and population residuals can be found by substituting [8.1.5] into [8.1.7]:

$$\hat{\mathbf{u}} = \mathbf{M_X}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \mathbf{M_X u}. \qquad [8.1.11]$$

The difference between the *OLS* estimate **b** and the true population parameter $\boldsymbol{\beta}$ is found by substituting [8.1.5] into [8.1.6]:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}\boldsymbol{\beta} + \mathbf{u}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}. \qquad [8.1.12]$$

The fit of an *OLS* regression is sometimes described in terms of the sample multiple correlation coefficient, or R^2. The *uncentered* R^2 (denoted R_u^2) is defined as the sum of squares of the fitted values ($\mathbf{x}_t'\mathbf{b}$) of the regression as a fraction of the sum of squares of y:

$$R_u^2 \equiv \frac{\sum_{t=1}^{T} (\mathbf{b}'\mathbf{x}_t\mathbf{x}_t'\mathbf{b})}{\sum_{t=1}^{T} y_t^2} = \frac{\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}. \qquad [8.1.13]$$

If the only explanatory variable in the regression were a constant term ($\mathbf{x}_t = 1$), then the fitted value for each observation would just be the sample mean $\bar{y}$ and the sum of squares of the fitted values would be $T\bar{y}^2$. This sum of squares is often compared with the sum of squares when a vector of variables $\mathbf{x}_t$ is included in the regression. The *centered* R^2 (denoted R_c^2) is defined as

$$R_c^2 \equiv \frac{\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - T\bar{y}^2}{\mathbf{y}'\mathbf{y} - T\bar{y}^2}. \qquad [8.1.14]$$

Most regression software packages report the centered R^2 rather than the uncentered R^2. If the regression includes a constant term, then R_c^2 must be between zero and unity. However, if the regression does not include a constant term, then R_c^2 can be negative.

The Classical Regression Assumptions

Statistical inference requires assumptions about the properties of the explanatory variables $\mathbf{x}_t$ and the population residuals u_t. The simplest case to analyze is the following.

Assumption 8.1: *(a) $\mathbf{x}_t$ is a vector of deterministic variables (for example, $\mathbf{x}_t$ might include a constant term and deterministic functions of t); (b) u_t is i.i.d. with mean 0 and variance σ^2; (c) u_t is Gaussian.*

To highlight the role of each of these assumptions, we first note the implications of Assumption 8.1(a) and (b) alone and then comment on the added implications that follow from (c).

Properties of the Estimated OLS *Coefficient Vector Under Assumption 8.1(a) and (b)*

In vector form, Assumption 8.1(b) could be written $E(\mathbf{u}) = \mathbf{0}$ and $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}_T$.

Taking expectations of [8.1.12] and using these conditions establishes that **b** is unbiased,

$$E(\mathbf{b}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[E(\mathbf{u})] = \boldsymbol{\beta}, \qquad [8.1.15]$$

with variance-covariance matrix given by

$$\begin{aligned} E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[E(\mathbf{u}\mathbf{u}')]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \qquad [8.1.16]$$

The *OLS* coefficient estimate $\mathbf{b}$ is unbiased and is a linear function of $\mathbf{y}$. The *Gauss-Markov theorem* states that the variance-covariance matrix of any alternative estimator of $\boldsymbol{\beta}$, if that estimator is also unbiased and a linear function of $\mathbf{y}$, differs from the variance-covariance matrix of $\mathbf{b}$ by a positive semidefinite matrix.[1] This means that an inference based on $\mathbf{b}$ about any linear combination of the elements of $\boldsymbol{\beta}$ will have a smaller variance than the corresponding inference based on any alternative linear unbiased estimator. The Gauss-Markov theorem thus establishes the optimality of the *OLS* estimate within a certain limited class.

Properties of the Estimated Coefficient Vector Under Assumption 8.1(a) Through (c)

When $\mathbf{u}$ is Gaussian, [8.1.12] implies that $\mathbf{b}$ is Gaussian. Hence, the preceding results imply

$$\mathbf{b} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \qquad [8.1.17]$$

It can further be shown that under Assumption 8.1(a) through (c), no unbiased estimator of $\boldsymbol{\beta}$ is more efficient than the *OLS* estimator $\mathbf{b}$.[2] Thus, with Gaussian residuals, the *OLS* estimator is optimal.

Properties of Estimated Residual Variance Under Assumption 8.1(a) and (b)

The *OLS* estimate of the variance of the disturbances σ^2 is

$$s^2 = RSS/(T-k) = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-k) = \mathbf{u}'\mathbf{M}_\mathbf{X}'\mathbf{M}_\mathbf{X}\mathbf{u}/(T-k) \qquad [8.1.18]$$

for $\mathbf{M}_\mathbf{X}$ the matrix in [8.1.8]. Recalling that $\mathbf{M}_\mathbf{X}$ is symmetric and idempotent, [8.1.18] becomes

$$s^2 = \mathbf{u}'\mathbf{M}_\mathbf{X}\mathbf{u}/(T-k). \qquad [8.1.19]$$

Also, since $\mathbf{M}_\mathbf{X}$ is symmetric, there exists a $(T \times T)$ matrix $\mathbf{P}$ such that[3]

$$\mathbf{M}_\mathbf{X} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' \qquad [8.1.20]$$

and

$$\mathbf{P}'\mathbf{P} = \mathbf{I}_T, \qquad [8.1.21]$$

where $\boldsymbol{\Lambda}$ is a $(T \times T)$ matrix with the eigenvalues of $\mathbf{M}_\mathbf{X}$ along the principal diagonal and zeros elsewhere. Note from [8.1.9] that $\mathbf{M}_\mathbf{X}\mathbf{v} = \mathbf{0}$ if $\mathbf{v}$ should be given by one of the k columns of $\mathbf{X}$. Assuming that the columns of $\mathbf{X}$ are linearly independent, the k columns of $\mathbf{X}$ thus represent k different eigenvectors of $\mathbf{M}_\mathbf{X}$ each associated

[1]See, for example, Theil (1971, pp. 119–20).

[2]See, for example, Theil (1971, pp. 390–91).

[3]See, for example, O'Nan (1976, p. 296).

with an eigenvalue equal to zero. Also from [8.1.8], $\mathbf{M_X v} = \mathbf{v}$ for any vector $\mathbf{v}$ that is orthogonal to the columns of $\mathbf{X}$ (that is, any vector $\mathbf{v}$ such that $\mathbf{X'v} = \mathbf{0}$); $(T - k)$ such vectors that are linearly independent can be found, associated with $(T - k)$ eigenvalues equal to unity. Thus, $\mathbf{\Lambda}$ contains k zeros and $(T - k)$ 1s along its principal diagonal. Notice from [8.1.20] that

$$\begin{aligned} \mathbf{u'M_X u} &= \mathbf{u'P\Lambda P'u} \\ &= (\mathbf{P'u})'\mathbf{\Lambda}(\mathbf{P'u}) \\ &= \mathbf{w'\Lambda w} \\ &= w_1^2\lambda_1 + w_2^2\lambda_2 + \cdots + w_T^2\lambda_T, \end{aligned} \tag{8.1.22}$$

where

$$\mathbf{w} \equiv \mathbf{P'u}.$$

Furthermore,

$$E(\mathbf{ww'}) = E(\mathbf{P'uu'P}) = \mathbf{P'}E(\mathbf{uu'})\mathbf{P} = \sigma^2\mathbf{P'P} = \sigma^2\mathbf{I}_T.$$

Thus, the elements of $\mathbf{w}$ are uncorrelated, with mean zero and variance σ^2. Since k of the λ's are zero and the remaining $T - k$ are unity, [8.1.22] becomes

$$\mathbf{u'M_X u} = w_1^2 + w_2^2 + \cdots + w_{T-k}^2. \tag{8.1.23}$$

Furthermore, each w_t^2 has expectation σ^2, so that

$$E(\mathbf{u'M_X u}) = (T - k)\sigma^2,$$

and from [8.1.19], s^2 gives an unbiased estimate of σ^2:

$$E(s^2) = \sigma^2.$$

Properties of Estimated Residual Variance Under Assumption 8.1(a) Through (c)

When u_t is Gaussian, w_t is also Gaussian and expression [8.1.23] is the sum of squares of $(T - k)$ independent $N(0, \sigma^2)$ variables. Thus,

$$RSS/\sigma^2 = \mathbf{u'M_X u}/\sigma^2 \sim \chi^2(T - k). \tag{8.1.24}$$

Again, it is possible to show that under Assumption 8.1(a) through (c), no other unbiased estimator of σ^2 has a smaller variance than does s^2.[4]

Notice also from [8.1.11] and [8.1.12] that $\mathbf{b}$ and $\hat{\mathbf{u}}$ are uncorrelated:

$$E[\hat{\mathbf{u}}(\mathbf{b} - \boldsymbol{\beta})'] = E[\mathbf{M_X uu'X(X'X)^{-1}}] = \sigma^2\mathbf{M_X X(X'X)^{-1}} = \mathbf{0}. \tag{8.1.25}$$

Under Assumption 8.1(a) through (c), both $\mathbf{b}$ and $\hat{\mathbf{u}}$ are Gaussian, so that absence of correlation implies that $\mathbf{b}$ and $\hat{\mathbf{u}}$ are independent. This means that $\mathbf{b}$ and s^2 are independent.

t Tests About $\boldsymbol{\beta}$ Under Assumption 8.1(a) Through (c)

Suppose that we wish to test the null hypothesis that β_i, the ith element of $\boldsymbol{\beta}$, is equal to some particular value β_i^0. The *OLS* t statistic for testing this null hypothesis is given by

$$t = \frac{(b_i - \beta_i^0)}{\hat{\sigma}_{b_i}} = \frac{(b_i - \beta_i^0)}{s(\xi^{ii})^{1/2}}, \tag{8.1.26}$$

[4]See Rao (1973, p. 319).

where ξ^{ii} denotes the row i, column i element of $(\mathbf{X}'\mathbf{X})^{-1}$ and $\hat{\sigma}_{b_i} \equiv \sqrt{s^2\xi^{ii}}$ is the standard error of the OLS estimate of the ith coefficient. The magnitude in [8.1.26] has an exact t distribution with $T - k$ degrees of freedom so long as $\mathbf{x}_t$ is deterministic and u_t is i.i.d. Gaussian. To verify this claim, note from [8.1.17] that under the null hypothesis, $b_i \sim N(\beta_i^0, \sigma^2\xi^{ii})$, meaning that $(b_i - \beta_i^0)/\sqrt{\sigma^2\xi^{ii}} \sim N(0, 1)$. Thus, if [8.1.26] is written as

$$t = \frac{(b_i - \beta_i^0)/\sqrt{\sigma^2\xi^{ii}}}{\sqrt{s^2/\sigma^2}},$$

the numerator is $N(0, 1)$ while from [8.1.24] the denominator is the square root of a $\chi^2\,(T - k)$ variable divided by its degrees of freedom. Recalling [8.1.25], the numerator and denominator are independent, confirming the exact t distribution claimed for [8.1.26].

F Tests About $\boldsymbol{\beta}$ Under Assumption 8.1(a) Through (c)

More generally, suppose we want a joint test of m different linear restrictions about $\boldsymbol{\beta}$, as represented by

$$H_0\colon \mathbf{R}\boldsymbol{\beta} = \mathbf{r}. \tag{8.1.27}$$

Here $\mathbf{R}$ is a known $(m \times k)$ matrix representing the particular linear combinations of $\boldsymbol{\beta}$ about which we entertain hypotheses and $\mathbf{r}$ is a known $(m \times 1)$ vector of the values that we believe these linear combinations take on. For example, to represent the simple hypothesis $\beta_i = \beta_i^0$ used previously, we would have $m = 1$, $\mathbf{R}$ a $(1 \times k)$ vector with unity in the ith position and zeros elsewhere, and $\mathbf{r}$ the scalar β_i^0. As a second example, consider a regression with $k = 4$ explanatory variables and the joint hypothesis that $\beta_1 + \beta_2 = 1$ and $\beta_3 = \beta_4$. In this case, $m = 2$ and

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad \mathbf{r} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{8.1.28}$$

Notice from [8.1.17] that under H_0,

$$\mathbf{Rb} \sim N(\mathbf{r}, \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'). \tag{8.1.29}$$

A *Wald test* of H_0 is based on the following result.

Proposition 8.1: *Consider an $(n \times 1)$ vector $\mathbf{z} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega}$ nonsingular. Then $\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z} \sim \chi^2(n)$.*

For the scalar case $(n = 1)$, observe that if $z \sim N(0, \sigma^2)$, then $(z/\sigma) \sim N(0, 1)$ and $z^2/\sigma^2 \sim \chi^2(1)$, as asserted by the proposition.

To verify Proposition 8.1 for the vector case, since $\boldsymbol{\Omega}$ is symmetric, there exists a matrix $\mathbf{P}$, as in [8.1.20] and [8.1.21], such that $\boldsymbol{\Omega} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ and $\mathbf{P}'\mathbf{P} = \mathbf{I}_n$ with $\boldsymbol{\Lambda}$ containing the eigenvalues of $\boldsymbol{\Omega}$. Since $\boldsymbol{\Omega}$ is positive definite, the diagonal elements of $\boldsymbol{\Lambda}$ are positive. Then

$$\begin{aligned} \mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z} &= \mathbf{z}'(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}')^{-1}\mathbf{z} \\ &= \mathbf{z}'[\mathbf{P}']^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{P}^{-1}\mathbf{z} \\ &= [\mathbf{P}^{-1}\mathbf{z}]'\boldsymbol{\Lambda}^{-1}\mathbf{P}^{-1}\mathbf{z} \\ &= \mathbf{w}'\boldsymbol{\Lambda}^{-1}\mathbf{w} \\ &= \sum_{i=1}^{n} w_i^2/\lambda_i, \end{aligned} \tag{8.1.30}$$

where $\mathbf{w} \equiv \mathbf{P}^{-1}\mathbf{z}$. Notice that $\mathbf{w}$ is Gaussian with mean zero and variance

$$E(\mathbf{ww}') = E(\mathbf{P}^{-1}\mathbf{zz}'[\mathbf{P}']^{-1}) = \mathbf{P}^{-1}\mathbf{\Omega}[\mathbf{P}']^{-1} = \mathbf{P}^{-1}\mathbf{P\Lambda P}'[\mathbf{P}']^{-1} = \mathbf{\Lambda}.$$

Thus [8.1.30] is the sum of squares of n independent Normal variables, each divided by its variance λ_i. It accordingly has a $\chi^2(n)$ distribution, as claimed.

Applying Proposition 8.1 directly to [8.1.29], under H_0,

$$(\mathbf{Rb} - \mathbf{r})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{r}) \sim \chi^2(m). \qquad [8.1.31]$$

Replacing σ^2 with the estimate s^2 and dividing by the number of restrictions gives the Wald form of the *OLS F* test of a linear hypothesis:

$$F = (\mathbf{Rb} - \mathbf{r})'[s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{r})/m. \qquad [8.1.32]$$

Note that [8.1.32] can be written

$$F = \frac{(\mathbf{Rb} - \mathbf{r})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{r})/m}{[RSS/(T - k)]/\sigma^2}.$$

The numerator is a $\chi^2(m)$ variable divided by its degrees of freedom, while the denominator is a $\chi^2(T - k)$ variable divided by its degrees of freedom. Again, since $\mathbf{b}$ and $\hat{\mathbf{u}}$ are independent, the numerator and denominator are independent of each other. Hence, [8.1.32] has an exact $F(m, T - k)$ distribution under H_0 when $\mathbf{x}_t$ is nonstochastic and u_t is i.i.d. Gaussian.

Notice that the t test of the simple hypothesis $\beta_i = \beta_i^0$ is a special case of the general formula [8.1.32], for which

$$F = (b_i - \beta_i^0)[s^2\xi^{ii}]^{-1}(b_i - \beta_i^0). \qquad [8.1.33]$$

This is the square of the t statistic in [8.1.26]. Since an $F(1, T - k)$ variable is just the square of a $t(T - k)$ variable, the identical answer results from (1) calculating [8.1.26] and using t tables to find the probability of so large an absolute value for a $t(T - k)$ variable, or (2) calculating [8.1.33] and using F tables to find the probability of so large a value for an $F(1, T - k)$ variable.

A Convenient Alternative Expression for the F *Test*

It is often straightforward to estimate the model in [8.1.1] subject to the restrictions in [8.1.27]. For example, to impose a constraint $\beta_1 = \beta_1^0$ on the first element of $\boldsymbol{\beta}$, we could just do an ordinary least squares regression of $y_t - \beta_1^0 x_{1t}$ on $x_{2t}, x_{3t}, \ldots, x_{kt}$. The resulting estimates $b_2^*, b_3^*, \ldots, b_k^*$ minimize $\sum_{t=1}^T [(y_t - \beta_1^0 x_{1t}) - b_2^* x_{2t} - b_3^* x_{3t} - \cdots - b_k^* x_{kt}]^2$ with respect to $b_2^*, b_3^*, \ldots, b_k^*$ and thus minimize the residual sum of squares [8.1.2] subject to the constraint that $\beta_1 = \beta_1^0$. Alternatively, to impose the constraint in [8.1.28], we could regress $y_t - x_{2t}$ on $(x_{1t} - x_{2t})$ and $(x_{3t} + x_{4t})$:

$$y_t - x_{2t} = \beta_1(x_{1t} - x_{2t}) + \beta_3(x_{3t} + x_{4t}) + u_t.$$

The *OLS* estimates b_1^* and b_3^* minimize

$$\begin{aligned}&\sum_{t=1}^T [(y_t - x_{2t}) - b_1^*(x_{1t} - x_{2t}) - b_3^*(x_{3t} + x_{4t})]^2 \\ &\qquad = \sum_{t=1}^T [y_t - b_1^* x_{1t} - (1 - b_1^*)x_{2t} - b_3^* x_{3t} - b_3^* x_{4t}]^2\end{aligned} \qquad [8.1.34]$$

and thus minimize [8.1.2] subject to [8.1.28].

Whenever the constraints in [8.1.27] can be imposed through a simple *OLS* regression on transformed variables, there is an easy way to calculate the F statistic

[8.1.32] just by comparing the residual sum of squares for the constrained and unconstrained regressions. The following result is established in Appendix 8.A at the end of this chapter.

Proposition 8.2: *Let* $\mathbf{b}$ *denote the unconstrained OLS estimate [8.1.6] and let* RSS_1 *be the residual sum of squares resulting from using this estimate:*

$$RSS_1 = \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\mathbf{b})^2. \qquad [8.1.35]$$

Let $\mathbf{b}^*$ *denote the constrained OLS estimate and* RSS_0 *the residual sum of squares from the constrained OLS estimation:*

$$RSS_0 = \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\mathbf{b}^*)^2. \qquad [8.1.36]$$

Then the Wald form of the OLS F test of a linear hypothesis [8.1.32] can equivalently be calculated as

$$F = \frac{(RSS_0 - RSS_1)/m}{RSS_1/(T - k)}. \qquad [8.1.37]$$

Expressions [8.1.37] and [8.1.32] will generate exactly the same number, regardless of whether the null hypothesis and the model are valid or not.

For example, suppose the sample size is $T = 50$ observations and the null hypothesis is $\beta_3 = \beta_4 = 0$ in an OLS regression with $k = 4$ explanatory variables. First regress y_t on x_{1t}, x_{2t}, x_{3t}, x_{4t} and call the residual sum of squares from this regression RSS_1. Next, regress y_t on just x_{1t} and x_{2t} and call the residual sum of squares from this restricted regression RSS_0. If

$$\frac{(RSS_0 - RSS_1)/2}{RSS_1/(50 - 4)}$$

is greater than 3.20 (the 5% critical value for an $F(2, 46)$ random variable), then the null hypothesis should be rejected.

8.2. Ordinary Least Squares Under More General Conditions

The previous section analyzed the regression model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$$

under the maintained Assumption 8.1 ($\mathbf{x}_t$ is deterministic and u_t is i.i.d. Gaussian). We will hereafter refer to this assumption as "case 1." This section generalizes this assumption to describe specifications likely to arise in time series analysis. Some of the key results are summarized in Table 8.1.

Case 2. Error Term i.i.d. Gaussian and Independent of Explanatory Variables

Consider the case in which $\mathbf{X}$ is stochastic but completely independent of $\mathbf{u}$.

Assumption 8.2:[5] *(a)* $\mathbf{x}_t$ *stochastic and independent of* u_s *for all* t, s; *(b)* $u_t \sim$ *i.i.d.* $N(0, \sigma^2)$.

[5] This could be replaced with the assumption $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$ with all the results to follow unchanged.

Many of the results for deterministic regressors continue to apply for this case. For example, taking expectations of [8.1.12] and exploiting the independence assumption,

$$E(\mathbf{b}) = \boldsymbol{\beta} + \{E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\}\{E(\mathbf{u})\} = \boldsymbol{\beta}, \qquad [8.2.1]$$

so that the *OLS* coefficient remains unbiased.

The distribution of test statistics for this case can be found by a two-step procedure. The first step evaluates the distribution conditional on $\mathbf{X}$; that is, it treats $\mathbf{X}$ as deterministic just as in the earlier analysis. The second step multiplies by the density of $\mathbf{X}$ and integrates over $\mathbf{X}$ to find the true unconditional distribution. For example, [8.1.17] implies that

$$\mathbf{b}|\mathbf{X} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \qquad [8.2.2]$$

If this density is multiplied by the density of $\mathbf{X}$ and integrated over $\mathbf{X}$, the result is no longer a Gaussian distribution; thus, $\mathbf{b}$ is non-Gaussian under Assumption 8.2. On the other hand, [8.1.24] implies that

$$RSS|\mathbf{X} \sim \sigma^2 \cdot \chi^2(T - k).$$

But this density is the same for all $\mathbf{X}$. Thus, when we multiply the density of $RSS|\mathbf{X}$ by the density of $\mathbf{X}$ and integrate, we will get exactly the same density. Hence, [8.1.24] continues to give the correct unconditional distribution for Assumption 8.2.

The same is true for the t and F statistics in [8.1.26] and [8.1.32]. Conditional on $\mathbf{X}$, $(b_i - \beta_i^0)/[\sigma(\xi^{ii})^{1/2}] \sim N(0, 1)$ and s/σ is the square root of an independent $[1/(T - k)]\cdot\chi^2(T - k)$ variable. Hence, conditional on $\mathbf{X}$, the statistic in [8.1.26] has a $t(T - k)$ distribution. Since this is true for any $\mathbf{X}$, when we multiply by the density of $\mathbf{X}$ and integrate over $\mathbf{X}$ we obtain the same distribution.

Case 3. Error Term i.i.d. Non-Gaussian and Independent of Explanatory Variables

Next consider the following specification.

Assumption 8.3: *(a) $\mathbf{x}_t$ stochastic and independent of u_s for all t, s; (b) u_t non-Gaussian but i.i.d. with mean zero, variance σ^2, and $E(u_t^4) = \mu_4 < \infty$; (c) $E(\mathbf{x}_t\mathbf{x}_t') = \mathbf{Q}_t$, a positive definite matrix with $(1/T)\Sigma_{t=1}^T \mathbf{Q}_t \rightarrow \mathbf{Q}$, a positive definite matrix; (d) $E(x_{it}x_{jt}x_{lt}x_{mt}) < \infty$ for all i, j, l, m, and t; (e) $(1/T)\Sigma_{t=1}^T(\mathbf{x}_t\mathbf{x}_t') \xrightarrow{p} \mathbf{Q}$.*

Since result [8.2.1] required only the independence assumption, $\mathbf{b}$ continues to be unbiased in this case. However, for hypothesis tests, the small-sample distributions of s^2 and the t and F statistics are no longer the same as when the population residuals are Gaussian. To justify the usual *OLS* inference rules, we have to appeal to asymptotic results, for which purpose Assumption 8.3 includes conditions (c) through (e). To understand these conditions, note that if $\mathbf{x}_t$ is covariance-stationary, then $E(\mathbf{x}_t\mathbf{x}_t')$ does not depend on t. Then $\mathbf{Q}_t = \mathbf{Q}$ for all t and condition (e) simply requires that $\mathbf{x}_t$ be ergodic for second moments. Assumption 8.3 also allows more general processes in that $E(\mathbf{x}_t\mathbf{x}_t')$ might be different for different t, so long as the limit of $(1/T)\Sigma_{t=1}^T E(\mathbf{x}_t\mathbf{x}_t')$ can be consistently estimated by $(1/T)\Sigma_{t=1}^T(\mathbf{x}_t\mathbf{x}_t')$.

TABLE 8.1
Properties of *OLS* Estimates and Test Statistics Under Various Assumptions

	Coefficient **b**	*Variance* s^2	*t statistic*	*F statistic*
Case 1	unbiased $\mathbf{b} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$	unbiased $(T - k)s^2/\sigma^2 \sim \chi^2(T - k)$	exact $t(T - k)$	exact $F(m, T - k)$
Case 2	unbiased non-Gaussian	unbiased $(T - k)s^2/\sigma^2 \sim \chi^2(T - k)$	exact $t(T - k)$	exact $F(m, T - k)$
Case 3	unbiased $\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$	unbiased $\sqrt{T}(s_T^2 - \sigma^2) \xrightarrow{L} N(0, \mu_4 - \sigma^4)$	$t_T \xrightarrow{L} N(0, 1)$	$mF_T \xrightarrow{L} \chi^2(m)$
Case 4	biased $\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$	biased $\sqrt{T}(s_T^2 - \sigma^2) \xrightarrow{L} N(0, \mu_4 - \sigma^4)$	$t_T \xrightarrow{L} N(0, 1)$	$mF_T \xrightarrow{L} \chi^2(m)$

Regression model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, $\mathbf{b}$ is given by [8.1.6], s^2 by [8.1.18], t statistic by [8.1.26], and F statistic by [8.1.32]; μ_4 denotes $E(u_t^4)$.
Case 1: $\mathbf{X}$ nonstochastic, $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$.
Case 2: $\mathbf{X}$ stochastic, $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$, $\mathbf{X}$ independent of $\mathbf{u}$.
Case 3: $\mathbf{X}$ stochastic, $\mathbf{u} \sim$ non-Gaussian $(\mathbf{0}, \sigma^2\mathbf{I}_T)$, $\mathbf{X}$ independent of $\mathbf{u}$, $T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} \mathbf{Q}$.
Case 4: Stationary autoregression with independent errors, $\mathbf{Q}$ given by [8.2.27].

To describe the asymptotic results, we denote the *OLS* estimator [8.1.3] by $\mathbf{b}_T$ to emphasize that it is based on a sample of size T. Our interest is in the behavior of $\mathbf{b}_T$ as T becomes large. We first establish that the *OLS* coefficient estimator is consistent under Assumption 8.3, that is, that $\mathbf{b}_T \xrightarrow{p} \boldsymbol{\beta}$.

Note that [8.1.12] implies

$$\begin{aligned}\mathbf{b}_T - \boldsymbol{\beta} &= \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t u_t\right] \\ &= \left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t u_t\right].\end{aligned} \qquad [8.2.3]$$

Consider the first term in [8.2.3]. Assumption 8.3(e) and Proposition 7.1 imply that

$$\left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \xrightarrow{p} \mathbf{Q}^{-1}. \qquad [8.2.4]$$

Considering next the second term in [8.2.3], notice that $\mathbf{x}_t u_t$ is a martingale difference sequence with variance-covariance matrix given by

$$E(\mathbf{x}_t u_t \mathbf{x}_t' u_t) = \{E(\mathbf{x}_t\mathbf{x}_t')\}\cdot\sigma^2,$$

which is finite. Thus, from Example 7.11,

$$\left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t u_t\right] \xrightarrow{p} \mathbf{0}. \qquad [8.2.5]$$

Applying Example 7.2 to [8.2.3] through [8.2.5],

$$\mathbf{b}_T - \boldsymbol{\beta} \xrightarrow{p} \mathbf{Q}^{-1}\cdot\mathbf{0} = \mathbf{0},$$

verifying that the *OLS* estimator is consistent.

Next turn to the asymptotic distribution of $\mathbf{b}$. Notice from [8.2.3] that

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) = \left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^{T} \mathbf{x}_t u_t\right]. \qquad [8.2.6]$$

We saw in [8.2.4] that the first term converges in probability to $\mathbf{Q}^{-1}$. The second term is $\sqrt{T}$ times the sample mean of $\mathbf{x}_t u_t$, where $\mathbf{x}_t u_t$ is a martingale difference sequence with variance $\sigma^2\cdot E(\mathbf{x}_t\mathbf{x}_t') = \sigma^2\mathbf{Q}_t$ and $(1/T)\Sigma_{t=1}^{\infty}\sigma^2\mathbf{Q}_t \to \sigma^2\mathbf{Q}$. Notice that under Assumption 8.3 we can apply Proposition 7.9:

$$\left[(1/\sqrt{T})\sum_{t=1}^{T} \mathbf{x}_t u_t\right] \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{Q}). \qquad [8.2.7]$$

Combining [8.2.6], [8.2.4], and [8.2.7], we see as in Example 7.5 that

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, [\mathbf{Q}^{-1}\cdot(\sigma^2\mathbf{Q})\cdot\mathbf{Q}^{-1}]) = N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}). \qquad [8.2.8]$$

In other words, we can act as if

$$\mathbf{b}_T \approx N(\boldsymbol{\beta}, \sigma^2\mathbf{Q}^{-1}/T), \qquad [8.2.9]$$

where the symbol $\approx$ means "is approximately distributed." Recalling Assumption 8.3(e), in large samples $\mathbf{Q}$ should be close to $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$. Thus $\mathbf{Q}^{-1}/T$ should be close to $[\Sigma_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t']^{-1} = (\mathbf{X}_T'\mathbf{X}_T)^{-1}$ for $\mathbf{X}_T$ the same $(T \times k)$ matrix that was represented in [8.1.5] simply by $\mathbf{X}$ (again, the subscript T is added at this point to emphasize that the dimensions of this matrix depend on T). Thus, [8.2.9] can be approximated by

$$\mathbf{b}_T \approx N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}_T'\mathbf{X}_T)^{-1}).$$

This, of course, is the same result obtained in [8.1.17], which assumed Gaussian disturbances. With non-Gaussian disturbances the distribution is not exact, but provides an increasingly good approximation as the sample size grows.

Next, consider consistency of the variance estimate s_T^2. Notice that the population residual sum of squares can be written

$$\begin{aligned}
&(\mathbf{y}_T - \mathbf{X}_T\boldsymbol{\beta})'(\mathbf{y}_T - \mathbf{X}_T\boldsymbol{\beta}) \\
&\quad = (\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T + \mathbf{X}_T\mathbf{b}_T - \mathbf{X}_T\boldsymbol{\beta})'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T + \mathbf{X}_T\mathbf{b}_T - \mathbf{X}_T\boldsymbol{\beta}) \qquad [8.2.10] \\
&\quad = (\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T) + (\mathbf{X}_T\mathbf{b}_T - \mathbf{X}_T\boldsymbol{\beta})'(\mathbf{X}_T\mathbf{b}_T - \mathbf{X}_T\boldsymbol{\beta}),
\end{aligned}$$

where cross-product terms have vanished, since

$$(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'\mathbf{X}_T(\mathbf{b}_T - \boldsymbol{\beta}) = 0,$$

by the *OLS* orthogonality condition [8.1.10]. Dividing [8.2.10] by T,

$$\begin{aligned}
&(1/T)(\mathbf{y}_T - \mathbf{X}_T\boldsymbol{\beta})'(\mathbf{y}_T - \mathbf{X}_T\boldsymbol{\beta}) \\
&\qquad = (1/T)(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T) + (1/T)(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{X}_T'\mathbf{X}_T(\mathbf{b}_T - \boldsymbol{\beta}),
\end{aligned}$$

or

$$\begin{aligned}
&(1/T)(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T) \\
&\qquad = (1/T)(\mathbf{u}_T'\mathbf{u}_T) - (\mathbf{b}_T - \boldsymbol{\beta})'(\mathbf{X}_T'\mathbf{X}_T/T)(\mathbf{b}_T - \boldsymbol{\beta}). \qquad [8.2.11]
\end{aligned}$$

Now, $(1/T)(\mathbf{u}_T'\mathbf{u}_T) = (1/T)\Sigma_{t=1}^T u_t^2$, where $\{u_t^2\}$ is an i.i.d. sequence with mean σ^2. Thus, by the law of large numbers,

$$(1/T)(\mathbf{u}_T'\mathbf{u}_T) \xrightarrow{p} \sigma^2.$$

For the second term in [8.2.11], we have $(\mathbf{X}_T'\mathbf{X}_T/T) \xrightarrow{p} \mathbf{Q}$ and $(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{p} \mathbf{0}$, and so, from Proposition 7.1,

$$(\mathbf{b}_T - \boldsymbol{\beta})'(\mathbf{X}_T'\mathbf{X}_T/T)(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{p} \mathbf{0}'\mathbf{Q}\mathbf{0} = 0.$$

Substituting these results into [8.2.11],

$$(1/T)(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T) \xrightarrow{p} \sigma^2. \qquad [8.2.12]$$

Now, [8.2.12] describes an estimate of the variance, which we denote $\hat{\sigma}_T^2$:

$$\hat{\sigma}_T^2 \equiv (1/T)(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T). \qquad [8.2.13]$$

The *OLS* estimator given in [8.1.18],

$$s_T^2 = [1/(T - k)](\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T)'(\mathbf{y}_T - \mathbf{X}_T\mathbf{b}_T), \qquad [8.2.14]$$

differs from $\hat{\sigma}_T^2$ by a term that vanishes as $T \to \infty$,

$$s_T^2 = a_T \cdot \hat{\sigma}_T^2,$$

where $a_T \equiv [T/(T - k)]$ with $\lim_{T\to\infty} a_T = 1$. Hence, from Proposition 7.1,

$$\operatorname{plim} s_T^2 = 1 \cdot \sigma^2,$$

establishing consistency of s_T^2.

To find the asymptotic distribution of s_T^2, consider first $\sqrt{T}(\hat{\sigma}_T^2 - \sigma^2)$. From [8.2.11], this equals

$$\begin{aligned}\sqrt{T}(\hat{\sigma}_T^2 - \sigma^2) = (1/\sqrt{T})(\mathbf{u}_T'\mathbf{u}_T) - \sqrt{T}\sigma^2 \\ - \sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})'(\mathbf{X}_T'\mathbf{X}_T/T)(\mathbf{b}_T - \boldsymbol{\beta}).\end{aligned} \qquad [8.2.15]$$

But

$$(1/\sqrt{T})(\mathbf{u}_T'\mathbf{u}_T) - \sqrt{T}\sigma^2 = (1/\sqrt{T}) \sum_{t=1}^{T} (u_t^2 - \sigma^2),$$

where $\{u_t^2 - \sigma^2\}$ is a sequence of i.i.d. variables with mean zero and variance $E(u_t^2 - \sigma^2)^2 = E(u_t^4) - 2\sigma^2 E(u_t^2) + \sigma^4 = \mu_4 - \sigma^4$. Hence, by the central limit theorem,

$$(1/\sqrt{T})(\mathbf{u}_T'\mathbf{u}_T) - \sqrt{T}\sigma^2 \xrightarrow{L} N(0, (\mu_4 - \sigma^4)). \qquad [8.2.16]$$

For the last term in [8.2.15], we have $\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$, $(\mathbf{X}_T'\mathbf{X}_T/T) \xrightarrow{p} \mathbf{Q}$, and $(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{p} \mathbf{0}$. Hence,

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})'(\mathbf{X}_T'\mathbf{X}_T/T)(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{p} 0. \qquad [8.2.17]$$

Putting [8.2.16] and [8.2.17] into [8.2.15], we conclude

$$\sqrt{T}(\hat{\sigma}_T^2 - \sigma^2) \xrightarrow{L} N(0, (\mu_4 - \sigma^4)). \qquad [8.2.18]$$

To see that s_T^2 has this same limiting distribution, notice that

$$\begin{aligned}\sqrt{T}(s_T^2 - \sigma^2) - \sqrt{T}(\hat{\sigma}_T^2 - \sigma^2) &= \sqrt{T}\{[T/(T - k)]\hat{\sigma}_T^2 - \hat{\sigma}_T^2\} \\ &= [(k\sqrt{T})/(T - k)]\hat{\sigma}_T^2.\end{aligned}$$

But $\lim_{T\to\infty} [(k\sqrt{T})/(T - k)] = 0$, establishing that

$$\sqrt{T}(s_T^2 - \sigma^2) - \sqrt{T}(\hat{\sigma}_T^2 - \sigma^2) \xrightarrow{p} 0 \cdot \sigma^2 = 0$$

and hence, from Proposition 7.3(a),

$$\sqrt{T}(s_T^2 - \sigma^2) \xrightarrow{L} N(0, (\mu_4 - \sigma^4)). \qquad [8.2.19]$$

Notice that if we are relying on asymptotic justifications for test statistics, theory offers us no guidance for choosing between s^2 and $\hat{\sigma}^2$ as estimates of σ^2, since they have the same limiting distribution.

Next consider the asymptotic distribution of the *OLS* t test of the null hypothesis $\beta_i = \beta_i^0$,

$$t_T = \frac{(b_{iT} - \beta_i^0)}{s_T\sqrt{\xi_T^{ii}}} = \frac{\sqrt{T}(b_{iT} - \beta_i^0)}{s_T\sqrt{T\xi_T^{ii}}}, \qquad [8.2.20]$$

where ξ_T^{ii} denotes the row i, column i element of $(\mathbf{X}_T'\mathbf{X}_T)^{-1}$. We have seen that $\sqrt{T}(b_{i,T} - \beta^0) \xrightarrow{L} N(0, \sigma^2 q^{ii})$, where q^{ii} denotes the row i, column i element of $\mathbf{Q}^{-1}$. Similarly, $T\xi_T^{ii}$ is the row i, column i element of $(\mathbf{X}_T'\mathbf{X}_T/T)^{-1}$ and converges in probability to q^{ii}. Also, $s_T \xrightarrow{p} \sigma$. Hence, the t statistic [8.2.20] has a limiting distribution that is the same as a $N(0, \sigma^2 q^{ii})$ variable divided by $\sqrt{\sigma^2 q^{ii}}$; that is,

$$t_T \xrightarrow{L} N(0, 1). \qquad [8.2.21]$$

Now, under the more restrictive conditions of Assumption 8.2, we saw that t_T would have a t distribution with $(T - k)$ degrees of freedom. Recall that a t variable with N degrees of freedom has the distribution of the ratio of a $N(0, 1)$ variable to the square root of $(1/N)$ times an independent $\chi^2(N)$ variable. But a $\chi^2(N)$ variable in turn is the sum of N squares of independent $N(0, 1)$ variables.

Thus, letting Z denote a $N(0, 1)$ variable, a t variable with N degrees of freedom has the same distribution as

$$t_N = \frac{Z}{\{(Z_1^2 + Z_2^2 + \cdots + Z_N^2)/N\}^{1/2}}.$$

By the law of large numbers,

$$(Z_1^2 + Z_2^2 + \cdots + Z_N^2)/N \xrightarrow{p} E(Z_t^2) = 1,$$

and so $t_N \xrightarrow{L} N(0, 1)$. Hence, the critical value for a t variable with N degrees of freedom will be arbitrarily close to that for a $N(0, 1)$ variable as N becomes large. Even though the statistic calculated in [8.2.20] does not have an exact $t(T - k)$ distribution under Assumption 8.3, if we treat it as if it did, then we will not be far wrong if our sample is sufficiently large.

The same is true of [8.1.32], the F test of m different restrictions:

$$\begin{aligned} F_T &= (\mathbf{R}\mathbf{b}_T - \mathbf{r})'[s_T^2\mathbf{R}(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}_T - \mathbf{r})/m \\ &= \sqrt{T}(\mathbf{R}\mathbf{b}_T - \mathbf{r})'[s_T^2\mathbf{R}(\mathbf{X}_T'\mathbf{X}_T/T)^{-1}\mathbf{R}']^{-1}\sqrt{T}(\mathbf{R}\mathbf{b}_T - \mathbf{r})/m. \end{aligned} \qquad [8.2.22]$$

Here $s_T^2 \xrightarrow{p} \sigma^2$, $\mathbf{X}_T'\mathbf{X}_T/T \xrightarrow{p} \mathbf{Q}$, and, under the null hypothesis,

$$\begin{aligned} \sqrt{T}(\mathbf{R}\mathbf{b}_T - \mathbf{r}) &= [\mathbf{R}\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})] \\ &\xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{R}\mathbf{Q}^{-1}\mathbf{R}'). \end{aligned}$$

Hence, under the null hypothesis,

$$m \cdot F_T \xrightarrow{p} [\mathbf{R}\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})]'[\sigma^2\mathbf{R}\mathbf{Q}^{-1}\mathbf{R}']^{-1}[\mathbf{R}\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})].$$

This is a quadratic function of a Normal vector of the type described by Proposition 8.1, from which

$$m \cdot F_T \xrightarrow{L} \chi^2(m).$$

Thus an asymptotic inference can be based on the approximation

$$(\mathbf{R}\mathbf{b}_T - \mathbf{r})'[s_T^2\mathbf{R}(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}_T - \mathbf{r}) \approx \chi^2(m). \qquad [8.2.23]$$

This is known as the *Wald form* of the *OLS* χ^2 test.

As in the case of the t and limiting Normal distributions, viewing [8.2.23] as $\chi^2(m)$ and viewing [8.2.22] as $F(m, T - k)$ asymptotically amount to the same test. Recall that an $F(m, N)$ variable is a ratio of a $\chi^2(m)$ variable to an independent $\chi^2(N)$ variable, each divided by its degrees of freedom. Thus, if Z_i denotes a $N(0, 1)$ variable and X a $\chi^2(m)$ variable,

$$F_{m,N} = \frac{X/m}{(Z_1^2 + Z_2^2 + \cdots + Z_N^2)/N}.$$

For the denominator,

$$(Z_1^2 + Z_2^2 + \cdots + Z_N^2)/N \xrightarrow{p} E(Z_t^2) = 1,$$

implying

$$F_{m,N} \underset{N\to\infty}{\xrightarrow{L}} X/m.$$

Hence, comparing [8.2.23] with a $\chi^2(m)$ critical value or comparing [8.2.22] with an $F(m, T - k)$ critical value will result in the identical test for sufficiently large T (see Exercise 8.2).

For a given sample of size T, the small-sample distribution (the t or F distribution) implies wider confidence intervals than the large-sample distribution (the

Normal or χ^2 distribution). Even when the justification for using the t or F distribution is only asymptotic, many researchers prefer to use the t or F tables rather than the Normal or χ^2 tables on the grounds that the former are more conservative and may represent a better approximation to the true small-sample distribution.

If we are relying only on the asymptotic distribution, the Wald test statistic [8.2.23] can be generalized to allow a test of a nonlinear set of restrictions on $\boldsymbol{\beta}$. Consider a null hypothesis consisting of m separate nonlinear restrictions of the form $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$ where $\mathbf{g}: \mathbb{R}^k \to \mathbb{R}^m$ and $\mathbf{g}(\cdot)$ has continuous first derivatives. Result [8.2.8] and Proposition 7.4 imply that

$$\sqrt{T}[\mathbf{g}(\mathbf{b}_T) - \mathbf{g}(\boldsymbol{\beta}_0)] \xrightarrow{L} \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]\mathbf{z},$$

where $\mathbf{z} \sim N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$ and

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

denotes the $(m \times k)$ matrix of derivatives of $\mathbf{g}(\cdot)$ with respect to $\boldsymbol{\beta}$, evaluated at the true value $\boldsymbol{\beta}_0$. Under the null hypothesis that $\mathbf{g}(\boldsymbol{\beta}_0) = \mathbf{0}$, it follows from Proposition 8.1 that

$$\{\sqrt{T}\cdot\mathbf{g}(\mathbf{b}_T)\}'\left\{\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]\sigma^2\mathbf{Q}^{-1}\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]'\right\}^{-1}\{\sqrt{T}\cdot\mathbf{g}(\mathbf{b}_T)\} \xrightarrow{L} \chi^2(m).$$

Recall that $\mathbf{Q}$ is the plim of $(1/T)(\mathbf{X}_T'\mathbf{X}_T)$. Since $\partial\mathbf{g}/\partial\boldsymbol{\beta}'$ is continuous and since $\mathbf{b}_T \xrightarrow{p} \boldsymbol{\beta}_0$, it follows from Proposition 7.1 that

$$\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\mathbf{b}_T}\right] \xrightarrow{p} \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right].$$

Hence a set of m nonlinear restrictions about $\boldsymbol{\beta}$ of the form $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$ can be tested with the statistic

$$\{\mathbf{g}(\mathbf{b}_T)\}'\left\{\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\mathbf{b}_T}\right]s_T^2(\mathbf{X}_T'\mathbf{X}_T)^{-1}\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\mathbf{b}_T}\right]'\right\}^{-1}\{\mathbf{g}(\mathbf{b}_T)\} \xrightarrow{L} \chi^2(m).$$

Note that the Wald test for linear restrictions [8.2.23] can be obtained as a special case of this more general formula by setting $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{R}\boldsymbol{\beta} - \mathbf{r}$.

One disadvantage of the Wald test for nonlinear restrictions is that the answer one obtains can be different depending on how the restrictions $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$ are parameterized. For example, the hypotheses $\beta_1 = \beta_2$ and $\beta_1/\beta_2 = 1$ are equivalent, and asymptotically a Wald test based on either parameterization should give the same answer. However, in a particular finite sample the answers could be quite different. In effect, the nonlinear Wald test approximates the restriction $\mathbf{g}(\mathbf{b}_T) = \mathbf{0}$ by the linear restriction

$$\mathbf{g}(\boldsymbol{\beta}_0) + \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right](\mathbf{b}_T - \boldsymbol{\beta}_0) = \mathbf{0}.$$

Some care must be taken to ensure that this linearization is reasonable over the range of plausible values for $\boldsymbol{\beta}$. See Gregory and Veall (1985), Lafontaine and White (1986), and Phillips and Park (1988) for further discussion.

Case 4. Estimating Parameters for an Autoregression

Consider now estimation of the parameters of a pth-order autoregression by *OLS*.

Assumption 8.4: The regression model is

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \qquad [8.2.24]$$

with roots of $(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = 0$ *outside the unit circle and with* $\{\varepsilon_t\}$ *an i.i.d. sequence with mean zero, variance* σ^2, *and finite fourth moment* μ_4.

An autoregression has the form of the standard regression model $y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$ with $\mathbf{x}_t' = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ and $u_t = \varepsilon_t$. Note, however, that an autoregression cannot satisfy condition (a) of Assumption 8.2 or 8.3. Even though u_t is independent of $\mathbf{x}_t$ under Assumption 8.4, it will not be the case that u_t is independent of $\mathbf{x}_{t+1}$. Without this independence, none of the small-sample results for case 1 applies. Specifically, even if ε_t is Gaussian, the *OLS* coefficient $\mathbf{b}$ gives a biased estimate of $\boldsymbol{\beta}$ for an autoregression, and the standard t and F statistics can only be justified asymptotically.

However, the asymptotic results for case 4 are the same as for case 3 and are derived in essentially the same way. To adapt the earlier notation, suppose that the sample consists of $T + p$ observations on y_t, numbered $(y_{-p+1}, y_{-p+2}, \ldots, y_0, y_1, \ldots, y_T)$; *OLS* estimation will thus use observations 1 through T. Then, as in [8.2.6],

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) = \left[(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{x}_t u_t\right]. \qquad [8.2.25]$$

The first term in [8.2.25] is

$$\left[(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}$$
$$= \begin{bmatrix} 1 & T^{-1}\Sigma y_{t-1} & T^{-1}\Sigma y_{t-2} & \cdots & T^{-1}\Sigma y_{t-p} \\ T^{-1}\Sigma y_{t-1} & T^{-1}\Sigma y_{t-1}^2 & T^{-1}\Sigma y_{t-1}y_{t-2} & \cdots & T^{-1}\Sigma y_{t-1}y_{t-p} \\ T^{-1}\Sigma y_{t-2} & T^{-1}\Sigma y_{t-2}y_{t-1} & T^{-1}\Sigma y_{t-2}^2 & \cdots & T^{-1}\Sigma y_{t-2}y_{t-p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ T^{-1}\Sigma y_{t-p} & T^{-1}\Sigma y_{t-p}y_{t-1} & T^{-1}\Sigma y_{t-p}y_{t-2} & \cdots & T^{-1}\Sigma y_{t-p}^2 \end{bmatrix}^{-1},$$

where Σ denotes summation over $t = 1$ to T. The elements in the first row or column are of the form $T^{-1}\Sigma y_{t-j}$ and converge in probability to $\mu = E(y_t)$, by Proposition 7.5. Other elements are of the form $T^{-1}\Sigma y_{t-i}y_{t-j}$, which, from [7.2.14], converges in probability to

$$E(y_{t-i}y_{t-j}) = \gamma_{|i-j|} + \mu^2.$$

Hence

$$\left[(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1} \xrightarrow{p} \mathbf{Q}^{-1} \qquad [8.2.26]$$

where

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & \mu & \mu & \cdots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \cdots & \gamma_{p-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \cdots & \gamma_{p-2} + \mu^2 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \cdots & \gamma_0 + \mu^2 \end{bmatrix}. \qquad [8.2.27]$$

For the second term in [8.2.25], observe that $\mathbf{x}_t u_t$ is a martingale difference sequence with positive definite variance-covariance matrix given by

$$E(\mathbf{x}_t u_t u_t \mathbf{x}_t') = E(u_t^2) \cdot E(\mathbf{x}_t \mathbf{x}_t') = \sigma^2 \mathbf{Q}.$$

Using an argument similar to that in Example 7.15, it can be shown that

$$\left[(1/\sqrt{T}) \sum_{t=1}^{T} \mathbf{x}_t u_t \right] \xrightarrow{L} N(\mathbf{0}, \sigma^2 \mathbf{Q}) \qquad [8.2.28]$$

(see Exercise 8.3). Substituting [8.2.26] and [8.2.28] into [8.2.25],

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}). \qquad [8.2.29]$$

It is straightforward to verify further that $\mathbf{b}_T$ and s_T^2 are consistent for this case. From [8.2.26], the asymptotic variance-covariance matrix of $\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})$ can be estimated consistently by $s_T^2(\mathbf{X}_T'\mathbf{X}_T/T)^{-1}$, meaning that standard t and F statistics that treat $\mathbf{b}_T$ as if it were $N(\boldsymbol{\beta}, s_T^2(\mathbf{X}_T'\mathbf{X}_T)^{-1})$ will yield asymptotically valid tests of hypotheses about the coefficients of an autoregression.

As a special case of [8.2.29], consider *OLS* estimation of a first-order autoregression,

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

with $|\phi| < 1$. Then $\mathbf{Q}$ is the scalar $E(y_{t-1}^2) = \gamma_0$, the variance of an $AR(1)$ process. We saw in Chapter 3 that this is given by $\sigma^2/(1 - \phi^2)$. Hence, for $\hat{\phi}$ the *OLS* coefficient,

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2},$$

result [8.2.29] implies that

$$\sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{L} N(0, \sigma^2 \cdot [\sigma^2/(1 - \phi^2)]^{-1}) = N(0, 1 - \phi^2). \qquad [8.2.30]$$

If more precise results than the asymptotic approximation in equation [8.2.29] are desired, the exact small-sample distribution of $\hat{\phi}_T$ can be calculated in either of two ways. If the errors in the autoregression [8.2.24] are $N(0, \sigma^2)$, then for any specified numerical value for $\phi_1, \phi_2, \ldots, \phi_p$ and c the exact small-sample distribution can be calculated using numerical routines developed by Imhof (1961); for illustrations of this method, see Evans and Savin (1981) and Flavin (1983). An alternative is to approximate the small-sample distribution by *Monte Carlo* methods. Here the idea is to use a computer to generate pseudo-random variables $\varepsilon_1, \ldots, \varepsilon_T$, each distributed $N(0, \sigma^2)$ from numerical algorithms such as that described in Kinderman and Ramage (1976). For fixed starting values $y_{-p+1}, \ldots, y_1$, the values for $y_1, y_2, \ldots, y_T$ can then be calculated by iterating on [8.2.24].[6] One then estimates the parameters of [8.2.24] with an *OLS* regression on this artificial sample. A new sample is generated for which a new *OLS* regression is estimated. By performing, say, 10,000 such regressions, an estimate of the exact small-sample distribution of the *OLS* estimates can be obtained.

For the case of a first-order autoregression, it is known from such calculations that $\hat{\phi}_T$ is downward-biased in small samples, with the bias becoming more severe as ϕ approaches unity. For example, for a sample of size $T = 25$ generated by [8.2.24] with $p = 1$, $c = 0$, and $\phi = 1$, the estimate $\hat{\phi}_T$ based on *OLS* estimation of [8.2.24] (with a constant term included) will be less than the true value of 1 in 95% of the samples, and will even fall below 0.6 in 10% of the samples.[7]

Case 5. Errors Gaussian with Known Variance-Covariance Matrix

Next consider the following case.

Assumption 8.5: *(a)* $\mathbf{x}_t$ *stochastic; (b) conditional on the full matrix* $\mathbf{X}$, *the vector* $\mathbf{u}$ *is* $N(\mathbf{0}, \sigma^2\mathbf{V})$; *(c)* $\mathbf{V}$ *is a known positive definite matrix.*

When the errors for different dates have different variances but are uncorrelated with each other (that is, $\mathbf{V}$ is diagonal), then the errors are said to exhibit *heteroskedasticity*. For $\mathbf{V}$ nondiagonal, the errors are said to be *autocorrelated*. Writing the variance-covariance matrix as the product of some scalar σ^2 and a matrix $\mathbf{V}$ is a convention that will help simplify the algebra and interpretation for some examples of heteroskedasticity and autocorrelation. Note again that Assumption 8.5(b) could not hold for an autoregression, since conditional on $\mathbf{x}_{t+1} = (1, y_t, y_{t-1}, \ldots, y_{t-p+1})'$ and $\mathbf{x}_t$, the value of u_t is known with certainty.

Recall from [8.1.12] that

$$(\mathbf{b} - \boldsymbol{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}.$$

Taking expectations conditional on $\mathbf{X}$,

$$E[(\mathbf{b} - \boldsymbol{\beta})|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\cdot E(\mathbf{u}) = \mathbf{0},$$

and by the law of iterated expectations,

$$E(\mathbf{b} - \boldsymbol{\beta}) = E_{\mathbf{X}}\{E[(\mathbf{b} - \boldsymbol{\beta})|\mathbf{X}]\} = \mathbf{0}.$$

Hence, the *OLS* coefficient estimate is unbiased.

The variance of $\mathbf{b}$ conditional on $\mathbf{X}$ is

$$\begin{aligned} E\{[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})']|\mathbf{X}\} &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]|\mathbf{X}\} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \qquad [8.2.31]$$

Thus, conditional on $\mathbf{X}$,

$$\mathbf{b}|\mathbf{X} \sim N\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right).$$

[6] Alternatively, one can generate the initial values for y with a draw from the appropriate unconditional distribution. Specifically, generate a $(p \times 1)$ vector $\mathbf{v} \sim N(\mathbf{0}, \mathbf{I}_p)$ and set $(y_{-p+1}, \ldots, y_0)' = \mu\cdot\mathbf{1} + \mathbf{P}\cdot\mathbf{v}$, where $\mu = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$, $\mathbf{1}$ denotes a $(p \times 1)$ vector of 1s, and $\mathbf{P}$ is the Cholesky factor such that $\mathbf{P}\cdot\mathbf{P}' = \boldsymbol{\Gamma}$ for $\boldsymbol{\Gamma}$ the $(p \times p)$ matrix whose columns stacked in a $(p^2 \times 1)$ vector comprise the first column of the matrix $\sigma^2[\mathbf{I}_{p^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1}$, where $\mathbf{F}$ is the $(p \times p)$ matrix defined in equation [1.2.3] in Chapter 1.

[7] These values can be inferred from Table B.5.

Unless $\mathbf{V} = \mathbf{I}_T$, this is not the same variance matrix as in [8.1.17], so that the *OLS* t statistic [8.1.26] does not have the interpretation as a Gaussian variable divided by an estimate of its standard deviation. Thus [8.1.26] will not have a $t(T - k)$ distribution in small samples, nor will it even asymptotically be $N(0, 1)$. A valid test of the hypothesis that $\beta_i = \beta_i^0$ for case 5 would be based not on [8.1.26] but rather on

$$t^* = \frac{(b_i - \beta_i^0)}{s\sqrt{d_{ii}}}, \tag{8.2.32}$$

where d_{ii} denotes the row i, column i element of $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. This statistic will be asymptotically $N(0, 1)$.

Although one could form an inference based on [8.2.32], in this case in which $\mathbf{V}$ is known, a superior estimator and test procedure are described in Section 8.3. First, however, we consider a more general case in which $\mathbf{V}$ is of unknown form.

Case 6. Errors Serially Uncorrelated but with General Heteroskedasticity

It may be possible to design asymptotically valid tests even in the presence of heteroskedasticity of a completely unknown form. This point was first observed by Eicker (1967) and White (1980) and extended to time series regressions by Hansen (1982) and Nicholls and Pagan (1983).

Assumption 8.6: *(a) $\mathbf{x}_t$ stochastic, including perhaps lagged values of y; (b) $\mathbf{x}_t u_t$ is a martingale difference sequence; (c) $E(u_t^2\mathbf{x}_t\mathbf{x}_t') = \mathbf{\Omega}_t$, a positive definite matrix, with $(1/T)\Sigma_{t=1}^T\mathbf{\Omega}_t$ converging to the positive definite matrix $\mathbf{\Omega}$ and $(1/T)\Sigma_{t=1}^T u_t^2\mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} \mathbf{\Omega}$; (d) $E(u_t^4 x_{it}x_{jt}x_{lt}x_{mt}) < \infty$ for all i, j, l, m, and t; (e) plims of $(1/T)\Sigma_{t=1}^T u_t x_{it}\mathbf{x}_t\mathbf{x}_t'$ and $(1/T)\Sigma_{t=1}^T x_{it}x_{jt}\mathbf{x}_t\mathbf{x}_t'$ exist and are finite for all i and j and $(1/T)\Sigma_{t=1}^T\mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} \mathbf{Q}$, a nonsingular matrix.*

Assumption 8.6(b) requires u_t to be uncorrelated with its own lagged values and with current and lagged values of $\mathbf{x}$. Although the errors are presumed to be serially uncorrelated, Assumption 8.6(c) allows a broad class of conditional heteroskedasticity for the errors. As an example of such heteroskedasticity, consider a regression with a single i.i.d. explanatory variable x_t with $E(x_t^2) = \mu_2$ and $E(x_t^4) = \mu_4$. Suppose that the variance of the residual for date t is given by $E(u_t^2|x_t) = a + bx_t^2$. Then $E(u_t^2x_t^2) = E_x[E(u_t^2|x_t)\cdot x_t^2] = E_x[(a + bx_t^2)\cdot x_t^2] = a\mu_2 + b\mu_4$. Thus, $\Omega_t = a\mu_2 + b\mu_4 = \Omega$ for all t. By the law of large numbers, $(1/T)\Sigma_{t=1}^T u_t^2x_t^2$ will converge to the population moment Ω. Assumption 8.6(c) allows more general conditional heteroskedasticity in that $E(u_t^2x_t^2)$ might be a function of t, provided that the time average of $(u_t^2x_t^2)$ converges. Assumption 8.6(d) and (e) impose bounds on higher moments of $\mathbf{x}$ and u.

Consistency of $\mathbf{b}$ is established using the same arguments as in case 3. The asymptotic variance is found from writing

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) = \left[(1/T)\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^T \mathbf{x}_t u_t\right].$$

Assumption 8.6(e) ensures that

$$\left[(1/T)\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$$

for some nonsingular matrix $\mathbf{Q}$. Similarly, $\mathbf{x}_t u_t$ satisfies the conditions of Proposition 7.9, from which

$$\left[(1/\sqrt{T}) \sum_{t=1}^{T} \mathbf{x}_t u_t\right] \xrightarrow{L} N(\mathbf{0}, \mathbf{\Omega}).$$

The asymptotic distribution of the *OLS* estimate is thus given by

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}). \qquad [8.2.33]$$

White's proposal was to estimate the asymptotic variance matrix consistently by substituting $\hat{\mathbf{Q}}_T = (1/T)\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t'$ and $\hat{\mathbf{\Omega}}_T = (1/T)\Sigma_{t=1}^T \hat{u}_t^2\mathbf{x}_t\mathbf{x}_t'$ into [8.2.33], where $\hat{u}_t$ denotes the *OLS* residual [8.1.4]. The following result is established in Appendix 8.A to this chapter.

Proposition 8.3: *With heteroskedasticity of unknown form satisfying Assumption 8.6, the asymptotic variance-covariance matrix of the OLS coefficient vector can be consistently estimated by*

$$\hat{\mathbf{Q}}_T^{-1}\hat{\mathbf{\Omega}}_T\hat{\mathbf{Q}}_T^{-1} \xrightarrow{p} \mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}. \qquad [8.2.34]$$

Recalling [8.2.33], the *OLS* estimate $\mathbf{b}_T$ can be treated as if

$$\mathbf{b}_T \approx N(\boldsymbol{\beta}, \hat{\mathbf{V}}_T/T)$$

where

$$\begin{aligned}\hat{\mathbf{V}}_T &= \hat{\mathbf{Q}}_T^{-1}\hat{\mathbf{\Omega}}_T\hat{\mathbf{Q}}_T^{-1} \\ &= (\mathbf{X}_T'\mathbf{X}_T/T)^{-1}\left[(1/T) \sum_{t=1}^{T} \hat{u}_t^2\mathbf{x}_t\mathbf{x}_t'\right](\mathbf{X}_T'\mathbf{X}_T/T)^{-1} \qquad [8.2.35] \\ &= T\cdot(\mathbf{X}_T'\mathbf{X}_T)^{-1}\left[\sum_{t=1}^{T} \hat{u}_t^2\mathbf{x}_t\mathbf{x}_t'\right](\mathbf{X}_T'\mathbf{X}_t)^{-1}.\end{aligned}$$

The square root of the row i, column i element of $\hat{\mathbf{V}}_T/T$ is known as a *heteroskedasticity-consistent standard error* for the *OLS* estimate b_i. We can, of course, also use $(\hat{\mathbf{V}}_T/T)$ to test a joint hypothesis of the form $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where $\mathbf{R}$ is an $(m \times k)$ matrix summarizing m separate hypotheses about $\boldsymbol{\beta}$. Specifically,

$$(\mathbf{R}\mathbf{b}_T - \mathbf{r})'[\mathbf{R}(\hat{\mathbf{V}}_T/T)\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}_T - \mathbf{r}) \qquad [8.2.36]$$

has the same asymptotic distribution as

$$[\sqrt{T}(\mathbf{R}\mathbf{b}_T - \mathbf{r})]'(\mathbf{R}\mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}\mathbf{R}')^{-1}[\sqrt{T}(\mathbf{R}\mathbf{b}_T - \mathbf{r})],$$

which, from [8.2.33], is a quadratic form of an asymptotically Normal $(m \times 1)$ vector $\sqrt{T}(\mathbf{R}\mathbf{b}_T - \mathbf{r})$ with weighting matrix the inverse of its variance-covariance matrix, $(\mathbf{R}\mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}\mathbf{R}')$. Hence, [8.2.36] has an asymptotic χ^2 distribution with m degrees of freedom.

It is also possible to develop an estimate of the asymptotic variance-covariance matrix of $\mathbf{b}_T$ that is robust with respect to both heteroskedasticity and autocorrelation:

$$\begin{aligned}&(\hat{\mathbf{V}}_T/T) \\ &= (\mathbf{X}_T'\mathbf{X}_T)^{-1}\Bigg[\sum_{t=1}^{T} \hat{u}_t^2\mathbf{x}_t\mathbf{x}_t' \\ &\quad + \sum_{v=1}^{q}\left[1 - \frac{v}{q+1}\right]\sum_{t=v+1}^{T}(\mathbf{x}_t\hat{u}_t\hat{u}_{t-v}\mathbf{x}_{t-v}' + \mathbf{x}_{t-v}\hat{u}_{t-v}\hat{u}_t\mathbf{x}_t')\Bigg](\mathbf{X}_T'\mathbf{X}_T)^{-1}.\end{aligned}$$

Here q is a parameter representing the number of autocorrelations used to approximate the dynamics for u_t. The square root of the row i, column i element of $(\hat{\mathbf{V}}_T/T)$ is known as the Newey-West (1987) heteroskedasticity- and autocorrelation-consistent standard error for the *OLS* estimator. The basis for this expression and alternative ways to calculate heteroskedasticity- and autocorrelation-consistent standard errors will be discussed in Chapter 10.

8.3. *Generalized Least Squares*

The previous section evaluated *OLS* estimation under a variety of assumptions, including $E(\mathbf{uu}') \neq \sigma^2\mathbf{I}_T$. Although *OLS* can be used in this last case, generalized least squares (*GLS*) is usually preferred.

GLS *with Known Covariance Matrix*

Let us reconsider data generated according to Assumption 8.5, under which $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$ with $\mathbf{V}$ a known $(T \times T)$ matrix. Since $\mathbf{V}$ is symmetric and positive definite, there exists a nonsingular $(T \times T)$ matrix L such that[8]

$$\mathbf{V}^{-1} = \mathbf{L}'\mathbf{L}. \qquad [8.3.1]$$

Imagine transforming the population residuals $\mathbf{u}$ by L:

$$\underset{(T\times 1)}{\tilde{\mathbf{u}}} \equiv \mathbf{Lu}.$$

This would generate a new set of residuals $\tilde{\mathbf{u}}$ with mean $\mathbf{0}$ and variance conditional on $\mathbf{X}$ given by

$$E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}'|\mathbf{X}) = \mathbf{L}\cdot E(\mathbf{uu}'|\mathbf{X})\mathbf{L}' = \mathbf{L}\sigma^2\mathbf{V}\mathbf{L}'.$$

But $\mathbf{V} = [\mathbf{V}^{-1}]^{-1} = [\mathbf{L}'\mathbf{L}]^{-1}$, meaning

$$E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}'|\mathbf{X}) = \sigma^2\mathbf{L}[\mathbf{L}'\mathbf{L}]^{-1}\mathbf{L}' = \sigma^2\mathbf{I}_T. \qquad [8.3.2]$$

We can thus take the matrix equation that characterizes the basic regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

and premultiply both sides by L:

$$\mathbf{Ly} = \mathbf{LX}\boldsymbol{\beta} + \mathbf{Lu},$$

to produce a new regression model

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{u}}, \qquad [8.3.3]$$

where

$$\tilde{\mathbf{y}} \equiv \mathbf{Ly} \qquad \tilde{\mathbf{X}} \equiv \mathbf{LX} \qquad \tilde{\mathbf{u}} \equiv \mathbf{Lu} \qquad [8.3.4]$$

with $\tilde{\mathbf{u}}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$. Hence, the transformed model [8.3.3] satisfies Assumption 8.2, meaning that all the results for that case apply to [8.3.3]. Specifically, the estimator

$$\tilde{\mathbf{b}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = (\mathbf{X}'\mathbf{L}'\mathbf{LX})^{-1}\mathbf{X}'\mathbf{L}'\mathbf{Ly} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad [8.3.5]$$

[8]We know that there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{V} = \mathbf{PP}'$ and so $\mathbf{V}^{-1} = [\mathbf{P}']^{-1}\mathbf{P}^{-1}$. Take $\mathbf{L} = \mathbf{P}^{-1}$ to deduce [8.3.1].

is Gaussian with mean $\boldsymbol{\beta}$ and variance $\sigma^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ conditional on $\mathbf{X}$ and is the minimum-variance unbiased estimator conditional on $\mathbf{X}$. The estimator [8.3.5] is known as the *generalized least squares* (*GLS*) estimator. Similarly,

$$\tilde{s}^2 = [1/(T - k)] \sum_{t=1}^{T} (\tilde{y}_t - \tilde{\mathbf{x}}_t'\tilde{\mathbf{b}})^2 \tag{8.3.6}$$

has an exact $[\sigma^2/(T - k)]\cdot\chi^2(T - K)$ distribution under Assumption 8.5, while

$$(\mathbf{R}\tilde{\mathbf{b}} - \mathbf{r})'[\tilde{s}^2\mathbf{R}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\tilde{\mathbf{b}} - \mathbf{r})/m$$

has an exact $F(m, T - k)$ distribution under the null hypothesis $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

We now discuss several examples to make these ideas concrete.

Heteroskedasticity

A simple case to analyze is one for which the variance of u_t is presumed to be proportional to the square of one of the explanatory variables for that equation, say, x_{1t}^2:

$$E(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \sigma^2 \begin{bmatrix} x_{11}^2 & 0 & \cdots & 0 \\ 0 & x_{12}^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & x_{1T}^2 \end{bmatrix} = \sigma^2\mathbf{V}.$$

Then it is easy to see that

$$\mathbf{L} = \begin{bmatrix} 1/|x_{11}| & 0 & \cdots & 0 \\ 0 & 1/|x_{12}| & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1/|x_{1T}| \end{bmatrix}$$

satisfies conditions [8.3.1] and [8.3.2]. Hence, if we regress $y_t/|x_{1t}|$ on $\mathbf{x}_t/|x_{1t}|$, all the standard *OLS* output from the regression will be valid.

Autocorrelation

As a second example, consider

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{8.3.7}$$

where $|\rho| < 1$ and ε_t is Gaussian white noise with variance σ^2. Then

$$E(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \frac{\sigma^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix} = \sigma^2\mathbf{V}. \tag{8.3.8}$$

Notice from expression [5.2.18] that the matrix

$$\mathbf{L} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} \qquad [8.3.9]$$

satisfies [8.3.1]. The *GLS* estimates are found from an *OLS* regression of $\tilde{\mathbf{y}} = \mathbf{Ly}$ on $\tilde{\mathbf{X}} = \mathbf{LX}$; that is, regress $y_1\sqrt{1-\rho^2}$ on $x_1\sqrt{1-\rho^2}$ and $y_t - \rho y_{t-1}$ on $\mathbf{x}_t - \rho\mathbf{x}_{t-1}$ for $t = 2, 3, \ldots, T$.

GLS *and Maximum Likelihood Estimation*

Assumption 8.5 asserts that $\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$. Hence, the log of the likelihood of $\mathbf{y}$ conditioned on $\mathbf{X}$ is given by

$$(-T/2)\log(2\pi) - (1/2)\log|\sigma^2\mathbf{V}| - (1/2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{V})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \qquad [8.3.10]$$

Notice that [8.3.1] can be used to write the last term in [8.3.10] as

$$\begin{aligned} -(1/2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{V})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) & \\ &= -[1/(2\sigma^2)](\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{L}'\mathbf{L})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= -[1/(2\sigma^2)](\mathbf{Ly} - \mathbf{LX}\boldsymbol{\beta})'(\mathbf{Ly} - \mathbf{LX}\boldsymbol{\beta}) \\ &= -[1/(2\sigma^2)](\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}). \end{aligned} \qquad [8.3.11]$$

Similarly, the middle term in [8.3.10] can be written as in [5.2.24]:

$$-(1/2)\log|\sigma^2\mathbf{V}| = -(T/2)\log(\sigma^2) + \log|\det(\mathbf{L})|, \qquad [8.3.12]$$

where $|\det(\mathbf{L})|$ denotes the absolute value of the determinant of $\mathbf{L}$. Substituting [8.3.11] and [8.3.12] into [8.3.10], the conditional log likelihood can be written as

$$-(T/2)\log(2\pi) - (T/2)\log(\sigma^2) + \log|\det(\mathbf{L})| - [1/(2\sigma^2)](\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}). \qquad [8.3.13]$$

Thus, the log likelihood is maximized with respect to $\boldsymbol{\beta}$ by an *OLS* regression of $\tilde{\mathbf{y}}$ on $\tilde{\mathbf{X}}$,[9] meaning that the *GLS* estimate [8.3.5] is also the maximum likelihood estimate under Assumption 8.5.

The *GLS* estimate $\tilde{\mathbf{b}}$ is still likely to be reasonable even if the residuals $\mathbf{u}$ are non-Gaussian. Specifically, the residuals of the transformed regression [8.3.3] have mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}_T$, and so this regression satisfies the conditions of the Gauss-Markov theorem—even if the residuals are non-Gaussian, $\tilde{\mathbf{b}}$ will have minimum variance (conditional on $\mathbf{X}$) among the class of all unbiased estimators that are linear functions of $\mathbf{y}$. Hence, maximization of [8.3.13], or quasi-maximum likelihood estimation, may offer a useful estimating principle even for non-Gaussian $\mathbf{u}$.

GLS *When the Variance Matrix of Residuals Must Be Estimated from the Data*

Up to this point we have been assuming that the elements of $\mathbf{V}$ are known a priori. More commonly, $\mathbf{V}$ is posited to be of a particular form $\mathbf{V}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a

[9]This assumes that the parameters of $\mathbf{L}$ do not involve $\boldsymbol{\beta}$, as is implied by Assumption 8.5.

vector of parameters that must be estimated from the data. For example, with first-order serial correlation of residuals as in [8.3.7], $\mathbf{V}$ is the matrix in [8.3.8] and $\boldsymbol{\theta}$ is the scalar ρ. As a second example, we might postulate that the variance of observation t depends on the explanatory variables according to

$$E(u_t^2|\mathbf{x}_t) = \sigma^2(1 + \alpha_1 x_{1t}^2 + \alpha_2 x_{2t}^2),$$

in which case $\boldsymbol{\theta} = (\alpha_1, \alpha_2)'$.

Our task is then to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ jointly from the data. One approach is to use as estimates the values of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ that maximize [8.3.13]. Since one can always form [8.3.13] and maximize it numerically, this approach has the appeal of offering a single rule to follow whenever $E(\mathbf{uu}'|\mathbf{X})$ is not of the simple form $\sigma^2\mathbf{I}_T$. However, other, simpler estimators can also have desirable properties.

It often turns out to be the case that

$$\sqrt{T}(\mathbf{X}_T'[\mathbf{V}_T(\hat{\boldsymbol{\theta}}_T)]^{-1}\mathbf{X}_T)^{-1}(\mathbf{X}_T'[\mathbf{V}_T(\hat{\boldsymbol{\theta}}_T)]^{-1}\mathbf{y}_T) \xrightarrow{p} \sqrt{T}(\mathbf{X}_T'[\mathbf{V}_T(\boldsymbol{\theta}_0)]^{-1}\mathbf{X}_T)^{-1}(\mathbf{X}_T'[\mathbf{V}_T(\boldsymbol{\theta}_0)]^{-1}\mathbf{y}_T),$$

where $\mathbf{V}_T(\boldsymbol{\theta}_0)$ denotes the true variance of errors and $\hat{\boldsymbol{\theta}}_T$ is any consistent estimate of $\boldsymbol{\theta}$. Moreover, a consistent estimate of $\boldsymbol{\theta}$ can often be obtained from a simple analysis of *OLS* residuals. Thus, an estimate coming from a few simple *OLS* and *GLS* regressions can have the same asymptotic distribution as the maximum likelihood estimator. Since regressions are much easier to implement than numerical maximization, the simpler estimates are often used.

Estimation with First-Order Autocorrelation of Regression Residuals and No Lagged Endogenous Variables

We illustrate these issues by considering a regression whose residuals follow the $AR(1)$ process [8.3.7]. For now we maintain the assumption that $\mathbf{u}|\mathbf{X}$ has mean zero and variance $\sigma^2\mathbf{V}(\rho)$, noting that this rules out lagged endogenous variables; that is, we assume that $\mathbf{x}_t$ is uncorrelated with u_{t-s}. The following subsection comments on the importance of this assumption. Recalling that the determinant of a lower triangular matrix is just the product of the terms on the principal diagonal, we see from [8.3.9] that $\det(\mathbf{L}) = \sqrt{1-\rho^2}$. Thus, the log likelihood [8.3.13] for this case is

$$\begin{aligned} -(T/2)\log(2\pi) &- (T/2)\log(\sigma^2) + (1/2)\log(1-\rho^2) \\ &- [(1-\rho^2)/(2\sigma^2)](y_1 - \mathbf{x}_1'\boldsymbol{\beta})^2 \\ &- [1/(2\sigma^2)]\sum_{t=2}^{T}[(y_t - \mathbf{x}_t'\boldsymbol{\beta}) - \rho(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})]^2. \end{aligned} \quad [8.3.14]$$

One approach, then, is to maximize [8.3.14] numerically with respect to $\boldsymbol{\beta}$, ρ, and σ^2. The reader may recognize [8.3.14] as the exact log likelihood function for an $AR(1)$ process (equation [5.2.9]) with $(y_t - \mu)$ replaced by $(y_t - \mathbf{x}_t'\boldsymbol{\beta})$.

Just as in the $AR(1)$ case, simpler estimates (with the same asymptotic distribution) are obtained if we condition on the first observation, seeking to maximize

$$\begin{aligned} -[(T-1)/2]\log(2\pi) &- [(T-1)/2]\log(\sigma^2) \\ &- [1/(2\sigma^2)]\sum_{t=2}^{T}[(y_t - \mathbf{x}_t'\boldsymbol{\beta}) - \rho(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})]^2. \end{aligned} \quad [8.3.15]$$

If we knew the value of ρ, then the value of $\boldsymbol{\beta}$ that maximizes [8.3.15] could be found by an *OLS* regression of $(y_t - \rho y_{t-1})$ on $(\mathbf{x}_t - \rho\mathbf{x}_{t-1})$ for $t = 2, 3, \ldots,$

T (call this regression A). Conversely, if we knew the value of $\boldsymbol{\beta}$, then the value of ρ that maximizes [8.3.15] would be found by an *OLS* regression of $(y_t - \mathbf{x}_t'\boldsymbol{\beta})$ on $(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})$ for $t = 2, 3, \ldots, T$ (call this regression B). We can thus start with an initial guess for ρ (often $\rho = 0$), and perform regression A to get an initial estimate of $\boldsymbol{\beta}$. For $\rho = 0$, this initial estimate of $\boldsymbol{\beta}$ would just be the *OLS* estimate $\mathbf{b}$. This estimate of $\boldsymbol{\beta}$ can be used in regression B to get an updated estimate of ρ, for example, by regressing the *OLS* residual $\hat{u}_t = y_t - \mathbf{x}_t'\mathbf{b}$ on its own lagged value. This new estimate of ρ can be used to repeat the two regressions. Zigzagging back and forth between A and B is known as the *iterated Cochrane-Orcutt* method and will converge to a local maximum of [8.3.15].

Alternatively, consider the estimate of ρ that results from the first iteration alone,

$$\hat{\rho} = \frac{(1/T)\sum_{t=1}^{T} \hat{u}_{t-1}\hat{u}_t}{(1/T)\sum_{t=1}^{T} \hat{u}_{t-1}^2}, \qquad [8.3.16]$$

where $\hat{u}_t = y_t - \mathbf{x}_t'\mathbf{b}$ and $\mathbf{b}$ is the *OLS* estimate of $\boldsymbol{\beta}$. To simplify expressions, we have renormalized the number of observations in the original sample to $T + 1$, denoted $y_0, y_1, \ldots, y_T$, so that T observations are used in the conditional maximum likelihood estimation. Notice that

$$\hat{u}_t = (y_t - \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\beta}'\mathbf{x}_t - \mathbf{b}'\mathbf{x}_t) = u_t + (\boldsymbol{\beta} - \mathbf{b})'\mathbf{x}_t,$$

allowing the numerator of [8.3.16] to be written

$$\begin{aligned}
(1/T)\sum_{t=1}^{T} \hat{u}_t\hat{u}_{t-1} \\
&= (1/T)\sum_{t=1}^{T} [u_t + (\boldsymbol{\beta} - \mathbf{b})'\mathbf{x}_t][u_{t-1} + (\boldsymbol{\beta} - \mathbf{b})'\mathbf{x}_{t-1}] \\
&= (1/T)\sum_{t=1}^{T} (u_t u_{t-1}) + (\boldsymbol{\beta} - \mathbf{b})'(1/T)\sum_{t=1}^{T} (u_t\mathbf{x}_{t-1} + u_{t-1}\mathbf{x}_t) \qquad [8.3.17] \\
&\quad + (\boldsymbol{\beta} - \mathbf{b})'\left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_{t-1}'\right](\boldsymbol{\beta} - \mathbf{b}).
\end{aligned}$$

As long as $\mathbf{b}$ is a consistent estimate of $\boldsymbol{\beta}$ and boundedness conditions ensure that plims of $(1/T)\Sigma_{t=1}^{T} u_t\mathbf{x}_{t-1}$, $(1/T)\Sigma_{t=1}^{T} u_{t-1}\mathbf{x}_t$, and $(1/T)\Sigma_{t=1}^{T} \mathbf{x}_t\mathbf{x}_{t-1}'$ exist, then

$$\begin{aligned}
(1/T)\sum_{t=1}^{T} \hat{u}_t\hat{u}_{t-1} &\xrightarrow{p} (1/T)\sum_{t=1}^{T} u_t u_{t-1} \\
&= (1/T)\sum_{t=1}^{T} (\varepsilon_t + \rho u_{t-1})u_{t-1} \qquad [8.3.18] \\
&\xrightarrow{p} \rho\cdot\text{Var}(u).
\end{aligned}$$

Similar analysis establishes that the denominator of [8.3.16] converges in probability to $\text{Var}(u)$, so that $\hat{\rho} \xrightarrow{p} \rho$.

If u_t is uncorrelated with $\mathbf{x}_s$ for $s = t - 1$, t, and $t + 1$, one can make the stronger claim that an estimate of ρ based on an autoregression of the *OLS* residuals $\hat{u}_t$ (expression [8.3.16]) has the same asymptotic distribution as an estimate of ρ based on the true population residuals u_t. Specifically, if $\text{plim}[(1/T)\Sigma_{t=1}^{T} u_t\mathbf{x}_{t-1}] =$

$\text{plim}[(1/T)\Sigma_{t=1}^{T} u_{t-1}\mathbf{x}_t] = \mathbf{0}$, then multiplying [8.3.17] by $\sqrt{T}$, we find

$$
\begin{aligned}
&(1/\sqrt{T}) \sum_{t=1}^{T} \hat{u}_t \hat{u}_{t-1} \\
&\quad = (1/\sqrt{T}) \sum_{t=1}^{T} (u_t u_{t-1}) + \sqrt{T}(\boldsymbol{\beta} - \mathbf{b})'(1/T) \sum_{t=1}^{T} (u_t \mathbf{x}_{t-1} + u_{t-1}\mathbf{x}_t) \\
&\qquad + \sqrt{T}(\boldsymbol{\beta} - \mathbf{b})'\left[(1/T) \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_{t-1}'\right](\boldsymbol{\beta} - \mathbf{b}) \\
&\quad \xrightarrow{p} (1/\sqrt{T}) \sum_{t=1}^{T} (u_t u_{t-1}) + \sqrt{T}(\boldsymbol{\beta} - \mathbf{b})'\mathbf{0} \\
&\qquad + \sqrt{T}(\boldsymbol{\beta} - \mathbf{b})' \,\text{plim}\left[(1/T) \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_{t-1}'\right]\mathbf{0} \\
&\quad = (1/\sqrt{T}) \sum_{t=1}^{T} (u_t u_{t-1}).
\end{aligned}
\tag{8.3.19}
$$

Hence,

$$
\sqrt{T}\left[\frac{(1/T) \sum_{t=1}^{T} \hat{u}_{t-1}\hat{u}_t}{(1/T) \sum_{t=1}^{T} \hat{u}_{t-1}^2}\right] \xrightarrow{p} \sqrt{T}\left[\frac{(1/T) \sum_{t=1}^{T} u_{t-1}u_t}{(1/T) \sum_{t=1}^{T} u_{t-1}^2}\right]. \tag{8.3.20}
$$

The *OLS* estimate of ρ based on the population residuals would have an asymptotic distribution given by [8.2.30]:

$$
\sqrt{T}\left[\frac{(1/T) \sum_{t=1}^{T} \hat{u}_{t-1}\hat{u}_t}{(1/T) \sum_{t=1}^{T} \hat{u}_{t-1}^2} - \rho\right] \xrightarrow{L} N(0, (1 - \rho^2)). \tag{8.3.21}
$$

Result [8.3.20] implies that an estimate of ρ has the same asymptotic distribution when based on any consistent estimate of $\boldsymbol{\beta}$. If the Cochrane-Orcutt iterations are stopped after just one evaluation of $\hat{\rho}$, the resulting estimate of ρ has the same asymptotic distribution as the estimate of ρ emerging from any subsequent step of the iteration.

The same also turns out to be true of the *GLS* estimate $\tilde{\mathbf{b}}$.

Proposition 8.4: *Suppose that Assumption 8.5(a) and (b) holds with* $\mathbf{V}$ *given by [8.3.8] and* $|\rho| < 1$. *Suppose in addition that* $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t u_s \xrightarrow{p} \mathbf{0}$ *for all s and that* $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ *and* $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t\mathbf{x}_{t-1}'$ *have finite plims. Then the GLS estimate* $\tilde{\mathbf{b}}$ *constructed from* $\mathbf{V}(\hat{\rho})$ *for* $\hat{\rho}$ *given by [8.3.16] has the same asymptotic distribution as* $\tilde{\mathbf{b}}$ *constructed from* $\mathbf{V}(\rho)$ *for the true value of* ρ.

Serial Correlation with Lagged Endogenous Variables

An *endogenous variable* is a variable that is correlated with the regression error term u_t. Many of the preceding results about serially correlated errors no longer hold if the regression contains lagged endogenous variables. For example, consider estimation of

$$y_t = \beta y_{t-1} + \gamma x_t + u_t, \qquad [8.3.22]$$

where u_t follows an $AR(1)$ process as in [8.3.7]. Since (1) u_t is correlated with u_{t-1} and (2) u_{t-1} is correlated with y_{t-1}, it follows that u_t is correlated with the explanatory variable y_{t-1}. Accordingly, it is not the case that $\text{plim}[(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t u_t] = \mathbf{0}$, the key condition required for consistency of the OLS estimator $\mathbf{b}$. Hence, $\hat{\rho}$ in [8.3.16] is not a consistent estimate of ρ.

If one nevertheless iterates on the Cochrane-Orcutt procedure, then the algorithm will converge to a *local* maximum of [8.3.15]. However, the resulting GLS estimate $\tilde{\mathbf{b}}$ need not be a consistent estimate of $\boldsymbol{\beta}$. Notwithstanding, the *global* maximum of [8.3.15] should provide a consistent estimate of $\boldsymbol{\beta}$. By experimenting with start-up values for iterated Cochrane-Orcutt other than $\rho = 0$, one should find this global maximum.[10]

A simple estimate of ρ that is consistent in the presence of lagged endogenous variables was suggested by Durbin (1960). Multiplying [8.3.22] by $(1 - \rho L)$ gives

$$y_t = (\rho + \beta)y_{t-1} - \rho\beta y_{t-2} + \gamma x_t - \rho\gamma x_{t-1} + \varepsilon_t. \qquad [8.3.23]$$

This is a restricted version of the regression model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 x_t + \alpha_4 x_{t-1} + \varepsilon_t, \qquad [8.3.24]$$

where the four regression coefficients $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ are restricted to be nonlinear functions of three underlying parameters (ρ, β, γ). Minimization of the sum of squared ε's in [8.3.23] is equivalent to maximum likelihood estimation conditioning on the first two observations. Moreover, the error term in equation [8.3.24] is uncorrelated with the explanatory variables, and so the α's can be estimated consistently by OLS estimation of [8.3.24]. Then $-\hat{\alpha}_4/\hat{\alpha}_3$ provides a consistent estimate of ρ despite the presence of lagged endogenous variables in [8.3.24].

Even if consistent estimates of ρ and $\boldsymbol{\beta}$ are obtained, Durbin (1970) emphasized that with lagged endogenous variables it will still not be the case that an estimate of ρ based on $(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}})$ has the same asymptotic distribution as an estimate based on $(y_t - \mathbf{x}_t'\boldsymbol{\beta})$. To see this, note that if $\mathbf{x}_t$ contains lagged endogenous variables, then [8.3.19] would no longer be valid. If $\mathbf{x}_t$ includes y_{t-1}, for example, then $\mathbf{x}_t$ and u_{t-1} will be correlated and $\text{plim}[(1/T)\Sigma_{t=1}^{T}u_{t-1}\mathbf{x}_t] \neq \mathbf{0}$, as was assumed in arriving at [8.3.19]. Hence, [8.3.20] will not hold when $\mathbf{x}_t$ includes lagged endogenous variables. Again, an all-purpose procedure that will work is to maximize the log likelihood function [8.3.15] numerically.

Higher-Order Serial Correlation[11]

Consider next the case when the distribution of $\mathbf{u}|\mathbf{X}$ can be described by a pth-order autoregression,

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t.$$

[10]See Betancourt and Kelejian (1981).

[11]This discussion is based on Harvey (1981, pp. 204–6).

The log likelihood conditional on $\mathbf{X}$ for this case becomes

$$\begin{aligned}
-(T/2)\log(2\pi) &- (T/2)\log(\sigma^2) - (1/2)\log|\mathbf{V}_p| \\
&- [1/(2\sigma^2)](\mathbf{y}_p - \mathbf{X}_p\boldsymbol{\beta})'\mathbf{V}_p^{-1}(\mathbf{y}_p - \mathbf{X}_p\boldsymbol{\beta}) \\
&- [1/(2\sigma^2)] \sum_{t=p+1}^{T} \Big[(y_t - \mathbf{x}_t'\boldsymbol{\beta}) - \rho_1(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta}) \\
&\qquad - \rho_2(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta}) - \cdots - \rho_p(y_{t-p} - \mathbf{x}_{t-p}'\boldsymbol{\beta})\Big]^2,
\end{aligned} \tag{8.3.25}$$

where the $(p \times 1)$ vector $\mathbf{y}_p$ denotes the first p observations on y, $\mathbf{X}_p$ is the $(p \times k)$ matrix of explanatory variables associated with these first p observations, and $\sigma^2\mathbf{V}_p$ is the $(p \times p)$ variance-covariance matrix of $(\mathbf{y}_p|\mathbf{X}_p)$. The row i, column j element of $\sigma^2\mathbf{V}_p$ is given by $\gamma_{|i-j|}$ for γ_k the kth autocovariance of an $AR(p)$ process with autoregressive parameters $\rho_1, \rho_2, \ldots, \rho_p$ and innovation variance σ^2. Letting $\mathbf{L}_p$ denote a $(p \times p)$ matrix such that $\mathbf{L}_p'\mathbf{L}_p = \mathbf{V}_p^{-1}$, GLS can be obtained by regressing $\tilde{\mathbf{y}}_p = \mathbf{L}_p\mathbf{y}_p$ on $\tilde{\mathbf{X}}_p = \mathbf{L}_p\mathbf{X}_p$ and $\tilde{y}_t = y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} - \cdots - \rho_p y_{t-p}$ on $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \rho_1\mathbf{x}_{t-1} - \rho_2\mathbf{x}_{t-2} - \cdots - \rho_p\mathbf{x}_{t-p}$ for $t = p+1, p+2, \ldots, T$. Equation [8.3.14] is a special case of [8.3.25] with $p = 1$, $\mathbf{V}_p = 1/(1-\rho^2)$, and $\mathbf{L}_p = \sqrt{1-\rho^2}$.

If we are willing to condition on the first p observations, the task is to choose $\boldsymbol{\beta}$ and $\rho_1, \rho_2, \ldots, \rho_p$ so as to minimize

$$\sum_{t=p+1}^{T} \Big[(y_t - \mathbf{x}_t'\boldsymbol{\beta}) - \rho_1(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta}) - \rho_2(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta}) - \cdots - \rho_p(y_{t-p} - \mathbf{x}_{t-p}'\boldsymbol{\beta})\Big]^2.$$

Again, in the absence of lagged endogenous variables we can iterate as in Cochrane-Orcutt, first taking the ρ_i's as given and regressing $\tilde{y}_t$ on $\tilde{\mathbf{x}}_t$, and then taking $\boldsymbol{\beta}$ as given and regressing $\hat{u}_t$ on $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p}$.

Any covariance-stationary process for the errors can always be approximated by a finite autoregression, provided that the order of the approximating autoregression (p) is sufficiently large. Amemiya (1973) demonstrated that by letting p go to infinity at a slower rate than the sample size T, this iterated GLS estimate will have the same asymptotic distribution as would the GLS estimate for the case when $\mathbf{V}$ is known. Alternatively, if theory implies an $ARMA(p, q)$ structure for the errors with p and q known, one can find exact or approximate maximum likelihood estimates by adapting the methods in Chapter 5, replacing μ in the expressions in Chapter 5 with $\mathbf{x}_t'\boldsymbol{\beta}$.

Further Remarks on Heteroskedasticity

Heteroskedasticity can arise from a variety of sources, and the solution depends on the nature of the problem identified. Using logs rather than levels of variables, allowing the explanatory variables to enter nonlinearly in the regression equation, or adding previously omitted explanatory variables to the regression may all be helpful. Judge, Griffiths, Hill, and Lee (1980) discussed a variety of solutions when the heteroskedasticity is thought to be related to the explanatory variables. In time series regressions, the explanatory variables themselves exhibit dynamic behavior, and such specifications then imply a dynamic structure for the conditional variance. An example of such a model is the autoregressive conditional heteroskedasticity specification of Engle (1982). Dynamic models of heteroskedasticity will be discussed in Chapter 21.

APPENDIX 8.A. *Proofs of Chapter 8 Propositions*

■ **Proof of Proposition 8.2.** The restricted estimate $\mathbf{b}^*$ that minimizes [8.1.2] subject to [8.1.27] can be calculated using the Lagrangean:

$$J = (1/2) \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2 + \boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}). \qquad [8.A.1]$$

Here $\boldsymbol{\lambda}$ denotes an $(m \times 1)$ vector of Lagrange multipliers; λ_i is associated with the constraint represented by the ith row of $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$. The term $\frac{1}{2}$ is a normalizing constant to simplify the expressions that follow. The constrained minimum is found by setting the derivative of [8.A.1] with respect to $\boldsymbol{\beta}$ equal to zero:[12]

$$\begin{aligned}\frac{\partial J}{\partial \boldsymbol{\beta}'} &= (1/2) \sum_{t=1}^{T} 2(y_t - \mathbf{x}_t'\boldsymbol{\beta}) \frac{\partial (y_t - \mathbf{x}_t'\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{\lambda}'\mathbf{R} \\ &= -\sum_{t=1}^{T} (y_t - \boldsymbol{\beta}'\mathbf{x}_t)\mathbf{x}_t' + \boldsymbol{\lambda}'\mathbf{R} = \mathbf{0}',\end{aligned}$$

or

$$\mathbf{b}^{*\prime} \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t' = \sum_{t=1}^{T} y_t\mathbf{x}_t' - \boldsymbol{\lambda}'\mathbf{R}.$$

Taking transposes,

$$\begin{aligned}\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]\mathbf{b}^* &= \sum_{t=1}^{T} \mathbf{x}_t y_t - \mathbf{R}'\boldsymbol{\lambda} \\ \mathbf{b}^* &= \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \left[\sum_{t=1}^{T} \mathbf{x}_t y_t\right] - \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \mathbf{R}'\boldsymbol{\lambda} \qquad [8.A.2] \\ &= \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda},\end{aligned}$$

where $\mathbf{b}$ denotes the unrestricted *OLS* estimate. Premultiplying [8.A.2] by $\mathbf{R}$ (and recalling that $\mathbf{b}^*$ satisfies $\mathbf{R}\mathbf{b}^* = \mathbf{r}$),

$$\mathbf{R}\mathbf{b} - \mathbf{r} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}$$

or

$$\boldsymbol{\lambda} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r}). \qquad [8.A.3]$$

Substituting [8.A.3] into [8.A.2],

$$\mathbf{b} - \mathbf{b}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r}). \qquad [8.A.4]$$

Notice from [8.A.4] that

$$\begin{aligned}(\mathbf{b} - \mathbf{b}^*)'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \mathbf{b}^*) &= \{(\mathbf{R}\mathbf{b} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\}(\mathbf{X}'\mathbf{X}) \\ &\quad \times \{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r})\} \\ &= (\mathbf{R}\mathbf{b} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'] \qquad [8.A.5] \\ &\quad \times [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r}) \\ &= (\mathbf{R}\mathbf{b} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r}).\end{aligned}$$

Thus, the magnitude in [8.1.32] is numerically identical to

$$F = \frac{(\mathbf{b} - \mathbf{b}^*)'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}^*)/m}{s^2} = \frac{(\mathbf{b} - \mathbf{b}^*)'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}^*)/m}{RSS_1/(T - k)}.$$

Comparing this with [8.1.37], we will have completed the demonstration of the equivalence of [8.1.32] with [8.1.37] if it is the case that

$$RSS_0 - RSS_1 = (\mathbf{b} - \mathbf{b}^*)'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \mathbf{b}^*). \qquad [8.A.6]$$

[12]We have used the fact that $\partial\mathbf{x}_t'\boldsymbol{\beta}/\partial\boldsymbol{\beta}' = \mathbf{x}_t'$. See the Mathematical Review (Appendix A) at the end of the book on the use of derivatives with respect to vectors.

Now, notice that

$$\begin{aligned} RSS_0 &= (\mathbf{y} - \mathbf{Xb}^*)'(\mathbf{y} - \mathbf{Xb}^*) \\ &= (\mathbf{y} - \mathbf{Xb} + \mathbf{Xb} - \mathbf{Xb}^*)'(\mathbf{y} - \mathbf{Xb} + \mathbf{Xb} - \mathbf{Xb}^*) \\ &= (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) + (\mathbf{b} - \mathbf{b}^*)'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}^*), \end{aligned} \qquad [8.A.7]$$

where the cross-product term has vanished, since $(\mathbf{y} - \mathbf{Xb})'\mathbf{X} = \mathbf{0}$ by the least squares property [8.1.10]. Equation [8.A.7] states that

$$RSS_0 = RSS_1 + (\mathbf{b} - \mathbf{b}^*)'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}^*), \qquad [8.A.8]$$

confirming [8.A.6]. ■

■ **Proof of Proposition 8.3.** Assumption 8.6(e) guarantees that $\hat{\mathbf{Q}}_T \xrightarrow{p} \mathbf{Q}$, so the issue is whether $\hat{\mathbf{\Omega}}_T$ gives a consistent estimate of $\mathbf{\Omega}$. Define $\mathbf{\Omega}_T^* \equiv (1/T)\Sigma_{t=1}^T u_t^2\mathbf{x}_t\mathbf{x}_t'$, noting that $\mathbf{\Omega}_T^*$ converges in probability to $\mathbf{\Omega}$ by Assumption 8.6(c). Thus, if we can show that $\hat{\mathbf{\Omega}}_T - \mathbf{\Omega}_T^* \xrightarrow{p} \mathbf{0}$, then $\hat{\mathbf{\Omega}}_T \xrightarrow{p} \mathbf{\Omega}$. Now,

$$\hat{\mathbf{\Omega}}_T - \mathbf{\Omega}_T^* = (1/T)\sum_{t=1}^T (\hat{u}_t^2 - u_t^2)\mathbf{x}_t\mathbf{x}_t'. \qquad [8.A.9]$$

But

$$\begin{aligned} (\hat{u}_t^2 - u_t^2) &= (\hat{u}_t + u_t)(\hat{u}_t - u_t) \\ &= [(y_t - \mathbf{b}_T'\mathbf{x}_t) + (y_t - \boldsymbol{\beta}'\mathbf{x}_t)][(y_t - \mathbf{b}_T'\mathbf{x}_t) - (y_t - \boldsymbol{\beta}'\mathbf{x}_t)] \\ &= [2(y_t - \boldsymbol{\beta}'\mathbf{x}_t) - (\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t][-(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t] \\ &= -2u_t(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t + [(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t]^2, \end{aligned}$$

allowing [8.A.9] to be written as

$$\hat{\mathbf{\Omega}}_T - \mathbf{\Omega}_T^* = (-2/T)\sum_{t=1}^T u_t(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t(\mathbf{x}_t\mathbf{x}_t') + (1/T)\sum_{t=1}^T [(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t]^2(\mathbf{x}_t\mathbf{x}_t'). \qquad [8.A.10]$$

The first term in [8.A.10] can be written

$$(-2/T)\sum_{t=1}^T u_t(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t(\mathbf{x}_t\mathbf{x}_t') = -2\sum_{i=1}^k (b_{iT} - \beta_i)\left[(1/T)\sum_{t=1}^T u_t x_{it}(\mathbf{x}_t\mathbf{x}_t')\right]. \qquad [8.A.11]$$

The second term in [8.A.11] has a finite plim by Assumption 8.6(e), and $(b_{iT} - \beta_i) \xrightarrow{p} 0$ for each i. Hence, the probability limit of [8.A.11] is zero.

Turning next to the second term in [8.A.10],

$$(1/T)\sum_{t=1}^T [(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{x}_t]^2(\mathbf{x}_t\mathbf{x}_t') = \sum_{i=1}^k\sum_{j=1}^k (b_{iT} - \beta_i)(b_{jT} - \beta_j)\left[(1/T)\sum_{t=1}^T x_{it}x_{jt}(\mathbf{x}_t\mathbf{x}_t')\right],$$

which again has plim zero. Hence, from [8.A.10],

$$\hat{\mathbf{\Omega}}_T - \mathbf{\Omega}_T^* \xrightarrow{p} \mathbf{0}. \quad ■$$

■ **Proof of Proposition 8.4.** Recall from [8.2.6] that

$$\begin{aligned} \sqrt{T}(\tilde{\mathbf{b}}_T - \boldsymbol{\beta}) &= \left[(1/T)\sum_{t=1}^T \tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^T \tilde{\mathbf{x}}_t\tilde{u}_t\right] \\ &= \left[(1/T)\sum_{t=1}^T (\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})'\right]^{-1} \\ &\quad \times \left[(1/\sqrt{T})\sum_{t=1}^T (\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(u_t - \hat{\rho}u_{t-1})\right]. \end{aligned} \qquad [8.A.12]$$

We will now show that $[(1/T)\Sigma_{t=1}^T(\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})']$ has the same plim as $[(1/T)\Sigma_{t=1}^T(\mathbf{x}_t - \rho\mathbf{x}_{t-1})(\mathbf{x}_t - \rho\mathbf{x}_{t-1})']$ and that $[(1/\sqrt{T})\Sigma_{t=1}^T(\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(u_t - \hat{\rho}u_{t-1})]$ has the same asymptotic distribution as $[(1/\sqrt{T})\Sigma_{t=1}^T(\mathbf{x}_t - \rho\mathbf{x}_{t-1})(u_t - \rho u_{t-1})]$.

Consider the first term in [8.A.12]:

$$
\begin{aligned}
(1/T) \sum_{t=1}^{T} (\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})'
&= (1/T) \sum_{t=1}^{T} [\mathbf{x}_t - \rho\mathbf{x}_{t-1} + (\rho - \hat{\rho})\mathbf{x}_{t-1}][\mathbf{x}_t - \rho\mathbf{x}_{t-1} + (\rho - \hat{\rho})\mathbf{x}_{t-1}]' \\
&= (1/T) \sum_{t=1}^{T} (\mathbf{x}_t - \rho\mathbf{x}_{t-1})(\mathbf{x}_t - \rho\mathbf{x}_{t-1})' \\
&\quad + (\rho - \hat{\rho})\cdot(1/T) \sum_{t=1}^{T} (\mathbf{x}_t - \rho\mathbf{x}_{t-1})\mathbf{x}'_{t-1} \\
&\quad + (\rho - \hat{\rho})\cdot(1/T) \sum_{t=1}^{T} \mathbf{x}_{t-1}(\mathbf{x}_t - \rho\mathbf{x}_{t-1})' \\
&\quad + (\rho - \hat{\rho})^2\cdot(1/T) \sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}'_{t-1}.
\end{aligned} \qquad [8.A.13]
$$

But $(\rho - \hat{\rho}) \xrightarrow{p} 0$, and the plims of $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}$ and $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t\mathbf{x}'_{t-1}$ are assumed to exist. Hence [8.A.13] has the same plim as $(1/T)\Sigma_{t=1}^{T}(\mathbf{x}_t - \rho\mathbf{x}_{t-1})(\mathbf{x}_t - \rho\mathbf{x}_{t-1})'$.

Consider next the second term in [8.A.12]:

$$
\begin{aligned}
(1/\sqrt{T}) \sum_{t=1}^{T} (\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(u_t - \hat{\rho}u_{t-1})
&= (1/\sqrt{T}) \sum_{t=1}^{T} [\mathbf{x}_t - \rho\mathbf{x}_{t-1} + (\rho - \hat{\rho})\mathbf{x}_{t-1}][u_t - \rho u_{t-1} + (\rho - \hat{\rho})u_{t-1}] \qquad [8.A.14] \\
&= (1/\sqrt{T}) \sum_{t=1}^{T} (\mathbf{x}_t - \rho\mathbf{x}_{t-1})(u_t - \rho u_{t-1}) \\
&\quad + \sqrt{T}(\rho - \hat{\rho})\cdot\left[(1/T) \sum_{t=1}^{T} \mathbf{x}_{t-1}(u_t - \rho u_{t-1})\right] \\
&\quad + \sqrt{T}(\rho - \hat{\rho})\cdot\left[(1/T) \sum_{t=1}^{T} (\mathbf{x}_t - \rho\mathbf{x}_{t-1})u_{t-1}\right] \\
&\quad + \sqrt{T}(\rho - \hat{\rho})^2\cdot\left[(1/T) \sum_{t=1}^{T} \mathbf{x}_{t-1}u_{t-1}\right].
\end{aligned}
$$

But [8.3.21] established that $\sqrt{T}(\rho - \hat{\rho})$ converges in distribution to a stable random variable. Since $\text{plim}[(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t u_s] = \mathbf{0}$, the last three terms in [8.A.14] vanish asymptotically. Hence,

$$
(1/\sqrt{T}) \sum_{t=1}^{T} (\mathbf{x}_t - \hat{\rho}\mathbf{x}_{t-1})(u_t - \hat{\rho}u_{t-1}) \xrightarrow{p} (1/\sqrt{T}) \sum_{t=1}^{T} (\mathbf{x}_t - \rho\mathbf{x}_{t-1})(u_t - \rho u_{t-1}),
$$

which was to be shown.

Chapter 8 Exercises

8.1. Show that the uncentered R_u^2 [8.1.13] can equivalently be written as

$$
R_u^2 = 1 - \left[\left(\sum_{t=1}^{T} \hat{u}_t^2\right) \div \left(\sum_{t=1}^{T} y_t^2\right)\right]
$$

for $\hat{u}_t$ the *OLS* sample residual [8.1.4]. Show that the centered R_c^2 can be written as

$$
R_c^2 = 1 - \left[\left(\sum_{t=1}^{T} \hat{u}_t^2\right) \div \left(\sum_{t=1}^{T} (y_t - \bar{y})^2\right)\right].
$$

8.2. Consider a null hypothesis H_0 involving $m = 2$ linear restrictions on $\boldsymbol{\beta}$. How large a sample size T is needed before the 5% critical value based on the Wald form of the *OLS* F test of H_0 is within 1% of the critical value of the Wald form of the *OLS* χ^2 test of H_0?

8.3. Derive result [8.2.28].

8.4. Consider a covariance-stationary process given by

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence with mean zero, variance σ^2, and finite fourth moment and where $\Sigma_{j=0}^{\infty} |\psi_j| < \infty$. Consider estimating a pth-order autoregression by *OLS*:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t.$$

Show that the *OLS* coefficients give consistent estimates of the population parameters that characterize the linear projection of y_t on a constant and p of its lags—that is, the coefficients give consistent estimates of the parameters $c, \phi_1, \ldots, \phi_p$ defined by

$$\hat{E}(y_t | y_{t-1}, y_{t-2}, \ldots, y_{t-p}) = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}$$

(HINT: Recall that $c, \phi_1, \ldots, \phi_p$ are characterized by equation [4.3.6]).

Chapter 8 References

Amemiya, Takeshi. 1973. "Generalized Least Squares with an Estimated Autocovariance Matrix." *Econometrica* 41:723–32.

Anderson, T. W. 1971. *The Statistical Analysis of Time Series*. New York: Wiley.

Betancourt, Roger, and Harry Kelejian. 1981. "Lagged Endogenous Variables and the Cochrane-Orcutt Procedure." *Econometrica* 49:1073–78.

Brillinger, David R. 1981. *Time Series: Data Anaysis and Theory*, expanded ed. San Francisco: Holden-Day.

Durbin, James. 1960. "Estimation of Parameters in Time-Series Regression Models." *Journal of the Royal Statistical Society* Series B, 22:139–53.

———. 1970. "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables." *Econometrica* 38:410–21.

Eicker, F. 1967. "Limit Theorems for Regressions with Unequal and Dependent Errors." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 59–62. Berkeley: University of California Press.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:987–1007.

Evans, G. B. A., and N. E. Savin. 1981. "Testing for Unit Roots: 1. *Econometrica* 49:753–79.

Flavin, Marjorie A. 1983. "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence." *Journal of Political Economy* 91:929–56.

Gregory, Allan W., and Michael R. Veall. 1985. "Formulating Wald Tests of Nonlinear Restrictions." *Econometrica* 53:1465–68.

Hansen, Lars P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50:1029–54.

Harvey, A. C. 1981. *The Econometric Analysis of Time Series*. New York: Wiley.

Hausman, Jerry A., and William E. Taylor. 1983. "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation." *Econometrica* 51:1527–49.

Imhof, J. P. 1961. "Computing the Distribution of Quadratic Forms in Normal Variables." *Biometrika* 48:419–26.

Judge, George G., William E. Griffiths, R. Carter Hill, and Tsoung-Chao Lee. 1980. *The Theory and Practice of Econometrics*. New York: Wiley.

Kinderman, A. J., and J. G. Ramage. 1976. "Computer Generation of Normal Random Variables." *Journal of the American Statistical Association* 71:893–96.

Lafontaine, Francine, and Kenneth J. White. 1986. "Obtaining Any Wald Statistic You Want." *Economics Letters* 21:35–40.

Maddala, G. S. 1977. *Econometrics*. New York: McGraw-Hill.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55:703–8.

Nicholls, D. F., and A. R. Pagan. 1983. "Heteroscedasticity in Models with Lagged Dependent Variables." *Econometrica* 51:1233–42.

O'Nan, Michael. 1976. *Linear Algebra*, 2d ed. New York: Harcourt Brace Jovanovich.

Phillips, P. C. B., and Joon Y. Park. 1988. "On the Formulation of Wald Tests of Nonlinear Restrictions." *Econometrica* 56:1065–83.

Rao, C. Radhakrishna. 1973. *Linear Statistical Inference and Its Applications*, 2d ed. New York: Wiley.

Theil, Henri. 1971. *Principles of Econometrics*. New York: Wiley.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48:817–38.

———. 1984. *Asymptotic Theory for Econometricians*. Orlando, Fla.: Academic Press.