

3

THE CLASSICAL LINEAR REGRESSION MODEL

3.1 TEXTBOOKS AS CATALOGS

In chapter 2 we learned that many of the estimating criteria held in high regard by econometricians (such as best unbiasedness and minimum mean square error) are characteristics of an estimator's sampling distribution. These characteristics cannot be determined unless a set of repeated samples can be taken or hypothesized; to take or hypothesize these repeated samples, knowledge of the way in which the observations are generated is necessary. Unfortunately, an estimator does not have the same characteristics for all ways in which the observations can be generated. This means that in some estimating situations a particular estimator has desirable properties but in other estimating situations it does *not* have desirable properties. Because there is no "superestimator" having desirable properties in all situations, for each estimating problem (i.e., for each different way in which the observations can be generated) the econometrician must determine anew which estimator is preferred. An econometrics textbook can be characterized as a catalog of which estimators are most desirable in what estimating situations. Thus, a researcher facing a particular estimating problem simply turns to the catalog to determine which estimator is most appropriate for him or her to employ in that situation. The purpose of this chapter is to explain how this catalog is structured.

The cataloging process described above is centered around a standard estimating situation referred to as the *classical linear regression model* (CLR model). It happens that in this standard situation the OLS estimator is considered the optimal estimator. This model consists of five assumptions concerning the way in which the data are generated. By changing these assumptions in one way or another, different estimating situations are created, in many of which the OLS estimator is no longer considered to be the optimal estimator. Most econometric problems can be characterized as situations in which one (or more) of these five assumptions is

violated in a particular way. The catalog works in a straightforward way: the estimating situation is modeled in the general mold of the CLR model and then the researcher pinpoints the way in which this situation differs from the standard situation as described by the CLR model (i.e., finds out which assumption of the CLR model is violated in this problem); he or she then turns to the textbook (catalog) to see whether the OLS estimator retains its desirable properties, and if not what alternative estimator should be used. Because econometricians often are not certain of whether the estimating situation they face is one in which an assumption of the CLR model is violated, the catalog also includes a listing of techniques useful in testing whether or not the CLR model assumptions are violated.

3.2 THE FIVE ASSUMPTIONS

The CLR model consists of five basic assumptions about the way in which the observations are generated.

(1) The *first assumption* of the CLR model is that the dependent variable can be calculated as a linear function of a specific set of independent variables, plus a disturbance term. The unknown coefficients of this linear function form the vector β and are assumed to be constants. Several violations of this assumption, called specification errors, are discussed in chapter 6:

- (a) *wrong regressors* – the omission of relevant independent variables or the inclusion of irrelevant independent variables;
- (b) *nonlinearity* – when the relationship between the dependent and independent variables is not linear;
- (c) *changing parameters* – when the parameters (β) do not remain constant during the period in which data were collected.

(2) The *second assumption* of the CLR model is that the expected value of the disturbance term is zero; i.e., the mean of the distribution from which the disturbance term is drawn is zero. Violation of this assumption leads to the *biased intercept* problem, discussed in chapter 7.

(3) The *third assumption* of the CLR model is that the disturbance terms all have the same variance and are not correlated with one another. Two major econometric problems, as discussed in chapter 8, are associated with violations of this assumption:

- (a) *heteroskedasticity* – when the disturbances do not all have the same variance;

- (b) *autocorrelated errors* – when the disturbances are correlated with one another.

(4) The *fourth assumption* of the CLR model is that the observations on the independent variable can be considered fixed in repeated samples; i.e., it is possible to repeat the sample with the same independent variable values. Three important econometric problems, discussed in chapters 9 and 10, correspond to violations of this assumption:

- (a) *errors in variables* – errors in measuring the independent variables;
- (b) *autoregression* – using a lagged value of the dependent variable as an independent variable;
- (c) *simultaneous equation estimation* – situations in which the dependent variables are determined by the simultaneous interaction of several relationships.

(5) The *fifth assumption* of the CLR model is that the number of observations is greater than the number of independent variables and that there are no exact linear relationships between the independent variables. Although this is viewed as an assumption for the general case, for a specific case it can easily be checked, so that it need not be assumed. The problem of *multicollinearity* (two or more independent variables being approximately linearly related in the sample data) is associated with this assumption. This is discussed in chapter 11.

All this is summarized in table 3.1, which presents these five assumptions of the CLR model, shows the appearance they take when dressed in mathematical notation, and lists the econometric problems most closely associated with violations of these assumptions. Later chapters in this book comment on the meaning and significance of these assumptions, note implications of their violation for the OLS estimator, discuss ways of determining whether or not they are violated, and suggest new estimators appropriate to situations in which one of these assumptions must be replaced by an alternative assumption. Before we move on to this, however, more must be said about the character of the OLS estimator in the context of the CLR model, because of the central role it plays in the econometrician's "catalog."

3.3 THE OLS ESTIMATOR IN THE CLR MODEL

The central role of the OLS estimator in the econometrician's catalog is that of a standard against which all other estimators are compared. The

Table 3.1 The assumptions of the CLR model

Assumption	Mathematical expression		Violations	Chapter in which discussed
	Bivariate	Multivariate		
(1) Dependent variable a linear function of a specific set of independent variables, plus a disturbance	$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, $t = 1, \dots, T$	$Y = X\beta + \varepsilon$	Wrong regressors Nonlinearity Changing parameters	6
(2) Expected value of disturbance term is zero	$E\varepsilon_t = 0$, for all t	$E\varepsilon = 0$	Biased intercept	7
(3) Disturbances have uniform variance and are uncorrelated	$E\varepsilon_t \varepsilon_r = 0, t \neq r$ $= \sigma^2, t = r$	$E\varepsilon \varepsilon' = \sigma^2 I$	Heteroskedasticity Autocorrelated errors	8
(4) Observations on independent variables can be considered fixed in repeated samples	x_t fixed in repeated samples	X fixed in repeated samples	Errors in variables Autoregression Simultaneous equations	9 10
(5) No exact linear relationships between independent variables and more observations than independent variables	$\sum_{t=1}^T (x_t - \bar{x})^2 \neq 0$	Rank of $X = K \leq T$	Perfect multicollinearity	11

The mathematical terminology is explained in the technical notes to this section. The notation is as follows: Y is a vector of observations on the dependent variable; X is a matrix of observations on the independent variables; ε is a vector of disturbances; σ^2 is the variance of the disturbances; I is the identity matrix; K is the number of independent variables; T is the number of observations.

reason for this is that the OLS estimator is extraordinarily popular. This popularity stems from the fact that, in the context of the CLR model, the OLS estimator has a large number of desirable properties, making it the overwhelming choice for the "optimal" estimator when the estimating problem is accurately characterized by the CLR model. This is best illustrated by looking at the eight criteria listed in chapter 2 and determining how the OLS estimator rates on these criteria in the context of the CLR model.

- (1) *Computational cost* Because of the popularity of the OLS estimator, many packaged computer routines exist, and for simple cases hand-held calculators can be used to perform the required calculations quickly. (Some hand-held calculators have OLS estimation built in.) Whenever the functional form being estimated is linear, as it is in the CLR model, the OLS estimator involves very little computational cost.
- (2) *Least squares* Because the OLS estimator is designed to minimize the sum of squared residuals, it is automatically "optimal" on this criterion.
- (3) *Highest R²* Because the OLS estimator is optimal on the least squares criterion, it will automatically be optimal on the highest R² criterion.
- (4) *Unbiasedness* The assumptions of the CLR model can be used to show that the OLS estimator β^{OLS} is an unbiased estimator of β .
- (5) *Best unbiasedness* In the CLR model β^{OLS} is a linear estimator; i.e., it can be written as a linear function of the errors. As noted earlier, it is unbiased. Among all linear unbiased estimators of β , it can be shown (in the context of the CLR model) to have the "smallest" variance-covariance matrix. Thus the OLS estimator is the BLUE in the CLR model. If we add the additional assumption that the disturbances are distributed normally (creating the CNLR model – the *classical normal linear regression model*), it can be shown that the OLS estimator is the best unbiased estimator (i.e., best among all unbiased estimators, not just linear unbiased estimators).
- (6) *Mean square error* It is not the case that the OLS estimator is the minimum mean square error estimator in the CLR model. Even among linear estimators, it is possible that a substantial reduction in variance can be obtained by adopting a slightly biased estimator. This is the OLS estimator's weakest point; chapters 11 and 12 discuss several estimators whose appeal lies in the possibility that they may beat OLS on the MSE criterion.
- (7) *Asymptotic criteria* Because the OLS estimator in the CLR model is unbiased, it is also unbiased in samples of infinite size and thus is asymptotically unbiased. It can also be shown that the variance-

covariance matrix of β^{OLS} goes to zero as the sample size goes to infinity, so that β^{OLS} is also a consistent estimator of β . Further, in the CNLR model it is asymptotically efficient.

- (8) *Maximum likelihood* It is impossible to calculate the maximum likelihood estimator given the assumptions of the CLR model, because these assumptions do not specify the functional form of the distribution of the disturbance terms. However, if the disturbances are assumed to be distributed normally (the CNLR model), it turns out that β^{MLE} is identical to β^{OLS} .

Thus, whenever the estimating situation can be characterized by the CLR model, the OLS estimator meets practically all of the criteria econometricians consider relevant. It is no wonder, then, that this estimator has become so popular. It is in fact *too* popular: it is often used, without justification, in estimating situations that are not accurately represented by the CLR model. If some of the CLR model assumptions do not hold, many of the desirable properties of the OLS estimator no longer hold. If the OLS estimator does not have the properties that are thought to be of most importance, an alternative estimator must be found. Before moving to this aspect of our examination of econometrics, however, we will spend a chapter discussing some concepts of and problems in inference, to provide a foundation for later chapters.

4

INTERVAL ESTIMATION AND HYPOTHESIS TESTING

4.1 INTRODUCTION

In addition to estimating parameters, econometricians often wish to construct confidence intervals for their estimates and test hypotheses concerning parameters. To strengthen the perspective from which violations of the CLR model are viewed in the following chapters, this chapter provides a brief discussion of these principles of inference in the context of traditional applications found in econometrics.

Under the null hypothesis most test statistics have a distribution that is tabulated in appendices at the back of statistics books, the most common of which are the standard normal, the t , the chi-square, and the F distributions. In small samples the applicability of all these distributions depends on the errors in the CLR model being normally distributed, something that is not one of the CLR model assumptions. For situations in which the errors are not distributed normally, it turns out that in most cases a traditional test statistic has an asymptotic distribution equivalent to one of these tabulated distributions; with this as justification, testing/interval estimation proceeds in the usual way, ignoring the small sample bias. For expository purposes, this chapter's discussion of inference is couched in terms of the classical normal linear regression (CNLR) model, in which the assumptions of the CLR model are augmented by assuming that the errors are distributed normally.

4.2 TESTING A SINGLE HYPOTHESIS: THE t TEST

Hypothesis tests on and interval estimates of single parameters are straightforward applications of techniques familiar to all students of elementary statistics. In the CNLR model the OLS estimator β^{OLS} generates estimates

that are distributed joint-normally in repeated samples. This means that $\beta_1^{\text{OLS}}, \beta_2^{\text{OLS}}, \dots, \beta_k^{\text{OLS}}$ are all connected to one another (through their covariances). In particular, this means that β_3^{OLS} , say, is distributed normally with mean β_3 (since the OLS estimator is unbiased) and variance $V(\beta_3^{\text{OLS}})$ equal to the third diagonal element of the variance-covariance matrix of β^{OLS} . The square root of $V(\beta_3^{\text{OLS}})$ is the standard deviation of β_3^{OLS} . Using the normal table and this standard deviation, interval estimates can be constructed and hypotheses can be tested.

A major drawback to this procedure is that the variance-covariance matrix of β^{OLS} is not usually known (because σ^2 , the variance of the disturbances, which appears in the formula for this variance-covariance matrix, is not usually known). Estimating σ^2 by s^2 , as discussed in the technical notes to section 3.3, allows an estimate of this matrix to be created. The square root of the third diagonal element of this matrix is the standard error of β_3^{OLS} , an estimate of the standard deviation of β_3^{OLS} . With this estimate the t -table can be used in place of the normal table to test hypotheses or construct interval estimates.

The use of such t tests, as they are called, is so common that most packaged computer programs designed to compute the OLS estimators (designed to run OLS regressions) have included in their output a number called the t statistic for each parameter estimate. This gives the value of the parameter estimate divided by its estimated standard deviation (the standard error). This value can be compared directly to critical values in the t -table to test the hypothesis that that parameter is equal to zero. In some research reports, this t statistic is printed in parentheses underneath the parameter estimates, creating some confusion because sometimes the standard errors appear in this position. (A negative number in parentheses would have to be a t value, so that this would indicate that these numbers were t values rather than standard errors.)

4.3 TESTING A JOINT HYPOTHESIS: THE F TEST

Suppose that a researcher wants to test the joint hypothesis that, say, the fourth and fifth elements of β are equal to 1.0 and 2.0, respectively. That is, he wishes to test the hypothesis that the sub-vector

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix}$$

is equal to the vector

$$\begin{bmatrix} 1.0 \\ 2.0 \end{bmatrix}$$

This is a different question from the two separate questions of whether β_4 is equal to 1.0 and whether β_5 is equal to 2.0. It is possible, for example, to accept the hypothesis that β_4 is equal to 1.0 and also to accept the hypothesis that β_5 is equal to 2.0, but to *reject* the joint hypothesis that

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix}$$

is equal to

$$\begin{bmatrix} 1.0 \\ 2.0 \end{bmatrix}$$

The purpose of this section is to explain how the F test is used to test such joint hypotheses. The following section explains how a difference between results based on separate tests and joint tests could arise.

The F statistic for testing a set of J linear constraints in a regression with K parameters (including the intercept) and T observations takes the generic form

$$\frac{[SSE(\text{constrained}) - SSE(\text{unconstrained})]/J}{SSE(\text{unconstrained})/(T - K)}$$

where the degrees of freedom for this F statistic are J and $T - K$. This generic form is worth memorizing – it is extremely useful for structuring F tests for a wide variety of special cases, such as Chow tests (chapter 6) and tests involving dummy variables (chapter 14).

When the constraints are true, because of the error term they will not be satisfied exactly by the data, so the SSE will increase when the constraints are imposed – minimization subject to constraints will not be as successful as minimization without constraints. But if the constraints are true the per-constraint increase in SSE should not be large, relative to the influence of the error term. The numerator has the “per-constraint” change in SSE due to imposing the constraints and the denominator has the “per-error” contribution to SSE . (The minus K in this expression corrects for degrees of freedom, explained in the general notes.) If their ratio is “too big” we would be reluctant to believe that it happened by chance, concluding that it must have happened because the constraints are false.

High values of this F statistic thus lead us to reject the null hypothesis that the constraints are true.

How does one find the constrained SSE ? A constrained regression is run to obtain the constrained SSE . The easiest example is the case of constraining a coefficient to be equal to zero – just run the regression omitting that coefficient’s variable. To run a regression constraining β_4^{OLS} to be 1.0 and β_5^{OLS} to be 2.0, subtract 1.0 times the fourth regressor and 2.0 times the fifth regressor from the dependent variable and regress this new, constructed dependent variable on the remaining regressors. In general, to incorporate a linear restriction into a regression, use the restriction to solve out one of the parameters, and rearrange the resulting equation to form a new regression involving constructed variables. An explicit example is given in the general notes.

4.4 INTERVAL ESTIMATION FOR A PARAMETER VECTOR

Interval estimation in the multidimensional case is best illustrated by a two-dimensional example. Suppose that the sub-vector

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix}$$

is of interest. The OLS estimate of this sub-vector is shown as the point in the center of the rectangle in figure 4.1. Using the t -table and the square root of the fourth diagonal term in the estimated variance–covariance matrix of β^{OLS} , a 95% confidence interval can be constructed for β_4 . This is shown in figure 4.1 as the interval from A to B ; β_4^{OLS} lies halfway between A and B . Similarly, a 95% confidence interval can be constructed for β_5 ; it is shown in figure 4.1 as the interval from C to D and is drawn larger than the interval AB to reflect an assumed larger standard error for β_5^{OLS} .

An interval estimate for the sub-vector

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix}$$

is a *region* or area that, when constructed in repeated samples, covers the true value (β_4, β_5) in, say, 95% of the samples. Furthermore, this region should for an efficient estimate be the smallest such region possible. A natural region to choose for this purpose is the rectangle formed by the individual interval estimates, as shown in figure 4.1. If β_4^{OLS} and β_5^{OLS} have

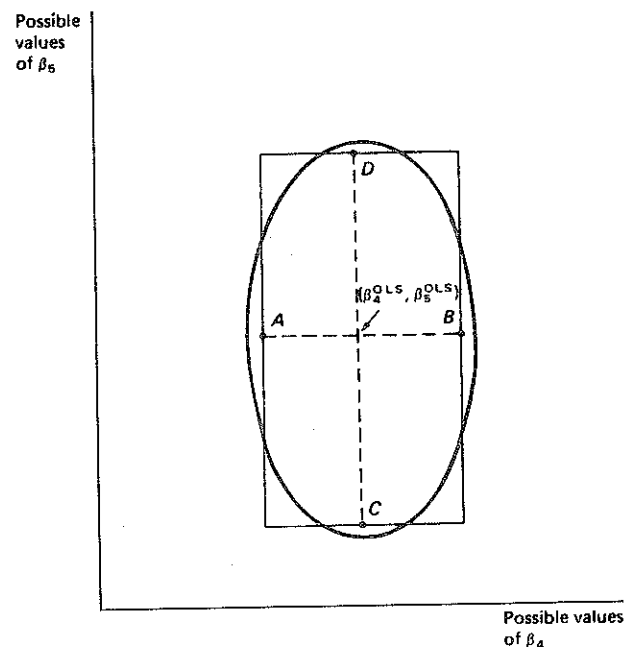


Figure 4.1 A confidence region with zero covariance

zero covariance, then in repeated sampling rectangles calculated in this fashion will cover the unknown point (β_4, β_5) in $0.95 \times 0.95 = 90.25\%$ of the samples. (In repeated samples the probability is 0.95 that the β_4 confidence interval covers β_4 , as is the probability that the β_5 confidence interval covers β_5 ; thus the probability for both β_4 and β_5 to be covered simultaneously is 0.95×0.95 .)

Evidently, this rectangle is not "big" enough to serve as a 95% joint confidence interval. Where should it be enlarged? Because the region must be kept as small as possible, the enlargement must come in those parts that have the greatest chance of covering (β_4, β_5) in repeated samples. The corners of the rectangle will cover (β_4, β_5) in a repeated sample whenever β_4^{OLS} and β_5^{OLS} are simultaneously a long way from their mean values of β_4 and β_5 . The probability in repeated samples of having these two unlikely events occur simultaneously is very small. Thus the areas just outside the rectangle near the points A, B, C, and D are more likely to cover (β_4, β_5) in repeated samples than are the areas just outside the corners of the rectangle: the rectangle should be made bigger near the points A, B, C, and D. Further thought suggests that the areas just outside the points A, B, C, and D are more likely, in repeated samples, to cover

(β_4, β_5) than the areas just *inside* the corners of the rectangle. Thus the total region could be made smaller by chopping a lot of area off the corners and extending slightly the areas near the points A, B, C, and D. In fact, the F statistic described earlier allows the econometrician to derive the confidence region as an ellipse, as shown in figure 4.1.

The ellipse in figure 4.1 represents the case of zero covariance between β_4^{OLS} and β_5^{OLS} . If β_4^{OLS} and β_5^{OLS} have a positive covariance (an estimate of this covariance is found in either the fourth column and fifth row or the fifth column and fourth row of the estimate of the variance-covariance matrix of β^{OLS}), whenever β_4^{OLS} is an overestimate of β_4 , β_5^{OLS} is likely to be an overestimate of β_5 , and whenever β_4^{OLS} is an underestimate of β_4 , β_5^{OLS} is likely to be an underestimate of β_5 . This means that the area near the top right-hand corner of the rectangle and the area near the bottom left-hand corner are no longer as unlikely to cover (β_4, β_5) in repeated samples; it also means that the areas near the top left-hand corner and bottom right-hand corner are even less likely to cover (β_4, β_5) . In this case the ellipse representing the confidence region is tilted to the right, as shown in figure 4.2. In the case of negative covariance between β_4^{OLS} and β_5^{OLS} , the ellipse is tilted to the left. In all cases, the ellipse remains centered on the point $(\beta_4^{\text{OLS}}, \beta_5^{\text{OLS}})$.

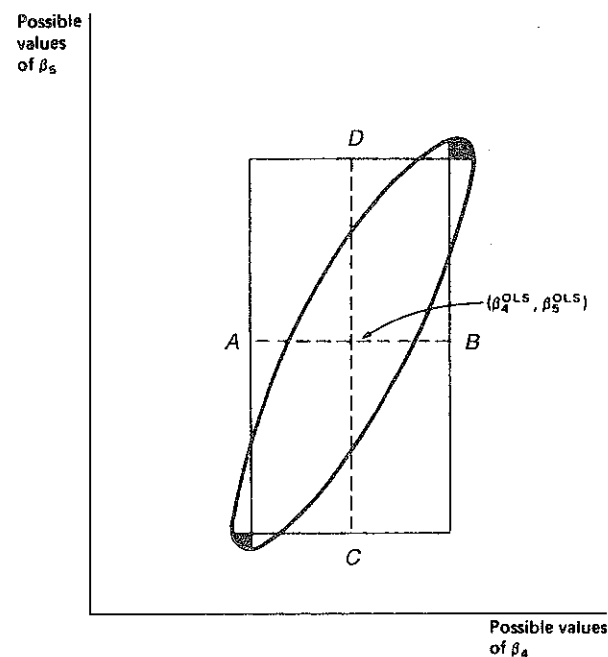


Figure 4.2 A confidence region with positive covariance

This two-dimensional example illustrates the possibility, mentioned earlier, of accepting two individual hypotheses but rejecting the corresponding joint hypothesis. Suppose the hypothesis is that $\beta_4 = 0$ and $\beta_5 = 0$, and suppose the point $(0,0)$ lies inside a corner of the rectangle in figure 4.1, but outside the ellipse. Testing the hypothesis $\beta_4 = 0$ using a t test concludes that β_4 is insignificantly different from zero (because the interval AB contains zero), and testing the hypothesis $\beta_5 = 0$ concludes that β_5 is insignificantly different from zero (because the interval CD contains zero). But testing the joint hypothesis

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

using an F test, concludes that

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix}$$

is significantly different from the zero vector because $(0,0)$ lies outside the ellipse. In this example one can confidently say that *at least one* of the two variables has a significant influence on the dependent variable, but one cannot with confidence assign that influence to either of the variables individually. The typical circumstance in which this comes about is in the case of multicollinearity (see chapter 11), in which independent variables are related so that it is difficult to tell which of the variables deserves credit for explaining variation in the dependent variable. Figure 4.2 is representative of the multicollinearity case.

In three dimensions the confidence region becomes a confidence volume and is represented diagrammatically by an ellipsoid. In higher dimensions diagrammatic representation is impossible, but the hypersurface corresponding to a critical value of the F statistic can be called a multi-dimensional ellipsoid.

4.5 LR, W, AND LM STATISTICS

The F test discussed above is applicable whenever we are testing linear restrictions in the context of the CNLR model. Whenever the problem cannot be cast into this mold – for example, if the restrictions are nonlinear, the model is nonlinear in the parameters, or the errors are distributed non-normally – this procedure is inappropriate and is usually replaced by one of three asymptotically equivalent tests. These are the *likelihood ratio*

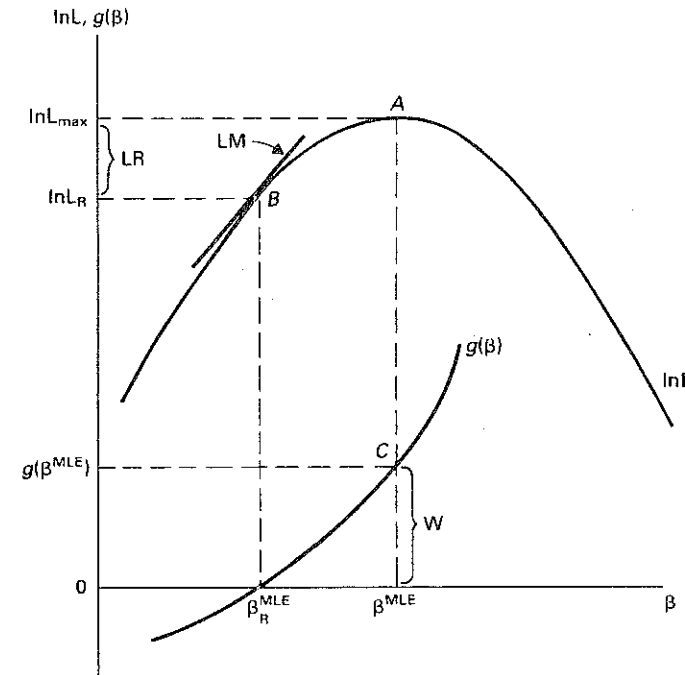


Figure 4.3 Explaining the LR, W, and LM statistics

(LR) test, the *Wald* (W) test, and the *Lagrange multiplier* (LM) test. The test statistics associated with these tests have unknown small-sample distributions, but are each distributed asymptotically as a chi-square (χ^2) with degrees of freedom equal to the number of restrictions being tested.

These three test statistics are based on three different rationales. Consider figure 4.3, in which the log-likelihood ($\ln L$) function is graphed as a function of β , the parameter being estimated. β^{MLE} is, by definition, the value of β at which $\ln L$ attains its maximum. Suppose the restriction being tested is written as $g(\beta) = 0$, satisfied at the value β_R^{MLE} where the function $g(\beta)$ cuts the horizontal axis:

- (1) *The LR test* If the restriction is true, then $\ln L_R$, the maximized value of $\ln L$ imposing the restriction, should not be *significantly* less than $\ln L_{\text{max}}$, the unrestricted maximum value of $\ln L$. The LR test tests whether $(\ln L_R - \ln L_{\text{max}})$ is significantly different from zero.
- (2) *The W test* If the restriction $g(\beta) = 0$ is true, then $g(\beta^{\text{MLE}})$ should not be *significantly* different from zero. The W test tests whether β^{MLE} (the unrestricted estimate of β) violates the restriction by a significant amount.

- (3) *The LM test* The log-likelihood function $\ln L$ is maximized at point A where the slope of $\ln L$ with respect to β is zero. If the restriction is true, then the slope of $\ln L$ at point B should not be *significantly* different from zero. The LM test tests whether the slope of $\ln L$, evaluated at the restricted estimate, is significantly different from zero.

When faced with three statistics with identical asymptotic properties, econometricians would usually choose among them on the basis of their small-sample properties, as determined by Monte Carlo studies. In this case, however, it happens that computational cost plays a dominant role in this respect. To calculate the LR statistic, both the restricted and the unrestricted estimates of β must be calculated. If neither is difficult to compute, then the LR test is computationally the most attractive of the three tests. To calculate the W statistic only the unrestricted estimate is required; if the restricted but not the unrestricted estimate is difficult to compute, owing to a nonlinear restriction, for example, the W test is computationally the most attractive. To calculate the LM statistic, only the restricted estimate is required; if the unrestricted but not the restricted estimate is difficult to compute – for example, when imposing the restriction transforms a nonlinear functional form into a linear functional form – the LM test is the most attractive.

4.6 BOOTSTRAPPING

Testing hypotheses exploits knowledge of the sampling distributions of test statistics when the null is true, and constructing confidence intervals requires knowledge of estimators' sampling distributions. Unfortunately, this "knowledge" is often questionable, or unavailable, for a variety of reasons:

- (1) Assumptions made concerning the distribution of the error term may be false. For example, the error may not be distributed normally, or even approximately normally, as is often assumed.
- (2) Algebraic difficulties in calculating the characteristics of a sampling distribution often cause econometricians to undertake such derivations assuming that the sample size is very large. The resulting "asymptotic" results may not be close approximations for the sample size of the problem at hand.
- (3) For some estimating techniques, such as minimizing the median squared error, even asymptotic algebra cannot produce formulas for estimator variances.

- (4) A researcher may obtain an estimate by undertaking a series of tests, the results of which lead eventually to adoption of a final estimation formula. This search process makes it impossible to derive algebraically the character of the sampling distribution.

One way of dealing with these problems is to perform a Monte Carlo study: data are simulated to mimic the process thought to be generating the data, the estimate or test statistic is calculated and this process is repeated several thousand times to allow computation of the character of the sampling distribution of the estimator or test statistic. To tailor the Monte Carlo study to the problem at hand, initial parameter estimates are used as the "true" parameter values, and the actual values of the explanatory variables are employed as the "fixed in repeated samples" values of the explanatory variables. But this tailoring is incomplete because in the Monte Carlo study the errors must be drawn from a known distribution such as the normal. This is a major drawback of the traditional Monte Carlo methodology in this context.

The bootstrap is a special Monte Carlo procedure which circumvents this problem. It does so by assuming that the unknown distribution of the error term can be adequately approximated by a discrete distribution that gives equal weight to each of the residuals from the original estimation. Assuming a reasonable sample size, in typical cases most of the residuals should be small in absolute value, so that although each residual is given equal weight (and thus is equally likely to be chosen in random draws from this distribution), small residuals predominate, causing random draws from this distribution to produce small values much more frequently than large values. This procedure, which estimates sampling distributions by using only the original data (and so "pulls itself up by its own bootstraps") has proved to be remarkably successful. In effect it substitutes computing power, the price of which has dramatically decreased, for theorem-proving, whose price has held constant or even increased as we have adopted more complicated estimation procedures.

The bootstrap begins by estimating the model in question and saving the residuals. It performs a Monte Carlo study, using the estimated parameter values as the "true" parameter values and the actual values of the explanatory variables as the fixed explanatory variable values. During this Monte Carlo study errors are drawn, with replacement, from the set of original residuals. In this way account is taken of the unknown distribution of the true errors. This "residual-based" technique is only appropriate whenever each error is equally likely to be drawn for each observation. If this is not the case, an alternative bootstrapping method is employed. See the general notes to this section for further discussion.

explained below. Nonetheless, some, such as Verbeek (2000), believe that its computational simplicity overcomes its unreliability.

- It is noted in appendix A that there are three different ways of estimating the information matrix. This implies that there are three different ways of estimating the variance-covariance matrix needed for calculating the W and LM tests. In general, the OPG variant is inferior to the alternatives and should be avoided; see, for example, Bera and McKenzie (1986). Unfortunately, however, some of the computationally attractive ways of calculating the LM statistic implicitly have built into them the OPG calculation for the variance-covariance matrix of the MLE, causing the size of the resulting LM statistic to be too large. In particular, versions of the LM test that are calculated as the explained sum of squares from regressing a column of ones on first derivatives are suspect. Davidson and MacKinnon (1983) suggest an alternative way of calculating the LM statistic for a wide variety of applications, through running what they call a *double-length regression* (DLR), which retains the computational attractiveness of the OPG variant of the LM test, but avoids its shortcomings. Godfrey (1988, pp. 82–4) has a good discussion. See also Davidson and MacKinnon (1988). Davidson and MacKinnon (1993, pp. 492–502) is a good textbook exposition.

4.6 Bootstrapping

- When drawing OLS residuals for bootstrapping they should be adjusted upwards by multiplying by the square root of $n/(n-k)$ to account for the fact that although the OLS residuals are unbiased estimates of the errors, they underestimate their absolute value.
- To find the sampling distribution of a test statistic on the null hypothesis, the null-hypothesis parameter values should be used when creating Monte Carlo repeated samples. In general, as with all Monte Carlo studies, every effort should be made to create the bootstrap samples in a way that incorporates all known facets of the data-generating process.
- The bootstrap should investigate the sampling distribution of an “asymptotically pivotal” statistic, a statistic whose sampling distribution does not depend on the true values of the parameters. For example, rather than estimating the sampling distribution of a parameter estimate, the sampling distribution of the associated t statistic should be estimated. The sampling distribution of the t statistic can be used indirectly to produce confidence intervals, rather than calculating confidence intervals directly using the sampling distribution of the parameter estimate.
- Davidson and MacKinnon (2000) suggest a means of determining how many bootstraps to calculate for testing purposes. Efron (1987) suggests that estimation of bias and variance requires only about 200, but estimation of confidence intervals, and thus use for hypothesis testing, requires about 2,000.

5

SPECIFICATION

5.1 INTRODUCTION

At one time econometricians tended to assume that the model provided by economic theory represented accurately the real-world mechanism generating the data, and viewed their role as one of providing “good” estimates for the key parameters of that model. If any uncertainty was expressed about the model specification, there was a tendency to think in terms of using econometrics to “find” the real-world data-generating mechanism. Both these views of econometrics are obsolete. It is now generally acknowledged that econometric models are “false” and that there is no hope, or pretense, that through them “truth” will be found. Feldstein’s (1982, p. 829) remarks are typical of this view: “in practice all econometric specifications are necessarily ‘false’ models. . . . The applied econometrician, like the theorist, soon discovers from experience that a useful model is not one that is ‘true’ or ‘realistic’ but one that is parsimonious, plausible and informative.” This is echoed by an oft-quoted remark attributed to George Box – “All models are wrong, but some are useful” – and another from Theil (1971, p. vi): “Models are to be used, but not to be believed.”

In light of this recognition, econometricians have been forced to articulate more clearly what econometric models are, one view being that they “are simply rough guides to understanding” (Quah, 1995, p. 1596). There is some consensus that models are metaphors, or windows, through which researchers view the observable world, and that their adoption depends not upon whether they can be deemed “true” but rather upon whether they can be said to (a) correspond to the facts, and (b) be useful. Econometric specification analysis therefore is a means of formalizing what is meant by “corresponding to the facts” and “being useful,” thereby defining what is meant by a “correctly specified model.” From this perspective econometric analysis becomes much more than estimation and inference in the context of a given model; in conjunction with economic theory, it plays a

crucial, preliminary role of searching for and evaluating a model, leading ultimately to its acceptance or rejection.

Econometrics textbooks are mainly devoted to the exposition of econometrics for estimation and inference in the context of a given model for the data-generating process. The more important problem of specification of this model is not given much attention, for three main reasons. First, specification is not easy. In the words of Hendry and Richard (1983, p. 112), "the data generation process is complicated, data are scarce and of uncertain relevance, experimentation is uncontrolled and available theories are highly abstract and rarely uncontroversial." Second, most econometricians would agree that specification is an innovative/imaginative process that cannot be taught: "Even with a vast arsenal of diagnostics, it is very hard to write down rules that can be used to guide a data analysis. So much is really subjective and subtle. . . . A great deal of what we teach in applied statistics is *not* written down, let alone in a form suitable for formal encoding. It is just simply 'lore'" (Welch, 1986, p. 405). And third, there is no accepted "best" way of going about finding a correct specification.

There is little that can be done about items one and two above; they must be lived with. Item three, however, is worthy of further discussion: regardless of how difficult a specification problem, or how limited a researcher's powers of innovation/imagination, an appropriate methodology should be employed when undertaking empirical work. Considerable controversy exists within the profession regarding what is the most appropriate methodology, however. The purpose of this chapter is to discuss this issue; it should be viewed as a prelude to the examination in chapter 6 of specific violations of the first assumption of the CLR model, and as background to the presentation in chapter 21 of rules guiding applied econometricians.

5.2 THREE METHODOLOGIES

Until about the mid-1970s, econometricians were too busy doing econometrics to worry about the principles that were or should be guiding empirical research. Sparked by the predictive failure of large-scale econometric models, and fueled by dissatisfaction with the gap between how econometrics was taught and how it was applied by practitioners, the profession began to examine with a critical eye the way in which econometric models were specified. This chapter is in part a summary of the state of this ongoing methodological debate. At considerable risk of oversimplification, three main approaches to the specification problem are described below in stylized form.

(1) AVERAGE ECONOMIC REGRESSION (AER)

This approach describes what is thought to be the usual way in which empirical work in economics is undertaken. The researcher begins with a specification that is viewed as being known to be correct, with data being used primarily to determine the orders of magnitude of a small number of unknown parameters. Significant values of diagnostic test statistics, such as the Durbin-Watson statistic, are initially interpreted as suggesting estimation problems that should be dealt with by adopting more sophisticated estimation methods, rather than as pointing to a misspecification of the chosen model. If these more sophisticated methods fail to "solve" the problem, the researcher then conducts "specification" tests, hunting for an alternative specification that is "better," using age-old criteria such as correct signs, high R^2 's, and significant t values on coefficients "known" to be nonzero. Thus in the AER approach the data ultimately do play a role in the specification, despite the researcher's initial attitude regarding the validity of the theoretical specification. This role may be characterized as proceeding from a simple model and "testing up" to a specific more general model.

(2) TEST, TEST, TEST (TTT)

This approach uses econometrics to discover which models of the economy are tenable, and to test rival views. To begin, the initial specification is made more general than the researcher expects the specification ultimately chosen to be, and testing of various restrictions, such as sets of coefficients equal to the zero vector, is undertaken to simplify this general specification; this testing can be characterized as "testing down" from a general to a more specific model. Following this, the model is subjected to a battery of diagnostic, or misspecification, tests, hunting for signs that the model is misspecified. (Note the contrast with AER "specification" tests, which hunt for specific alternative specifications.) A significant diagnostic, such as a small DW value, is interpreted as pointing to a model misspecification rather than as pointing to a need for more sophisticated estimation methods. The model is continually respecified until a battery of diagnostic tests allows a researcher to conclude that the model is satisfactory on several specific criteria (discussed in the general notes), in which case it is said to be "congruent" with the evidence.

(3) FRAGILITY ANALYSIS

The specification ultimately arrived at by the typical AER or TTT search may be inappropriate because its choice is sensitive to the initial specification investigated, the order in which tests were undertaken, type I and

type II errors, and innumerable prior beliefs of researchers concerning the parameters that subtly influence decisions taken (through the exercise of innovation/imagination) throughout the specification process. It may, however, be the case that the different possible specifications that could have arisen from the AER or the TTT approaches would all lead to the same conclusion with respect to the purpose for which the study was undertaken, in which case why worry about the specification? This is the attitude towards specification adopted by the third approach. Suppose that the purpose of the study is to estimate the coefficients of some "key" variables. The first step of this approach, after identifying a general family of models, is to undertake an "extreme bounds analysis," in which the coefficients of the key variables are estimated using all combinations of included/excluded "doubtful" variables. If the resulting range of estimates is too wide for comfort, an attempt is made to narrow this range by conducting a "fragility analysis." A Bayesian method (see chapter 13) is used to incorporate non-sample information into the estimation, but in such a way as to allow for a range of this Bayesian information, corresponding to the range of such information that will surely characterize the many researchers interested in this estimation. This range of information will produce a range of estimates of the parameters of interest; a narrow range ("sturdy" estimates) implies that the data at hand yield useful information, but if this is not the case ("fragile" estimates), it must be concluded that inferences from these data are too fragile to be believed.

Which is the best of these three general approaches? There is no agreement that one of these methodologies is unequivocally the best to employ; each has faced criticism, a general summary of which is provided below.

(1) The AER is the most heavily criticized, perhaps because it reflects most accurately what researchers actually do. It is accused of using econometrics merely to illustrate assumed-known theories. The attitude that significant diagnostics reflect estimation problems rather than specification errors is viewed in an especially negative light, even by those defending this approach. "Testing up" invites bias because tests do not maintain their specified type I error if the model used for estimation does not contain the "true" data generating process as a special case. The *ad hoc* use of extraneous information (as opposed to its incorporation through formal means), such as the "right" signs on coefficient estimates, is deplored, especially by those with a Bayesian bent. The use of statistics such as R^2 , popular with those following this methodology, is frowned upon. Perhaps most frustrating to critics is the lack of a well-defined structure and set of criteria for this approach; there is never an adequate description of the path taken to the ultimate specification.

(2) The TTT methodology is also criticized for failing in practice to provide an adequate description of the path taken to the ultimate specification, primarily due to the role played in specification by innovation/imagination. This corresponds to an underlying suspicion that practitioners using this methodology find it necessary to use many of the *ad hoc* rules of thumb followed in the AER approach. The impossibility of estimating a completely general model to begin a TTT analysis implies that this method must necessarily adopt some form of testing up. The heavy reliance on testing in this methodology raises fears of a proliferation of type I errors (creating pre-test bias, discussed in section 12.4 of chapter 12), exacerbated by the small degrees of freedom due to the very general initial specification and by the fact that many of these tests have only asymptotic justification. Controlling type I errors when testing down requires adoption of a lower α value for the tests, but this is not routinely done. Requiring specifications to pass formal tests may conflict with model usefulness.

(3) Objections to fragility analysis usually come from those not comfortable with the Bayesian approach, even though care has been taken to make it palatable to non-Bayesians. Such objections are theological in nature and not likely to be resolved. There is vagueness regarding how large a range of parameter estimates has to be to conclude that it is fragile; attempts to formalize this lead to measures comparable to the test statistics this approach seeks to avoid. The methodology never does lead to the adoption of a specific specification, something that researchers find unsatisfactory. There is no scope for the general family of models initially chosen to be changed in the light of what the data have to say. Many researchers find Bayesian prior formulation both difficult and alien. Some object that this analysis too often concludes that results are fragile.

5.3 GENERAL PRINCIPLES FOR SPECIFICATION

Although the controversy over econometric methodology may never be resolved, the debate has been fruitful in that some general principles have emerged to guide model specification.

- (1) Economic theory should be the foundation of and guiding force in a specification search. Notwithstanding this, using the data to help create a "more informed" economic theory can be of considerable value, so long as the researcher does not use the same data to test the theory.
- (2) Models whose residuals do not test as insignificantly different from white noise (random errors) should be initially viewed as reflecting

a misspecification, not as needing a special estimation procedure, a view that too many researchers are prone to take.

- (3) Although "testing down" carries less bias than "testing up," beginning with a completely general model is not feasible. Consequently, in practice a blend of testing up and testing down needs to be used. A simple specification is proposed, and a general variant of that specification is tested down. On the basis of what has been learned during this process, a more complicated model may be proposed, and this process repeated.
- (4) Tests of misspecification are better undertaken by testing simultaneously for several misspecifications rather than testing one-by-one for these misspecifications. By such an "overtesting" technique one avoids the problem of one type of misspecification adversely affecting a test for some other type of misspecification. This approach helps to deflect the common criticism that such tests rely for their power on aspects of the maintained hypothesis about which little is known.
- (5) Models should routinely be exposed to a battery of misspecification diagnostic tests before being accepted, but in keeping with (1) above, subject to the qualification that the resulting model tells a clean economic story. Of particular importance is that a subset of the data should be set aside before model specification and estimation, so that these tests can include tests for predicting extra-sample observations.
- (6) Researchers should be obliged to show that their model encompasses rival models, in the sense that it can predict what results would be obtained were one to run the regression suggested by a rival model. The chosen model should be capable of explaining the data and of explaining the successes and failures of rival models in accounting for the same data.
- (7) Bounds on the range of results corresponding to different reasonable specifications should be reported (a "sensitivity" analysis), rather than just providing the results of the specification eventually adopted, and the path taken to the selection of this specification should be fully reported.

5.4 MISSPECIFICATION TESTS/DIAGNOSTICS

Despite the protestations of fragility analysis advocates, testing has come to play a more and more prominent role in econometric work. Thanks to the ingenuity of econometric theorists, and the power of asymptotic algebra, an extremely large number of tests has been developed, seemingly

catering to practitioners' every possible need, but at the same time courting confusion because of unknown small-sample properties, suspicions of low power, and often-conflicting prescriptions. It is not possible in this book to discuss all or even a majority of these tests. The more prominent among them are discussed briefly in later chapters when it is relevant to do so; before moving on to these chapters, however, it may be useful to have an overview of tests used for specification purposes. They fall into several categories.

- (1) *Omitted variable (OV) tests* F and t tests for zero restrictions on (or, more generally, linear combinations of) the parameters, as discussed in chapter 4, are commonly used for specification purposes. Several more complicated tests, such as Hausman tests, can be reformulated as OV tests in an artificial regression, greatly simplifying testing.
- (2) *RESET tests* RESET tests, discussed in chapter 6, are used to test for whether unknown variables have been omitted from a regression specification, and are not to be confused with OV tests that test for zero coefficients on known variables. They can also be used to detect a misspecified functional form.
- (3) *Tests for functional form* Two types of tests for functional form are available, as discussed in chapter 6. The first type, such as tests based on recursive residuals and the rainbow test, does not specify a specific alternative functional form. For the second type, functional form is tested by testing a restriction on a more general functional form, such as a Box-Cox transformation.
- (4) *Tests for structural change* In this category fall tests for parameter constancy, discussed in chapter 6, such as Chow tests, cusum and cusum-of-squares tests, and predictive failure (or post-sample prediction) tests. Chapter 18 refers to additional tests developed for use in modern time series analysis.
- (5) *Tests for outliers* These tests, among which are included tests for normality, are sometimes used as general tests for misspecification. Examples are the Jarque-Bera test, the Shapiro-Wilk test, the Cook outlier test, and the use of the DFITS measure (discussed in chapter 20).
- (6) *Tests for nonspherical errors* These are tests for various types of serial correlation and heteroskedasticity, discussed in chapter 8. Examples are the Durbin-Watson test, the Breusch-Godfrey test, Durbin's h and m tests, the Goldfeld-Quandt test, the Breusch-Pagan test, and the White test.
- (7) *Tests for exogeneity* These tests, often referred to as Hausman tests, test for contemporaneous correlation between regressors and the

error. They are discussed in chapter 9 (testing for measurement error) and chapter 10 (testing for simultaneous equation bias).

- (8) *Data transformation tests* These tests, which do not have any specific alternative hypothesis, are considered variants of the Hausman test. Examples are the grouping test and the differencing test.
- (9) *Non-nested tests* When testing rival models that are not nested, as might arise when testing for encompassing, non-nested tests must be employed. Examples are the non-nested F test and the J test.
- (10) *Conditional moment tests* These tests are based on a very general testing methodology which in special cases gives rise to most of the tests listed above. Beyond serving as a unifying framework for existing tests, the value of this testing methodology is that it suggests how specification tests can be undertaken in circumstances in which alternative tests are difficult to construct.

Categorizing tests in this way is awkward, for several reasons.

- (1) Such a list will inevitably be incomplete. For example, it could be expanded to incorporate tests for specification encountered in more advanced work. Should there be categories for unit root and cointegration tests (see chapter 18), identification tests (see chapter 10), and selection bias tests (see chapter 16), for example? What about Bayesian "tests"?
- (2) It is common for practitioners to use a selection criterion, such as the Akaike information criterion, or adjusted R^2 , to aid in model specification, particularly for determining things like the number of lags to include. Should this methodology be classified as a test?
- (3) These categories are not mutually exclusive. There are non-nested variants of tests for nonspherical errors and of functional form tests, some tests for functional form are just variants of tests for structural break, and the RESET test is a special case of an OV test, for example.
- (4) Tests take different forms. Some are LM tests, some are LR tests, and some are W tests. Some use F -tables, some use t -tables, some use χ^2 -tables, and some require their own special tables. Some are exact tests, whereas some rest on an asymptotic justification.
- (5) Some tests are "specification" tests, involving a specific alternative, whereas others are "misspecification" tests, with no specific alternative.

This last distinction is particularly relevant for this chapter. A prominent feature of the list of general principles given earlier is the use of misspecification tests, the more common of which are often referred to as diagnostics. These tests are designed to detect an inadequate specification

(as opposed to "specification" tests, which examine the validity of a specific alternative). There have been calls for researchers to submit their models to misspecification tests as a matter of course, and it is becoming common for computer packages automatically to print out selected diagnostics.

Of the tests listed above, several fall into the misspecification category. Possibly the most prominent are the nonspherical error tests. As stressed in chapter 8, a significant value for the DW statistic could be due to several misspecifications (an omitted variable, a dynamic misspecification, or an incorrect functional form), not just to autocorrelated errors, the usual conclusion drawn by those following the AER methodology. The same is true of tests for heteroskedasticity. As noted in chapter 6, significant values of RESET could be due to an incorrect functional form, and tests for structural break and the first type of functional form test statistic could be significant because of a structural break, an omitted variable, or an incorrect functional form. So these tests should be viewed as misspecification tests. Outliers could arise from a variety of specification errors, so they also can be classified as misspecification tests.

It could be argued that the misspecification tests mentioned in the preceding paragraph are to some extent specification tests because they can be associated with one or more specific classes of alternatives that have inspired their construction. Because of this they are discussed in later chapters when that class of alternative is addressed. Three of the tests listed above, however, are sufficiently general in nature that there is no obvious alternative specification determining where they should appear in later chapters. These are data transformation tests, non-nested tests, and conditional moment tests.

Data transformation tests The idea behind data transformation tests is that if the null hypothesis of a linear functional form with a set of specific explanatory variables is correct, then estimating with raw data should yield coefficient estimates very similar to those obtained from using linearly transformed data. If the two sets of estimated coefficients are not similar, one can conclude that the null hypothesis is not correct, but one cannot draw any conclusion about what dimension of that null hypothesis is incorrect, since many different misspecifications could have given rise to this discrepancy. Choosing a specific transformation, and formalizing what is meant by "very similar," produces a test statistic. Fortunately, as explained in the technical notes, data transformation tests have been shown to be equivalent to OV tests, greatly simplifying their application.

Non-nested tests Two models are non-nested (or "separate") if one cannot be obtained from the other by imposing a restriction. The importance of this distinction is that in this circumstance it is not possible to follow the usual testing methodology, namely to employ a test of the restriction

as a specification test. Non-nested hypothesis tests provide a means of testing the specification of one model by exploiting the supposed "falsity" of other models. A model chosen to play the role of the "other" model need not be an alternative model under consideration, but this is usually the case. If the null model is the "correct" model, then the "other" model should not be able to explain anything beyond that explained by the null model. Formalizing this, as explained in the technical notes, produces a non-nested hypothesis test, on the basis of which the null can be either rejected or not rejected/accepted. If the former is the case, then one cannot conclude that the "other" model should be accepted – the role of the "other" model in this exercise is simply to act as a standard against which to measure the performance of the null. (This is what makes this test a misspecification test, rather than a specification test.) If one wants to say something about the "other" model, then the roles of the two hypotheses must be reversed, with the "other" model becoming the null, and the test repeated. Note that in this testing procedure it is possible to reject both models or to accept both models.

Conditional moment tests These tests are undertaken by selecting a function of the data and parameters that under a correct specification should be zero, computing this function for each observation (evaluated at the MLEs), taking the average over all the observations, and testing this average against zero. The function used for this purpose is usually a moment or a conditional moment (such as the product of an exogenous variable and the residual), explaining why these tests are called *moment (M)* or *conditional moment (CM)* tests. The test would be formed by creating an estimate of the variance-covariance matrix of this average and using a Wald test formula. Its main appeal is that in some circumstances it is easier to formulate appropriate moment conditions than to derive alternative tests.

5.5 R² AGAIN

The coefficient of determination, R², is often used in specification searches. Because it is so frequently abused by practitioners, an extension of our earlier (section 2.4) discussion of this statistic is warranted.

It is noted in the general notes to section 4.3 that the *F* test statistic could be interpreted in terms of R² and changes in R². Whether or not a set of extra independent variables belongs in a relationship depends on whether or not, by adding the extra regressors, the R² statistic increases significantly. This suggests that, when one is trying to determine which independent variable should be included in a relationship, one should search for the highest R².

This rule would lead to the choice of a relationship with too many regressors (independent variables) in it, because the addition of a regressor cannot cause the R² statistic to fall (for the same reason that the addition of a regressor cannot cause the minimized sum of squared residuals to become larger – minimizing without the restriction that the extra regressor must be ignored gives at least as low a minimand as when the restriction is imposed). Correcting the R² statistic for degrees of freedom solves this problem. The R² statistic adjusted to account for degrees of freedom is called the "adjusted R²" or " \bar{R}^2 " and is now reported by most packaged computer regression programs, and by practically all researchers, in place of the unadjusted R².

Adding another regressor changes the degrees of freedom associated with the measures that make up the R² statistic. If an additional regressor accounts for very little of the unexplained variation in the dependent variable, \bar{R}^2 falls (whereas R² rises). Thus, only if \bar{R}^2 rises should an extra variable be seriously considered for inclusion in the set of independent variables. This suggests that econometricians should search for the "best" set of independent variables by determining which potential set of independent variables produces the highest \bar{R}^2 . This procedure is valid only in the sense that the "correct set" of independent variables will produce, on average in repeated samples, a higher \bar{R}^2 than will any "incorrect" set of independent variables.

Another common use of the R² statistic is in the context of measuring the relative importance of different independent variables in determining the dependent variable. Textbooks present several ways of decomposing the R² statistic into component parts, each component being identified with one independent variable and used as a measure of the relative importance of that independent variable in the regression. Unfortunately, none of these partitions of R² is meaningful unless it happens that the independent variables are uncorrelated with one another in the sample at hand. (This happens only by experimental design or by extraordinary luck, economists almost never being in a position to affect either.) In the typical case in which the independent variables are correlated in the sample, these suggested partitionings are not meaningful because: (a) they can no longer be legitimately allocated to the independent variables; (b) they no longer add up to R²; or (c) they do add up to R² but contain negative as well as positive terms.

The main reason for this can be explained as follows. Suppose there are only two independent variables, and they are correlated in the sample. Two correlated variables can be thought of as having, between them, three sorts of variation: variation unique to the first variable, variation unique to the second variable, and variation common to both variables. (When the variables are uncorrelated, this third type of variation does not

exist.) Each of the three types of variation in this set of two variables "explains" some of the variation in the dependent variable. The basic problem is that no one can agree how to divide the explanatory power of the common variation between the two independent variables. If the dependent variable is regressed on both independent variables, the resulting R^2 reflects the explanatory power of all three types of independent variable variation. If the dependent variable is regressed on only one independent variable, variation unique to the other variable is removed and the resulting R^2 reflects the explanatory power of the other two types of independent variable variation. Thus, if one independent variable is removed, the remaining variable gets credit for *all* of the common variation. If the second independent variable were reinstated and the resulting increase in R^2 were used to measure the influence of this second variable, this variable would get credit for *none* of the common variation. Thus it would be illegitimate to measure the influence of an independent variable either by its R^2 in a regression of the dependent variable on only that independent variable, or by the addition to R^2 when that independent variable is added to a set of regressors. This latter measure clearly depends on the order in which the independent variables are added. Such procedures, and others like them, can only be used when the independent variables are uncorrelated in the sample. The use of breakdowns of the R^2 statistic in this context should be avoided.