6.3.5 INTRINSICALLY LINEAR MODELS

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

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

or

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

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

This fact leads us to a useful aspect of intrinsically linear models; they have an "invariance property." Using the nonlinear least squares procedure described in ~~Section 7.3.2~~ the next chapter, we could estimate α and β_2 directly by minimizing the sum of squares function;

$$\text{Minimize with respect to } (\alpha,\beta_2)\text{: } S(\alpha,\beta_2) = \sum_{i=1}^{n} \left(\ln Y_i - \ln \alpha - \beta_2 \ln X_i\right)^2. \quad \cancel{(7\text{-}3)} \quad (6\text{-}8)$$

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

$$\text{Minimize with respect to } (\beta_1,\beta_2)\text{: } S(\beta_1,\beta_2) = \sum_{i=1}^{n} \left(y_i - \beta_1 - \beta_2 x_i\right)^2, \quad \cancel{(7\text{-}4)} \quad (6\text{-}9)$$

is simple to solve with the least squares estimator we have used up to this point. The invariance feature that applies is that the two sets of results will be numerically identical; we will get the identical result from estimating α using ~~(7-3)~~ (6-8) and from using $\exp(\beta_1)$ from (6-9) ~~(7-4)~~. By exploiting this result, we can broaden the definition of linearity and include some additional cases that might otherwise be quite complex.

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. $\hat{\alpha} = e^{\beta}$

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.

6.1

DEFINITION 6.1 Intrinsic Linearity

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

Example ~~6.4~~ 6.7 Intrinsically Linear Regression

In Section ~~16~~.6.4 14, 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. Because this value is a nonlinear function of the estimated parameters, we use the delta method to estimate the standard error. Using the data from that example,[9] the least squares estimates of β_1 and β_2 (with standard errors in parentheses) are -4.1431 (23.734) and 2.4261 (1.5915). The estimated covariance is -36.979. The estimate of β is $-4.1431/2.4261 = -1.7077$. We estimate the sampling variance of $\hat{\beta}$ with

FN 9 sup 9

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

$$= 8.6889^2.$$

TB 6.6

Table 6.~~5~~6 compares the least squares and maximum likelihood estimates of the parameters. The lower standard errors for the maximum likelihood estimates result from the inefficient (equal) weighting given to the observations by the least squares procedure. The gamma distribution is highly skewed. In addition, we know from our results in Appendix C that this distribution is an exponential family. We found for the gamma distribution that the sufficient statistics for this density were $\Sigma_i y_i$ and $\Sigma_i \ln y_i$. The least squares estimator does not use the second of these, whereas an efficient estimator will.

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

6-28

6.6

TABLE ~~6.5~~ 8 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~~ 2.345	3.151	~~0.663~~ 0.794

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, and 1967).~~

Example 6.5 CES Production Function

The constant elasticity of substitution production function may be written

$$\ln y = \ln \gamma - \frac{\nu}{\rho} \ln[\delta K^{-\rho} + (1-\delta)L^{-\rho}] + \varepsilon. \tag{6-9}$$

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

$$\begin{aligned} \ln y &= \ln \gamma + \nu\delta \ln K + \nu(1-\delta)\ln L + \rho\nu\delta(1-\delta)\left\{-\tfrac{1}{2}[\ln K - \ln L]^2\right\} + \varepsilon' \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon', \end{aligned} \tag{6-10}$$

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

$$\begin{aligned} &\beta_1 = \ln \gamma, \quad \beta_2 = \nu\delta, \quad \beta_3 = \nu(1-\delta), \quad \beta_4 = \rho\nu\delta(1-\delta), \\ &\gamma = e^{\beta_1}, \quad \delta = \beta_2/(\beta_2+\beta_3), \quad \nu = \beta_2+\beta_3, \quad \rho = \beta_4(\beta_2+\beta_3)/(\beta_2\beta_3). \end{aligned} \tag{6-11}$$

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 (6-11) are the same as those we would obtain had we found the nonlinear least squares estimates of (6-10) directly. As Kmenta shows, however, they are not the same as the nonlinear least squares estimates of (6-9) due to the use of the Taylor series approximation to get to (6-10). 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

$$y_i = \beta_1(\theta)x_{i1} + \beta_2(\theta)x_{i2} + \cdots + \beta_K(\theta)x_{ik} + \varepsilon_i \tag{6-~~12~~9}$$

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,

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

6-29

TABLE 6.5 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 (θ) are said to be **exactly identified** in terms of the parameters of the linear model β. An excellent example is provided by Kmenta (1986, p. 515, and 1967).

KT

Example 6.~~5~~8 CES Production Function

The constant elasticity of substitution production function may be written

$$\ln y = \ln \gamma - \frac{\nu}{\rho}\ln[\delta K^{-\rho} + (1-\delta)L^{-\rho}] + \varepsilon. \qquad (6\text{-}\cancel{9}10)$$

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

$$\begin{aligned}\ln y &= \ln\gamma + \nu\delta\ln K + \nu(1-\delta)\ln L + \rho\nu\delta(1-\delta)\left\{-\tfrac{1}{2}[\ln K - \ln L]^2\right\} + \varepsilon' \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon',\end{aligned} \qquad (6\text{-}\cancel{10}11)$$

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

$$\begin{aligned}&\beta_1 = \ln\gamma, \quad \beta_2 = \nu\delta, \quad \beta_3 = \nu(1-\delta), \quad \beta_4 = \rho\nu\delta(1-\delta), \\ &\gamma = e^{\beta_1}, \quad \delta = \beta_2/(\beta_2+\beta_3), \quad \nu = \beta_2+\beta_3, \quad \rho = \beta_4(\beta_2+\beta_3)/(\beta_2\beta_3).\end{aligned} \qquad (6\text{-}\cancel{11}12)$$

6-12

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 (6-11) are the same as those we would obtain had we found the nonlinear least squares estimates of (6-10) directly. As Kmenta shows, however, they are not the same as the nonlinear least squares estimates of (6-9) due to the use of the Taylor series approximation to get to (6-10). We would use the delta method to construct the estimated asymptotic covariance matrix for the estimates of $\theta' = [\gamma, \delta, \nu, \rho]$. The derivatives matrix is 6-11 6-10

(6-11)

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

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

Not all models of the form

$$y_i = \beta_1(\theta)x_{i1} + \beta_2(\theta)x_{i2} + \cdots + \beta_K(\theta)x_{ik} + \varepsilon_i \qquad (6\text{-}1\cancel{2}3)$$

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,

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

6-30

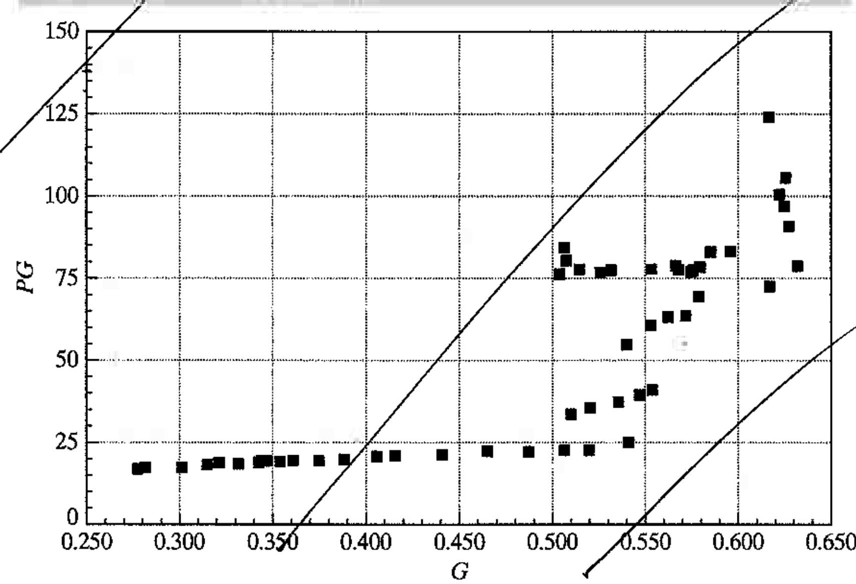
is nonlinear. The reason is that if we write it in the form of (6-1~~2~~3), we fail to account for the condition that β_4 equals $\beta_2\beta_3$, which is a **nonlinear restriction.** In this model, the three parameters α, β, and γ are **overidentified** in terms of the four parameters β_1, β_2, β_3, and β_4. Unrestricted least squares estimates of β_2, β_3, and β_4 can be used to obtain two estimates of each of the underlying parameters, and there is no assurance that these will be the same. Models that are not intrinsically linear are treated in Chapter 7.

KT

6.4 MODELING AND TESTING FOR A STRUCTURAL BREAK

One of the more common applications of the F test is in tests of **structural change.**[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 or all of the regression coefficients are different in different subsets of the data. To analyze a number of examples, we will revisit the data on the U.S. gasoline market that we examined in Examples 2.3, 4.4, 4.7, and 4.8. As Figure 6.5 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.[9] It seems unlikely that the same regression model would apply to both periods.

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



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

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

6-31

is nonlinear. The reason is that if we write it in the form of (6-12), 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 β_1, β_2, β_3, and β_4. Unrestricted least squares estimates of β_2, β_3, and β_4 can be used to obtain two estimates of each of the underlying parameters, and there is no assurance that these will be the same.

6.4 MODELING AND TESTING FOR A STRUCTURAL BREAK

FN 10

FIG 6.5

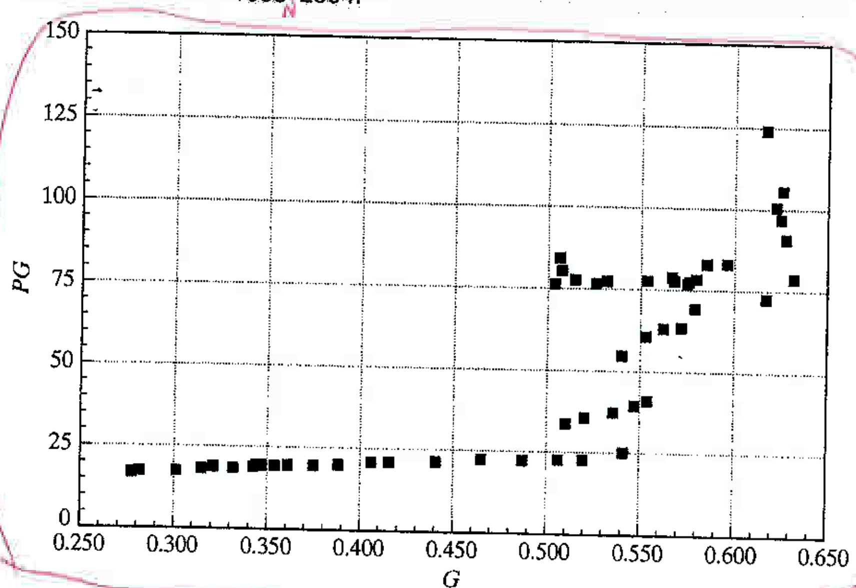
FN 11

AU: KT "structural change" was KT earlier in chap. KT again here?

One of the more common applications of the F test is in tests of **structural change.**[8] 10 In specifying a regression model, we assume that its assumptions apply to all the observations in our sample. It is straightforward, however, to test the hypothesis that some or all of the regression coefficients are different in different subsets of the data. To analyze a number of examples, we will revisit the data on the U.S. gasoline market that we examined in Examples 2.3, 4.4, 4.7, and 4.8. As Figure 6.5 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.[9] 11 It seems unlikely that the same regression model would apply to both periods.

4.2, 4.4, 4.8 and 4.9

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



10 [8]This test is often labeled a **Chow test,** in reference to Chow (1960). KT

11 [9]The observed data will doubtless reveal similar disruption in 2006.

6-32

6.4.1 DIFFERENT PARAMETER VECTORS

The gasoline consumption data span two very different periods. Up to 1973, fuel was plentiful and world prices for gasoline had been stable or falling for at least two decades. The embargo of 1973 marked a transition in this market, marked by shortages, rising prices, and intermittent turmoil. It is possible that the entire relationship described by our regression model changed in 1974. To test this as a hypothesis, we could proceed as follows: Denote the first 21 years of the data in $\mathbf{y}$ and $\mathbf{X}$ as $\mathbf{y}_1$ and $\mathbf{X}_1$ and the remaining years as $\mathbf{y}_2$ and $\mathbf{X}_2$. An unrestricted regression that allows the coefficients to be different in the two periods is

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix}. \tag{6-13}$$

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

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1'\mathbf{y}_1\\ \mathbf{X}_2'\mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix}, \tag{6-14}$$

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 squares from the two separate regressions:

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

The restricted coefficient vector can be obtained in two ways. Formally, the restriction $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ is $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, where $\mathbf{R} = [\mathbf{I} : -\mathbf{I}]$ and $\mathbf{q} = \mathbf{0}$. The general result given earlier can be applied directly. An easier way to proceed is to build the restriction directly into the model. If the two coefficient vectors are the same, then (6-13) 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 (5-6), where J, the number of restrictions, is the number of columns in $\mathbf{X}_2$ and the denominator degrees of freedom is $n_1 + n_2 - 2k$.

(5-16)

6.4.2 INSUFFICIENT OBSERVATIONS

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

6-33

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

$$F[n_2, n_1 - K] = \frac{(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}_1'\mathbf{e}_1)/n_2}{\mathbf{e}_1'\mathbf{e}_1/(n_1 - K)}. \qquad \text{(6-1}\cancel{5}\text{6)}$$

FN 12

Note that the numerator degrees of freedom is n_2, not K.[~~10~~ 12] This test has been labeled the Chow **predictive test** because it is equivalent to extending the restricted model to the shorter subperiod and basing the test on the prediction errors of the model in this latter period.

6.4.3 CHANGE IN A SUBSET OF COEFFICIENTS

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

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

and

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

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

AU: Do you want text callout for TB 6.7 here or on msp 6-36?

TB 6.7

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 6.7 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~~ 12] 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

$$\mathbf{X} = \begin{bmatrix} \mathbf{i}_{\text{pre}} & \mathbf{Z}_{\text{pre}} & \mathbf{0} & \mathbf{0} & \mathbf{W}_{\text{pre}} \\ \mathbf{0} & \mathbf{0} & \mathbf{i}_{\text{post}} & \mathbf{Z}_{\text{post}} & \mathbf{W}_{\text{post}} \end{bmatrix}. \qquad (6\text{-}16)$$

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

6.4.4 TESTS OF STRUCTURAL BREAK WITH UNEQUAL VARIANCES

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

If the sample size is reasonably large, then we have a test that is valid whether or not the disturbance variances are the same. Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ 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,

FN 13

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

Under the null hypothesis, the Wald statistic,

$$W = (\hat{\theta}_1 - \hat{\theta}_2)'(\hat{\mathbf{V}}_1 + \hat{\mathbf{V}}_2)^{-1}(\hat{\theta}_1 - \hat{\theta}_2), \qquad (6\text{-}17)$$

FN 14

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.

6 -35

Ohtani and Kobayashi (1986) have devised a "bounds" test that gives a partial remedy for the problem.[15]

FN 15

KT

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.

Example 6.6 The World Health Report

The 2000 version of the World Health Organization's (WHO) *World Health Report* contained a major country-by-country inventory of the world's health care systems. [World Health Organization (2000). See also http://www.who.int/whr/en/.] The book documented years of research and thousands of pages of material. Among the most controversial and most publicly debated (and excoriated) parts of the report was a single chapter that described a comparison of the delivery of health care by 191 countries—nearly all of the world's population. [Evans et al. (2000a,b). See, e.g., Hilts (2000) for reporting in the popular press.] The study examined the efficiency of health care delivery on two measures: the standard one that is widely studied, (disability adjusted) life expectancy (DALE), and an innovative new measure created by the authors that was a composite of five outcomes (COMP) and that accounted for efficiency and fairness in delivery. The regression-style modeling, which was done in the setting of a frontier model (see Section 14.6.3), related health care attainment to two major inputs, education and (per capita) health care expenditure. The residuals were analyzed to obtain the country comparisons.

The data in Appendix Table F6.2 were used by the researchers at WHO for the study. (They used a panel of data for the years 1993 to 1997. We have extracted the 1997 data for this example.) The WHO data have been used by many researchers in subsequent analyses. [See, e.g., Hollingsworth and Wildman (2002), Gravelle et al. (2002), and Greene (2004).] The regression model used by the WHO contained DALE or COMP on the left-hand side and health care expenditure, education, and education squared on the right. Greene (2004) added a number of additional variables such as per capita GDP, a measure of the distribution of income, and World Bank measures of government effectiveness and democratization of the political structure.

Among the controversial aspects of the study was the fact that the model aggregated countries of vastly different characteristics. A second striking aspect of the results, suggested in Hilts (2000) and documented in Greene (2004), was that, in fact, the "efficient" countries in the study were the 30 relatively wealthy OECD members, while the rest of the world on average fared much more poorly. We will pursue that aspect here with respect to DALE. Analysis of COMP is left as an exercise. Table 6.6 presents estimates of the regression models for DALE for the pooled sample, the OECD countries, and the non-OECD countries, respectively. Superficially, there do not appear to be very large differences across the two subgroups. We first tested the joint significance of the additional variables, income distribution (GINI), per capita GDP, etc. For each group, the F statistic is $[(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e})/7]/[\mathbf{e}'\mathbf{e}/(n-11)]$. These F statistics are shown in the last row of the table. The critical values for F[7,184] (all), F[7,23] (OECD), and F[7,154] (non-OECD) are 2.060, 2.442, and 2.070, respectively. We conclude that the additional explanatory variables are significant contributors to the fit for the non-OECD countries (and for all countries), but not for the OECD countries. Finally, to conduct

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

Example 6.9 Structural Break in the Gasoline Market

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

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

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

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

minus

TB 6.7

minus

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

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

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

minus

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

minus

once again. (See Section 6.4.2 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:

minus

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

minus

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

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

minus

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

minus

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

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

TABLE 6.7 Gasoline Consumption Functions

Coefficients	*1953–2004*	*Pooled*	*Preshock*	*Postshock*
Constant	−26.6787	−24.9009	−22.1647	
Constant		−24.8167	0	−15.3283
ln *Income/Pop*	1.6250	1.4562	.8482	0.3739
ln *PG*	−0.05392	−0.1132	−0.03227	−0.1240
ln *PNC*	−0.08343	−0.1044	0.6988	−0.001146
ln *PUC*	−0.08467	−0.08646	−0.2905	−0.02167
Year	−0.01393	−0.009232	0.01006	0.004492
R^2	0.9649	0.9683	0.9975	0.9529
Standard error	0.04709	0.04524	0.01161	0.01689
Sum of squares	0.101997	0.092082	0.00202244	0.007127899

Example 6.10 The World Health Report

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

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

Au: Has OECD been spelled out before? If not, spell it out here.

Among the controversial aspects of the study was the fact that the model aggregated countries of vastly different characteristics. A second striking aspect of the results, suggested in Hilts (2000) and documented in Greene (2004), was that, in fact, the "efficient" countries in the study were the 30 relatively wealthy OECD members, while the rest of the world on average fared much more poorly. We will pursue that aspect here with respect to DALE. Analysis of COMP is left as an exercise. Table 6.8 presents estimates of the regression models for DALE for the pooled sample, the OECD countries, and the non-OECD countries, respectively. Superficially, there do not appear to be very large differences across the two subgroups. We first tested the joint significance of the additional variables, income distribution (GINI), per capita GDP, etc. For each group, the F statistic is $[(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e})/7]/[\mathbf{e}'\mathbf{e}/(n-11)]$. These F statistics are shown in the last row of the table. The critical values for F[7,180] (all), F[7,19] (OECD), and F[7,150] (non-OECD) are 2.061, 2.543, and 2.071, respectively. We conclude that the additional explanatory variables are significant contributors to the fit for the non-OECD countries (and for all countries), but not for the OECD countries. Finally, to conduct the structural change test of OECD vs. non-OECD, we computed

and so on

Au: OK to spell out "etc." in text?

$$F[11,169] = \frac{[7757/007 - (69.74428 + 7378.598)]/11}{(69.74428 + 7378.598)/(191-11-11)} = 0.637.$$

The 95 percent critical value for $F[11,169]$ is 1.846. So, we do not reject the hypothesis that the regression model is the same for the two groups of countries. The Wald statistic in (6-17) tells a different story. The statistic is 35.221. The 95 percent critical value from the chi-squared table with 11 degrees of freedom is 19.675. On this basis, we would reject the hypothesis that the two coefficient vectors are the same.

6-39

TABLE 6.8 Regression Results for Life Expectancy

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

align columns on decimal point

~~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.~~

6.4.5 PREDICTIVE TEST OF MODEL STABILITY

The hypothesis test defined in (6-1~~5~~) in Section 6.4.2 is equivalent to $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_1$ in the "model"
6

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

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

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

This formulation states that

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

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

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

where $\mathbf{b}_1$ is the least squares slopes based on the first T_1 observations and $\mathbf{c}_2$ is $\mathbf{y}_2 - \mathbf{X}_2\mathbf{b}_1$. The covariance matrix for the full set of estimates is s^2 times the bracketed matrix. The two subvectors of residuals in this regression are $\mathbf{e}_1 = \mathbf{y}_1 - \mathbf{X}_1\mathbf{b}_1$ and $\mathbf{e}_2 = \mathbf{y}_2 - (\mathbf{X}_2\mathbf{b}_1 + \mathbf{I}\mathbf{c}_2) = \mathbf{0}$, so the sum of squared residuals in this least squares regression is just $\mathbf{e}_1'\mathbf{e}_1$. This is the same sum of squares as appears in (6-15). 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 (6-15). This implies

(6-16)

6-41

that the F statistic for testing the null hypothesis in this model is precisely that which appeared earlier in (6-18), which suggests why the test is labeled the "predictive test."

6

6.5 SUMMARY AND CONCLUSIONS

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

Key Terms and Concepts

- Binary variable
- Chow test
- Dummy variable
- Dummy variable trap
- Exactly identified
- Identification condition
- Interaction term
- Intrinsically linear
- Knots
- Loglinear model
- Marginal effect
- Nonlinear restriction
- Overidentified
- Piecewise continuous
- Predictive test
- Qualification indices
- Response
- Semilog model
- Spline
- Structural change
- Threshold effect
- Time profile
- Treatment
- Wald test

Exercises

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

	1954	*1955*	*1956*	*1957*	*1958*	*1959*	*1960*	*All*
Observations	65	55	87	95	103	87	78	570
e′e	104	88	206	144	199	308	211	1425

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

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

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,

6-42

that the F statistic for testing the null hypothesis in this model is precisely that which appeared earlier in (6-15), which suggests why the test is labeled the "predictive test."

6.5 SUMMARY AND CONCLUSIONS

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

Au: The KTs "Semilog model" and "Threshold effect" were not bold KTs in text. Mark bold in text or delete from list

Key Terms and Concepts

- Binary variable
- Chow test
- Dummy variable
- Dummy variable trap
- Exactly identified
- Identification condition
- Interaction term
- Intrinsically linear
- Knots
- Loglinear model
- Marginal effect
- Nonlinear restriction
- Overidentified
- Piecewise continuous
- Predictive test
- Qualification indices
- Response
- Semilog model
- Spline
- Structural change
- Threshold effect
- Time profile
- Treatment
- Wald test

Exercises

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

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

Au: Exercise titles in earlier chap. were bold Roman. Here also?

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

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

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

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

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 $\mathbf{y} = a\mathbf{i} + \mathbf{X}\mathbf{b} + c\mathbf{d} + \mathbf{e}$, so $\mathbf{q} = \mathbf{y} - c\mathbf{d} - \mathbf{e}$. From here on, we drop the constant term. Thus, in the regression in (3) you are regressing $[\mathbf{y} - c\mathbf{d} - \mathbf{e}]$ on $\mathbf{y}$ and $\mathbf{d}$.

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

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

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

Consider the simple errors in the variables model,

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

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

a. Assume that x^*, u, and ε are all normally distributed with means μ^*, 0, and 0, variances σ_*^2, σ_u^2, and σ_ε^2, and zero covariances. Obtain the probability limits of the least squares estimators of α and β.

6-44

b. As an alternative, consider regressing x on a constant and y, and then computing the reciprocal of the estimate. Obtain the probability limit of this estimator.
c. Do the "direct" and "reverse" estimators bound the true coefficient?

4. *Reverse regression continued*. Suppose that the model in Exercise 3 is extended to $y = \beta x^* + \gamma d + \varepsilon$, $x = x^* + u$. For convenience, we drop the constant term. Assume that x^*, ε, and u are independent normally distributed with zero means. Suppose that d is a random variable that takes the values one and zero with probabilities π and $1 - \pi$ in the population and is independent of all other variables in the model. To put this formulation in context, the preceding model (and variants of it) have appeared in the literature on discrimination. We view y as a "wage" variable, x^* as "qualifications," and x as some imperfect measure such as education. The dummy variable d is membership ($d = 1$) or nonmembership ($d = 0$) in some protected class. The hypothesis of discrimination turns on $\gamma < 0$ versus $\gamma \geq 0$.
   a. What is the probability limit of c, the least squares estimator of γ, in the least squares regression of y on x and d? [Hints: The independence of x^* and d is important. Also, plim $\mathbf{d}'\mathbf{d}/n = \text{Var}[d] + E^2[d] = \pi(1 - \pi) + \pi^2 = \pi$. This minor modification does not affect the model substantively, but it greatly simplifies the algebra.] Now suppose that x^* and d are not independent. In particular, suppose that $E[x^* \mid d = 1] = \mu^1$ and $E[x^* \mid d = 0] = \mu^0$. Repeat the derivation with this assumption.
   b. Consider, instead, a regression of x on y and d. What is the probability limit of the coefficient on d in this regression? Assume that x^* and d are independent.
   c. Suppose that x^* and d are not independent, but γ is, in fact, less than zero. Assuming that both preceding equations still hold, what is estimated by $(\bar{y} \mid d = 1) - (\bar{y} \mid d = 0)$? What does this quantity estimate if γ does equal zero?

Applications

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

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

   a. Compute the full regression by least squares and report your results. Based on your results, what is the estimate of the marginal value, in \$/hour, of an additional year of education, for someone who has 12 years of education when all other variables are at their means and *Broken home* = 0?
   b. We are interested in possible nonlinearities in the effect of education on ln *Wage*. (Koop and Tobias focused on experience. As before, we are not attempting to replicate their results.) A histogram of the education variable shows values from 9 to 20, a huge spike at 12 years (high school graduation) and, perhaps surprisingly, a second at 15—intuition would have anticipated it at 16. Consider aggregating

6-45

the education variable into a set of dummy variables:

$$HS = 1 \text{ if } Educ \leq 12, 0 \text{ otherwise}$$

$$Col = 1 \text{ if } Educ > 12 \text{ and } Educ \leq 16, 0 \text{ otherwise}$$

$$Grad = 1 \text{ if } Educ > 16, 0 \text{ otherwise.}$$

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

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

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

that is,

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

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

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

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

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

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

4

6-46 / End 6

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

b. In the aforementioned study, Solow states:

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

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

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

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

4. Data on the number of incidents of wave damage to a sample of ships, with the type of ship and the period when it was constructed, are given in Table 6.8. 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." 6.9

TB 6.9

TABLE 6.8 Ship Damage Incidents (9)

	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).

✓