7

FUNCTIONAL FORM AND STRUCTURAL CHANGE

7.1 INTRODUCTION

In this chapter, we are concerned with the functional form of the regression model. Many different types of functions are "linear" by the definition considered in Section 2.3.1. By using different transformations of the dependent and independent variables, dummy variables and different arrangements of functions of variables, a wide variety of models can be constructed that are all estimable by linear least squares. Section 7.2 considers using binary variables to accommodate nonlinearities in the model. Section 7.3 broadens the class of models that are linear in the parameters. Sections 7.4 and 7.5 then examine the issue of specifying and testing for change in the underlying model that generates the data, under the heading of **structural change.**

7.2 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.

7.2.1 BINARY VARIABLES IN REGRESSION

Dummy variables are usually used in regression equations that also contain other quantitative variables. In the earnings equation in Example 4.3, we included a variable *Kids* to indicate whether there were children in the household under the assumption that for many married women, this fact is a significant consideration in labor supply behavior. The results shown in Example 7.1 appear to be consistent with this hypothesis.

Example 7.1 Dummy Variable in an Earnings Equation

Table 7.1 following reproduces the estimated earnings equation in Example 4.3. The variable Kids is a dummy variable, which equals one if there are children under 18 in the household and zero otherwise. Since this is a **semilog equation,** the value of $-.35$ for the coefficient is an extremely large effect, that 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 which affect 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.

TABLE 7.1 Estimated Earnings Equation

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

Sum of squared residuals:	599.4582		
Standard error of the regression:	1.19044		
R^2 based on 428 observations	0.040995		
Variable	***Coefficient***	***Standard Error***	***t Ratio***
Constant	3.24009	1.7674	1.833
Age	0.20056	0.08386	2.392
Age^2	−0.0023147	0.00098688	−2.345
Education	0.067472	0.025248	2.672
Kids	−0.35119	0.14753	−2.380

In recent applications, researchers in many fields have studied the effects of **treatment** on some kind of **response.** Examples include the effect of college on, lifetime income, sex differences in labor supply behavior as in Example 7.1, and in salary structures in industries, and in pre- versus postregime shifts in macroeconomic models, to name but a few. These examples can all be formulated in regression models involving a single dummy variable:

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

One of the important issues in policy analysis concerns measurement of such treatment effects when the dummy variable results from an individual participation decision. For example, in studies of the effect of job training programs on post-training earnings, the "treatment dummy" might be measuring the latent motivation and initiative of the participants rather than the effect of the program, itself. We will revisit this subject in Section 22.4.

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 20.6.5.) 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).

7.2.2 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:

$$C_t = \beta x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \delta_4 D_{t4} + \varepsilon_t. \quad \textbf{(7-1)}$$

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

7.2.3 SEVERAL GROUPINGS

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

$$y_{it} = \alpha + \beta x_{it} + \delta_i + \theta_t + \varepsilon_{it}. \quad \textbf{(7-2)}$$

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 7.2 Analysis of Covariance

The data in Appendix Table F7.1 were used in a study of efficiency in production of airline services in Greene (1997b). The airline industry has been a favorite subject of study [e.g., Schmidt and Sickles (1984); Sickles, Good, and Johnson (1986)], partly because of interest in this rapidly changing market in a period of deregulation and partly because of an abundance of large, high-quality data sets collected by the (no longer existent) Civil Aeronautics Board. The original data set consisted of 25 firms observed yearly for 15 years (1970 to 1984), a "balanced panel." Several of the firms merged during this period and several others experienced strikes, which reduced the number of complete observations substantially. Omitting these and others because of missing data on some of the variables left a group of 10 full

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

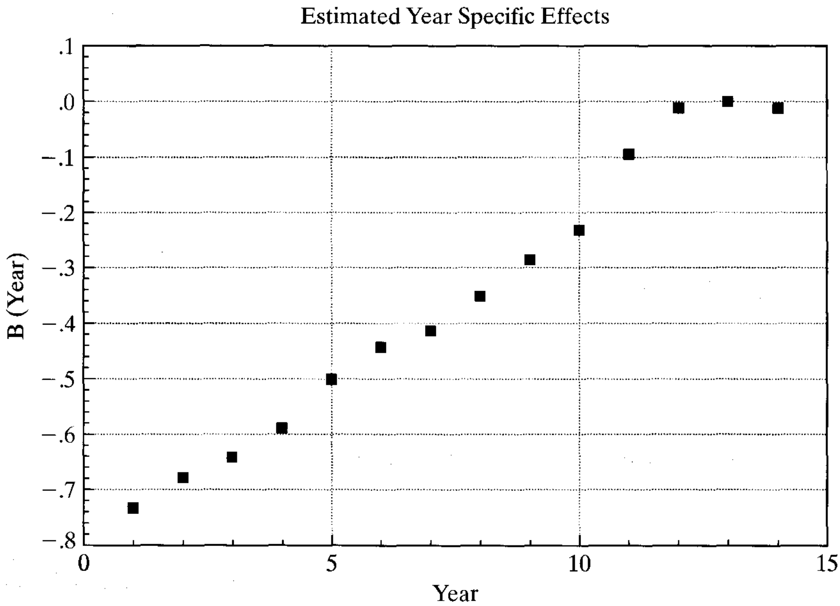


FIGURE 7.1 Estimated Year Dummy Variable Coefficients.

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

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

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

$$\ln C_{i,t} = 13.56 + .8866 \ln Q_{i,t} + 0.01261 \ln^2 Q_{i,t} + 0.1281 \ln P_{fi,t} - 0.8855\, LF_{i,t} + \text{time effects} + \text{firm effects}.$$

The year effects display a revealing pattern, as shown in Figure 7.1. 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 7.2 presents the sums of squares from the four regressions. The F statistic for the hypothesis that there are no firm specific effects is 65.94, which is highly significant. The statistic for the time effects is only 2.61, which is larger than the critical value

TABLE 7.2 *F* tests for Firm and Year Effects

Model	*Sum of Squares*	*Parameters*	*F*	*Deg.Fr.*
Full Model	0.17257	24	—	
Time Effects	1.03470	19	65.94	[5,66]
Firm Effects	0.26815	10	2.61	[14,66]
No Effects	1.27492	5	22.19	[19,66]

of 1.84, but perhaps less so than Figure 7.1 might have suggested. In the absence of the year specific dummy variables, the year specific effects are probably largely absorbed by the price of fuel.

7.2.4 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, income $= \beta_1 + \beta_2\, age + \beta_3 E + \varepsilon$. The difficulty with this approach is that it assumes that the increment in income at each threshold is the same; β_3 is the difference between income with a Ph.D. and a master's and between a master's and a bachelor's degree. This is unlikely and unduly restricts the regression. A more flexible model would use three (or four) binary variables, one for each level of education. Thus, we would write

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

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

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

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

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

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.

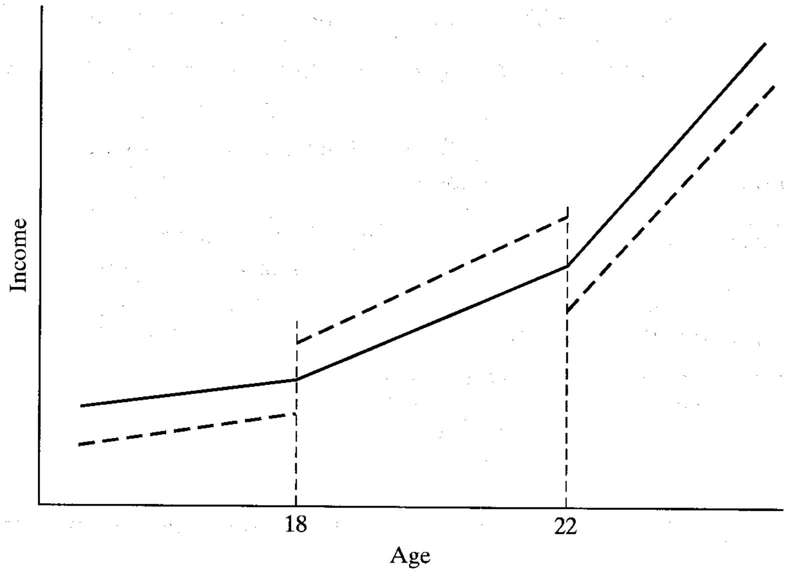


FIGURE 7.2 Spline Function.

7.2.5 SPLINE 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 7.2. Based on the discussion in the preceding paragraph, we could fit such a regression model just by dividing the sample into three subsamples. However, this would neglect the continuity of the proposed function. The result would appear more like the dotted figure than the continuous function we had in mind. Restricted regression and what is known as a **spline** function can be used to achieve the desired effect.[2]

The function we wish to estimate is

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

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

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

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

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

$$\text{income} = \beta_1 + \beta_2 \text{ age} + \gamma_1 d_1 + \delta_1 d_1 \text{ age} + \gamma_2 d_2 + \delta_2 d_2 \text{ age} + \varepsilon. \tag{7-3}$$

This relationship is the dashed function in Figure 7.2. The slopes in the three segments are β_2, $\beta_2 + \delta_1$, and $\beta_2 + \delta_1 + \delta_2$. To make the function **piecewise continuous,** we require that the segments join at the knots—that is,

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

and

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

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

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

Doing likewise for the second and inserting these in (7-3), we obtain

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

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

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

and

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

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

7.3 NONLINEARITY IN THE VARIABLES

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

$$\begin{aligned} g(y) &= \beta_1 f_1(\mathbf{z}) + \beta_2 f_2(\mathbf{z}) + \cdots + \beta_K f_K(\mathbf{z}) + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon \\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon. \end{aligned} \tag{7-4}$$

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

7.3.1 FUNCTIONAL FORMS

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

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

In this model, the coefficients are elasticities:

$$\left(\frac{\partial y}{\partial x_k}\right)\left(\frac{x_k}{y}\right) = \frac{\partial \ln y}{\partial \ln x_k} = \beta_k. \tag{7-5}$$

In the loglinear equation, measured changes are in proportional or percentage terms; β_k measures the percentage change in y associated with a one percent change in x_k. This removes the units of measurement of the variables from consideration in using the regression model. An alternative approach sometimes taken is to measure the variables and associated changes in standard deviation units. If the data are "standardized" before estimation using $x_{ik}^* = (x_{ik} - \bar{x}_k)/s_k$ and likewise for y, then the least squares regression coefficients measure changes in standard deviation units rather than natural 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_y/s_k produces the same result.

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

$$\ln y = \beta_1 + \beta_2 x + \varepsilon. \tag{7-6}$$

We used this form in the investment equation in Section 6.2,

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

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 $-.005$ in Table 6.1 suggests that over the full estimation period, after accounting for all other factors, the average rate of growth of investment was $-.5$ percent per year.

The coefficients in the semilog model are partial- or semi-elasticities; in (7-6), 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 7.1. The coefficient on *Kids* of $-.35$ suggests that all else equal, earnings are approximately 35 percent less when there are children in the household.

The quadratic earnings equation in Example 7.1 shows another use of nonlinearities in the variables. Using the results in Example 7.1, we find that for a woman with 12 years of schooling and children in the household, the age-earnings profile appears as in Figure 7.3. This figure suggests an important question in this framework. It is tempting to conclude that Figure 7.3 shows the earnings trajectory of a person at different ages, but that is not what the data provide. The model is based on a cross section, and what it displays is the earnings of different people of different ages. How this profile relates to the expected earnings path of one individual is a different, and complicated question.

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

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

In this model,

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

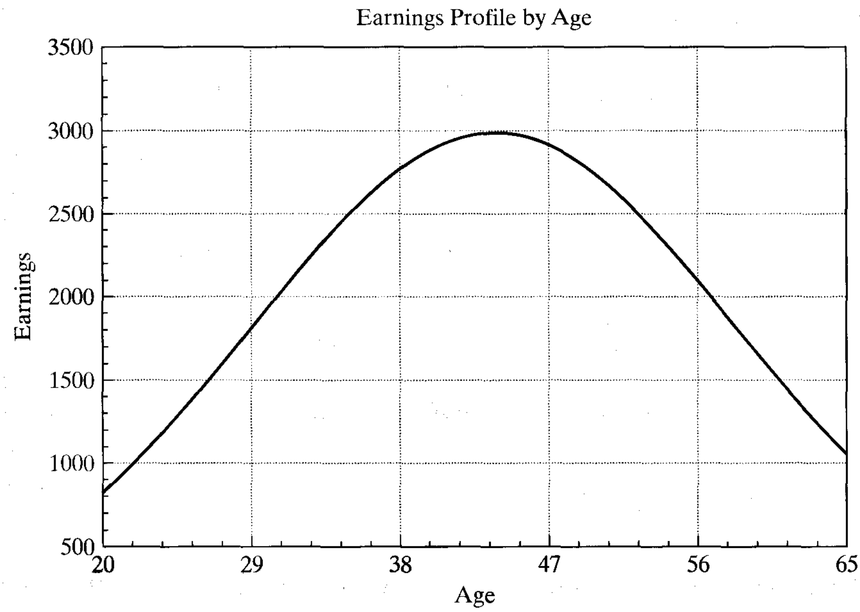


FIGURE 7.3 Age-Earnings Profile.

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

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

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

7.3.2 IDENTIFYING NONLINEARITY

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

Example 7.3 Functional Form for a Nonlinear Cost Function

In a celebrated study of economies of scale in the U.S. electric power industry, Nerlove (1963) analyzed the production costs of 145 American electric 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 one 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 14.) Thus, the firm's objective was cost minimization subject to the constraint of the production function. This can be formulated as a Lagrangean problem,

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

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

$$P_K K + P_L L + P_F F = C(Q, P_K, P_L, P_F) = rAQ^{1/r} P_K^{\alpha_K/r} P_L^{\alpha_L/r} P_F^{\alpha_F/r} e^{\varepsilon_i/r}$$

or

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

where $\beta_q = 1/(\alpha_K + \alpha_L + \alpha_F)$ is now the parameter of interest and $\beta_j = \alpha_j/r$, $j = K, L, F$.[3] 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.[4] 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 7.3.

---

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

[4]In the context of the econometric model, the restriction has a testable implication by the definition in Chapter 6. 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 14. Fortunately, in this instance, the hypothesis is not rejected. (It is in the application in Chapter 14.)

TABLE 7.3 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	0.594	−0.0085	0.952
	(0.0174)	(0.205)	(0.191)	
Group 1	0.398	0.641	−0.093	0.512
Group 2	0.668	0.105	0.364	0.635
Group 3	0.931	0.408	0.249	0.571
Group 4	0.915	0.472	0.133	0.871
Group 5	1.045	0.604	−0.295	0.920

Initial estimates of the parameters of the cost function are shown in the top row of Table 7.3. The hypothesis of constant returns to scale can be firmly rejected. The t ratio is $(0.721 - 1)/0.0174 = -16.03$, so we conclude that this estimate is significantly less than one or, by implication, r is significantly greater than one. 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. [See (5-31a,b). Christensen and Greene's (1976) estimator of this model with these data produced a positive estimate. See Section 14.3.1.]

The striking pattern of the residuals shown in Figure 7.4[5] and some thought about the implied form of the production function suggested that something was missing from the model.[6] In theory, the estimated model implies a continually declining average cost curve, which in turn implies persistent economies of scale at all levels of output. This conflicts with the textbook notion of a U-shaped average cost curve and appears implausible for the data. Note the three clusters of residuals in the figure. Two approaches were used to analyze the model.

By sorting the sample into five groups on the basis of output and fitting separate regressions to each group, Nerlove fit a piecewise loglinear model. The results are given in the lower rows of Table 7.3, where the firms in the successive groups are progressively larger. The results are persuasive that the (log)-linear cost function is inadequate. The output coefficient that rises toward and then crosses 1.0 is consistent with a U-shaped cost curve as surmised earlier.

A second approach was to expand the cost function to include a quadratic term in log output. This approach corresponds to a much more general model and produced the result given in Table 7.4. Again, a simple t test strongly suggests that increased generality is called for; $t = 0.117/0.012 = 9.75$. The output elasticity in this quadratic model is $\beta_q + 2\gamma_{qq} \log Q$.[7] There are economies of scale when this value is less than one and constant returns to scale when it equals one. Using the two values given in the table (0.151 and 0.117, respectively), we find that this function does, indeed, produce a U shaped average cost curve with minimum at $\log_{10} Q = (1 - 0.151)/(2 \times 0.117) = 3.628$, or $Q = 4248$, which was roughly in the middle of the range of outputs for Nerlove's sample of firms.

[5]The residuals are created as deviations of predicted total cost from actual, so they do not sum to zero.

[6]A Durbin–Watson test of correlation among the residuals (see Section 12.5.1) 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.

[7]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.

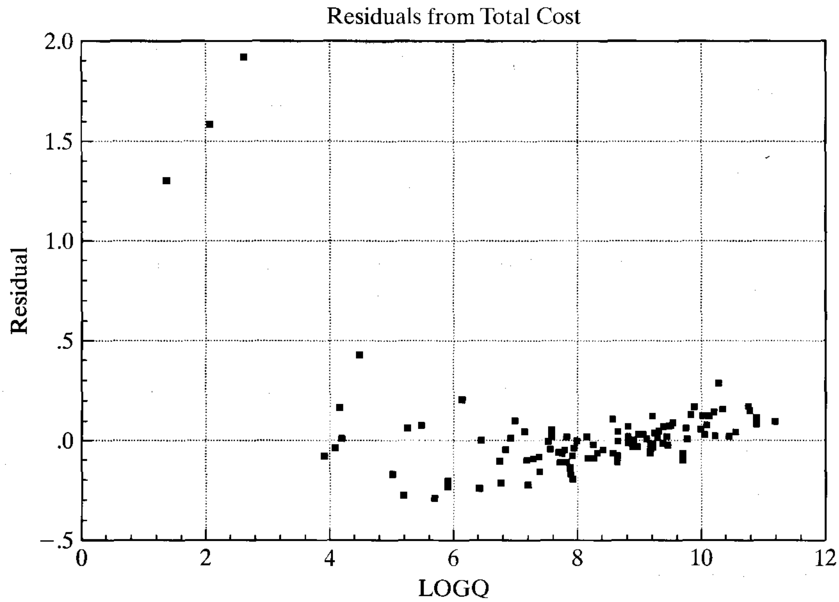


FIGURE 7.4 Residuals from Predicted Cost.

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 7.4 did somewhat underestimate the range of outputs in which unit costs of production would continue to decline. They also redid the study using a sample of 123 firms from 1970, and found similar results. In the latter sample, however, it appeared that many firms had expanded rapidly enough to exhaust the available economies of scale. We will revisit the 1970 data set in a study of efficiency in Section 17.6.4.

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.

7.3.3 INTRINSIC LINEARITY AND IDENTIFICATION

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

$$\ln Y_i = \ln \alpha + \beta_2 \ln X_i + \varepsilon_i \tag{7-8}$$

TABLE 7.4 Log-Quadratic Cost Function (Standard Errors in Parentheses)

	$\log Q$	$\log^2 Q$	$\log(P_L/P_F)$	$\log(P_K/P_F)$	R^2
All firms	0.151 (0.062)	0.117 (0.012)	0.498 (0.161)	−0.062 (0.151)	0.95

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 estimate of α is suggested.

This fact leads us to a second aspect of intrinsically linear models. Maximum likelihood estimators have an "invariance property." In the classical normal regression model, the maximum likelihood estimator of σ is the square root of the maximum likelihood estimator of σ^2. Under some conditions, least squares estimators have the same property. By exploiting this, we can broaden the definition of linearity and include some additional cases that might otherwise be quite complex.

DEFINITION 7.1 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 7.4 Intrinsically Linear Regression

In Section 17.5.4, we will estimate the parameters of the model

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

by maximum likelihood. 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. Since 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, 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 7.5 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.

TABLE 7.5 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	2.403	3.151	0.663

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 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).

Example 7.5 CES Production Function

The constant elasticity of substitution production function may be written

$$\ln y = \ln \gamma - \frac{\nu}{\rho} \ln[\delta K^{-\rho} + (1-\delta)L^{-\rho}] + \varepsilon. \tag{7-9}$$

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

$$\begin{aligned} \ln y &= \ln \gamma + \nu\delta \ln K + \nu(1-\delta)\ln L + \rho\nu\delta(1-\delta)\left\{-\tfrac{1}{2}[\ln K - \ln L]^2\right\} + \varepsilon' \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon', \end{aligned} \tag{7-10}$$

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

$$\begin{aligned} &\beta_1 = \ln \gamma, \quad \beta_2 = \nu\delta, \quad \beta_3 = \nu(1-\delta), \quad \beta_4 = \rho\nu\delta(1-\delta), \\ &\gamma = e^{\beta_1}, \quad \delta = \beta_2/(\beta_2+\beta_3), \quad \nu = \beta_2 + \beta_3, \quad \rho = \beta_4(\beta_2+\beta_3)/(\beta_2\beta_3). \end{aligned} \tag{7-11}$$

Estimates of $\beta_1, \beta_2, \beta_3$, and β_4 can be computed by least squares. The estimates of γ, δ, ν, and ρ obtained by the second row of (7-11) are the same as those we would obtain had we found the nonlinear least squares estimates of (7-10) directly. As Kmenta shows, however, they are not the same as the nonlinear least squares estimates of (7-9) due to the use of the Taylor series approximation to get to (7-10). 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

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

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

$$y_i = \beta_1(\theta)x_{i1} + \beta_2(\theta)x_{i2} + \cdots + \beta_K(\theta)x_{ik} + \varepsilon_i \tag{7-12}$$

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 (7-12), 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 $\beta_1, \beta_2, \beta_3$, 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.

7.4 MODELING AND TESTING FOR A STRUCTURAL BREAK

One of the more common applications of the F test is in tests of **structural change.**[8] 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 of or all 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[9] that we examined in Example 2.3. As Figure 7.5 following 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. It seems unlikely that the same regression model would apply to both periods.

7.4.1 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 (at least for a decade or so), 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 14 years of the data in $\mathbf{y}$ and $\mathbf{X}$ as $\mathbf{y}_1$ and $\mathbf{X}_1$ and the remaining years as $\mathbf{y}_2$ and $\mathbf{X}_2$. An unrestricted regression that allows the coefficients to be different in the two periods is

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix}. \tag{7-13}$$

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

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1'\mathbf{y}_1\\ \mathbf{X}_2'\mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix}, \tag{7-14}$$

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

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

[9] The data are listed in Appendix Table A6.1.

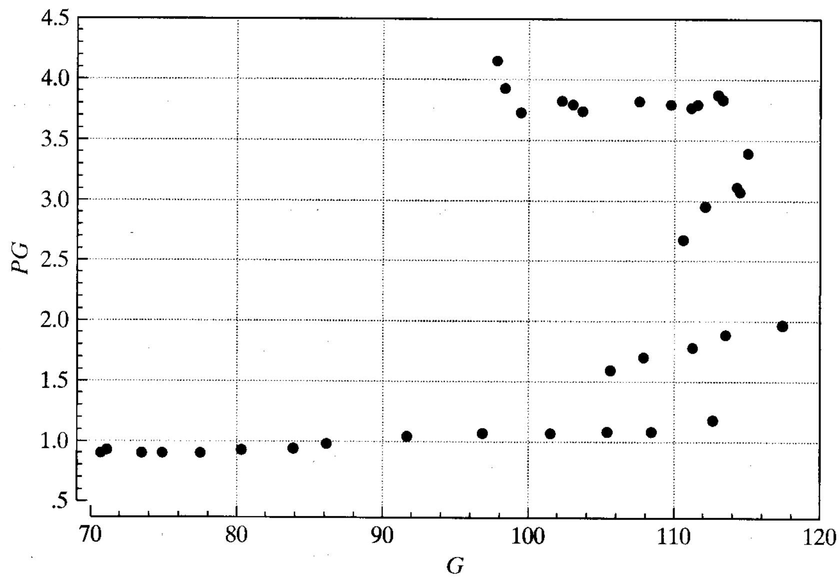


FIGURE 7.5 Gasoline Price and Per Capita Consumption, 1960–1995.

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 (7-13) 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 (6-6), where J, the number of restrictions, is the number of columns in $\mathbf{X}_2$ and the denominator degrees of freedom is $n_1 + n_2 - 2k$.

7.4.2 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. Since 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

$$F[n_2, n_1 - K] = \frac{(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}_1'\mathbf{e}_1)/n_2}{\mathbf{e}_1'\mathbf{e}_1/(n_1 - K)}. \tag{7-15}$$

Note that the numerator degrees of freedom is n_2, not K.[10] 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. We will have a closer look at that result in Section 7.5.3.

7.4.3 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 log-linear 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 $\mathbf{X}$ 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 Table 7.6 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 $\mathbf{W}$

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

denote the variables whose coefficients are thought to have remained constant. Then, the regressor matrix in the constrained regression would appear as

$$\mathbf{X} = \begin{bmatrix} \mathbf{i}_{\text{pre}} & \mathbf{Z}_{\text{pre}} & \mathbf{0} & \mathbf{0} & \mathbf{W}_{\text{pre}} \\ \mathbf{0} & \mathbf{0} & \mathbf{i}_{\text{post}} & \mathbf{Z}_{\text{post}} & \mathbf{W}_{\text{post}} \end{bmatrix}. \tag{7-16}$$

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

7.4.4 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 smaller 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,[11] with asymptotic covariance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$. Then, under the null hypothesis that the true parameters are the same,

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

Under the null hypothesis, the Wald statistic,

$$W = (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)'(\hat{\mathbf{V}}_1 + \hat{\mathbf{V}}_2)^{-1}(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2), \tag{7-17}$$

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.[12] 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.

---

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

[12] 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.

Ohtani and Kobayashi (1986) have devised a "bounds" test that gives a partial remedy for the problem.[13]

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. We will return to this analysis of structural breaks in time-series models in Section 7.5.4.

7.5 TESTS OF MODEL STABILITY

The tests of structural change described in Section 7.4 assume that the process underlying the data is stable up to a known transition point, where it makes a discrete change to a new, but thereafter stable, structure. In our gasoline market, that might be a reasonable assumption. In many other settings, however, the change to a new regime might be more gradual and less obvious. In this section, we will examine two tests that are based on the idea that a regime change might take place slowly, and at an unknown point in time, or that the regime underlying the observed data might simply not be stable at all.

7.5.1 HANSEN'S TEST

Hansen's (1992) test of model stability is based on a cumulative sum of the least squares residuals. From the least squares normal equations, we have

$$\sum_{t=1}^{T} \mathbf{x}_t e_t = 0 \quad \text{and} \quad \sum_{t=1}^{T}\left(e_t^2 - \frac{\mathbf{e}'\mathbf{e}}{n}\right) = 0.$$

Let the vector $\mathbf{f}_t$ be the $(K+1)\times 1$ tth observation in this pair of sums. Then, $\sum_{t=1}^{T}\mathbf{f}_t = \mathbf{0}$. Let the sequence of partial sums be $\mathbf{s}_t = \sum_{r=1}^{t}\mathbf{f}_r$, so $\mathbf{s}_T = \mathbf{0}$. Finally, let $\mathbf{F} = T\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t'$ and $\mathbf{S} = \sum_{t=1}^{T}\mathbf{s}_t\mathbf{s}_t'$. Hansen's test statistic can be computed simply as $H = \text{tr}(\mathbf{F}^{-1}\mathbf{S})$. Large values of H give evidence against the hypothesis of model stability. The logic of Hansen's test is that if the model is stable through the T periods, then the cumulative sums in $\mathbf{S}$ will not differ greatly from those in $\mathbf{F}$. Note that the statistic involves both the regression and the variance. The distribution theory underlying this nonstandard test statistic is much more complicated than the computation. Hansen provides asymptotic critical values for the test of model constancy which vary with the number of coefficients in the model. A few values for the 95 percent significance level are 1.01 for $K = 2$, 1.90 for $K = 6$, 3.75 for $K = 15$, and 4.52 for $K = 19$.

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

7.5.2 RECURSIVE RESIDUALS AND THE CUSUMS TEST

Example 7.6 shows a test of structural change based essentially on the model's ability to predict correctly outside the range of the observations used to estimate it. A similar logic underlies an alternative test of model stability proposed by Brown, Durbin, and Evans (1975) based on **recursive residuals.** The technique is appropriate for time-series data and might be used if one is uncertain about when a structural change might have taken place. The null hypothesis is that the coefficient vector $\boldsymbol{\beta}$ is the same in every period; the alternative is simply that it (or the disturbance variance) is not. The test is quite general in that it does not require a prior specification of when the structural change takes place. The cost, however, is that the power of the test is rather limited compared with that of the Chow test.[14]

Suppose that the sample contains a total of T observations.[15] The tth recursive residual is the ex post prediction error for y_t when the regression is estimated using only the first $t-1$ observations. Since it is computed for the next observation beyond the sample period, it is also labeled a **one step ahead prediction error;**

$$e_t = y_t - \mathbf{x}_t'\mathbf{b}_{t-1},$$

where $\mathbf{x}_t$ is the vector of regressors associated with observation y_t and $\mathbf{b}_{t-1}$ is the least squares coefficients computed using the first $t-1$ observations. The forecast variance of this residual is

$$\sigma_{ft}^2 = \sigma^2\left[1 + \mathbf{x}_t'(\mathbf{X}_{t-1}'\mathbf{X}_{t-1})^{-1}\mathbf{x}_t\right]. \qquad \textbf{(7-18)}$$

Let the rth scaled residual be

$$w_r = \frac{e_r}{\sqrt{1 + \mathbf{x}_r'(\mathbf{X}_{r-1}'\mathbf{X}_{r-1})^{-1}\mathbf{x}_r}}. \qquad \textbf{(7-19)}$$

Under the hypothesis that the coefficients remain constant during the full sample period, $w_r \sim N[0, \sigma^2]$ and is independent of w_s for all $s \neq r$. Evidence that the distribution of w_r is changing over time weighs against the hypothesis of model stability.

One way to examine the residuals for evidence of instability is to plot $w_r/\hat{\sigma}$ (see below) simply against the date. Under the hypothesis of the model, these residuals are uncorrelated and are approximately normally distributed with mean zero and standard deviation 1. Evidence that these residuals persistently stray outside the error bounds -2 and $+2$ would suggest model instability. (Some authors and some computer packages plot e_r instead, in which case the error bounds are $\pm 2\hat{\sigma}\sqrt{1 + \mathbf{x}_r'(\mathbf{X}_{r-1}'\mathbf{X}_{r-1})^{-1}\mathbf{x}_r}$.

The **CUSUM test** is based on the cumulated sum of the residuals:

$$W_t = \sum_{r=K+1}^{r=t} \frac{w_r}{\hat{\sigma}}, \qquad \textbf{(7-20)}$$

where $\hat{\sigma}^2 = (T-K-1)^{-1}\sum_{r=K+1}^{T}(w_r - \bar{w})^2$ and $\bar{w} = (T-K)^{-1}\sum_{r=K+1}^{T} w_r$. Under

[14]The test is frequently criticized on this basis. The Chow test, however, is based on a rather definite piece of information, namely, when the structural change takes place. If this is not known or must be estimated, then the advantage of the Chow test diminishes considerably.

[15]Since we are dealing explicitly with time-series data at this point, it is convenient to use T instead of n for the sample size and t instead of i to index observations.

the null hypothesis, W_t has a mean of zero and a variance approximately equal to the number of residuals being summed (because each term has variance 1 and they are independent). The test is performed by plotting W_t against t. Confidence bounds for the sum are obtained by plotting the two lines that connect the points $[K, \pm a(T-K)^{1/2}]$ and $[T, \pm 3a(T-K)^{1/2}]$. Values of a that correspond to various significance levels can be found in their paper. Those corresponding to 95 percent and 99 percent are 0.948 and 1.143, respectively. The hypothesis is rejected if W_t strays outside the boundaries.

Example 7.6 Structural Break in the Gasoline Market

The previous Figure 7.5 shows a plot of prices and quantities in the U.S. gasoline market from 1960 to 1995. The first 13 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:

$$\ln(G/pop)_t = \beta_1 + \beta_2 \ln(I/pop) + \beta_3 \ln P_{Gt} + \beta_4 \ln P_{NCt} + \beta_5 \ln P_{UCt} + \beta_6 t + \varepsilon_t.$$

The three prices in the equation are for G, new cars, and used cars. I/pop is per capita income, and G/pop is per capita gasoline consumption. Regression results for four functional forms are shown in Table 7.6. Using the data for the entire sample, 1960 to 1995, and for the two subperiods, 1960 to 1973 and 1974 to 1995, 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

$$F[6, 24] = \frac{(0.02521877 - 0.000652271 - 0.004662163)/6}{(0.000652271 + 0.004662163)/(14 + 22 - 12)} = 14.958.$$

The tabled critical value is 2.51, 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 36 observations to fit the model, the sum of squares is $\mathbf{e}_*'\mathbf{e}_* = 0.02521877$. When the $n_1 = 4$ observations for 1974, 1975, 1980 and 1981 are removed from the sample, the sum of squares falls to $\mathbf{e}'\mathbf{e} = 0.01968599$. The F statistic is 1.817. Since the tabled critical value for $F[4, 32 - 6]$ is 2.72, 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 36 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

TABLE 7.6 Gasoline Consumption Equations

Coefficients	*1960–1995*	*Pooled*	*Preshock*	*Postshock*
Constant	24.6718	21.2630	−51.1812	
Constant		21.3403		20.4464
ln I/pop	1.95463	1.83817	0.423995	1.01408
ln PG	−0.115530	−0.178004	0.0945467	−0.242374
ln PNC	0.205282	0.209842	0.583896	0.330168
ln PUC	−0.129274	−0.128132	−0.334619	−0.0553742
Year	−0.019118	−0.168618	0.0263665	−0.0126170
R^2	0.968275	0.978142	0.998033	0.920642
Standard error	0.02897572	0.02463767	0.00902961	0.017000
Sum of squares	0.02521877	0.0176034	0.000652271	0.004662163

year indicated and zero in all 35 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 0.01968599, so the F statistic for testing the joint hypothesis that the four coefficients are zero is $F[4, 36-10] = \{[(0.02518777 - 0.01968599)/4]/[0.01968599/(36-10)]\} = 1.817$, once again. (See Section 7.4.2 for discussion of this test.)

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;

$$F[5, 24] = \frac{(0.0176034 - 0.000652271 - 0.004662163)/5}{(0.000652271 + 0.004662163)/(14 + 22 - 12)} = 11.099.$$

The tabled critical value is 2.62, 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

$$F[3, 24] = \frac{(0.00802099 - 0.000652271 - 0.004662163)/3}{(0.000652271 + 0.004662163)/(14 + 22 - 12)} = 4.086.$$

(The restricted regression is not shown.) The critical value from the F table is 3.01, so this hypothesis is rejected as well. Note, however, that this value is far smaller than those we obtained previously. The P-value for this value is 0.981, so, in fact, at the 99 percent significance level, we would not have rejected the hypothesis. 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 (7-17) for the regression results in Table 7.6 gives a value of 128.6673. The 5 percent critical value from the chi-squared table for 6 degrees of freedom is 12.59. So, on the basis of the Wald test, we would reject the hypothesis that the same coefficient vector applies in the two subperiods 1960 to 1973 and 1974 to 1995. We should note that the Wald statistic is valid only in large samples, and our samples of 14 and 22 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. Hansen's test is not consistent with this result; using the computations outlined earlier, we obtain a value of $H = 1.7249$. Since the critical value is 1.90, the hypothesis of model stability is now not rejected.

Figure 7.6 shows the CUSUM test for the gasoline market. The results here are more or less consistent with the preceding results. The figure does suggest a structural break, though at 1984, not at 1974 or 1980 when we might have expected it.

7.5.3 PREDICTIVE TEST

The hypothesis test defined in (7-15) in Section 7.4.2 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

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

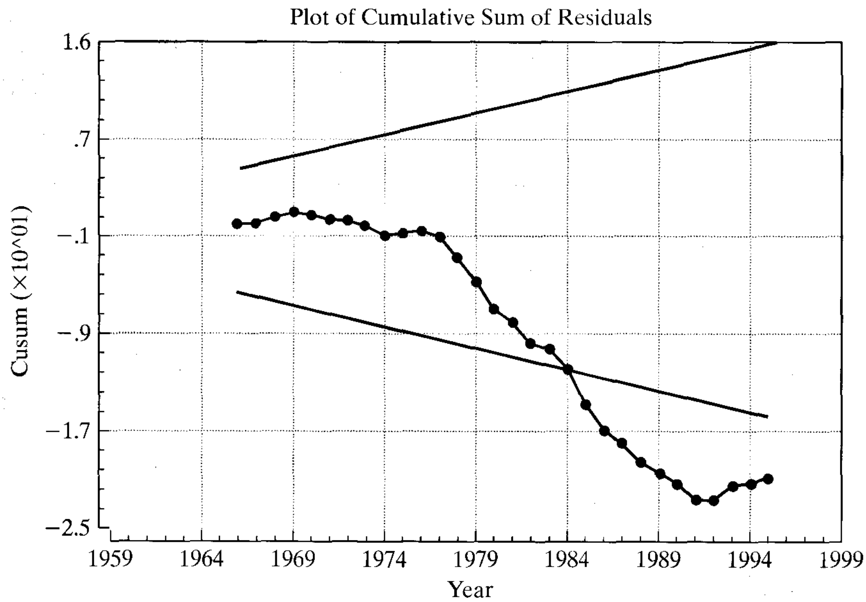


FIGURE 7.6 CUSUM Test.

This formulation states that

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

Since 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;

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

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 (7-15). 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 (7-15). This implies that the F statistic for testing the null hypothesis in this model is precisely that which appeared earlier in (7-15), which suggests why the test is labeled the "predictive test."

7.5.4 UNKNOWN TIMING OF THE STRUCTURAL BREAK[16]

The testing procedures described in this section all assume that the point of the structural break is known. When this corresponds to a discrete historical event, this is a reasonable assumption. But, in some applications, the timing of the break may be unknown. The Chow and Wald tests become useless at this point. The CUSUMS test is a step in the right direction for this situation, but, as noted by a number of authors [e.g., Andrews (1993)] it has serious power problems. Recent research has provided several strategies for testing for structural change when the change point is unknown.

In Section 7.4 we considered a test of parameter equality in two populations. The natural approach suggested there was a comparison of two separately estimated parameter vectors based on the Wald criterion,

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

where 1 and 2 denote the two populations. An alternative approach to the testing procedure is based on a likelihood ratio-like statistic,

$$\lambda = h[(L_1 + L_2), L]$$

where $L_1 + L_2$ is the log likelihood function (or other estimation criterion) under the alternative hypothesis of model instability (structural break) and L is the log likelihood for the pooled estimator based on the null hypothesis of stability and h is the appropriate function of the values, such as $h(a, b) = -2(b - a)$ for maximum likelihood estimation. A third approach, based on the Lagrange multiplier principle, will be developed below. There is a major problem with this approach; the split between the two subsamples must be known in advance. In the time series application we will examine in this section, the problem to be analyzed is that of determining whether a model can be claimed to be stable through a sample period $t = 1, \ldots, T$ against the alternative hypothesis that the structure changed at some *unknown* time t^*. Knowledge of the sample split is crucial for the tests suggested above, so some new results are called for.

We suppose that the *model* $E[\mathbf{m}(y_t, \mathbf{x}_t \mid \boldsymbol{\beta})] = \mathbf{0}$ is to be estimated by GMM using T observations. The model is stated in terms of a moment condition, but we intend for this to include estimation by maximum likelihood, or linear or nonlinear least squares. As noted earlier, all these cases are included. Assuming GMM just provides us a convenient way to analyze all the cases at the same time. The hypothesis to be investigated is as follows: Let $[\pi T] = T_1$ denote the integer part of πT where $0 < \pi < 1$. Thus, this is a proportion π of the sample observations, and defines subperiod 1, $t = 1, \ldots, T_1$. Under the null hypothesis, the model $E[\mathbf{m}(y_t, \mathbf{x}_t \mid \boldsymbol{\beta})] = \mathbf{0}$ is stable for the entire sample period. Under the alternative hypothesis, the model $E[\mathbf{m}(y_t, \mathbf{x}_t \mid \boldsymbol{\beta}_1)] = \mathbf{0}$ applies to

[16]The material in this section is more advanced than that in the discussion thus far. It may be skipped at this point with no loss in continuity. Since this section relies heavily on GMM estimation methods, you may wish to read Chapter 18 before continuing.

observations $1, \dots, [\pi T]$ and model $E[\mathbf{m}(y_t, \mathbf{x}_t \mid \boldsymbol{\beta}_2)] = \mathbf{0}$ applies to the remaining $T - [\pi T]$ observations.[17] This describes a nonstandard sort of hypothesis test since under the null hypothesis, the 'parameter' of interest, π, is not even part of the model. Andrews and Ploberger (1994) denote this a "nuisance parameter [that] is present only under the alternative."

Suppose π were known. Then, the optimal GMM estimator for the first subsample would be obtained by minimizing with respect to the parameters $\boldsymbol{\beta}_1$ the criterion function

$$\begin{aligned} q_1(\pi) &= \bar{\mathbf{m}}_1'(\pi \mid \boldsymbol{\beta}_1)[\text{Est.Asy. Var}\sqrt{[\pi T]}\bar{\mathbf{m}}_1'(\pi \mid \boldsymbol{\beta}_1)]^{-1}\bar{\mathbf{m}}_1(\pi \mid \boldsymbol{\beta}_1) \\ &= \bar{\mathbf{m}}_1'(\pi \mid \boldsymbol{\beta}_1)[\mathbf{W}_1(\pi)]^{-1}\bar{\mathbf{m}}_1(\pi \mid \boldsymbol{\beta}_1) \end{aligned}$$

where

$$\bar{\mathbf{m}}_1(\pi \mid \boldsymbol{\beta}_1) = \frac{1}{[\pi T]}\sum_{t=1}^{[\pi T]} \mathbf{m}_t(y_t, \mathbf{x}_t \mid \boldsymbol{\beta}_1).$$

The asymptotic covariance (weighting) matrix will generally be computed using a first round estimator in

$$\hat{\mathbf{W}}_1(\pi) = \frac{1}{[\pi T]}\sum_{t=1}^{[\pi T]} \mathbf{m}_t(\pi \mid \hat{\boldsymbol{\beta}}_1^0)\mathbf{m}_t'(\pi \mid \hat{\boldsymbol{\beta}}_1^0). \tag{7-21}$$

In this time-series setting, it would be natural to accommodate serial correlation in this estimator. Following Hall and Sen (1999), the counterpart to the Newey-West (1987a) estimator (see Section 11.3) would be

$$\hat{\mathbf{W}}_1(\pi) = \hat{\mathbf{W}}_{1,0}(\pi) + \sum_{j=1}^{B(T)} w_{j,T}\left[\hat{\mathbf{W}}_{1,j}(\pi) + \hat{\mathbf{W}}_{1,j}'(\pi)\right]$$

where $\hat{\mathbf{W}}_{1,0}(\pi)$ is given in (7-21) and

$$\hat{\mathbf{W}}_{1,j}(\pi) = \frac{1}{[\pi T]}\sum_{t=j+1}^{[\pi T]} \mathbf{m}_t(\pi \mid \hat{\boldsymbol{\beta}}_1^0)\mathbf{m}_{t-j}'(\pi \mid \hat{\boldsymbol{\beta}}_1^0).$$

$B(T)$ is the bandwidth, chosen to be $O(T^{1/4})$—this is the L in (10-16) and (12-17)—and $w_{j,T}$ is the kernel. Newey and West's value for this is the Bartlett kernel, $[1 - j/(1 + B(T))]$. (See, also, Andrews (1991), Hayashi (2000, pp. 408–409) and the end of Section C.3.) The asymptotic covariance matrix for the GMM estimator would then be computed using

$$\text{Est.Asy. Var}[\hat{\boldsymbol{\beta}}_1] = \frac{1}{[\pi T]}\left[\hat{\bar{\mathbf{G}}}_1'(\pi)\hat{\mathbf{W}}_1^{-1}(\pi)\hat{\bar{\mathbf{G}}}_1(\pi)\right]^{-1} = \hat{\mathbf{V}}_1$$

[17] Andrews (1993), on which this discussion draws heavily, allows for some of the parameters to be assumed to be constant throughout the sample period. This adds some complication to the algebra involved in obtaining the estimator, since with this assumption, efficient estimation requires joint estimation of the parameter vectors, whereas our formulation allows GMM estimation to proceed with separate subsamples when needed. The essential results are the same.

where

$$\bar{\hat{\mathbf{G}}}_1(\pi) = \frac{1}{[\pi T]} \sum_{t=1}^{[\pi T]} \frac{\partial \mathbf{m}_t(\pi \mid \hat{\boldsymbol{\beta}}_1)}{\partial \hat{\boldsymbol{\beta}}_1'}.$$

Estimators for the second sample are found by changing the summations to $[\pi T] + 1, \ldots T$ and for the full sample by summing from 1 to T.

Still assuming that π is known, the three standard test statistics for testing the null hypothesis of model constancy against the alternative of structural break at $[\pi T]$ would be as follows: The Wald statistic is

$$W_T(\pi) = [\hat{\boldsymbol{\beta}}_1(\pi) - \hat{\boldsymbol{\beta}}_2(\pi)]'\{\hat{\mathbf{V}}_1(\pi) + \hat{\mathbf{V}}_2(\pi)\}^{-1}[\hat{\boldsymbol{\beta}}_1(\pi) - \hat{\boldsymbol{\beta}}_2(\pi)],$$

[See Andrews and Fair (1988).] There is a small complication with this result in this time-series context. The two subsamples are generally not independent so the additive result above is not quite appropriate. Asymptotically, the number of observations close to the switch point, if there is one, becomes small, so this is only a finite sample problem. The likelihood ratio-like statistic would be

$$LR_T(\pi) = [q_1(\pi \mid \hat{\boldsymbol{\beta}}_1) + q_2(\pi \mid \hat{\boldsymbol{\beta}}_2)] - [q_1(\pi \mid \hat{\boldsymbol{\beta}}) + q_2(\pi \mid \hat{\boldsymbol{\beta}})]$$

where $\hat{\boldsymbol{\beta}}$ is based on the full sample. (This result makes use of our assumption that there are no common parameters so that the criterion for the full sample is the sum of those for the subsamples. With common parameters, it becomes slightly more complicated.) The Lagrange multiplier statistic is the most convenient of the three. All matrices with subscript "T" are based on the full sample GMM estimator. The weighting and derivative matrices are computed using the full sample. The moment equation is computed at the first subsample [though the sum is divided by T not $[\pi T]$—see Andrews (1993, eqn. (4.4)];

$$LM_T(\pi) = \frac{T}{\pi(1-\pi)} \bar{\mathbf{m}}_1(\pi \mid \hat{\boldsymbol{\beta}}_T)'\hat{\mathbf{V}}_T^{-1}\bar{\hat{\mathbf{G}}}_T\left[\bar{\hat{\mathbf{G}}}_T'\hat{\mathbf{V}}_T^{-1}\bar{\hat{\mathbf{G}}}_T\right]^{-1}\bar{\hat{\mathbf{G}}}_T'\hat{\mathbf{V}}_T^{-1}\bar{\mathbf{m}}_1(\pi \mid \hat{\boldsymbol{\beta}}_T).$$

The LM statistic is simpler, as it requires the model only to be estimated once, using the full sample. (Of course, this is a minor virtue. The computations for the full sample and the subsamples are the same, so the same amount of setup is required either way.) In each case, the statistic has a limiting chi-squared distribution with K degrees of freedom where K is the number of parameters in the model.

Since π is unknown, the preceding does not solve the problem posed at the outset. The CUSUMS and Hansen tests discussed in Section 7.5 were proposed for that purpose, but lack power and are generally for linear regression models. Andrews (1993) has derived the behavior of the test statistic obtained by computing the statistics suggested previously at the range of candidate values, that is the different partitionings of the sample say $\pi_0 = .15$ to $.85$, then retaining the maximum value obtained. These are the Sup $W_T(\pi)$, Sup $LR_T(\pi)$ and Sup $LM_T(\pi)$, respectively. Although for a given π, the statistics have limiting chi-squared distributions, obviously, the maximum does not. Tables of critical values obtained by Monte Carlo methods are provided in Andrews (1993). An interesting side calculation in the process is to plot the values of the test statistics. (See the following application.) Two alternatives to the supremum test are suggested by Andrews and Ploberger (1994) and Sowell (1996). The average statistics, Avg $W_T(\pi)$, Avg $LR_T(\pi)$ and Avg $LM_T(\pi)$ are computed by taking the sample average of the sequence of values over the R partitions of the sample from $\pi = \pi_0$ to $\pi = 1 - \pi_0$. The exponential statistics are computed as

$$\text{Exp } W_T(\pi) = \ln\left[\frac{1}{R}\sum_{r=1}^{R} \exp[.5W_T(\pi_r)]\right]$$

and likewise for the LM and LR statistics. Tables of critical values for a range of values of π_0 and K are provided by the authors.[18]

Not including the Hall and Sen approaches, the preceding provides nine different statistics for testing the hypothesis of parameter constancy—though Andrews and Ploberger (1994) suggest that the Exp LR and Avg LR versions are less than optimal. As the authors note, all are based on statistics which converge to chi-squared statistics. Andrews and Ploberger present some results to suggest that the exponential form may be preferable based on its power characteristics.

In principle the preceding suggests a maximum likelihood estimator of π (or T_1) if ML is used as the estimation method. Properties of the estimator are difficult to obtain, as shown in Bai (1997). Moreover, Bai's (1997) study based on least squares estimation of a linear model includes some surprising results that suggest that in the presence of multiple change points in a sample, the outcome of the Andrews and Ploberger tests may depend crucially on what time interval is examined.[19]

Example 7.7 Instability of the Demand for Money

We will examine the demand for money in some detail in Chapters 19 and 20. At this point, we will take a cursory look at a simple (and questionable) model

$$(m - p)_t = \alpha + \beta y_t + \gamma i_t + \varepsilon_t$$

where m, p, and y are the logs of the money supply (M1), the price level (CPI_U) and GDP, respectively, and i is the interest rate (90-day T-bill rate) in our data set. Quarterly data on these and several other macroeconomic variables are given in Appendix F5.1 for the quarters 1950.1 to 2000.4. We will apply the techniques described above to this money demand equation. The data span 204 quarters. We chose a window from 1957.3 (quarter 30) to 1993.3 (quarter 175), which correspond roughly to $\pi = .15$ to $\pi = .85$. The function is estimated by GMM using as instruments $\mathbf{z}_t = [1, i_t, i_{t-1}, y_{t-1}\, y_{t-2}]$. We will use a Newey–West estimator for the weighting matrix with $L = 204^{1/4} \approx 4$, so we will lose 4 additional

[18] An extension of the Andrews and Ploberger methods based on the overidentifying restrictions in the GMM estimator is developed in Hall and Sen (1999). Approximations to the critical values are given by Hansen (1997). Further results are given in Hansen (2000).

[19] Bai (1991), Bai, Lumsdaine and Stock (1999), Bai and Perron (1998a,b) and Bai (1997). "Estimation" of π or T_1 raises a peculiarity of this strand of literature. In many applications, the notion of a change point is tied to an historical event, such as a war or a major policy shift. For example, in Bai (1997, p. 557), a structural change in an estimated model of the relationship between T-bill rates and the Fed's discount rate is associated with a specific date, October 9, 1979, a date which marked the beginning of a change in Fed operating procedures. A second change date in his sample was associated with the end of that Fed policy regime while a third between these two had no obvious identity. In such a case, the idea of a fixed π requires some careful thought as to what is meant by $T \to \infty$. If the sampling process is defined to have a true origin in a physical history, wherever it is, then π cannot be fixed. As T increases, π must decline to zero and "estimation" of π makes no sense. Alternatively, if π really is meant to denote a specific proportion of the sample, but remains tied to an actual date, then presumably, increasing the sample size means shifting both origin and terminal in opposite directions, at the same rate. Otherwise, insisting that the regime switch occur at time πT has an implausible economic implication. Changing the orientation of the search to the change date, T_1, itself, does not remove the ambiguities. We leave the philosophical resolution of either interpretation to the reader. Andrews' (1993, p. 845) assessment of the situation is blunt: "[n]o optimality properties are known for the ML estimator of π."

TABLE 7.7 Results of Model Stability Tests

Statistic	*Maximum*	*Average*	*Average exp*
LM	10.43	4.42	3.31
Wald	11.85	4.57	3.67
LR	15.69	—	—
Critical Value	14.15[a]	4.22[b]	6.07[c]

[a]Andrews (1993), Table I, $p = 3, \pi_0 = 0.15$.
[b]Andrews and Ploberger (1994), Table II, $p = 3, \pi_0 = 0.15$.
[c]Andrews and Ploberger (1994), Table I, $p = 3, \pi_0 = 0.15$.

observations after the two lagged values in the instruments. Thus, the estimation sample is 1951.3 to 2000.4, a total of 197 observations.

The GMM estimator is precisely the instrumental variables estimator shown in Chapter 5. The estimated equation (with standard errors shown in parentheses) is

$$(m - p)_t = -1.824\,(0.166) + 0.306\,(0.0216)\, y_t - 0.0218\,(0.00252)\, i_t + e_t.$$

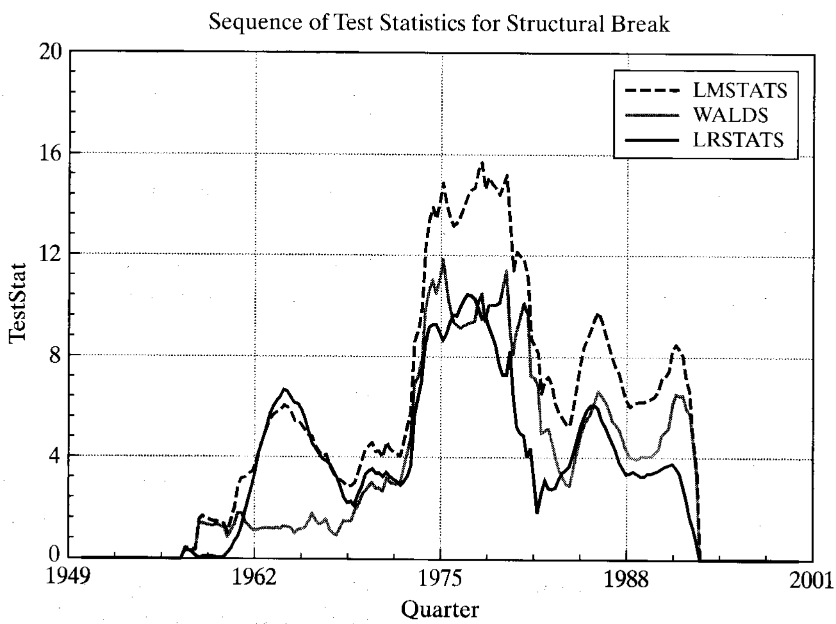
The Lagrange multiplier form of the test is particularly easy to carry out in this framework. The sample moment equations are

$$E\,[\bar{\mathbf{m}}_T] = E\left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_t(y_t - \mathbf{x}_t'\beta)\right] = \mathbf{0}.$$

The derivative matrix is likewise simple; $\bar{\mathbf{G}} = -(1/T)\mathbf{Z}'\mathbf{X}$. The results of the various testing procedures are shown in Table 7.7.

The results are mixed; some of the statistics reject the hypothesis while others do not. Figure 7.7 shows the sequence of test statistics. The three are quite consistent. If there is a structural break in these data, it occurs in the late 1970s. These results coincide with Bai's findings discussed in the preceding footnote.

FIGURE 7.7 Structural Change Test Statistics.



7.6 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. These tests rely on information about when (or how) the sample is to be partitioned for the test. In many time series cases, this is unknown. Tests designed for this more complex case were considered in Section 7.5.4.

Key Terms and Concepts

- Binary variable
- Chow test
- CUSUM test
- Dummy variable
- Dummy variable trap
- Exactly identified
- Hansen's test
- Identification condition
- Interaction term
- Intrinsically linear
- Knots
- Loglinear model
- Marginal effect
- Nonlinear restriction
- One step ahead prediction error
- Overidentified
- Piecewise continuous
- Predictive test
- Qualification indices
- Recursive residual
- Response
- Semilog model
- Spline
- Structural change
- Threshold effect
- Time profile
- Treatment
- Wald test

Exercises

1. 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 F7.2. 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.]
2. 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.

   a. Compute a scatter diagram of q/A against k.
   b. 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.]

c. 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.

3. 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
e′e	104	88	206	144	199	308	211	1425

4. *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

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

where

$\bar{y}_1$ = mean of y for observations with $d = 1$,
$\bar{y}$ = mean of y for all observations,
P = mean of d,
R^2 = coefficient of determination for (1),
r_{yd}^2 = 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.

5. *Reverse regression continued.* This and the next exercise continue the analysis of Exercise 4. In Exercise 4, 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_u^2, and σ_ε^2, 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?

6. *Reverse regression continued.* Suppose that the model in Exercise 5 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

TABLE 7.8 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).

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 $(\bar{y} \mid d=1) - (\bar{y} \mid d = 0)$? What does this quantity estimate if γ does equal zero?

7. Data on the number of incidents of damage to a sample of ships, with the type of ship and the period when it was constructed, are given in the Table 7.8. 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."