (output). Several clusters of residuals of the same sign suggested a need to reexamine the assumed functional form.

By sorting the sample into five groups on the basis of output and fitting separate regressions to each group, Nerlove fit a piecewise loglinear model. The results are given in the lower rows of Table 6.3, where the firms in the successive groups are progressively larger. The results are persuasive that the (log)linear cost function is inadequate. The output coefficient that rises toward and then crosses 1.0 is consistent with a U-shaped cost curve as surmised earlier.

A second approach was to expand the cost function to include a quadratic term in log output. This approach corresponds to a much more general model and produced the results given in Table 6.4. Again, a simple t test strongly suggests that increased generality is called for; $t = 0.051/0.00054 = 9.44$. The output elasticity in this quadratic model is $\beta_q + 2\gamma_{qq} \log Q$. There are economies of scale when this value is less than 1 and constant returns to scale when it equals one. Using the two values given in the table (0.152 and 0.0052, respectively), we find that this function does, indeed, produce a U-shaped average cost curve with minimum at $\ln Q = (1-0.152)/(2\times 0.051) = 8.31$, or $Q = 4079$, which is roughly in the middle of the range of outputs for Nerlove's sample of firms.

This study was updated by Christensen and Greene (1976). Using the same data but a more elaborate (translog) functional form and by simultaneously estimating the factor demands and the cost function, they found results broadly similar to Nerlove's. Their preferred functional form did suggest that Nerlove's generalized model in Table 6.4 did somewhat underestimate the range of outputs in which unit costs of production would continue to decline. They also redid the study using a sample of 123 firms from 1970, and found similar results. In the latter sample, however, it appeared that many firms had expanded rapidly enough to exhaust the available economies of scale. We will revisit the 1970 data set in a study of production costs in ~~Examples 14.5 and 14.8 and efficiency in Application 16.2.~~ Chapters 10 and 18.

The preceding example illustrates three useful tools in identifying and dealing with unspecified nonlinearity: analysis of residuals, the use of piecewise linear regression, and the use of polynomials to approximate the unknown regression function.

[8] Nerlove inadvertently measured economies of scale from this function as $1/(\beta_q + \delta \log Q)$, where β_q and δ are the coefficients on $\log Q$ and $\log^2 Q$. The correct expression would have been $1/[\partial \log C/\partial \log Q] = 1/[\beta_q + 2\delta \log Q]$. This slip was periodically rediscovered in several later papers.

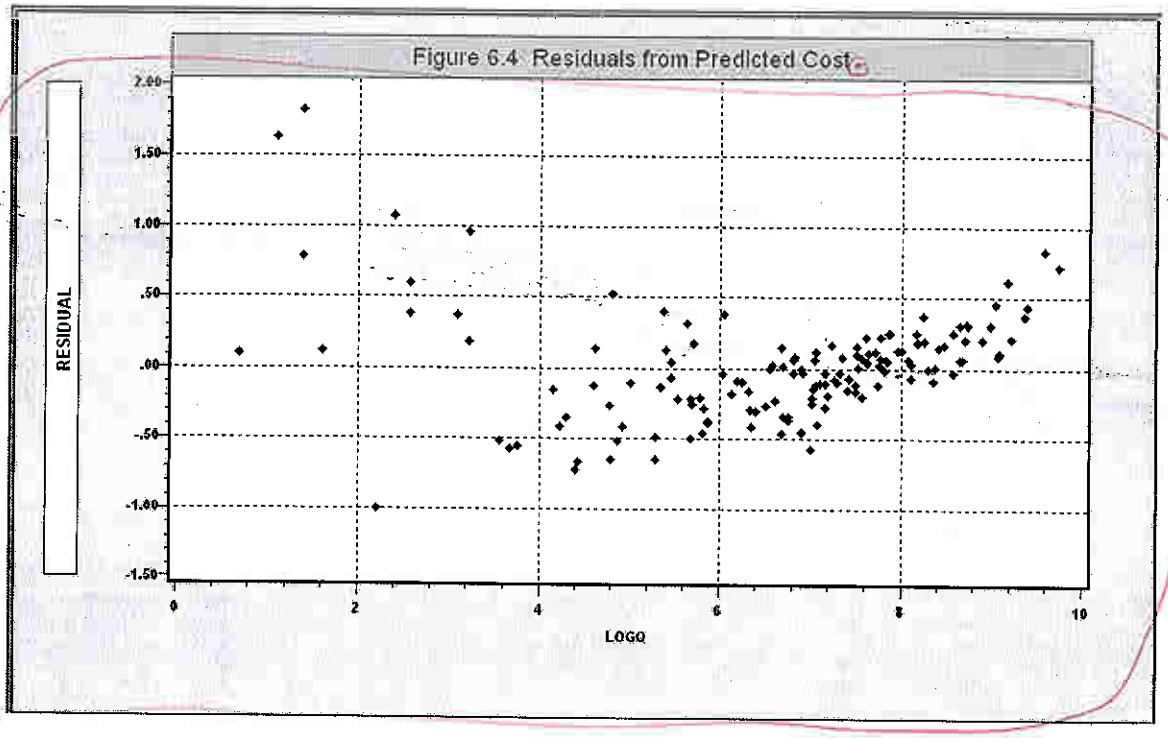

TABLE 6.4 Cobb-Douglas Cost Functions
(standard errors in parentheses)

	log Q	log P_L – log P_F	logP_K – log P_F	R^2
All firms	0.721	0.593	-0.0074	0.932
	(0.0174)	(0.205)	(0.191)	
Group 1	0.400	0.615	-0.081	0.513
Group 2	0.658	0.094	0.378	0.633
Group 3	0.938	0.402	0.250	0.573
Group 4	0.912	0.507	0.093	0.826
Group 5	1.044	0.603	-0.289	0.921

Note minus signs

TABLE 6.5 Log-Quadratic Cost Function
(standard errors in parentheses)

	log Q	$\log^2$ Q	log P_L – log P_F	logP_K – log P_F	R^2
All firms	0.152	0.051	0.481	0.074	0.96
	(0.062)	(0.0054)	(0.161)	(0.150)	