precisely as possible because the underlying theory is weak or because we do not have the right kind of data to test the model. As Davidson notes, "Because of the non-experimental nature of economics, we are never sure how the observed data were generated. The test of any hypothesis in economics always turns out to depend on additional assumptions necessary to specify a reasonably parsimonious model, which may or may not be justified."[14]

The practical question then is not why specification errors are made, for they generally are, but how to detect them. Once it is found that specification errors have been made, the remedies often suggest themselves. If, for example, it can be shown that a variable is inappropriately omitted from a model, the obvious remedy is to include that variable in the analysis, assuming, of course, the data on that variable are available.

In this section we discuss some tests that one may use to detect specification errors.

Detecting the Presence of Unnecessary Variables (Overfitting a Model)

Suppose we develop a k-variable model to explain a phenomenon:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad \textbf{(13.4.1)}$$

However, we are not totally sure that, say, the variable X_k really belongs in the model. One simple way to find this out is to test the significance of the estimated β_k with the usual t test: $t = \hat{\beta}_k / \text{se}\,(\hat{\beta}_k)$. But suppose that we are not sure whether, say, X_3 and X_4 legitimately belong in the model. This can be easily ascertained by the F test discussed in Chapter 8. Thus, detecting the presence of an irrelevant variable (or variables) is not a difficult task.

It is, however, very important to remember that in carrying out these tests of significance we have a specific model in mind. We accept that model as the **maintained hypothesis** or the "truth," however tentative it may be. Given that model, then, we can find out whether one or more regressors are really relevant by the usual t and F tests. But note carefully that we should not use the t and F tests to build a model *iteratively,* that is, we should not say that initially Y is related to X_2 only because $\hat{\beta}_2$ is statistically significant and then expand the model to include X_3 and decide to keep that variable in the model if $\hat{\beta}_3$ turns out to be statistically significant, and so on. This strategy of building a model is called the **bottom-up approach** (starting with a smaller model and expanding it as one goes along) or by the somewhat pejorative term, **data mining** (other names are **regression fishing, data grubbing, data snooping,** and **number crunching**).

[14]James Davidson, *Econometric Theory,* Blackwell Publishers, Oxford, U.K., 2000, p. 153.

The primary objective of data mining is to develop the "best" model after several diagnostic tests so that the model finally chosen is a "good" model in the sense that all the estimated coefficients have the "right" signs, they are statistically significant on the basis of the t and F tests, the R^2 value is reasonably high and the Durbin–Watson d has acceptable value (around 2), etc. The purists in the profession look down on the practice of data mining. In the words of William Pool, ". . . making an empirical regularity the foundation, rather than an implication of economic theory, is always dangerous."[15] One reason for "condemning" data mining is as follows.

Nominal versus True Level of Significance in the Presence of Data Mining. A danger of data mining that the unwary researcher faces is that the conventional levels of significance (α) such as 1, 5, or 10 percent are *not the true levels of significance*. Lovell has suggested that if there are c candidate regressors out of which k are finally selected ($k \leq c$) on the basis of data mining, then the true level of significance (α^*) is related to the nominal level of significance (α) as follows:[16]

$$\alpha^* = 1 - (1 - \alpha)^{c/k} \tag{13.4.2}$$

or approximately as

$$\alpha^* \approx (c/k)\alpha \tag{13.4.3}$$

For example, if $c = 15$, $k = 5$, and $\alpha = 5$ percent, from (13.4.3) the true level of significance is $(15/5)(5) = 15$ percent. Therefore, if a researcher data-mines and selects 5 out of 15 regressors and reports only the results of the condensed model at the nominal 5 percent level of significance and declares that the results are statistically significant, one should take this conclusion with a big grain of salt, for we know the (true) level of significance is in fact 15 percent. It should be noted that if $c = k$, that is, there is no data mining, the true and nominal levels of significance are the same. Of course, in practice most researchers report only the results of their "final" regression without necessarily telling about all the data mining, or **pretesting,** that has gone before.[17]

Despite some of its obvious drawbacks, there is increasing recognition, especially among applied econometricians, that the purist (i.e., non–data mining) approach to model building is not tenable. As Zaman notes:

> Unfortunately, experience with real data sets shows that such a [purist approach] is neither feasible nor desirable. It is not feasible because it is a rare economic

[15]William Pool, "Is Inflation Too Low," the *Cato Journal,* vol. 18, no. 3, Winter 1999, p. 456.
[16]M. Lovell, "Data Mining," *Review of Economics and Statistics,* vol. 65, 1983, pp. 1–12.
[17]For a detailed discussion of pretesting and the biases it can lead to, see Wallace, T. D., "Pretest Estimation in Regression: A Survey," *American Journal of Agricultural Economics,* vol. 59, 1977, pp. 431–443.

> theory which leads to a unique model. It is not desirable because a crucial aspect of learning from the data is learning what types of models are and are not supported by data. Even if, by rare luck, the initial model shows a good fit, it is frequently important to explore and learn the types of the models the data does or does not agree with.[18]

A similar view is expressed by Kerry Patterson who maintains that:

> This [data mining] approach suggests that economic theory and empirical specification interact rather than be kept in separate compartments.[19]

Instead of getting caught in the data mining versus the purist approach to model-building controversy, one can endorse the view expressed by Peter Kennedy:

> [that model specification] needs to be a well-thought-out combination of theory and data, and that testing procedures used in specification searches should be designed to minimize the costs of data mining. Examples of such procedures are setting aside data for out-of-sample prediction tests, adjusting significance levels [a la Lovell], and avoiding questionable criteria such as maximizing R^2.[20]

If we look at data mining in a broader perspective as a process of discovering empirical regularities that might suggest errors and/or omissions in (existing) theoretical models, it has a very useful role to play. To quote Kennedy again, "The art of the applied econometrician is to allow for data-driven theory while avoiding the considerable dangers in data mining."[21]

Tests for Omitted Variables and Incorrect Functional Form

In practice we are never sure that the model adopted for empirical testing is "the truth, the whole truth and nothing but the truth." On the basis of theory or introspection and prior empirical work, we develop a model that we believe captures the essence of the subject under study. We then subject the model to empirical testing. After we obtain the results, we begin the postmortem, keeping in mind the criteria of a good model discussed earlier. It is at this stage that we come to know if the chosen model is adequate. In determining model adequacy, we look at some broad features of the results, such as the $\bar{R}^2$ value, the estimated t ratios, the signs of the estimated coefficients in relation to their prior expectations, the Durbin–Watson statistic, and the like. If these diagnostics are reasonably good, we proclaim that the

[18]Asad Zaman, *Statistical Foundations for Econometric Techniques,* Academic Press, New York, 1996, p. 226.
[19]Kerry Patterson, *An Introduction to Applied Econometrics,* St. Martin's Press, New York, 2000, p. 10.
[20]Peter Kennedy, "Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics," unpublished manuscript.
[21]Kennedy, op. cit., p. 13.

chosen model is a fair representation of reality. By the same token, if the results do not look encouraging because the $\bar{R}^2$ value is too low or because very few coefficients are statistically significant or have the correct signs or because the Durbin–Watson d is too low, then we begin to worry about model adequacy and look for remedies: Maybe we have omitted an important variable, or have used the wrong functional form, or have not first-differenced the time series (to remove serial correlation), and so on. To aid us in determining whether model inadequacy is on account of one or more of these problems, we can use some of the following methods.

Examination of Residuals. As noted in Chapter 12, examination of the residuals is a good visual diagnostic to detect autocorrelation or heteroscedasticity. But these residuals can also be examined, especially in cross-sectional data, for model specification errors, such as omission of an important variable or incorrect functional form. If in fact there are such errors, a plot of the residuals will exhibit distinct patterns.

To illustrate, let us reconsider the cubic total cost of production function first considered in Chapter 7. Assume that the true total cost function is described as follows, where Y = total cost and X = output:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i \tag{13.4.4}$$

but a researcher fits the following quadratic function:

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \tag{13.4.5}$$

and another researcher fits the following linear function:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \tag{13.4.6}$$

Although we know that both researchers have made specification errors, for pedagogical purposes let us see how the estimated residuals look in the three models. (The cost-output data are given in Table 7.4.) Figure 13.1 speaks for itself: As we move from left to right, that is, as we approach the truth, not only are the residuals smaller (in absolute value) but also they do not exhibit the pronounced cyclical swings associated with the misfitted models.

The utility of examining the residual plot is thus clear: If there are specification errors, the residuals will exhibit noticeable patterns.

The Durbin–Watson *d* Statistic Once Again. If we examine the routinely calculated Durbin–Watson d in Table 13.1, we see that for the linear cost function the estimated d is 0.716, suggesting that there is positive "correlation" in the estimated residuals: for $n = 10$ and $k' = 1$, the 5 percent

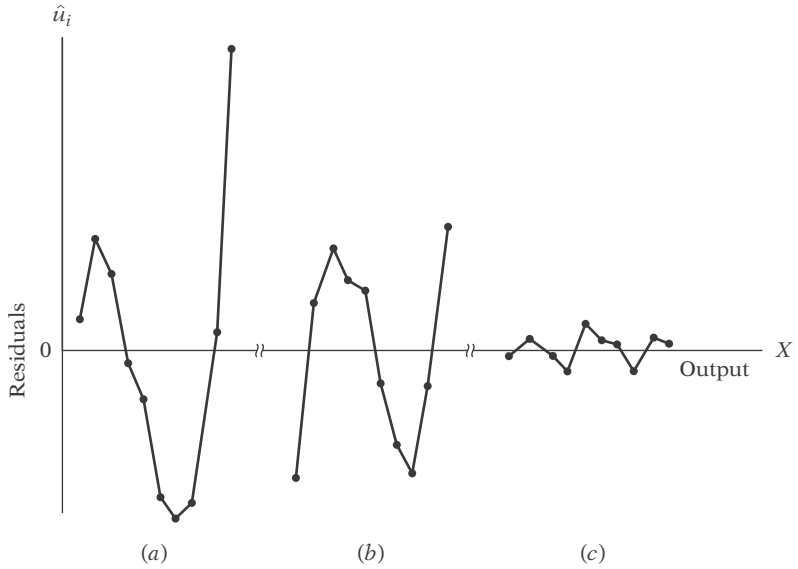

FIGURE 13.1 Residuals $\hat{u}_i$ from (*a*) linear, (*b*) quadratic, and (*c*) cubic total cost functions.

TABLE 13.1 ESTIMATED RESIDUALS FROM THE LINEAR, QUADRATIC, AND CUBIC TOTAL COST FUNCTIONS

Observation number	$\hat{u}_i$, linear model*	$\hat{u}_i$, quadratic model†	$\hat{u}_i$, cubic model**
1	6.600	−23.900	−0.222
2	19.667	9.500	1.607
3	13.733	18.817	−0.915
4	−2.200	13.050	−4.426
5	−9.133	11.200	4.435
6	−26.067	−5.733	1.032
7	−32.000	−16.750	0.726
8	−28.933	−23.850	−4.119
9	4.133	−6.033	1.859
10	54.200	23.700	0.022

* $\hat{Y}_i = 166.467 + 19.933X_i$; (19.021) (3.066) ; (8.752) (6.502) ; $R^2 = 0.8409$, $\bar{R}^2 = 0.8210$, $d = 0.716$

† $\hat{Y}_i = 222.383 - 8.0250X_i + 2.542X_i^2$; (23.488) (9.809) (0.869) ; (9.468) (−0.818) (2.925) ; $R^2 = 0.9284$, $\bar{R}^2 = 0.9079$, $d = 1.038$

** $\hat{Y}_i = 141.767 + 63.478X_i - 12.962X_i^2 + 0.939X_i^3$; (6.375) (4.778) (0.9856) (0.0592) ; (22.238) (13.285) (−13.151) (15.861) ; $R^2 = 0.9983$, $\bar{R}^2 = 0.9975$, $d = 2.70$

critical d values are $d_L = 0.879$ and $d_U = 1.320$. Likewise, the computed d value for the quadratic cost function is 1.038, whereas the 5 percent critical values are $d_L = 0.697$ and $d_U = 1.641$, indicating indecision. But if we use the modified d test (see Chapter 12), we can say that there is positive "correlation" in the residuals, for the computed d is less than d_U. For the cubic cost function, the true specification, the estimated d value does not indicate any positive "correlation" in the residuals.[22]

The observed positive "correlation" in the residuals when we fit the linear or quadratic model is not a measure of (first-order) serial correlation but of (model) specification error(s). The observed correlation simply reflects the fact that some variable(s) that belong in the model are included in the error term and need to be culled out from it and introduced in their own right as explanatory variables: If we exclude the X_i^3 from the cost function, then as (13.2.3) shows, the error term in the mis-specified model (13.2.2) is in fact $(u_{1i} + \beta_4 X_i^3)$ and it will exhibit a systematic pattern (e.g., positive autocorrelation) if X_i^3 in fact affects Y significantly.

To use the Durbin–Watson test for detecting model specification error(s), we proceed as follows:

1. From the assumed model, obtain the OLS residuals.

2. If it is believed that the assumed model is mis-specified because it excludes a relevant explanatory variable, say, Z from the model, order the residuals obtained in Step 1 according to increasing values of Z. *Note:* The Z variable could be one of the X variables included in the assumed model or it could be some function of that variable, such as X^2 or X^3.

3. Compute the d statistic from the residuals thus ordered by the usual d formula, namely,

$$d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

Note: The subscript t is the index of observation here and does not necessarily mean that the data are time series.

4. From the Durbin–Watson tables, if the estimated d value is significant, then one can accept the hypothesis of model mis-specification. If that turns out to be the case, the remedial measures will naturally suggest themselves.

In our cost example, the $Z\,(= X)$ variable (output) was already ordered.[23] Therefore, we do not have to compute the d statistic afresh. As we have seen, the d statistic for both the linear and quadratic cost functions suggests

[22] In the present context, a value of $d = 2$ will mean no specification error. (Why?)

[23] It does not matter if we order $\hat{u}_i$ according to X_i^2 or X_i^3 since these are functions of X_i, which is already ordered.

specification errors. The remedies are clear: Introduce the quadratic and cubic terms in the linear cost function and the cubic term in the quadratic cost function. In short, run the cubic cost model.

Ramsey's RESET Test. Ramsey has proposed a general test of specification error called RESET (regression specification error test).[24] Here we will illustrate only the simplest version of the test. To fix ideas, let us continue with our cost-output example and assume that the cost function is linear in output as

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \qquad \textbf{(13.4.6)}$$

where Y = total cost and X = output. Now if we plot the residuals $\hat{u}_i$ obtained from this regression against $\hat{Y}_i$, the estimated Y_i from this model, we get the picture shown in Figure 13.2. Although $\sum \hat{u}_i$ and $\sum \hat{u}_i \hat{Y}_i$ are necessarily zero

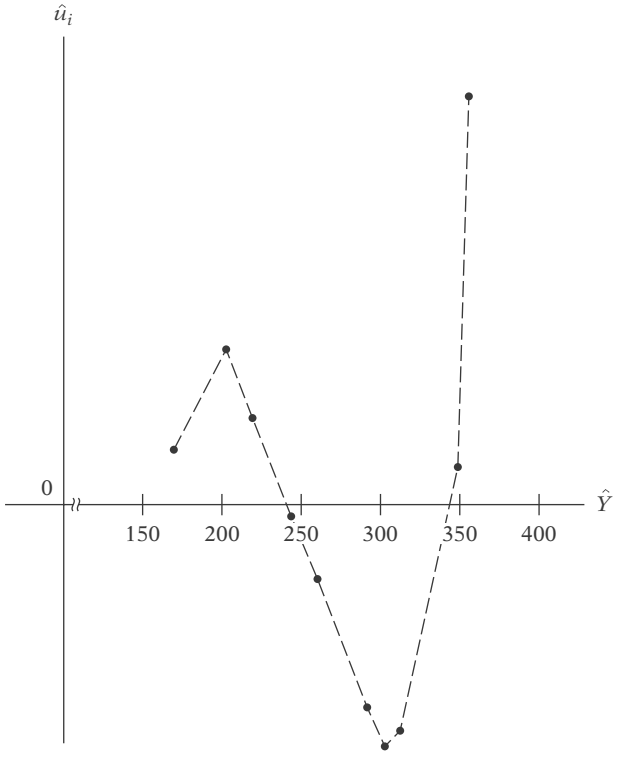

FIGURE 13.2 Residuals $\hat{u}_i$ and estimated Y from the linear cost function: $Y_i = \lambda_1 + \lambda_2 X_i + u_i$.

[24] J. B. Ramsey, "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society,* series B, vol. 31, 1969, pp. 350–371.

(why? see Chapter 3), the residuals in this figure show a pattern in which their mean changes systematically with $\hat{Y}_i$. This would suggest that if we introduce $\hat{Y}_i$ in some form as regressor(s) in (13.4.6), it should increase R^2. And if the increase in R^2 is statistically significant (on the basis of the F test discussed in Chapter 8), it would suggest that the linear cost function (13.4.6) was mis-specified. This is essentially the idea behind RESET. **The steps involved in RESET are as follows:**

1. From the chosen model, e.g., (13.4.6), obtain the estimated Y_i, that is, $\hat{Y}_i$.

2. Rerun (13.4.6) introducing $\hat{Y}_i$ in some form as an additional regressor(s). From Figure 13.2, we observe that there is a curvilinear relationship between $\hat{u}_i$ and $\hat{Y}_i$, suggesting that one can introduce $\hat{Y}_i^2$ and $\hat{Y}_i^3$ as additional regressors. Thus, we run

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + u_i \qquad \textbf{(13.4.7)}$$

3. Let the R^2 obtained from (13.4.7) be R^2_{new} and that obtained from (13.4.6) be R^2_{old}. Then we can use the F test first introduced in (8.5.18), namely,

$$F = \frac{\left(R^2_{\text{new}} - R^2_{\text{old}}\right)/\text{number of new regressors}}{\left(1 - R^2_{\text{new}}\right)/(n - \text{number of parameters in the new model})} \qquad \textbf{(8.5.18)}$$

to find out if the increase in R^2 from using (13.4.7) is statistically significant.

4. If the computed F value is significant, say, at the 5 percent level, one can accept the hypothesis that the model (13.4.6) is mis-specified.

Returning to our illustrative example, we have the following results (standard errors in parentheses):

$$\begin{aligned} \hat{Y}_i &= 166.467 + 19.933 X_i \\ &\quad (19.021) \quad (3.066) \qquad R^2 = 0.8409 \end{aligned} \qquad \textbf{(13.4.8)}$$

$$\begin{aligned} \hat{Y}_i = 2140.7223 &+ 476.6557 X_i - 0.09187 \hat{Y}_i^2 + 0.000119 \hat{Y}_i^3 \\ (132.0044) \quad &(33.3951) \quad (0.00620) \quad (0.0000074) \\ &R^2 = 0.9983 \end{aligned} \qquad \textbf{(13.4.9)}$$

Note: $\hat{Y}_i^2$ and $\hat{Y}_i^3$ in (13.4.9) are obtained from (13.4.8).

Now applying the F test we find

$$\begin{aligned} F &= \frac{(0.9983 - 0.8409)/2}{(1 - 0.9983)/(10 - 4)} \\ &= 284.4035 \end{aligned} \qquad \textbf{(13.4.10)}$$

The reader can easily verify that this F value is highly significant, indicating that the model (13.4.8) is mis-specified. Of course, we have reached the same conclusion on the basis of the visual examination of the residuals as well as the Durbin–Watson d value.

One advantage of RESET is that it is easy to apply, for it does not require one to specify what the alternative model is. But that is also its disadvantage because knowing that a model is mis-specified does not help us necessarily in choosing a better alternative.

Lagrange Multiplier (LM) Test for Adding Variables. This is an alternative to Ramsey's RESET test. To illustrate this test, we will continue with the preceding illustrative example.

If we compare the linear cost function (13.4.6) with the cubic cost function (13.4.4), the former is a *restricted version* of the latter (recall our discussion of **restricted least-squares** from Chapter 8). The restricted regression (13.4.6) assumes that the coefficients of the squared and cubed output terms are equal to zero. To test this, the LM test proceeds as follows:

1. Estimate the restricted regression (13.4.6) by OLS and obtain the residuals, $\hat{u}_i$.

2. If in fact the unrestricted regression (13.4.4) is the true regression, the residuals obtained in (13.4.6) should be related to the squared and cubed output terms, that is, X_i^2 and X_i^3.

3. This suggests that we regress the $\hat{u}_i$ obtained in Step 1 on all the regressors (including those in the restricted regression), which in the present case means

$$\hat{u}_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + \alpha_4 X_i^3 + v_i \qquad \textbf{(13.4.11)}$$

where v is an error term with the usual properties.

4. For large-sample size, Engle has shown that n (the sample size) times the R^2 estimated from the (auxiliary) regression (13.4.11) follows the chi-square distribution with df equal to the number of restrictions imposed by the restricted regression, two in the present example since the terms X_i^2 and X_i^3 are dropped from the model.[25] Symbolically, we write

$$nR^2 \underset{\text{asy}}{\sim} \chi^2_{(\text{number of restrictions})} \qquad \textbf{(13.4.12)}$$

where asy means asymptotically, that is, in large samples.

5. If the chi-square value obtained from (13.4.12) exceeds the critical chi-square value at the chosen level of significance, we reject the restricted regression. Otherwise, we do not reject it.

[25]R. F. Engle, "A General Approach to Lagrangian Multiplier Model Diagnostics," *Journal of Econometrics,* vol. 20, 1982, pp. 83–104.

For our example, the regression results are as follows:

$$\hat{Y}_i = 166.467 + 19.333X_i \tag{13.4.13}$$

where Y is total cost and X is output. The standard errors for this regression are already given in Table 13.1.

When the residuals from (13.4.13) are regressed as just suggested in Step 3, we obtain the following results:

$$\begin{aligned} \widehat{u}_i &= -24.7 \quad + 43.5443X_i - 12.9615X_i^2 + 0.9396X_i^3 \\ \text{se} &= \quad (6.375) \quad (4.779) \qquad (0.986) \qquad (0.059) \\ & \qquad\qquad\qquad\qquad\qquad\qquad R^2 = 0.9896 \end{aligned} \tag{13.4.14}$$

Although our sample size of 10 is by no means large, just to illustrate the LM mechanism, we obtain $nR^2 = (10)(0.9896) = 9.896$. From the chi-square table we observe that for 2 df the 1 percent critical chi-square value is about 9.21. Therefore, the observed value of 9.896 is significant at the 1 percent level, and our conclusion would be to reject the restricted regression (i.e., the linear cost function). We reached the similar conclusion on the basis of Ramsey's RESET test.

13.5 ERRORS OF MEASUREMENT

All along we have assumed implicitly that the dependent variable Y and the explanatory variables, the X's, are measured without any errors. Thus, in the regression of consumption expenditure on income and wealth of households, we assume that the data on these variables are "accurate"; they are not *guess estimates,* extrapolated, interpolated, or rounded off in any systematic manner, such as to the nearest hundredth dollar, and so on. Unfortunately, this ideal is not met in practice for a variety of reasons, such as nonresponse errors, reporting errors, and computing errors. Whatever the reasons, error of measurement is a potentially troublesome problem, for it constitutes yet another example of specification bias with the consequences noted below.

Errors of Measurement in the Dependent Variable *Y*

Consider the following model:

$$Y_i^* = \alpha + \beta X_i + u_i \tag{13.5.1}$$

where Y_i^* = permanent consumption expenditure[26]
X_i = current income
u_i = stochastic disturbance term

[26]This phrase is due to Milton Friedman. See also exercise 13.8.

Since Y_i^* is not directly measurable, we may use an observable expenditure variable Y_i such that

$$Y_i = Y_i^* + \varepsilon_i \tag{13.5.2}$$

where ε_i denote errors of measurement in Y_i^*. Therefore, instead of estimating (13.5.1), we estimate

$$\begin{aligned} Y_i &= (\alpha + \beta X_i + u_i) + \varepsilon_i \\ &= \alpha + \beta X_i + (u_i + \varepsilon_i) \\ &= \alpha + \beta X_i + v_i \end{aligned} \tag{13.5.3}$$

where $v_i = u_i + \varepsilon_i$ is a composite error term, containing the population disturbance term (which may be called the *equation error term*) and the measurement error term.

For simplicity assume that $E(u_i) = E(\varepsilon_i) = 0$, $\text{cov}\,(X_i, u_i) = 0$ (which is the assumption of the classical linear regression), and $\text{cov}\,(X_i, \varepsilon_i) = 0$; that is, the errors of measurement in Y_i^* are uncorrelated with X_i, and $\text{cov}\,(u_i, \varepsilon_i) = 0$; that is, the equation error and the measurement error are uncorrelated. With these assumptions, it can be seen that β estimated from either (13.5.1) or (13.5.3) will be an unbiased estimator of the true β (see exercise 13.7); that is, the errors of measurement in the dependent variable Y do not destroy the unbiasedness property of the OLS estimators. However, the variances and standard errors of β estimated from (13.5.1) and (13.5.3) will be different because, employing the usual formulas (see Chapter 3), we obtain

$$\text{Model (13.5.1):} \qquad \text{var}\,(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} \tag{13.5.4}$$

$$\begin{aligned} \text{Model (13.5.3):} \qquad \text{var}\,(\hat{\beta}) &= \frac{\sigma_v^2}{\sum x_i^2} \\ &= \frac{\sigma_u^2 + \sigma_\varepsilon^2}{\sum x_i^2} \end{aligned} \tag{13.5.5}$$

Obviously, the latter variance is larger than the former.[27] Therefore, **although the errors of measurement in the dependent variable still give unbiased estimates of the parameters and their variances, the estimated variances are now larger than in the case where there are no such errors of measurement.**

[27]But note that this variance is still unbiased because under the stated conditions the composite error term $v_i = u_i + \varepsilon_i$ still satisfies the assumptions underlying the method of least squares.

Errors of Measurement in the Explanatory Variable *X*

Now assume that instead of (13.5.1), we have the following model:

$$Y_i = \alpha + \beta X_i^* + u_i \tag{13.5.6}$$

where Y_i = current consumption expenditure
X_i^* = permanent income
u_i = disturbance term (equation error)

Suppose instead of observing X_i^*, we observe

$$X_i = X_i^* + w_i \tag{13.5.7}$$

where w_i represents errors of measurement in X_i^*. Therefore, instead of estimating (13.5.6), we estimate

$$\begin{aligned} Y_i &= \alpha + \beta(X_i - w_i) + u_i \\ &= \alpha + \beta X_i + (u_i - \beta w_i) \\ &= \alpha + \beta X_i + z_i \end{aligned} \tag{13.5.8}$$

where $z_i = u_i - \beta w_i$, a compound of equation and measurement errors.

Now even if we assume that w_i has zero mean, is serially independent, and is uncorrelated with u_i, we can no longer assume that the composite error term z_i is independent of the explanatory variable X_i because [assuming $E(z_i) = 0$]

$$\begin{aligned} \operatorname{cov}(z_i, X_i) &= E[z_i - E(z_i)][X_i - E(X_i)] \\ &= E(u_i - \beta w_i)(w_i) \qquad \text{using (13.5.7)} \\ &= E\left(-\beta w_i^2\right) \\ &= -\beta\sigma_w^2 \end{aligned} \tag{13.5.9}$$

Thus, the explanatory variable and the error term in (13.5.8) are correlated, which violates the crucial assumption of the classical linear regression model that the explanatory variable is uncorrelated with the stochastic disturbance term. If this assumption is violated, it can be shown that the *OLS estimators are not only biased but also inconsistent, that is, they remain biased even if the sample size n increases indefinitely.*[28]

[28] As shown in **App. A**, $\hat{\beta}$ is a consistent estimator of β if, as n increases indefinitely, the sampling distribution of $\hat{\beta}$ will ultimately collapse to the true β. Technically, this is stated as $\operatorname{plim}_{n\to\infty}\hat{\beta} = \beta$. As noted in **App. A**, consistency is a large-sample property and is often used to study the behavior of an estimator when its finite or small-sample properties (e.g., unbiasedness) cannot be determined.

For model (13.5.8), it is shown in Appendix 13A, Section 13A.3 that

$$\text{plim}\,\hat{\beta} = \beta\left[\frac{1}{1+\sigma_w^2/\sigma_{X^*}^2}\right] \tag{13.5.10}$$

where σ_w^2 and $\sigma_{X^*}^2$ are variances of w_i and X^*, respectively, and where plim $\hat{\beta}$ means the probability limit of β.

Since the term inside the brackets is expected to be less than 1 (why?), (13.5.10) shows that even if the sample size increases indefinitely, $\hat{\beta}$ will not converge to β. Actually, if β is assumed positive, $\hat{\beta}$ will underestimate β, that is, it is biased toward zero. Of course, if there are no measurement errors in X (i.e., $\sigma_w^2 = 0$), $\hat{\beta}$ will provide a consistent estimator of β.

Therefore, measurement errors pose a serious problem when they are present in the explanatory variable(s) because they make consistent estimation of the parameters impossible. Of course, as we saw, if they are present only in the dependent variable, the estimators remain unbiased and hence they are consistent too. If errors of measurement are present in the explanatory variable(s), what is the solution? The answer is not easy. At one extreme, we can assume that if σ_w^2 is small compared to $\sigma_{X^*}^2$, for all practical purposes we can "assume away" the problem and proceed with the usual OLS estimation. Of course, the rub here is that we cannot readily observe or measure σ_w^2 and $\sigma_{X^*}^2$ and therefore there is no way to judge their relative magnitudes.

One other suggested remedy is the use of **instrumental** or **proxy variables** that, although highly correlated with the original X variables, are uncorrelated with the equation and measurement error terms (i.e., u_i and w_i). If such proxy variables can be found, then one can obtain a consistent estimate of β. But this task is much easier said than done. In practice it is not easy to find good proxies; we are often in the situation of complaining about the bad weather without being able to do much about it. Besides, it is not easy to find out if the selected instrumental variable is in fact independent of the error terms u_i and w_i.

In the literature there are other suggestions to solve the problem.[29] But most of them are specific to the given situation and are based on restrictive assumptions. There is really no satisfactory answer to the measurement errors problem. That is why it is so crucial to measure the data as accurately as possible.

[29]See Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, *Advanced Econometric Methods,* Springer-Verlag, New York, 1984, pp. 273–277. See also Kennedy, op. cit., pp. 138–140, for a discussion of weighted regression as well as instrumental variables.

AN EXAMPLE

We conclude this section with an example constructed to highlight the preceding points.

Table 13.2 gives hypothetical data on true consumption expenditure Y^*, true income X^*, measured consumption Y, and measured income X. The table also explains how these variables were measured.[30]

Measurement Errors in the Dependent Variable *Y* Only

Based on the given data, the true consumption function is

$$\begin{aligned} \hat{Y}_i^* &= 25.00 + 0.6000X_i^* \\ &\quad (10.477) \quad (0.0584) \\ t &= (2.3861) \quad (10.276) \\ R^2 &= 0.9296 \end{aligned} \tag{13.5.11}$$

whereas, if we use Y_i instead of Y_i^*, we obtain

$$\begin{aligned} \hat{Y}_i &= 25.00 + 0.6000X_i^* \\ &\quad (12.218) \quad (0.0681) \\ t &= (2.0461) \quad (8.8118) \\ R^2 &= 0.9066 \end{aligned} \tag{13.5.12}$$

As these results show, and according to the theory, the estimated coefficients remain the same. The only effect of errors of measurement in the dependent variable is that the estimated standard errors of the coefficients tend to be larger [see (13.5.5)], which is clearly seen in (13.5.12). In passing, note that the regression coefficients in (13.5.11) and (13.5.12) are the same because the sample was generated to match the assumptions of the measurement error model.

Errors of Measurement in *X*

We know that the true regression is (13.5.11). Suppose now that instead of using X_i^*, we use X_i. (*Note:* In reality X_i^* is rarely observable.) The regression results are as follows:

$$\begin{aligned} \hat{Y}_i^* &= 25.992 + 0.5942X_i \\ &\quad (11.0810) \quad (0.0617) \\ t &= (2.3457) \quad (9.6270) \\ R^2 &= 0.9205 \end{aligned} \tag{13.5.13}$$

These results are in accord with the theory—when there are measurement errors in the explanatory variable(s), the estimated coefficients are biased. Fortunately, in this example the bias is rather small—from (13.5.10) it is evident that the bias depends on $\sigma_w^2/\sigma_{X^*}^2$, and in generating the data it was assumed that $\sigma_w^2 = 36$ and $\sigma_{X^*}^2 = 3667$, thus making the bias factor rather small, about 0.98 percent ($= 36/3667$).

We leave it to the reader to find out what happens when there are errors of measurement in both Y and X, that is, if we regress Y_i on X_i rather than Y_i^* on X_i^* (see exercise 13.23).

TABLE 13.2
HYPOTHETICAL DATA ON Y^* (TRUE CONSUMPTION EXPENDITURE), X^* (TRUE INCOME), Y (MEASURED CONSUMPTION EXPENDITURE), AND X (MEASURED INCOME); ALL DATA IN DOLLARS

Y^*	X^*	Y	X	ε	w	u
75.4666	80.00	67.6011	80.0940	−7.8655	0.0940	2.4666
74.9801	100.00	75.4438	91.5721	0.4636	−8.4279	−10.0199
102.8242	120.00	109.6956	112.1406	6.8714	2.1406	5.8242
125.7651	140.00	129.4159	145.5969	3.6509	5.5969	16.7651
106.5035	160.00	104.2388	168.5579	−2.2647	8.5579	−14.4965
131.4318	180.00	125.8319	171.4793	−5.5999	−8.5207	−1.5682
149.3693	200.00	153.9926	203.5366	4.6233	3.5366	4.3693
143.8628	220.00	152.9208	222.8533	9.0579	2.8533	−13.1372
177.5218	240.00	176.3344	232.9879	−1.1874	−7.0120	8.5218
182.2748	260.00	174.5252	261.1813	−7.7496	1.1813	1.2748

Note: The data on X^* are assumed to be given. In deriving the other variables the assumptions made were as follows: (1) $E(u_i) = E(\varepsilon_i) = E(w_i) = 0$; (2) cov (X, u) = cov (X, ε) = cov (u, ε) = cov (w, u) = cov $(\varepsilon, w) = 0$; (3) $\sigma_u^2 = 100$, $\sigma_s^2 = 36$, and $\sigma_w^2 = 36$; and (4) $Y_i^* = 25 + 0.6X_i^* + u_i$, $Y_i = Y_i^* + \varepsilon_i$, and $X_i = X_i^* + w_i$.

[30] I am indebted to Kenneth J. White for constructing this example. See his *Computer Handbook Using SHAZAM,* for use with Damodar Gujarati, *Basic Econometrics,* September 1985, pp. 117–121.

13.6 INCORRECT SPECIFICATION OF THE STOCHASTIC ERROR TERM

A common problem facing a researcher is the specification of the error term u_i that enters the regression model. Since the error term is not directly observable, there is no easy way to determine the form in which it enters the model. To see this, let us return to the models given in (13.2.8) and (13.2.9). For simplicity of exposition, we have assumed that there is no intercept in the model. We further assume that u_i in (13.2.8) is such that $\ln u_i$ satisfies the usual OLS assumptions.

If we assume that (13.2.8) is the "correct" model but estimate (13.2.9), what are the consequences? It is shown in Appendix 13.A, Section 13A.4, that if $\ln u_i \sim N(0,\ \sigma^2)$, then

$$u_i \sim \text{log normal}\left[e^{\sigma^2/2},\ e^{\sigma^2}\left(e^{\sigma^2}-1\right)\right] \tag{13.6.1}$$

as a result:

$$E(\hat{\alpha}) = \beta e^{\sigma^2/2} \tag{13.6.2}$$

where e is the base of the natural logarithm.

As you can see, $\hat{\alpha}$ is a biased estimator, as its average value is not equal to the true β.

We will have more to say about the specification of the stochastic error term in the chapter on nonlinear-in-the-parameter regression models.

13.7 NESTED VERSUS NON-NESTED MODELS

In carrying out specification testing, it is useful to distinguish between **nested and non-nested models.** To distinguish between the two, consider the following models:

$$\text{Model A:}\quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$

$$\text{Model B:}\quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

We say that Model B is nested in Model A because it is a special case of Model A: If we estimate Model A and test the hypothesis that $\beta_4 = \beta_5 = 0$ and do not reject it on the basis of, say, the F test,[31] Model A reduces to Model B. If we add variable X_4 to Model B, then Model A will reduce to Model B if β_5 is zero; here we will use the *t test* to test the hypothesis that the coefficient of X_5 is zero.

Without calling them such, the specification error tests that we have discussed previously and the restricted F test that we discussed in Chapter 8 are essentially tests of nested hypothesis.

[31]More generally, one can use the likelihood ratio test, or the Wald test or the Lagrange Multiplier test, which were discussed briefly in Chap. 8.

Now consider the following models:

$$\text{Model C:} \quad Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$$

$$\text{Model D:} \quad Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$$

where the X's and Z's are different variables. We say that Models C and D are **non-nested** because one cannot be derived as a special case of the other. In economics, as in other sciences, more than one competing theory may explain a phenomenon. Thus, the monetarists would emphasize the role of money in explaining changes in GDP, whereas the Keynesians may explain them by changes in government expenditure.

It may be noted here that one can allow Models C and D to contain regressors that are common to both. For example, X_3 could be included in Model D and Z_2 could be included in Model C. Even then these are non-nested models, because Model C does not contain Z_3 and Model D does not contain X_2.

Even if the same variables enter the model, the functional form may make two models non-nested. For example, consider the model:

$$\text{Model E:} \quad Y_i = \beta_1 + \beta_2 \ln Z_{2i} + \beta_3 \ln Z_{3i} + w_i$$

Models D and E are non-nested, as one cannot be derived as a special case of the other.

Since we already have looked at tests of nested models (t and F tests), in the following section we discuss some of the tests of non-nested models, which earlier we called model mis-specification errors.

13.8 TESTS OF NON-NESTED HYPOTHESES

According to Harvey,[32] there are two approaches to testing non-nested hypotheses: (1) the **discrimination approach,** where given two or more competing models, one chooses a model based on some criteria of goodness of fit, and (2) the **discerning approach** (my terminology) where, in investigating one model, we take into account information provided by other models. We consider these approaches briefly.

The Discrimination Approach

Consider Models C and D above. Since both models involve the same dependent variable, we can choose between two (or more) models based on some goodness-of-fit criterion, such as R^2 or adjusted R^2, which we have already discussed. But keep in mind that in comparing two or more models,

[32]Andrew Harvey, *The Econometric Analysis of Time Series,* 2d ed., The MIT Press, Cambridge, Mass., 1990, Chap. 5.

the regressand must be the same. Besides these criteria, there are other criteria that are also used. These include **Akaike's information criterion (AIC), Schwarz's information criterion (SIC),** and **Mallows's C_p criterion.** We discuss these criteria in Section 13.9. Most modern statistical software packages have one or more of these criteria built into their regression routines. In the last section of this chapter, we will illustrate these criteria using an extended example. On the basis of one or more of these criteria a model is finally selected that has the highest $\bar{R}^2$ or the lowest value of AIC or SIC, etc.

The Discerning Approach

The Non-Nested *F* Test or Encompassing *F* Test. Consider Models C and D introduced earlier. How do we choose between the two models? For this purpose suppose we estimate the following nested or *hybrid* model:

$$\text{Model F:} \qquad Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \lambda_4 Z_{2i} + \lambda_5 Z_{3i} + u_i$$

Notice that Model F *nests or encompasses* models C and D. But note that C is not nested in D and D is not nested in C, so they are non-nested models.

Now if Model C is correct, $\lambda_4 = \lambda_5 = 0$, whereas Model D is correct if $\lambda_2 = \lambda_3 = 0$. This testing can be done by the usual F test, hence the name non-nested F test.

However, there are problems with this testing procedure. *First,* if the X's and the Z's are highly correlated, then, as noted in the chapter on multicollinearity, it is quite likely that one or more of the λ's are individually statistically insignificant, although on the basis of the F test one can reject the hypothesis that all the slope coefficients are simultaneously zero. In this case, we have no way of deciding whether Model C or Model D is the correct model. *Second,* there is another problem. Suppose we choose Model C as the *reference hypothesis* or model, and find that all its coefficients are significant. Now we add Z_2 or Z_3 or both to the model and find, using the F test, that their incremental contribution to the explained sum of squares (ESS) is statistically insignificant. Therefore, we decide to choose Model C.

But suppose we had instead chosen Model D as the reference model and found that all its coefficients were statistically significant. But when we add X_2 or X_3 or both to this model, we find, again using the F test, that their incremental contribution to ESS is insignificant. Therefore, we would have chosen model D as the correct model. Hence, "the choice of the reference hypothesis could determine the outcome of the choice model,"[33] especially if severe multicollinearity is present in the competing regressors. *Finally,* the artificially nested model F may not have any economic meaning.

[33]Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, *Advanced Econometric Methods,* Springer Verlag, New York, 1984, p. 416.

AN ILLUSTRATIVE EXAMPLE: THE ST. LOUIS MODEL

To determine whether changes in nominal GNP can be explained by changes in the money supply (monetarism) or by changes in government expenditure (Keynesianism), we consider the following models:

$$\begin{aligned}\dot{Y}_t &= \alpha + \beta_0 \dot{M}_t + \beta_1 \dot{M}_{t-1} + \beta_2 \dot{M}_{t-2} + \beta_3 \dot{M}_{t-3} + \beta_4 \dot{M}_{t-4} + u_{1t} \\ &= \alpha + \sum_{i=0}^{4} \beta_i \dot{M}_{t-i} + u_{1t} \end{aligned} \tag{13.8.1}$$

$$\begin{aligned}\dot{Y}_t &= \gamma + \lambda_0 \dot{E}_t + \lambda_1 \dot{E}_{t-1} + \lambda_2 \dot{E}_{t-2} + \lambda_3 \dot{E}_{t-3} + \lambda_4 \dot{E}_{t-4} + u_{2t} \\ &= \gamma + \sum_{i=0}^{4} \lambda_i \dot{E}_{t-i} + u_{2t} \end{aligned} \tag{13.8.2}$$

where $\dot{Y}_t$ = rate of growth in nominal GNP at time t
$\dot{M}_t$ = rate of growth in the money supply (M_1 version) at time t
$\dot{E}_t$ = rate of growth in full, or high, employment government expenditure at time t

In passing, note that (13.8.1) and (13.8.2) are examples of **distributed lag models,** a topic thoroughly discussed in Chapter 17. For the time being, simply note that the effect of a unit change in the money supply or government expenditure on GNP is distributed over a period of time and is not instantaneous.

Since a priori it may be difficult to decide between the two competing models, let us enmesh the two models as shown below:

$$\dot{Y}_t = \text{constant} + \sum_{i=0}^{4} \beta_i \dot{M}_{t-i} + \sum_{i=0}^{4} \lambda_i \dot{E}_{t-i} + u_{3t} \tag{13.8.3}$$

This nested model is one form in which the famous (Federal Reserve Bank of) St. Louis model, a pro-monetary-school bank, has been expressed and estimated. The results of this model for the period 1953–I to 1976–IV for the United States are as follows (t ratios in parentheses):[34]

Coefficient	Estimate	Coefficient	Estimate	
β_0	0.40 (2.96)	λ_0	0.08 (2.26)	
β_1	0.41 (5.26)	λ_1	0.06 (2.52)	
β_2	0.25 (2.14)	λ_2	0.00 (0.02)	
β_3	0.06 (0.71)	λ_3	−0.06 (−2.20)	**(13.8.4)**
β_4	−0.05 (−0.37)	λ_4	−0.07 (−1.83)	
$\sum_{i=0}^{4} \beta_i$	1.06 (5.59)	$\sum_{i=0}^{4} \lambda_i$	0.03 (0.40)	
				$R^2 = 0.40$
				$d = 1.78$

What do these results suggest about the superiority of one model over the other? If we consider the cumulative effect of a unit change in $\dot{M}$ and $\dot{E}$ on $\dot{Y}$, we obtain, respectively, $\sum_{i=0}^{4} \beta_i = 1.06$ and $\sum_{i=0}^{4} \lambda_i = 0.03$, the former being statistically significant and the latter not. This comparison would tend to support the monetarist claim that it is changes in the money supply that determine changes in the (nominal) GNP. It is left as an exercise for the reader to evaluate critically this claim.

[34]See Keith M. Carlson, "Does the St. Louis Equation Now Believe in Fiscal Policy?" *Review, Federal Reserve Bank of St. Louis,* vol. 60, no. 2, February 1978, p. 17, table IV.

Davidson–MacKinnon *J* Test.[35] Because of the problems just listed in the non-nested *F* testing procedure, alternatives have been suggested. One is the *Davidson–MacKinnon J* test. To illustrate this test, suppose we want to compare hypothesis or Model C with hypothesis or Model D. The ***J* test** proceeds as follows:

1. We estimate Model D and from it we obtain the estimated Y values, $\hat{Y}_i^D$.

2. We add the predicted Y value in Step 1 as an additional regressor to Model C and estimate the following model:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 \hat{Y}_i^D + u_i \qquad \textbf{(13.8.5)}$$

where the $\hat{Y}_i^D$ values are obtained from Step 1. This model is an example of the **encompassing principle,** as in the Hendry methodology.

3. Using the t test, test the hypothesis that $\alpha_4 = 0$.

4. If the hypothesis that $\alpha_4 = 0$ is not rejected, we can accept (i.e., not reject) Model C as the true model because $\hat{Y}_i^D$ included in (13.8.5), which represent the influence of variables not included in Model C, have no additional explanatory power beyond that contributed by Model C. In other words, Model C *encompasses* Model D in the sense that the latter model does not contain any additional information that will improve the performance of Model C. By the same token, if the null hypothesis is rejected, Model C cannot be the true model (why?).

5. Now we reverse the roles of hypotheses, or Models C and D. We now estimate Model C first, use the estimated Y values from this model as regressor in (13.8.5), repeat Step 4, and decide whether to accept Model D over Model C. More specifically, we estimate the following model:

$$Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 \hat{Y}_i^C + u_i \qquad \textbf{(13.8.6)}$$

where $\hat{Y}_i^C$ are the estimated Y values from Model C. We now test the hypothesis that $\beta_4 = 0$. If this hypothesis is not rejected, we choose Model D over C. If the hypothesis that $\beta_4 = 0$ is rejected, choose C over D, as the latter does not improve over the performance of C.

Although it is intuitively appealing, the J test has some problems. Since the tests given in (13.8.5) and (13.8.6) are performed independently, we have the following likely outcomes:

	Hypothesis: $\alpha_4 = 0$	
Hypothesis: $\beta_4 = 0$	**Do not reject**	**Reject**
Do not reject	Accept both C and D	Accept D, reject C
Reject	Accept C, reject D	Reject both C and D

[35] R. Davidson and J. G. MacKinnon, "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica,* vol. 49, 1981, pp. 781–793.

As this table shows, we will not be able to get a clear answer if the *J* testing procedure leads to the acceptance or rejection of both models. In case both models are rejected, neither model helps us to explain the behavior of *Y*. Similarly, if both models are accepted, as Kmenta notes, "the data are apparently not rich enough to discriminate between the two hypotheses [models]."[36]

Another problem with the *J* test is that when we use the *t* statistic to test the significance of the estimated *Y* variable in models (13.8.5) and (13.8.6), the *t* statistic has the standard normal distribution only asymptotically, that is, in large samples. Therefore, the *J* test may not be very powerful (in the statistical sense) in small samples because it tends to reject the true hypothesis or model more frequently than it ought to.

AN ILLUSTRATIVE EXAMPLE

To illustrate the *J* test, consider the data given in Table 13.3. This table gives data on per capita personal consumption expenditure (PPCE) and per capita disposable personal income (PDPI), both measured in 1987 dollars, for the United States for the period 1970–1991. Now consider the following rival models:

$$\text{Model A:} \quad \text{PPCE}_t = \alpha_1 + \alpha_2 \text{PDPI}_t + \alpha_3 \text{PDPI}_{t-1} + u_t \qquad \textbf{(13.8.7)}$$

$$\text{Model B:} \quad \text{PPCE}_t = \beta_1 + \beta_2 \text{PDPI}_t + \beta_3 \text{PPCE}_{t-1} + u_t \qquad \textbf{(13.8.8)}$$

Model A states that PPCE depends on PDPI in the current and previous time period; this model is an example of what is known as the **distributed lag model** (see Chapter 17). Model B postulates that PPCE depends on current PDPI as well as PPCE in the previous time period; this model represents what is known as the **autoregressive model** (see Chapter 17). The

TABLE 13.3
PER CAPITA PERSONAL CONSUMPTION EXPENDITURE (PPCE) AND PER CAPITA PERSONAL DISPOSABLE INCOME (PDPI), 1987 DOLLARS, U.S., 1970–1991

Year	PPCE	PDPI	Year	PPCE	PDPI
1970	8,842	9,875	1981	10,770	12,156
1971	9,022	10,111	1982	10,782	12,146
1972	9,425	10,414	1983	11,179	12,349
1973	9,752	11,013	1984	11,617	13,029
1974	9,602	10,832	1985	12,015	13,258
1975	9,711	10,906	1986	12,336	13,552
1976	10,121	11,192	1987	12,568	13,545
1977	10,425	11,406	1988	12,903	13,890
1978	10,744	11,851	1989	13,029	14,005
1979	10,876	12,039	1990	13,044	14,068
1980	10,746	12,005	1991	12,824	13,886

Source: Economic Report of the President, 1993, Table B-5, p. 355.

(*Continued*)

[36]Jan Kmenta, op. cit., p. 597.

AN ILLUSTRATIVE EXAMPLE (*Continued*)

reason for introducing the lagged value of PPCE in this model is to reflect inertia or habit persistence.

The results of estimating these models separately were as follows:

$$\text{Model A:} \quad \widehat{\text{PPCE}}_t = -1{,}299.0536 + 0.9204\,\text{PDPI}_t + 0.0931\,\text{PDPI}_{t-1}$$
$$t = \quad (-4.0378) \quad (6.0178) \quad (0.6308) \qquad \textbf{(13.8.9)}$$
$$R^2 = 0.9888 \quad d = 0.8092$$

$$\text{Model B:} \quad \widehat{\text{PPCE}}_t = -841.8568 + 0.7117\,\text{PDPI}_t + 0.2954\,\text{PPCE}_{t-1}$$
$$t = \quad (-2.4137) \quad (5.4634) \quad (2.3681) \qquad \textbf{(13.8.10)}$$
$$R^2 = 0.9912 \quad d = 1.0144$$

If one were to choose between these two models on the basis of the discrimination approach, using, say, the highest R^2 criterion, one would choose (13.8.10); besides, in (13.8.10) both variables seem to be individually statistically significant, whereas in (13.8.9) only the current PDPI is statistically significant (but beware of the collinearity problem!).

But choosing (13.8.10) over (13.8.9) may not be appropriate because for predictive purposes there is not much difference in the two estimated R^2 values.

To apply the *J* test, suppose we assume Model A is the null hypothesis, that is, the maintained model, and Model B is the alternative hypothesis. Now following the *J* test steps discussed earlier we use the estimated PPCE values from model (13.8.10) as an additional regressor in Model A, giving the following outcome:

$$\widehat{\text{PPCE}}_t = 1{,}322.7958 - 0.7061\text{PDPI}_t - 0.4357\text{PDPI}_{t-1} + 2.1335\widehat{\text{PPCE}}_t^{\,B}$$
$$t = \quad (1.5896) \quad (-1.3958) \quad (-2.1926) \quad (3.3141) \qquad \textbf{(13.8.11)}$$
$$R^2 = 0.9932 \quad d = 1.7115$$

where $\widehat{\text{PPCE}}_t^{\,B}$ on the right side of (13.8.11) are the estimated PPCE values from model B, (13.8.10). Since the coefficient of this variable is statistically significant (at the two-tail 0.004 level), following the *J* test procedure, we have to reject Model A in favor of Model B.

Now assuming Model B as the maintained hypothesis and Model A as the alternative hypothesis, and following exactly the same procedure as before, we obtain the following results:

$$\widehat{\text{PPCE}}_t = -6{,}549.8659 + 5.1176\text{PDPI}_t + 0.6302\text{PPCE}_{t-1} - 4.6776\widehat{\text{PPCE}}_t^{\,A}$$
$$t = \quad (-2.4976) \quad (2.5424) \quad (3.4141) \quad (-2.1926) \qquad \textbf{(13.8.12)}$$
$$R^2 = 0.9920 \quad d = 1.7115$$

where $\widehat{\text{PPCE}}^{A}$ on the right side of (13.8.12) is obtained from the Model A, (13.8.9). But in this regression, the coefficient of $\widehat{\text{PPCE}}_t^{\,A}$ on the right side is also statistically significant (at the two-tail 0.0425 level). This result would suggest that we should now reject Model B in favor of Model A!

All this tells us is that neither model is particularly useful in explaining the behavior of per capita personal consumption expenditure in the United States over the period 1970–1991.

Of course, we have considered only two competing models. In reality, there may be more than two models. The *J* test procedure can be extended to multiple model comparisons, although the analysis can become quickly complex.

This example shows very vividly why the CLRM assumes that the regression model used in the analysis is correctly specified. Obviously it is very crucial in developing a model to pay very careful attention to the phenomenon being modeled.