6
FUNCTIONAL FORM AND STRUCTURAL CHANGE

6.1 INTRODUCTION

This chapter will complete our 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 process 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 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 5.2, 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.

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

Example 6.1 Dummy Variable in an Earnings Equation

Table 6.1 following reproduces the estimated earnings equation in Example 5.2. 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.

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 $E[y|x,d]$ asociated with the change in d is

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

Example 6.2 Value of a Signature

In Example 4.10 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. 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

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

TABLE 6.2 Estimated Equations for Log Price

$\ln price =$ [illegible] $\beta_2 \ln Area + \beta_3 aspect\ ratio + \beta_4 signature + \varepsilon$

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	
Variable	***Coefficient***	***Standard Error***	***t***	***Coefficient***	***Standard 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

(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

$$100\% \times (\Delta E[Price]/Price) = 100\%[\exp(1.26) - 1] = 252\%.$$

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 0.4417. For the percentage difference of 252%, we have an estimated standard error of 44.17%.

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 1,228.69 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 feature 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.

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 23.3.5.) 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

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

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:

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

Using the results from Section 3.3 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.

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 percent 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 percent 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 percent.

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.

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

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 6.4 Analysis of Covariance

The data in Appendix Table F6.1 were used in a study of efficiency in production of airline services in Greene (2007a). 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 observations, from which we have selected six for the examples to follow. We will fit a cost equation of the form

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

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

$$\begin{aligned} \ln C_{i,t} &= 13.56 + 0.8866 \ln Q_{i,t} + 0.0\text{[illegible]} \ln^2 Q_{i,t} + 0.1281 \ln P_{fi,t} - 0.8855\, LF_{i,t} \\ &\quad + \text{time effects} + \text{firm effects}. \end{aligned}$$

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).

We are interested in whether the firm effects, the time effects, both, or neither are statistically significant. Table 6.3 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 of 1.84, but perhaps less so than Figure 6.1 might have suggested. In the absence of the

TABLE 6.3 *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]

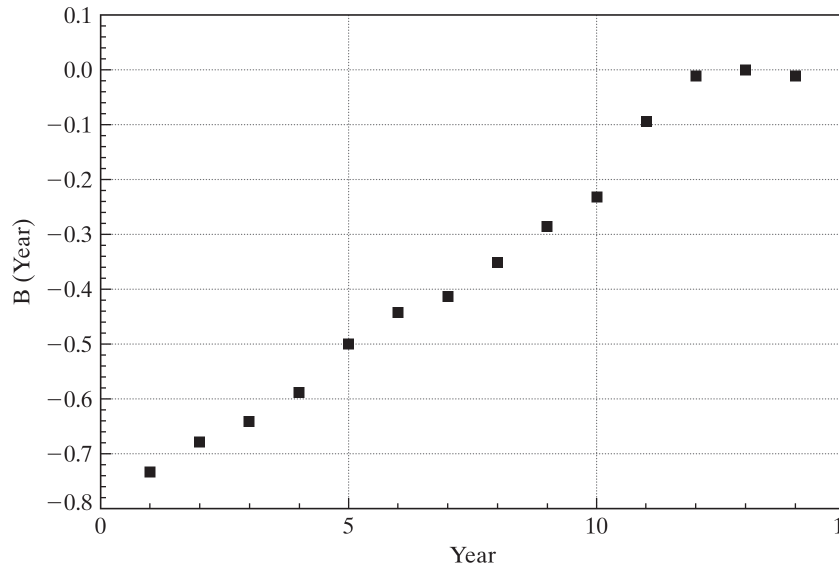

FIGURE 6.1 Estimated Year Dummy Variable Coefficients.

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

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

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, $\textit{income} = \beta_1 + \beta_2\, \textit{age} + \beta_3 E + \varepsilon$. 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

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

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 preversus 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 $\mathbf{x}$) 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 of the Tennessee STAR experiment in which school performance measures were observed before and after a policy dictated a 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, for example, school performance, pre- to postintervention; $b_2 = \bar{y}_1 - \bar{y}_0$.

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 de[illegible] 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 cholesterol 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 posttreatment. The impact is measured by the difference in differences,

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

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).

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 posttreatment observation on a group, where the outcome variable is y, with

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

In this model, T_t is a dummy variable that is zero in the pretreatment 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

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

[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).

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,

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

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

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 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 percent. The author examined the impact of this abrupt change in labor market conditions on wages and employment for nonimmigrants. 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

$$(Y_i) = 1 \text{ if they are unemployed and 0 if they are employed.}$$

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,

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

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,

$$\begin{aligned} E[y_i|\mathrm{M}, 79] &= \beta_{79} + \gamma_{\mathrm{M}} \quad \text{and} \quad E[y_i|\mathrm{M}, 81] = \beta_{81} + \gamma_{\mathrm{M}} + \delta \quad \text{for Miami,} \\ E[y_i|\mathrm{L}, 79] &= \beta_{79} + \gamma_{\mathrm{L}} \quad \text{and} \quad E[y_i|\mathrm{L}, 81] = \beta_{81} + \gamma_{\mathrm{L}} \qquad \text{for Los Angeles.} \end{aligned}$$

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

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

If both cities were exposed to migration,

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

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

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

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.

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

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

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

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

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

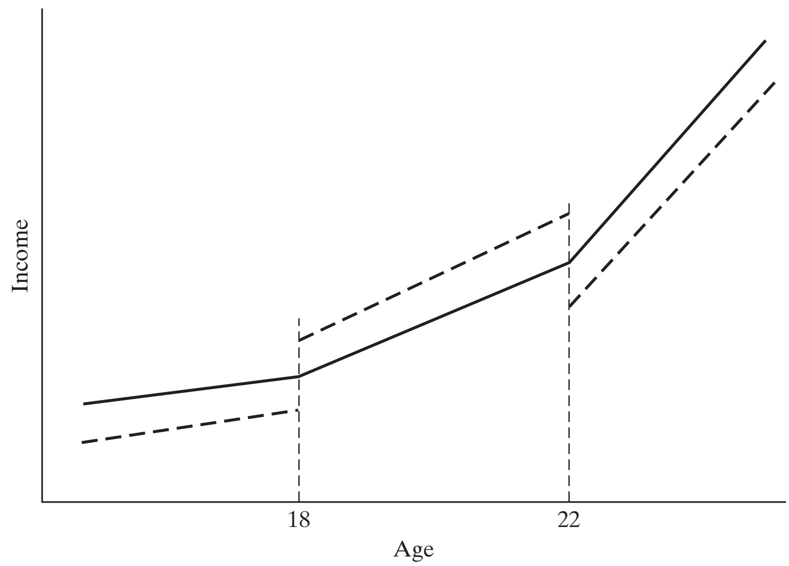

FIGURE 6.2 Spline Function.

regression and what is known as a **spline** function can be used to achieve the desired effect.[4]

The function we wish to estimate is

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

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

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

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

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

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

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 $\delta_2 = 0$.

6.3.2 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 $x_{ik}^* = (x_{ik} - \overline{x}_k)/s_k$ 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 5.2.2,

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

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 5.2 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.

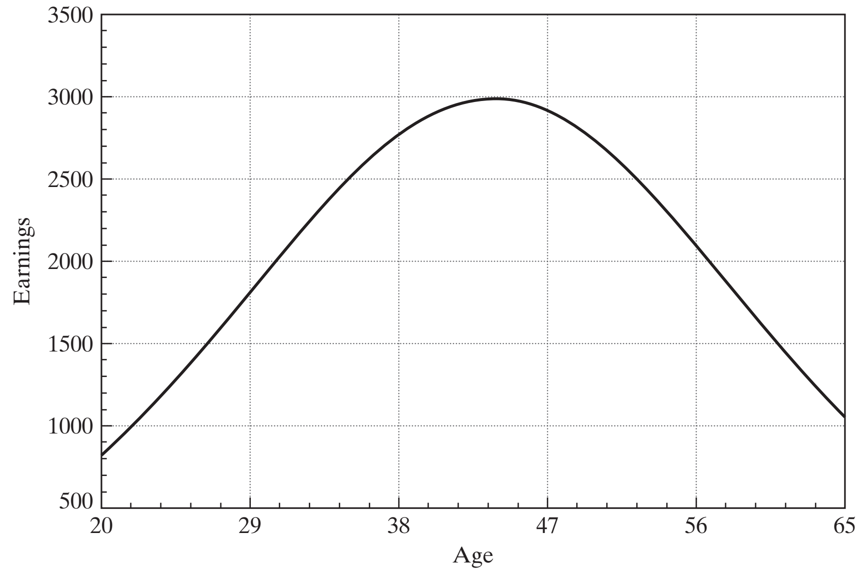

FIGURE 6.3 Age-Earnings Profile.

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 5.2. 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.

6.3.3 INTERACTION EFFECTS

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,

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

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

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

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.

6.3.4 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,

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

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

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

	log Q	log P_L − log P_F	log P_K − log P_F	R^2
All firms	0.721	0.593	−0.0085	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

linear) loglinear cost function,

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

or

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

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.[5] 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 6.4.[6]

Initial estimates of the parameters of the cost function are shown in the top row of Table 6.4. 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.[7] In theory, the estimated model implies a continually declining average cost curve,

[5]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.)

[6]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.

[7]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.

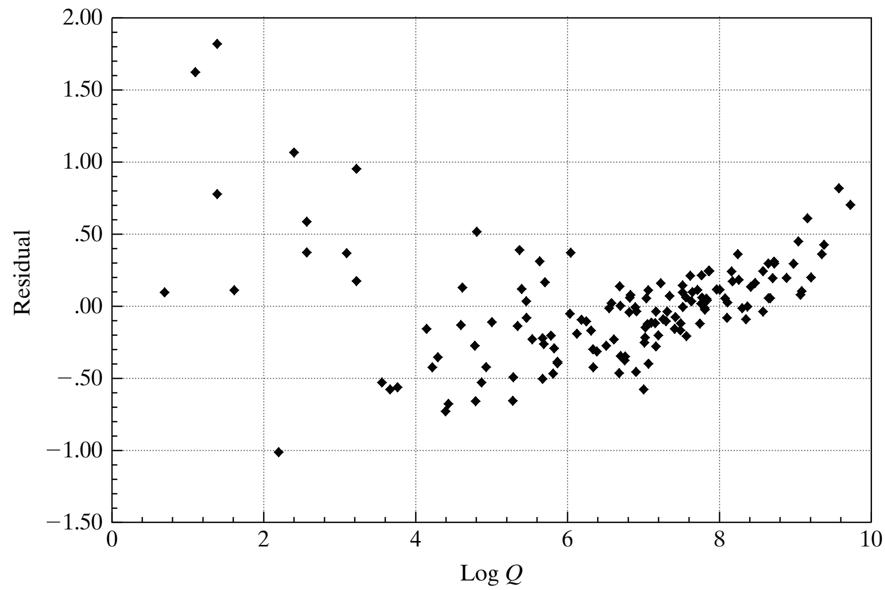

FIGURE 6.4 Residuals from Predicted Cost.

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 extend the model.

By sorting the sample into five groups of 29 firms 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 6.4, 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 6.5. 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 $\beta_q + 2\gamma_{qq} \log Q$.[8] There are economies of scale when this value is less than 1 and constant returns to scale when it equals 1. 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 6.5 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.

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

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

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 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.

6.3.5 INTRINSICALLY LINEAR MODELS

The loglinear model illustrates an intermediate case of a nonlinear regression model. The equation is **intrinsically linear**, however. By taking logs of $Y_i = \alpha X_i^{\beta_2} e^{\varepsilon_i}$, we obtain

$$\ln Y_i = \ln \alpha + \beta_2 \ln X_i + \varepsilon_i$$

or

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

Although this equation is linear in most respects, something has changed in that it is no longer linear in α. Written in terms of β_1, we obtain a fully linear model. But that may not be the form of interest. Nothing is lost, of course, since β_1 is just $\ln \alpha$. If β_1 can be estimated, then an obvious estimator of α is suggested, $\hat{\alpha} = \exp(b_1)$.

This fact leads us to a useful aspect of intrinsically linear models; they have an "invariance property." Using the nonlinear least squares procedure described in the next chapter, we could estimate α and β_2 directly by minimizing the sum of squares function;

$$\text{Minimize with respect to } (\alpha, \beta_2) : S(\alpha, \beta_2) = \sum_{i=1}^{n} (\ln Y_i - \ln \alpha - \beta_2 \ln X_i)^2. \tag{6-8}$$

This is a complicated mathematical problem because of the appearance the term $\ln \alpha$. However, the equivalent linear least squares problem,

$$\text{Minimize with respect to } (\beta_1, \beta_2) : S(\beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2, \tag{6-9}$$

is simple to solve with the least squares estimator we have used up to this point. The invariance feature that applies is that the two sets of results will be numerically identical; we will get the identical result from estimating α using (6-8) and from using $\exp(\beta_1)$ from (6-9). By exploiting this result, we can broaden the definition of linearity and include some additional cases that might otherwise be quite complex.

TABLE 6.6 Estimates of the Regression in a Gamma Model: Least Squares versus Maximum Likelihood

	β		ρ	
	Estimate	*Standard Error*	*Estimate*	*Standard Error*
Least squares	−1.708	8.689	2.426	1.592
Maximum likelihood	−4.719	2.345	3.151	0.794

DEFINITION 6.1 Intrinsic Linearity
In the classical linear regression model, if the K parameters $\beta_1, \beta_2, \ldots, \beta_K$ can be written as K one-to-one, possibly nonlinear functions of a set of K underlying parameters $\theta_1, \theta_2, \ldots, \theta_K$, then the model is intrinsically linear in θ.

Example 6.7 Intrinsically Linear Regression

In Section 14.6.4, we will estimate by maximum likelihood the parameters of the model

$$f(y \mid \beta, x) = \frac{(\beta + x)^{-\rho}}{\Gamma(\rho)} y^{\rho-1} e^{-y/(\beta+x)}.$$

In this model, $E[y \mid x] = (\beta\rho) + \rho x$, which suggests another way that we might estimate the two parameters. This function is an intrinsically linear regression model, $E[y \mid x] = \beta_1 + \beta_2 x$, in which $\beta_1 = \beta\rho$ and $\beta_2 = \rho$. We can estimate the parameters by least squares and then retrieve the estimate of β using b_1/b_2. Because this value is a nonlinear function of the estimated parameters, we use the delta method to estimate the standard error. Using the data from that example,[9] the least squares estimates of β_1 and β_2 (with standard errors in parentheses) are −4.1431 (23.734) and 2.4261 (1.5915). The estimated covariance is −36.979. The estimate of β is $-4.1431/2.4261 = -1.7077$. We estimate the sampling variance of $\hat{\beta}$ with

$$\begin{aligned}\text{Est. Var}[\hat{\beta}] &= \left(\frac{\partial \hat{\beta}}{\partial b_1}\right)^2 \widehat{\text{Var}}[b_1] + \left(\frac{\partial \hat{\beta}}{\partial b_2}\right)^2 \widehat{\text{Var}}[b_2] + 2\left(\frac{\partial \hat{\beta}}{\partial b_1}\right)\left(\frac{\partial \hat{\beta}}{\partial b_2}\right)\widehat{\text{Cov}}[b_1, b_2] \\ &= 8.6889^2.\end{aligned}$$

Table 6.6 compares the least squares and maximum likelihood estimates of the parameters. The lower standard errors for the maximum likelihood estimates result from the inefficient (equal) weighting given to the observations by the least squares procedure. The gamma distribution is highly skewed. In addition, we know from our results in Appendix C that this distribution is an exponential family. We found for the gamma distribution that the sufficient statistics for this density were $\Sigma_i y_i$ and $\Sigma_i \ln y_i$. The least squares estimator does not use the second of these, whereas an efficient estimator will.

The emphasis in intrinsic linearity is on "one to one." If the conditions are met, then the model can be estimated in terms of the functions $\beta_1, \ldots, \beta_K$, and the underlying parameters derived after these are estimated. The one-to-one correspondence is an **identification condition**. If the condition is met, then the underlying parameters of the

[9]The data are given in Appendix Table FC.1.

regression ($\boldsymbol{\theta}$) are said to be **exactly identified** in terms of the parameters of the linear model $\boldsymbol{\beta}$. An excellent example is provided by Kmenta (1986, p. 515, and 1967).

Example 6.8 CES Production Function

The constant elasticity of substitution production function may be written

$$\ln y = \ln \gamma - \frac{\nu}{\rho} \ln[\delta K^{-\rho} + (1-\delta)L^{-\rho}] + \varepsilon. \tag{6-10}$$

A Taylor series approximation to this function around the point $\rho = 0$ is

$$\begin{aligned} \ln y &= \ln \gamma + \nu\delta \ln K + \nu(1-\delta) \ln L + \rho\nu\delta(1-\delta)\left\{-\tfrac{1}{2}[\ln K - \ln L]^2\right\} + \varepsilon' \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon', \end{aligned} \tag{6-11}$$

where $x_1 = 1$, $x_2 = \ln K$, $x_3 = \ln L$, $x_4 = -\frac{1}{2}\ln^2(K/L)$, and the transformations are

$$\begin{aligned} &\beta_1 = \ln \gamma, \quad \beta_2 = \nu\delta, \quad \beta_3 = \nu(1-\delta), \quad \beta_4 = \rho\nu\delta(1-\delta), \\ &\gamma = e^{\beta_1}, \quad \delta = \beta_2/(\beta_2 + \beta_3), \quad \nu = \beta_2 + \beta_3, \quad \rho = \beta_4(\beta_2+\beta_3)/(\beta_2\beta_3). \end{aligned} \tag{6-12}$$

Estimates of β_1, β_2, β_3, and β_4 can be computed by least squares. The estimates of γ, δ, ν, and ρ obtained by the second row of (6-12) are the same as those we would obtain had we found the nonlinear least squares estimates of (6-11) directly. As Kmenta shows, however, they are not the same as the nonlinear least squares estimates of (6-10) due to the use of the Taylor series approximation to get to (6-11). We would use the delta method to construct the estimated asymptotic covariance matrix for the estimates of $\boldsymbol{\theta}' = [\gamma, \delta, \nu, \rho]$. The derivatives matrix is

$$\mathbf{C} = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}'} = \begin{bmatrix} e^{\beta_1} & 0 & 0 & 0 \\ 0 & \beta_3/(\beta_2+\beta_3)^2 & -\beta_2/(\beta_2+\beta_3)^2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -\beta_3\beta_4/\left(\beta_2^2\beta_3\right) & -\beta_2\beta_4/\left(\beta_2\beta_3^2\right) & (\beta_2+\beta_3)/(\beta_2\beta_3) \end{bmatrix}.$$

The estimated covariance matrix for $\hat{\boldsymbol{\theta}}$ is $\hat{\mathbf{C}}\,[s^2(\mathbf{X}'\mathbf{X})^{-1}]\hat{\mathbf{C}}'$.

Not all models of the form

$$y_i = \beta_1(\theta)x_{i1} + \beta_2(\theta)x_{i2} + \cdots + \beta_K(\theta)x_{ik} + \varepsilon_i \tag{6-13}$$

are intrinsically linear. Recall that the condition that the functions be one to one (i.e., that the parameters be exactly identified) was required. For example,

$$y_i = \alpha + \beta x_{i1} + \gamma x_{i2} + \beta\gamma x_{i3} + \varepsilon_i$$

is nonlinear. The reason is that if we write it in the form of (6-13), we fail to account for the condition that β_4 equals $\beta_2\beta_3$, which is a **nonlinear restriction**. In this model, the three parameters α, β, and γ are **overidentified** in terms of the four parameters β_1, β_2, β_3, and β_4. Unrestricted least squares estimates of β_2, β_3, and β_4 can be used to obtain two estimates of each of the underlying parameters, and there is no assurance that these will be the same. Models that are not intrinsically linear are treated in Chapter 7.

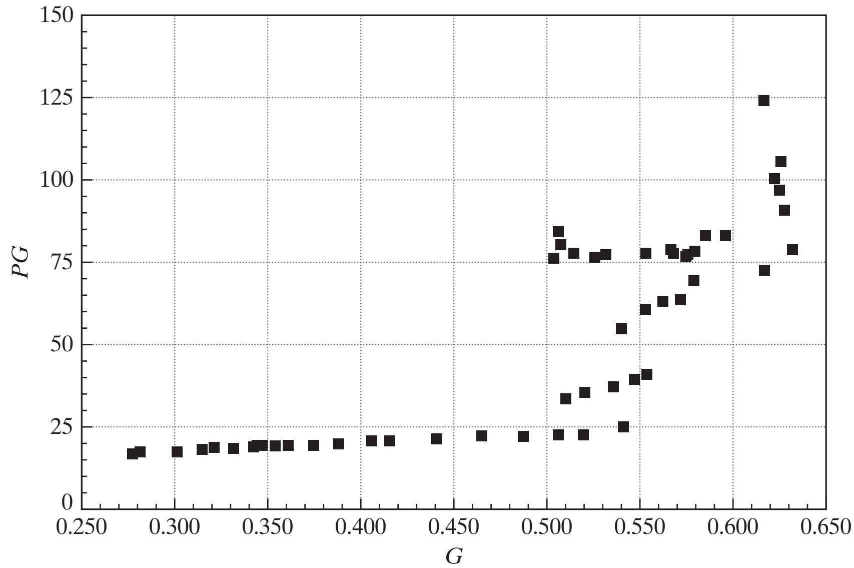

FIGURE 6.5 Gasoline Price and Per Capita Consumption, 1953–2004.

6.4 MODELING AND TESTING FOR A STRUCTURAL BREAK

One of the more common applications of the F test is in tests of **structural change**.[10] In specifying a regression model, we assume that its assumptions apply to all the observations in our sample. It is straightforward, however, to test the hypothesis that some or all of the regression coefficients are different in different subsets of the data. To analyze a number of examples, we will revisit the data on the U.S. gasoline market that we examined in Examples 2.3, 4.2, 4.4, 4.8 and 4.9. As Figure 6.5 suggests, this market behaved in predictable, unremarkable fashion prior to the oil shock of 1973 and was quite volatile thereafter. The large jumps in price in 1973 and 1980 are clearly visible, as is the much greater variability in consumption.[11] It seems unlikely that the same regression model would apply to both periods.

6.4.1 DIFFERENT PARAMETER VECTORS

The gasoline consumption data span two very different periods. Up to 1973, fuel was plentiful and world prices for gasoline had been stable or falling for at least two decades. The embargo of 1973 marked a transition in this market, marked by shortages, rising prices, and intermittent turmoil. It is possible that the entire relationship described by our regression model changed in 1974. To test this as a hypothesis, we could proceed as follows: Denote the first 21 years of the data in $\mathbf{y}$ and $\mathbf{X}$ as $\mathbf{y}_1$ and $\mathbf{X}_1$ and the remaining

[10] This test is often labeled a **Chow test,** in reference to Chow (1960).

[11] The observed data will doubtless reveal similar disruption in 2006.

years as $\mathbf{y}_2$ and $\mathbf{X}_2$. An unrestricted regression that allows the coefficients to be different in the two periods is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}. \tag{6-14}$$

Denoting the data matrices as $\mathbf{y}$ and $\mathbf{X}$, we find that the unrestricted least squares estimator is

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \tag{6-15}$$

which is least squares applied to the two equations separately. Therefore, the total sum of squared residuals from this regression will be the sum of the two residual sums of squares from the two separate regressions:

$$\mathbf{e}'\mathbf{e} = \mathbf{e}_1'\mathbf{e}_1 + \mathbf{e}_2'\mathbf{e}_2.$$

The restricted coefficient vector can be obtained in two ways. Formally, the restriction $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ is $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, where $\mathbf{R} = [\mathbf{I} : -\mathbf{I}]$ and $\mathbf{q} = \mathbf{0}$. The general result given earlier can be applied directly. An easier way to proceed is to build the restriction directly into the model. If the two coefficient vectors are the same, then (6-14) may be written

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix},$$

and the restricted estimator can be obtained simply by stacking the data and estimating a single regression. The residual sum of squares from this restricted regression, $\mathbf{e}_*'\mathbf{e}_*$, then forms the basis for the test. The test statistic is then given in (5-16), where J, the number of restrictions, is the number of columns in $\mathbf{X}_2$ and the denominator degrees of freedom is $n_1 + n_2 - 2k$.

6.4.2 INSUFFICIENT OBSERVATIONS

In some circumstances, the data series are not long enough to estimate one or the other of the separate regressions for a test of structural change. For example, one might surmise that consumers took a year or two to adjust to the turmoil of the two oil price shocks in 1973 and 1979, but that the market never actually fundamentally changed or that it only changed temporarily. We might consider the same test as before, but now only single out the four years 1974, 1975, 1980, and 1981 for special treatment. Because there are six coefficients to estimate but only four observations, it is not possible to fit the two separate models. Fisher (1970) has shown that in such a circumstance, a valid way to proceed is as follows:

1. Estimate the regression, using the full data set, and compute the restricted sum of squared residuals, $\mathbf{e}_*'\mathbf{e}_*$.
2. Use the longer (adequate) subperiod (n_1 observations) to estimate the regression, and compute the unrestricted sum of squares, $\mathbf{e}_1'\mathbf{e}_1$. This latter computation is done assuming that with only $n_2 < K$ observations, we could obtain a perfect fit and thus contribute zero to the sum of squares.

3. The F statistic is then computed, using

$$F[n_2, n_1 - K] = \frac{(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}_1'\mathbf{e}_1)/n_2}{\mathbf{e}_1'\mathbf{e}_1/(n_1 - K)}. \tag{6-16}$$

Note that the numerator degrees of freedom is n_2, not K.[12] This test has been labeled the Chow **predictive test** because it is equivalent to extending the restricted model to the shorter subperiod and basing the test on the prediction errors of the model in this latter period.

6.4.3 CHANGE IN A SUBSET OF COEFFICIENTS

The general formulation previously suggested lends itself to many variations that allow a wide range of possible tests. Some important particular cases are suggested by our gasoline market data. One possible description of the market is that after the oil shock of 1973, Americans simply reduced their consumption of gasoline by a fixed proportion, but other relationships in the market, such as the income elasticity, remained unchanged. This case would translate to a simple shift downward of the loglinear regression model or a reduction only in the constant term. Thus, the unrestricted equation has separate coefficients in the two periods, while the restricted equation is a pooled regression with separate constant terms. The regressor matrices for these two cases would be of the form

$$(\text{unrestricted})\ \mathbf{X}_U = \begin{bmatrix} \mathbf{i} & \mathbf{0} & \mathbf{W}_{\text{pre73}} & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & \mathbf{W}_{\text{post73}} \end{bmatrix}$$

and

$$(\text{restricted})\ \mathbf{X}_R = \begin{bmatrix} \mathbf{i} & \mathbf{0} & \mathbf{W}_{\text{pre73}} \\ \mathbf{0} & \mathbf{i} & \mathbf{W}_{\text{post73}} \end{bmatrix}.$$

The first two columns of $\mathbf{X}_U$ are dummy variables that indicate the subperiod in which the observation falls.

Another possibility is that the constant and one or more of the slope coefficients changed, but the remaining parameters remained the same. The results in Example 6.9 suggest that the constant term and the price and income elasticities changed much more than the cross-price elasticities and the time trend. The Chow test for this type of restriction looks very much like the one for the change in the constant term alone. Let $\mathbf{Z}$ denote the variables whose coefficients are believed to have changed, and let $\mathbf{W}$ denote the variables whose coefficients are thought to have remained constant. Then, the regressor matrix in the constrained regression would appear as

$$\mathbf{X} = \begin{bmatrix} \mathbf{i}_{\text{pre}} & \mathbf{Z}_{\text{pre}} & \mathbf{0} & \mathbf{0} & \mathbf{W}_{\text{pre}} \\ \mathbf{0} & \mathbf{0} & \mathbf{i}_{\text{post}} & \mathbf{Z}_{\text{post}} & \mathbf{W}_{\text{post}} \end{bmatrix}. \tag{6-17}$$

As before, the unrestricted coefficient vector is the combination of the two separate regressions.

[12]One way to view this is that only $n_2 < K$ coefficients are needed to obtain this perfect fit.

6.4.4 TESTS OF STRUCTURAL BREAK WITH UNEQUAL VARIANCES

An important assumption made in using the Chow test is that the disturbance variance is the same in both (or all) regressions. In the restricted model, if this is not true, the first n_1 elements of $\boldsymbol{\varepsilon}$ have variance σ_1^2, whereas the next n_2 have variance σ_2^2, and so on. The restricted model is, therefore, heteroscedastic, and our results for the classical regression model no longer apply. As analyzed by Schmidt and Sickles (1977), Ohtani and Toyoda (1985), and Toyoda and Ohtani (1986), it is quite likely that the actual probability of a type I error will be larger than the significance level we have chosen. (That is, we shall regard as large an F statistic that is actually less than the *appropriate* but unknown critical value.) Precisely how severe this effect is going to be will depend on the data and the extent to which the variances differ, in ways that are not likely to be obvious.

If the sample size is reasonably large, then we have a test that is valid whether or not the disturbance variances are the same. Suppose that $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are two consistent and asymptotically normally distributed estimators of a parameter based on independent samples,[13] with asymptotic covariance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$. Then, under the null hypothesis that the true parameters are the same,

$$\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2 \text{ has mean } \mathbf{0} \text{ and asymptotic covariance matrix } \mathbf{V}_1 + \mathbf{V}_2.$$

Under the null hypothesis, the Wald statistic,

$$W = (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)'(\hat{\mathbf{V}}_1 + \hat{\mathbf{V}}_2)^{-1}(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2), \tag{6-18}$$

has a limiting chi-squared distribution with K degrees of freedom. A test that the difference between the parameters is zero can be based on this statistic.[14] It is straightforward to apply this to our test of common parameter vectors in our regressions. Large values of the statistic lead us to reject the hypothesis.

In a small or moderately sized sample, the Wald test has the unfortunate property that the probability of a type I error is persistently larger than the critical level we use to carry it out. (That is, we shall too frequently reject the null hypothesis that the parameters are the same in the subsamples.) We should be using a larger critical value. Ohtani and Kobayashi (1986) have devised a "bounds" test that gives a partial remedy for the problem.[15]

It has been observed that the size of the **Wald test** may differ from what we have assumed, and that the deviation would be a function of the alternative hypothesis. There are two general settings in which a test of this sort might be of interest. For comparing two possibly different populations—such as the labor supply equations for men versus women—not much more can be said about the suggested statistic in the absence of specific information about the alternative hypothesis. But a great deal of work on this type of statistic has been done in the time-series context. In this instance, the nature of the alternative is rather more clearly defined.

[13]Without the required independence, this test and several similar ones will fail completely. The problem becomes a variant of the famous Behrens–Fisher problem.

[14]See Andrews and Fair (1988). The true size of this suggested test is uncertain. It depends on the nature of the alternative. If the variances are radically different, the assumed critical values might be somewhat unreliable.

[15]See also Kobayashi (1986). An alternative, somewhat more cumbersome test is proposed by Jayatissa (1977). Further discussion is given in Thursby (1982).

Example 6.9 Structural Break in the Gasoline Market

Figure 6.5 shows a plot of prices and quantities in the U.S. gasoline market from 1953 to 2004. The first 21 points are the layer at the bottom of the figure and suggest an orderly market. The remainder clearly reflect the subsequent turmoil in this market.

We will use the Chow tests described to examine this market. The model we will examine is the one suggested in Example 2.3, with the addition of a time trend:

$$\ln(G/Pop)_t = \beta_1 + \beta_2 \ln(Income/Pop)_t + \beta_3 \ln PG_t + \beta_4 \ln PNC_t + \beta_5 \ln PUC_t + \beta_6 t + \varepsilon_t.$$

The three prices in the equation are for G, new cars and used cars. *Income/Pop* is per capita Income, and *G/Pop* is per capita gasoline consumption. The time trend is computed as $t =$ Year -1952, so in the first period $t = 1$. Regression results for four functional forms are shown in Table 6.7. Using the data for the entire sample, 1953 to 2004, and for the two subperiods, 1953 to 1973 and 1974 to 2004, we obtain the three estimated regressions in the first and last two columns. The F statistic for testing the restriction that the coefficients in the two equations are the same is

$$F[6, 40] = \frac{(0.101997 - (0.00202244 + 0.007127899))/6}{(0.00202244 + 0.007127899)/(21 + 31 - 12)} = 67.645.$$

The tabled critical value is 2.336, so, consistent with our expectations, we would reject the hypothesis that the coefficient vectors are the same in the two periods. Using the full set of 52 observations to fit the model, the sum of squares is $\mathbf{e}^{*\prime}\mathbf{e}^{*} = 0.101997$. When the $n_2 = 4$ observations for 1974, 1975, 1980, and 1981 are removed from the sample, the sum of squares falls to $\mathbf{e}'\mathbf{e} = 0.0973936$. The F statistic is 0.496. Because the tabled critical value for $F[4, 48-6]$ is 2.594, we would not reject the hypothesis of stability. The conclusion to this point would be that although something has surely changed in the market, the hypothesis of a temporary disequilibrium seems not to be an adequate explanation.

An alternative way to compute this statistic might be more convenient. Consider the original arrangement, with all 52 observations. We now add to this regression four binary variables, Y1974, Y1975, Y1980, and Y1981. Each of these takes the value one in the single year indicated and zero in all 51 remaining years. We then compute the regression with the original six variables and these four additional dummy variables. The sum of squared residuals in this regression is 0.0973936 (precisely the same as when the four observations are deleted from the sample—see Exercise 7 in Chapter 3), so the F statistic for testing the joint hypothesis that the four coefficients are zero is

$$F[4, 42] = \frac{(0.101997 - 0.0973936)/4}{0.0973936/(52 - 6 - 4)} = 0.496$$

once again. (See Section 6.4.2 for discussion of this test.)

TABLE 6.7 Gasoline Consumption Functions

Coefficients	*1953–2004*	*Pooled*	*Preshock*	*Postshock*
Constant	−26.6787	−24.9009	−22.1647	
Constant		−24.8167		−15.3283
ln *Income/Pop*	1.6250	1.4562	0.8482	0.3739
ln *PG*	−0.05392	−0.1132	−0.03227	−0.1240
ln *PNC*	−0.08343	−0.1044	0.6988	−0.001146
ln *PUC*	−0.08467	−0.08646	−0.2905	−0.02167
Year	−0.01393	−0.009232	0.01006	0.004492
R^2	0.9649	0.9683	0.9975	0.9529
Standard error	0.04709	0.04524	0.01161	0.01689
Sum of squares	0.101997	0.092082	0.00202244	0.007127899

The F statistic for testing the restriction that the coefficients in the two equations are the same apart from the constant term is based on the last three sets of results in the table:

$$F[5, 40] = \frac{(0.092082 - (0.00202244 + 0.007127899))/5}{(0.00202244 + 0.007127899)/(21 + 31 - 12)} = 72.506.$$

The tabled critical value is 2.449, so this hypothesis is rejected as well. The data suggest that the models for the two periods are systematically different, beyond a simple shift in the constant term.

The F ratio that results from estimating the model subject to the restriction that the two automobile price elasticities and the coefficient on the time trend are unchanged is

$$F[3, 40] = \frac{(0.01441975 - (0.00202244 + 0.007127899))/3}{(0.00202244 + 0.007127899)/(52 - 6 - 6)} = 7.678.$$

(The restricted regression is not shown.) The critical value from the F table is 2.839, so this hypothesis is rejected as well. Note, however, that this value is far smaller than those we obtained previously. This fact suggests that the bulk of the difference in the models across the two periods is, indeed, explained by the changes in the constant and the price and income elasticities.

The test statistic in (6-18) for the regression results in Table 6.7 gives a value of 502.34. The 5 percent critical value from the chi-squared table for six degrees of freedom is 12.59. So, on the basis of the Wald test, we would once again reject the hypothesis that the same coefficient vector applies in the two subperiods 1953 to 1973 and 1974 to 2004. We should note that the Wald statistic is valid only in large samples, and our samples of 21 and 31 observations hardly meet that standard. We have tested the hypothesis that the regression model for the gasoline market changed in 1973, and on the basis of the F test (Chow test) we strongly rejected the hypothesis of model stability.

Example 6.10 The World Health Report

The 2000 version of the World Health Organization's (WHO) *World Health Report* contained a major country-by-country inventory of the world's health care systems. [World Health Organization (2000). See also http://www.who.int/whr/en/.] The book documented years of research and has thousands of pages of material. Among the most controversial and most publicly debated parts of the report was a single chapter that described a comparison of the delivery of health care by 191 countries—nearly all of the world's population. [Evans et al. (2000a,b). See, e.g., Hilts (2000) for reporting in the popular press.] The study examined the efficiency of health care delivery on two measures: the standard one that is widely studied, (disability adjusted) life expectancy (DALE), and an innovative new measure created by the authors that was a composite of five outcomes (COMP) and that accounted for efficiency and fairness in delivery. The regression-style modeling, which was done in the setting of a frontier model (see Chapter 18), related health care attainment to two major inputs, education and (per capita) health care expenditure. The residuals were analyzed to obtain the country comparisons.

The data in Appendix Table F6.3 were used by the researchers at the WHO for the study. (They used a panel of data for the years 1993 to 1997. We have extracted the 1997 data for this example.) The WHO data have been used by many researchers in subsequent analyses. [See, e.g., Hollingsworth and Wildman (2002), Gravelle et al. (2002), and Greene (2004).] The regression model used by the WHO contained DALE or COMP on the left-hand side and health care expenditure, education, and education squared on the right. Greene (2004) added a number of additional variables such as per capita GDP, a measure of the distribution of income, and World Bank measures of government effectiveness and democratization of the political structure.

Among the controversial aspects of the study was the fact that the model aggregated countries of vastly different characteristics. A second striking aspect of the results, suggested in Hilts (2000) and documented in Greene (2004), was that, in fact, the "efficient" countries in the study were the 30 relatively wealthy OECD members, while the rest of the world on average fared much more poorly. We will pursue that aspect here with respect to DALE. Analysis of COMP is left as an exercise. Table 6.8 presents estimates of the regression models for

TABLE 6.8 Regression Results for Life Expectancy

	All Countries		*OECD*		*Non-OECD*	
Constant	25.237	38.734	42.728	49.328	26.812	41.408
Health exp	0.00629	−0.00180	0.00268	0.00114	0.00955	−0.00178
Education	7.931	7.178	6.177	5.156	7.0433	6.499
Education^2	−0.439	−0.426	−0.385	−0.329	−0.374	−0.372
Gini coeff		−17.333		−5.762		−21.329
Tropic		−3.200		−3.298		−3.144
Pop. Dens.		−0.255e−4		0.000167		−0.425e−4
Public exp		−0.0137		−0.00993		−0.00939
PC GDP		0.000483		0.000108		0.000600
Democracy		1.629		−0.546		1.909
Govt. Eff.		0.748		1.224		0.786
R^2	0.6824	0.7299	0.6483	0.7340	0.6133	0.6651
Std. Err.	6.984	6.565	1.883	1.916	7.366	7.014
Sum of sq.	9121.795	7757.002	92.21064	69.74428	8518.750	7378.598
N	191		30		161	
GDP/Pop	6609.37		18199.07		4449.79	
F test		4.524		0.874		3.311

DALE for the pooled sample, the OECD countries, and the non-OECD countries, respectively. Superficially, there do not appear to be very large differences across the two subgroups. We first tested the joint significance of the additional variables, income distribution (GINI), per capita GDP, and so on. For each group, the F statistic is $[(\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e})/7]/[\mathbf{e}'\mathbf{e}/(n-11)]$. These F statistics are shown in the last row of the table. The critical values for F[7,180] (all), F[7,19] (OECD), and F[7,150] (non-OECD) are 2.061, 2.543, and 2.071, respectively. We conclude that the additional explanatory variables are significant contributors to the fit for the non-OECD countries (and for all countries), but not for the OECD countries. Finally, to conduct the structural change test of OECD vs. non-OECD, we compute

$$F[11, 169] = \frac{[7757.007 - (69.74428 + 7378.598)]/11}{(69.74428 + 7378.598)/(191 - 11 - 11)} = 0.637.$$

The 95 percent critical value for F[11,169] is 1.846. So, we do not reject the hypothesis that the regression model is the same for the two groups of countries. The Wald statistic in (6-18) tells a different story. The statistic is 35.221. The 95 percent critical value from the chi-squared table with 11 degrees of freedom is 19.675. On this basis, we would reject the hypothesis that the two coefficient vectors are the same.

6.4.5 PREDICTIVE TEST OF MODEL STABILITY

The hypothesis test defined in (6-16) in Section 6.4.2 is equivalent to $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_1$ in the "model"

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_1 + \varepsilon_t, \quad t = 1, \ldots, T_1$$
$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_2 + \varepsilon_t, \quad t = T_1 + 1, \ldots, T_1 + T_2.$$

(Note that the disturbance variance is assumed to be the same in both subperiods.) An alternative formulation of the model (the one used in the example) is

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{0}\\ \mathbf{X}_2 & \mathbf{I}\end{bmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix}.$$

This formulation states that

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_1 + \varepsilon_t, \qquad t = 1, \ldots, T_1$$
$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_2 + \gamma_t + \varepsilon_t, \quad t = T_1 + 1, \ldots, T_1 + T_2.$$

Because each γ_t is unrestricted, this alternative formulation states that the regression model of the first T_1 periods ceases to operate in the second subperiod (and, in fact, no systematic model operates in the second subperiod). A test of the hypothesis $\boldsymbol{\gamma} = \mathbf{0}$ in this framework would thus be a test of model stability. The least squares coefficients for this regression can be found by using the formula for the partitioned inverse matrix

$$\begin{aligned}
\begin{pmatrix}\mathbf{b}\\ \mathbf{c}\end{pmatrix} &= \begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2 & \mathbf{X}_2' \\ \mathbf{X}_2 & \mathbf{I}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1'\mathbf{y}_1 + \mathbf{X}_2'\mathbf{y}_2 \\ \mathbf{y}_2\end{bmatrix} \\
&= \begin{bmatrix}(\mathbf{X}_1'\mathbf{X}_1)^{-1} & -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_2' \\ -\mathbf{X}_2(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{I} + \mathbf{X}_2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_2'\end{bmatrix}\begin{bmatrix}\mathbf{X}_1'\mathbf{y}_1 + \mathbf{X}_2'\mathbf{y}_2 \\ \mathbf{y}_2\end{bmatrix} \\
&= \begin{pmatrix}\mathbf{b}_1 \\ \mathbf{c}_2\end{pmatrix}
\end{aligned}$$

where $\mathbf{b}_1$ is the least squares slopes based on the first T_1 observations and $\mathbf{c}_2$ is $\mathbf{y}_2 - \mathbf{X}_2\mathbf{b}_1$. The covariance matrix for the full set of estimates is s^2 times the bracketed matrix. The two subvectors of residuals in this regression are $\mathbf{e}_1 = \mathbf{y}_1 - \mathbf{X}_1\mathbf{b}_1$ and $\mathbf{e}_2 = \mathbf{y}_2 - (\mathbf{X}_2\mathbf{b}_1 + \mathbf{I}\mathbf{c}_2) = \mathbf{0}$, so the sum of squared residuals in this least squares regression is just $\mathbf{e}_1'\mathbf{e}_1$. This is the same sum of squares as appears in (6-16). The degrees of freedom for the denominator is $[T_1 + T_2 - (K + T_2)] = T_1 - K$ as well, and the degrees of freedom for the numerator is the number of elements in $\boldsymbol{\gamma}$ which is T_2. The restricted regression with $\boldsymbol{\gamma} = \mathbf{0}$ is the pooled model, which is likewise the same as appears in (6-16). This implies that the F statistic for testing the null hypothesis in this model is precisely that which appeared earlier in (6-16), which suggests why the test is labeled the "predictive test."

6.5 SUMMARY AND CONCLUSIONS

This chapter has discussed the functional form of the regression model. We examined the use of dummy variables and other transformations to build nonlinearity into the model. We then considered other nonlinear models in which the parameters of the nonlinear model could be recovered from estimates obtained for a linear regression. The final sections of the chapter described hypothesis tests designed to reveal whether the assumed model had changed during the sample period, or was different for different groups of observations.

Key Terms and Concepts

- Binary variable
- Chow test
- Control group
- Control observations
- Difference in differences
- Dummy variable
- Dummy variable trap
- Exactly identified
- Identification condition
- Interaction terms
- Intrinsically linear
- Knots
- Loglinear model
- Marginal effect
- Natural experiment

- Nonlinear restriction
- Overidentified
- Piecewise continuous
- Placebo effect
- Predictive test
- Qualification indices
- Response
- Semilog equation
- Spline
- Structural change
- Threshold effects
- Time profile
- Treatment
- Treatment group
- Wald test

Exercises

1. A regression model with $K = 16$ independent variables is fit using a panel of seven years of data. The sums of squares for the seven separate regressions and the pooled regression are shown below. The model with the pooled data allows a separate constant for each year. Test the hypothesis that the same coefficients apply in every year.

	1954	*1955*	*1956*	*1957*	*1958*	*1959*	*1960*	*All*
Observations	65	55	87	95	103	87	78	570
$\mathbf{e}'\mathbf{e}$	104	88	206	144	199	308	211	1425

2. *Reverse regression.* A common method of analyzing statistical data to detect discrimination in the workplace is to fit the regression

$$y = \alpha + \mathbf{x}'\boldsymbol{\beta} + \gamma d + \varepsilon, \tag{1}$$

where y is the wage rate and d is a dummy variable indicating either membership ($d = 1$) or nonmembership ($d = 0$) in the class toward which it is suggested the discrimination is directed. The regressors $\mathbf{x}$ include factors specific to the particular type of job as well as indicators of the qualifications of the individual. The hypothesis of interest is $H_0 : \gamma \geq 0$ versus $H_1 : \gamma < 0$. The regression seeks to answer the question, "In a given job, are individuals in the class ($d = 1$) paid less than equally qualified individuals not in the class ($d = 0$)?" Consider an alternative approach. Do individuals in the class in the same job as others, and receiving the same wage, uniformly have higher qualifications? If so, this might also be viewed as a form of discrimination. To analyze this question, Conway and Roberts (1983) suggested the following procedure:

1. Fit (1) by ordinary least squares. Denote the estimates a, $\mathbf{b}$, and c.
2. Compute the set of **qualification indices,**

$$\mathbf{q} = a\mathbf{i} + \mathbf{X}\mathbf{b}. \tag{2}$$

Note the omission of $c\mathbf{d}$ from the fitted value.

3. Regress $\mathbf{q}$ on a constant, $\mathbf{y}$ and $\mathbf{d}$. The equation is

$$\mathbf{q} = \alpha_* + \beta_*\mathbf{y} + \gamma_*\mathbf{d} + \varepsilon_*. \tag{3}$$

The analysis suggests that if $\gamma < 0$, $\gamma_* > 0$.

a. Prove that the theory notwithstanding, the least squares estimates c and c_* are related by

$$c_* = \frac{(\overline{y}_1 - \overline{y})(1 - R^2)}{(1 - P)\left(1 - r_{yd}^2\right)} - c, \tag{4}$$

where

$\overline{y}_1 =$ mean of y for observations with $d = 1$,
$\overline{y} =$ mean of y for all observations,
$P =$ mean of d,
$R^2 =$ coefficient of determination for (1),
$r^2_{yd} =$ squared correlation between y and d.

[*Hint*: The model contains a constant term. Thus, to simplify the algebra, assume that all variables are measured as deviations from the overall sample means and use a partitioned regression to compute the coefficients in (3). Second, in (2), use the result that based on the least squares results $\mathbf{y} = a\mathbf{i} + \mathbf{Xb} + c\mathbf{d} + \mathbf{e}$, so $\mathbf{q} = \mathbf{y} - c\mathbf{d} - \mathbf{e}$. From here on, we drop the constant term. Thus, in the regression in (3) you are regressing $[\mathbf{y} - c\mathbf{d} - \mathbf{e}]$ on $\mathbf{y}$ and $\mathbf{d}$.

b. Will the sample evidence necessarily be consistent with the theory? [*Hint*: Suppose that $c = 0$.]

A symposium on the Conway and Roberts paper appeared in the *Journal of Business and Economic Statistics* in April 1983.

3. *Reverse regression continued*. This and the next exercise continue the analysis of Exercise 2. In Exercise 2, interest centered on a particular dummy variable in which the regressors were accurately measured. Here we consider the case in which the crucial regressor in the model is measured with error. The paper by Kamlich and Polachek (1982) is directed toward this issue.

   Consider the simple errors in the variables model,

$$y = \alpha + \beta x^* + \varepsilon, \quad x = x^* + u,$$

   where u and ε are uncorrelated and x is the erroneously measured, observed counterpart to x^*.

   a. Assume that x^*, u, and ε are all normally distributed with means μ^*, 0, and 0, variances σ^2_*, σ^2_u, and σ^2_ε, and zero covariances. Obtain the probability limits of the least squares estimators of α and β.
   b. As an alternative, consider regressing x on a constant and y, and then computing the reciprocal of the estimate. Obtain the probability limit of this estimator.
   c. Do the "direct" and "reverse" estimators bound the true coefficient?

4. *Reverse regression continued*. Suppose that the model in Exercise 3 is extended to $y = \beta x^* + \gamma d + \varepsilon$, $x = x^* + u$. For convenience, we drop the constant term. Assume that x^*, ε, and u are independent normally distributed with zero means. Suppose that d is a random variable that takes the values one and zero with probabilities π and $1 - \pi$ in the population and is independent of all other variables in the model. To put this formulation in context, the preceding model (and variants of it) have appeared in the literature on discrimination. We view y as a "wage" variable, x^* as "qualifications," and x as some imperfect measure such as education. The dummy variable d is membership ($d = 1$) or nonmembership ($d = 0$) in some protected class. The hypothesis of discrimination turns on $\gamma < 0$ versus $\gamma \geq 0$.

   a. What is the probability limit of c, the least squares estimator of γ, in the least squares regression of y on x and d? [*Hints*: The independence of x^* and d is important. Also, plim $\mathbf{d}'\mathbf{d}/n = \text{Var}[d] + E^2[d] = \pi(1 - \pi) + \pi^2 = \pi$. This minor modification does not affect the model substantively, but it greatly simplifies the algebra.] Now suppose that x^* and d are not independent. In particular, suppose that $E[x^* \mid d = 1] = \mu^1$ and $E[x^* \mid d = 0] = \mu^0$. Repeat the derivation with this assumption.

b. Consider, instead, a regression of x on y and d. What is the probability limit of the coefficient on d in this regression? Assume that x^* and d are independent.

c. Suppose that x^* and d are not independent, but γ is, in fact, less than zero. Assuming that both preceding equations still hold, what is estimated by $(\overline{y} \mid d = 1) - (\overline{y} \mid d = 0)$? What does this quantity estimate if γ does equal zero?

Applications

1. In Application 1 in Chapter 3 and Application 1 in Chapter 5, we examined Koop and Tobias's data on wages, education, ability, and so on. We continue the analysis here. (The source, location and configuration of the data are given in the earlier application.) We consider the model

$$\begin{aligned}\ln \mathit{Wage} = {} & \beta_1 + \beta_2\,\mathit{Educ} + \beta_3\,\mathit{Ability} + \beta_4\,\mathit{Experience} \\ & + \beta_5\,\mathit{Mother's\ education} + \beta_6\,\mathit{Father's\ education} + \beta_7\,\mathit{Broken\ home} \\ & + \beta_8\,\mathit{Siblings} + \varepsilon.\end{aligned}$$

a. Compute the full regression by least squares and report your results. Based on your results, what is the estimate of the marginal value, in \$/hour, of an additional year of education, for someone who has 12 years of education when all other variables are at their means and *Broken home* $= 0$?

b. We are interested in possible nonlinearities in the effect of education on ln *Wage*. (Koop and Tobias focused on experience. As before, we are not attempting to replicate their results.) A histogram of the education variable shows values from 9 to 20, a huge spike at 12 years (high school graduation) and, perhaps surprisingly, a second at 15—intuition would have anticipated it at 16. Consider aggregating the education variable into a set of dummy variables:

$$\begin{aligned} \mathit{HS} &= 1 \text{ if } \mathit{Educ} \le 12, 0 \text{ otherwise} \\ \mathit{Col} &= 1 \text{ if } \mathit{Educ} > 12 \text{ and } \mathit{Educ} \le 16, 0 \text{ otherwise} \\ \mathit{Grad} &= 1 \text{ if } \mathit{Educ} > 16, 0 \text{ otherwise.} \end{aligned}$$

Replace *Educ* in the model with (*Col, Grad*), making high school (*HS*) the base category, and recompute the model. Report all results. How do the results change? Based on your results, what is the marginal value of a college degree? (This is actually the marginal value of having 16 years of education—in recent years, college graduation has tended to require somewhat more than four years on average.) What is the marginal impact on ln *Wage* of a graduate degree?

c. The aggregation in part b actually loses quite a bit of information. Another way to introduce nonlinearity in education is through the function itself. Add Educ^2 to the equation in part a and recompute the model. Again, report all results. What changes are suggested? Test the hypothesis that the quadratic term in the equation is not needed—that is, that its coefficient is zero. Based on your results, sketch a profile of log wages as a function of education.

d. One might suspect that the value of education is enhanced by greater ability. We could examine this effect by introducing an interaction of the two variables in the equation. Add the variable

$$Educ\_Ability = Educ \times Ability$$

to the base model in part a. Now, what is the marginal value of an additional year of education? The sample mean value of ability is 0.052374. Compute a confidence interval for the marginal impact on ln *Wage* of an additional year of education for a person of average ability.

e. Combine the models in c and d. Add both $Educ^2$ and *Educ_Ability* to the base model in part a and reestimate. As before, report all results and describe your findings. If we define "low ability" as less than the mean and "high ability" as greater than the mean, the sample averages are −0.798563 for the 7,864 low-ability individuals in the sample and +0.717891 for the 10,055 high-ability individuals in the sample. Using the formulation in part c, with this new functional form, sketch, describe, and compare the log wage profiles for low- and high-ability individuals.

2. (An extension of Application 1.) Here we consider whether different models as specified in Application 1 would apply for individuals who reside in "Broken homes." Using the results in Sections 6.4.1 and 6.4.4, test the hypothesis that the same model (not including the *Broken home* dummy variable) applies to both groups of individuals, those with *Broken home* = 0 and with *Broken home* = 1.

3. In Solow's classic (1957) study of technical change in the U.S. economy, he suggests the following aggregate production function: $q(t) = A(t)f[k(t)]$, where $q(t)$ is aggregate output per work hour, $k(t)$ is the aggregate capital labor ratio, and $A(t)$ is the technology index. Solow considered four static models, $q/A = \alpha + \beta \ln k$, $q/A = \alpha - \beta/k$, $\ln(q/A) = \alpha + \beta \ln k$, and $\ln(q/A) = \alpha + \beta/k$. Solow's data for the years 1909 to 1949 are listed in Appendix Table F6.4.

a. Use these data to estimate the α and β of the four functions listed above. (*Note*: Your results will not quite match Solow's. See the next exercise for resolution of the discrepancy.)

b. In the aforementioned study, Solow states:

> A scatter of q/A against k is shown in Chart 4. Considering the amount of a priori doctoring which the raw figures have undergone, the fit is remarkably tight. Except, that is, for the layer of points which are obviously too high. These maverick observations relate to the seven last years of the period, 1943–1949. From the way they lie almost exactly parallel to the main scatter, one is tempted to conclude that in 1943 the aggregate production function simply shifted.

Compute a scatter diagram of q/A against k and verify the result he notes above.

c. Estimate the four models you estimated in the previous problem including a dummy variable for the years 1943 to 1949. How do your results change? (*Note*: These results match those reported by Solow, although he did not report the coefficient on the dummy variable.)

d. Solow went on to surmise that, in fact, the data were fundamentally different in the years before 1943 than during and after. Use a Chow test to examine the difference in the two subperiods using your four functional forms. Note that with the dummy variable, you can do the test by introducing an interaction term between the dummy and whichever function of k appears in the regression. Use an F test to test the hypothesis.

4. Data on the number of incidents of wave damage to a sample of ships, with the type of ship and the period when it was constructed, are given in Table 6.9. There are five types of ships and four different periods of construction. Use F tests and dummy variable regressions to test the hypothesis that there is no significant "ship type effect" in the expected number of incidents. Now, use the same procedure to test whether there is a significant "period effect."

TABLE 6.9 Ship Damage Incidents

	Period Constructed			
Ship Type	***1960–1964***	***1965–1969***	***1970–1974***	***1975–1979***
A	0	4	18	11
B	29	53	44	18
C	1	1	2	1
D	0	0	11	4
E	0	7	12	1

Source: Data from McCullagh and Nelder (1983, p. 137).