SOME STATISTICAL PROPERTIES OF THE OLS ESTIMATOR

The expectation or mean vector of $\hat{\beta}$, and its dispersion matrix as well, may be found from the expression

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta+\varepsilon)\\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned} \tag{13}$$

The expectation is

$$\begin{aligned}E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon)\\ &= \beta.\end{aligned} \tag{14}$$

Thus $\hat{\beta}$ is an unbiased estimator. The deviation of $\hat{\beta}$ from its expected value is $\hat{\beta} - E(\hat{\beta}) = (X'X)^{-1}X'\varepsilon$. Therefore the dispersion matrix, which contains the variances and covariances of the elements of $\hat{\beta}$, is

$$\begin{aligned}D(\hat{\beta}) &= E\Big[\big\{\hat{\beta} - E(\hat{\beta})\big\}\big\{\hat{\beta} - E(\hat{\beta})\big\}'\Big]\\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}\\ &= \sigma^2(X'X)^{-1}.\end{aligned} \tag{15}$$

The Gauss–Markov theorem asserts that $\hat{\beta}$ is the unbiased linear estimator of least dispersion. This dispersion is usually characterised in terms of the variance of an arbitrary linear combination of the elements of $\hat{\beta}$, although it may also be characterised in terms of the determinant of the dispersion matrix $D(\hat{\beta})$. Thus

(16) If $\hat{\beta}$ is the ordinary least-squares estimator of β in the classical linear regression model, and if β^* is any other linear unbiased estimator of β, then $V(q'\beta^*) \geq V(q'\hat{\beta})$ where q is any constant vector of the appropriate order.

Proof. Since $\beta^* = Ay$ is an unbiased estimator, it follows that $E(\beta^*) = AE(y) = AX\beta = \beta$ which implies that $AX = I$. Now let us write $A = (X'X)^{-1}X' + G$. Then $AX = I$ implies that $GX = 0$. It follows that

$$\begin{aligned}D(\beta^*) &= AD(y)A'\\ &= \sigma^2\big\{(X'X)^{-1}X' + G\big\}\big\{X(X'X)^{-1} + G'\big\}\\ &= \sigma^2(X'X)^{-1} + \sigma^2 GG'\\ &= D(\hat{\beta}) + \sigma^2 GG'.\end{aligned} \tag{17}$$

Therefore, for any constant vector q of order k, there is the identity

$$\begin{aligned} V(q'\beta^*) &= q'D(\hat{\beta})q + \sigma^2 q'GG'q \\ &\geq q'D(\hat{\beta})q = V(q'\hat{\beta}); \end{aligned} \tag{18}$$

and thus the inequality $V(q'\beta^*) \geq V(q'\hat{\beta})$ is established.

CHARACTERISTIC ROOTS AND VECTORS OF A SYMMETRIC MATRIX

Let A be an $n \times n$ symmetric matrix such that $A = A'$, and imagine that the scalar λ and the vector x satisfy the equation $Ax = \lambda x$. Then λ is a characteristic root of A and x is a corresponding characteristic vector. We also refer to characteristic roots as latent roots or eigenvalues. The characteristic vectors are also called eigenvectors.

(11) The characteristic vectors corresponding to two distinct characteristic roots are orthogonal. Thus, if $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ with $\lambda_1 \neq \lambda_2$, then $x_1'x_2 = 0$.

Proof. Premultiplying the defining equations by x_2' and x_1' respectively, gives $x_2'Ax_1 = \lambda_1 x_2'x_1$ and $x_1'Ax_2 = \lambda_2 x_1'x_2$. But $A = A'$ implies that $x_2'Ax_1 = x_1'Ax_2$, whence $\lambda_1 x_2'x_1 = \lambda_2 x_1'x_2$. Since $\lambda_1 \neq \lambda_2$, it must be that $x_1'x_2 = 0$.

The characteristic vector corresponding to a particular root is defined only up to a factor of proportionality. For let x be a characteristic vector of A such that $Ax = \lambda x$. Then multiplying the equation by a scalar μ gives $A(\mu x) = \lambda(\mu x)$ or $Ay = \lambda y$; so $y = \mu x$ is another characteristic vector corresponding to λ.

(12) If $P = P' = P^2$ is a symmetric idempotent matrix, then its characteristic roots can take only the values of 0 and 1.

Proof. Since $P = P^2$, it follows that, if $Px = \lambda x$, then $P^2 x = \lambda x$ or $P(Px) = P(\lambda x) = \lambda^2 x = \lambda x$, which implies that $\lambda = \lambda^2$. This is possible only when $\lambda = 0, 1$.

The Diagonalisation of a Symmetric Matrix

Let A be an $n \times n$ symmetric matrix, and let $x_1, \ldots, x_n$ be a set of n linearly independent characteristic vectors corresponding to its roots $\lambda_1, \ldots, \lambda_n$. Then we can form a set of normalised vectors

$$c_1 = \frac{x_1}{\sqrt{x_1'x_1}}, \ \ldots, \ c_n = \frac{x_n}{\sqrt{x_n'x_n}},$$

which have the property that

$$c_i'c_j = \begin{cases} 0, & \text{if } i \neq j; \\ 1, & \text{if } i = j. \end{cases}$$

The first of these reflects the condition that $x_i'x_j = 0$. It follows that $C = [c_1, \ldots, c_n]$ is an orthonormal matrix such that $C'C = CC' = I$.

Now consider the equation $A[c_1, \ldots, c_n] = [\lambda_1 c_1, \ldots, \lambda_n c_n]$ which can also be written as $AC = C\Lambda$ where $\Lambda = \text{Diag}\{\lambda_1, \ldots, \lambda_n\}$ is the matrix with λ_i as its ith diagonal elements and with zeros in the non-diagonal positions. Postmultiplying the equation by C' gives $ACC' = A = C\Lambda C'$; and premultiplying by C' gives $C'AC = C'C\Lambda = \Lambda$. Thus $A = C\Lambda C'$ and $C'AC = \Lambda$; and C is effective in diagonalising A.

Let D be a diagonal matrix whose ith diagonal element is $1/\sqrt{\lambda_i}$ so that $D'D = \Lambda^{-1}$ and $D'\Lambda D = I$. Premultiplying the equation $C'AC = \Lambda$ by D' and postmultiplying it by D gives $D'C'ACD = D'\Lambda D = I$ or $TAT' = I$, where $T = D'C'$. Also, $T'T = CDD'C' = C\Lambda^{-1}C' = A^{-1}$. Thus we have shown that

(13) For any symmetric matrix $A = A'$, there exists a matrix T such that $TAT' = I$ and $T'T = A^{-1}$.

COCHRANE'S THEOREM: THE DECOMPOSITION OF A CHI-SQUARE

The standard test of an hypothesis regarding the vector β in the model $N(y; X\beta, \sigma^2 I)$ entails a multi-dimensional version of Pythagoras' Theorem. Consider the decomposition of the vector y into the systematic component and the residual vector. This gives

$$\begin{aligned} y &= X\hat{\beta} + (y - X\hat{\beta}) \quad \text{and} \\ y - X\beta &= (X\hat{\beta} - X\beta) + (y - X\hat{\beta}), \end{aligned} \tag{1}$$

where the second equation comes from subtracting the unknown mean vector $X\beta$ from both sides of the first. These equations can also be expressed in terms of the projector $P = X(X'X)^{-1}X'$ which gives $Py = X\hat{\beta}$ and $(I - P)y = y - X\hat{\beta} = e$. Using the definition $\varepsilon = y - X\beta$ within the second of the equations, we have

$$\begin{aligned} y &= Py + (I - P)y \qquad \text{and} \\ \varepsilon &= P\varepsilon + (I - P)\varepsilon. \end{aligned} \tag{2}$$

The reason for rendering the equation in this notation is that it enables us to envisage more clearly the Pythagorean relationship between the vectors. Thus, using the fact that $P = P' = P^2$ and the fact that $P'(I - P) = 0$, it can be established that

$$\text{(3)} \qquad \begin{aligned} \varepsilon'\varepsilon &= \varepsilon' P\varepsilon + \varepsilon'(I - P)\varepsilon \qquad \text{or} \\ \varepsilon'\varepsilon &= (X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) + (y - X\hat{\beta})'(y - X\hat{\beta}). \end{aligned}$$

The terms in these expressions represent squared lengths; and the vectors themselves form the sides of a right-angled triangle with $P\varepsilon$ at the base, $(I - P)\varepsilon$ as the vertical side and ε as the hypotenuse.

The usual test of an hypothesis regarding the elements of the vector β is based on the foregoing relationships. Imagine that the hypothesis postulates that the true value of the parameter vector is β_0. To test this notion, we compare the value of $X\beta_0$ with the estimated mean vector $X\hat{\beta}$. The test is a matter of assessing the proximity of the two vectors which is measured by the square of the distance which separates them. This would be given by $\varepsilon' P\varepsilon = (X\hat{\beta} - X\beta_0)'(X\hat{\beta} - X\beta_0)$ if the hypothesis were true. If the hypothesis is untrue and if $X\beta_0$ is remote from the true value of $X\beta$, then the distance is liable to be excessive. The distance can only be assessed in comparison with the variance σ^2 of the disturbance term or with an estimate thereof. Usually, one has to make do with the estimate of σ^2 which is provided by

$$\text{(4)} \qquad \begin{aligned} \hat{\sigma}^2 &= \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \\ &= \frac{\varepsilon'(I - P)\varepsilon}{T - k}. \end{aligned}$$

The numerator of this estimate is simply the squared length of the vector $e = (I - P)y = (I - P)\varepsilon$ which constitutes the vertical side of the right-angled triangle.

The test uses the result that

(5) If $y \sim N(X\beta, \sigma^2 I)$ and if $\hat{\beta} = (X'X)^{-1}X'y$, then

$$F = \left\{ \frac{(X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta)}{k} \middle/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \right\}$$

is distributed as an $F(k, T - k)$ statistic.

This result depends upon Cochrane's Theorem concerning the decomposition of a chi-square random variate. The following is a statement of the theorem which is attuned to our present requirements:

(6) Let $\varepsilon \sim N(0, \sigma^2 I_T)$ be a random vector of T independently and identically distributed elements. Also let $P = X(X'X)^{-1}X'$ be a

symmetric idempotent matrix, such that $P = P' = P^2$, which is constructed from a matrix X of order $T \times k$ with Rank$(X) = k$. Then

$$\frac{\varepsilon' P\varepsilon}{\sigma^2} + \frac{\varepsilon'(I-P)\varepsilon}{\sigma^2} = \frac{\varepsilon'\varepsilon}{\sigma^2} \sim \chi^2(T),$$

which is a chi-square variate of T degrees of freedom, represents the sum of two independent chi-square variates $\varepsilon' P\varepsilon/\sigma^2 \sim \chi^2(k)$ and $\varepsilon'(I-P)\varepsilon/\sigma^2 \sim \chi^2(T-k)$ of k and $T-k$ degrees of freedom respectively.

To prove this result, we begin by finding an alternative expression for the projector $P = X(X'X)^{-1}X'$. First consider the fact that $X'X$ is a symmetric positive-definite matrix. It follows that there exists a matrix transformation T such that $T(X'X)T' = I$ and $T'T = (X'X)^{-1}$. Therefore $P = XT'TX' = C_1C_1'$, where $C_1 = XT'$ is a $T \times k$ matrix comprising k orthonormal vectors such that $C_1'C_1 = I_k$ is the identity matrix of order k.

Now define C_2 to be a complementary matrix of $T-k$ orthonormal vectors. Then $C = [C_1, C_2]$ is an orthonormal matrix of order T such that

$$\begin{aligned} CC' &= C_1C_1' + C_2C_2' = I_T \qquad \text{and} \\ C'C &= \begin{bmatrix} C_1'C_1 & C_1'C_2 \\ C_2'C_1 & C_2'C_2 \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ 0 & I_{T-k} \end{bmatrix}. \end{aligned} \tag{7}$$

The first of these results allows us to set $I - P = I - C_1C_1' = C_2C_2'$. Now, if $\varepsilon \sim N(0, \sigma^2 I_T)$ and if C is an orthonormal matrix such that $C'C = I_T$, then it follows that $C'\varepsilon \sim N(0, \sigma^2 I_T)$. In effect, if ε is a normally distributed random vector with a density function which is centred on zero and which has spherical contours, and if C is the matrix of a rotation, then nothing is altered by applying the rotation to the random vector. On partitioning $C'\varepsilon$, we find that

$$\begin{bmatrix} C_1'\varepsilon \\ C_1'\varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 I_k & 0 \\ 0 & \sigma^2 I_{t-k} \end{bmatrix} \right), \tag{8}$$

which is to say that $C_1'\varepsilon \sim N(0, \sigma^2 I_k)$ and $C_2'\varepsilon \sim N(0, \sigma^2 I_{T-k})$ are independently distributed normal vectors. It follows that

$$\begin{aligned} \frac{\varepsilon' C_1C_1'\varepsilon}{\sigma^2} &= \frac{\varepsilon' P\varepsilon}{\sigma^2} \sim \chi^2(k) \quad \text{and} \\ \frac{\varepsilon' C_2C_2'\varepsilon}{\sigma^2} &= \frac{\varepsilon'(I-P)\varepsilon}{\sigma^2} \sim \chi^2(T-k) \end{aligned} \tag{9}$$

are independent chi-square variates. Since $C_1C_1' + C_2C_2' = I_T$, the sum of these two variates is

$$\frac{\varepsilon' C_1C_1'\varepsilon}{\sigma^2} + \frac{\varepsilon' C_2C_2'\varepsilon}{\sigma^2} = \frac{\varepsilon'\varepsilon}{\sigma^2} \sim \chi^2(T); \tag{10}$$

and thus the theorem is proved.

The statistic under (5) can now be expresed in the form of

$$F = \left\{ \frac{\varepsilon' P \varepsilon}{k} \Bigg/ \frac{\varepsilon'(I-P)\varepsilon}{T-k} \right\}. \tag{11}$$

This is manifestly the ratio of two chi-sqaure variates divided by their respective degrees of freedom; and so it has an F distribution with these degrees of freedom. This result provides the means for testing the hypothesis concerning the parameter vector β.

DYNAMIC REGRESSION MODELS: TRANSFER FUNCTIONS

Consider a simple dynamic model of the form

$$y(t) = \phi y(t-1) + x(t)\beta + \varepsilon(t). \tag{1}$$

With the use of the lag operator, we can rewrite this as

$$(1-\phi L)y(t) = \beta x(t) + \varepsilon(t) \tag{2}$$

or, equivalently, as

$$y(t) = \frac{\beta}{1-\phi L}x(t) + \frac{1}{1-\phi L}\varepsilon(t). \tag{3}$$

The latter is the so-called rational transfer-function form of the equation. We can replace the operator L within the transfer functions or filters associated with the signal sequence $x(t)$ and disturbance sequence $\varepsilon(t)$ by a complex number z. Then, for the transfer function associated with the signal, we get

$$\frac{\beta}{1-\phi z} = \beta\big\{1 + \phi z + \phi^2 z^2 + \cdots\big\}, \tag{4}$$

where the RHS comes from a familiar power-series expansion.

The sequence $\{\beta, \beta\phi, \beta\phi^2, \ldots\}$ of the coefficients of the expansion constitutes the impulse response of the transfer function. That is to to say, if we imagine that, on the input side, the signal is a unit-impulse sequence of the form

$$x(t) = \{\ldots, 0, 1, 0, 0, \ldots\}, \tag{5}$$

which has zero values at all but one instant, then its mapping through the transfer function would result in an output sequence of

$$r(t) = \{\ldots, 0, \beta, \beta\phi, \beta\phi^2, \ldots\}. \tag{6}$$

Another important concept is the step response of the filter. We may imagine that the input sequence is zero-valued up to a point in time when it assumes a constant unit value:

$$x(t) = \{\ldots, 0, 1, 1, 1, \ldots\}. \tag{7}$$

The mapping of this sequence through the transfer function would result in an output sequence of

$$s(t) = \{\ldots, 0, \beta, \beta + \beta\phi, \beta + \beta\phi + \beta\phi^2, \ldots\} \tag{8}$$

whose elements, from the point when the step occurs in $x(t)$, are simply the partial sums of the impulse-response sequence.

This sequence of partial sums $\{\beta, \beta+\beta\phi, \beta+\beta\phi+\beta\phi^2, \ldots\}$ is described as the step response. Given that $|\phi| < 1$, the step response converges to a value

$$\gamma = \frac{\beta}{1-\phi} \tag{9}$$

which is described as the steady-state gain or the long-term multiplier of the transfer function.

These various concepts apply to models of any order. Consider the equation

$$\alpha(L)y(t) = \beta(L)x(t) + \varepsilon(t), \tag{10}$$

where

$$\begin{aligned} \alpha(L) &= 1 + \alpha_1 L + \cdots + \alpha_p L^p \\ &= 1 - \phi_1 L - \cdots - \phi_p L^p, \\ \beta(L) &= 1 + \beta_1 L + \cdots + \beta_k L^k \end{aligned} \tag{11}$$

are polynomials of the lag operator. The transfer-function form of the model is simply

$$y(t) = \frac{\beta(L)}{\alpha(L)}x(t) + \frac{1}{\alpha(L)}\varepsilon(t), \tag{12}$$

The rational function associated with $x(t)$ has a series expansion

$$\begin{aligned} \frac{\beta(z)}{\alpha(z)} &= \omega(z) \\ &= \{\omega_0 + \omega_1 z + \omega_2 z^2 + \cdots\}; \end{aligned} \tag{13}$$

and the sequence of the coefficients of this expansion constitutes the impulse-response function. The partial sums of the coefficients constitute the step-response function. The gain of the transfer function is defined by

$$\gamma = \frac{\beta(1)}{\alpha(1)} = \frac{\beta_0 + \beta_1 + \cdots + \beta_k}{1 + \alpha_1 + \cdots + \alpha_p}. \tag{14}$$

The method of finding the coefficients of the series expansion of the transfer function in the general case can be illustrated by the second-order case:

$$\frac{\beta_0 + \beta_1 z}{1 - \phi_1 z - \phi_2 z^2} = \big\{\omega_0 + \omega_1 z + \omega_2 z^2 + \cdots\big\}. \tag{15}$$

We rewrite this equation as

$$\beta_0 + \beta_1 z = \big\{1 - \phi_1 z - \phi_2 z^2\big\}\big\{\omega_0 + \omega_1 z + \omega_2 z^2 + \cdots\big\}. \tag{16}$$

Then, by performing the multiplication on the RHS, and by equating the coefficients of the same powers of z on the two sides of the equation, we find that

$$\begin{aligned}
\beta_0 &= \omega_0, & \omega_0 &= \beta_0,\\
\beta_1 &= \omega_1 - \phi_1\omega_0, & \omega_1 &= \beta_1 + \phi_1\omega_0,\\
0 &= \omega_2 - \phi_1\omega_1 - \phi_2\omega_0, & \omega_2 &= \phi_1\omega_1 + \phi_2\omega_0,\\
&\ \ \vdots & &\ \ \vdots\\
0 &= \omega_n - \phi_1\omega_{n-1} - \phi_2\omega_{n-2}, & \omega_n &= \phi_1\omega_{n-1} + \phi_2\omega_{n-2}.
\end{aligned} \tag{17}$$

By examining this scheme, we are able to distinguish between the different roles which are played by the numerator parameters β_0, β_1 and the denominator parameters ϕ_1, ϕ_2. The parameters of the numerator serve as initial conditions for the process which generates the impulse response. The denominator parameters determine the dynamic nature of the impulse response.

Consider the case where the impulse response takes the form a damped sinusoid. This case arises when the roots of the equation $\alpha(z) = 1 - \phi z - \phi_2 z^2 = 0$ are a pair of conjugate complex numbers falling outside the unit circle—as they are bound to do if the response is to be a damped one. Then the parameters β_0 and β_1 are jointly responsible for the initial amplitude and for the phase of the sinusoid. The phase is the time lag which displaces the peak of the sinusoid so that it occurs after the starting time $t = 0$ of the response, which is where the peak of an undisplaced cosine response would occur.

The parameters ϕ_1 and ϕ_2, on the other hand, serve to determine the period of the sinusoidal fluctuations and the degree of damping, which is the rate at which the impulse response converges to zero.

It seems that all four parameters ought to be present in a model which aims at capturing any of the dynamic responses of which a second-order system is capable. To omit one of the numerator parameters of the model would be a mistake unless, for example, there is good reason to assume that the impulse response attains its maximum value at the starting time $t = 0$. We are rarely in the position to make such an assumption.

DYNAMIC REGRESSION MODELS: LAGGED DEPENDENT VARIABLES AND AUTOREGRESSIVE DISTURNANCES

Models with Lagged-Dependent Variables

The reactions of economic agents, such as consumers or investors, to changes in their environment resulting, for example, from changes in prices or incomes, are never instantaneous. The changes are likely to be distributed over time; and positions of equilibrium, if they are ever attained, are likely to be approached gradually.

The slowness to respond may be due to two factors. In the first place, there will be time delays in the transmission and the reception of the information upon which the agents base their actions. In the second place, costs will be entailed in the process of adapting to the new circumstances; and these costs are liable to be positively related to the speed and to the extent of the adjustments. For these reasons, it is appropriate to make some provision in econometric equations for dynamic responses which are distributed over time.

The easiest way of setting an econometric equation in motion is to introduce an element of feedback. This is done by including one or more lagged values of the dependent variable on the right-hand side of the equation to stand in the company of the other explanatory variables. It transpires that, if the current disturbance is unrelated to the lagged dependent variables, then the standard results concerning the consistency of the ordinary least-squares regression procedure retain their validity. This is despite the fact that we can no longer assert that the ordinary least-square estimates of the parameters are unbiased in finite samples.

If the current disturbances and the lagged-dependent variables which are included on the RHS of a dynamic regression equation are not unrelated, then resulting parameter estimates are liable to suffer from considerable biases. The biases are worst when the variance of the disturbance process is large relative to the variances of the explanatory variables.

The essential nature of the problem can be illustrated via a simple model which includes only a lagged dependent variable and which has no other explanatory variables. Imagine that the disturbances follow a first-order autoregressive process. Then there are two equations to be considered. The first of these is the regression equation

$$y(t) = y(t-1)\beta + \eta(t), \qquad \text{where} \quad |\beta| < 1, \tag{18}$$

and the second is the equation

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t), \qquad \text{where} \quad |\rho| < 1, \tag{19}$$

which describes the autoregressive disturbance process. Here $\varepsilon(t)$ stands for an unobservable white-noise process which generates a sequence of independently and identically distributed random variables which are assumed to be independent of the elements of $y(t)$ which precede them in time. The conditions on the parameters β and ρ are necessary to ensure the stability of the model. That is to say, they are necessary conditions for the attainment of a long-run equilibrium in the dynamic response.

Equations (18) and (19), it will be observed, have the same mathematical form. Using the lag operator L, we may rewrite them, in slightly different forms, as

$$(I - \beta L)y(t) = \eta(t) \quad \text{and} \quad \eta(t) = \frac{\varepsilon(t)}{I - \rho L}. \tag{20}$$

Combining the latter gives

$$(I - \rho L)(I - \beta L)y(t) = \bigl\{I - (\rho + \beta)L + \rho\beta L^2\bigr\}y(t) = \varepsilon(t). \tag{21}$$

What we have here is just a particular rendering of the equation

$$(I - \beta_1 L - \beta_2 L^2)y(t) = \varepsilon(t) \tag{22}$$

which relates to the regression of the sequence $y(t)$ on itself lagged by one and by two periods. The only restriction which is entailed by writing the equation in the form of (21) derives from the implication that ρ and β are real-valued coefficients. In the case of equation (22), the corresponding values λ_1 and λ_2, which would be obtained by factorising the the polynomial

$$1 + \beta_1 z + \beta_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z), \tag{23}$$

might be complex numbers. In that case, the two equations (21) and (22) would have different implications regarding their dynamic responses to the disturbances in $\varepsilon(t)$.

Now consider the effect of fitting a model with a single lagged value from the sequence $y(t)$ in the role of the explanatory variable. This can be described as the endeavour to estimate the parameter β of equation (18) by applying ordinary least-squares regression to the equation whilst overlooking the serially correlated nature of the disturbance sequence $\eta(t)$.

Both $y(t)$ and $\eta(t)$ are serially correlated sequences which are linked to each other via equation (18). Therefore the current elements of $\eta(t)$ will be correlated with both past, current and future values of $y(t)$. This means that the essential condition on which the consistency of the ordinary least-squares estimator depends is violated.

On substituting the expression $y_t = (\rho + \beta)y_{t-1} - \rho\beta y_{t-2} + \varepsilon_t$ into the regression formula, we derive the following expression for the estimate:

$$\begin{aligned}\hat{\beta} &= \frac{\sum y_{t-1}y_t}{\sum y_{t-1}^2}\\ &= (\rho+\beta)\frac{\sum y_{t-1}^2}{\sum y_{t-1}^2} - \rho\beta\frac{\sum y_{t-1}y_{t-2}}{\sum y_{t-1}^2} + \frac{\sum y_{t-1}\varepsilon_t}{\sum y_{t-1}^2}.\end{aligned} \tag{24}$$

It is straightforward to take limits in the expression as the sample size T increases. Let $\hat{\beta} \rightarrow \delta$ as $T \rightarrow \infty$. Then the equation above becomes the equation

$$\delta = (\beta + \rho) - \beta\rho\delta. \tag{25}$$

The final term on the RHS of (24) vanishes since, according to the assumptions, the elements of $\varepsilon(t)$ are uncorrelated with elements of $y(t)$ which precede them in time. Rearranging equation (25) gives the result that

$$\delta = \frac{\rho + \beta}{1 + \rho\beta}. \tag{26}$$

Notice that the expression for δ is symmetric with respect of ρ and β. However, we have tended to regard β as the regression parameter and ρ as the parameter of an autoregressive disturbance process. This distinction now appears to be false. However, if $y(t-1)$ on the RHS of equation (18) were standing in the company of another explanatory variable, say $x(t)$, then the distinction would be a valid one.

Now let us imagine, for the sake of argument, that $\rho \rightarrow 0$. Then it is clear that $\delta \rightarrow \beta$. Since the variance of the process $\eta(t)$ is related positively to the value of ρ, it can be said that the bias in β is directly related to the variance of the serially-correlated disturbance process. Exactly the same result obtains when $y(t-1)$ is accompanied in the regression equation by other explanatory variables.

DYMAMIC MOEDLS: ERROR CORRECTION FORMS, NONSTATIONARITY AND COINTERGRATED VARIABLES

Consider taking $y(t-1)$ from both sides of the equation under (1). This gives

$$\begin{aligned}\nabla y(t) = y(t) - y(t-1) &= (\phi - 1)y(t-1) + \beta x(t) + \varepsilon(t)\\ &= (1-\phi)\left\{\frac{\beta}{1-\phi}x(t) - y(t-1)\right\} + \varepsilon(t)\\ &= \lambda\big\{\gamma x(t) - y(t-1)\big\} + \varepsilon(t),\end{aligned} \tag{27}$$

where $\lambda = 1-\phi$ and where γ is the gain of the transfer function as defined under (9). This is the so-called error-correction form of the equation; and it indicates that the change in $y(t)$ is a function of the extent to which the proportions of the series $x(t)$ and $y(t-1)$ differs from those which would prevail in the steady state.

The error-correction form provides the basis for estimating the parameters of the model when the signal series $x(t)$ is trended or nonstationary. In such circumstances, it is easy to obtain an accurate estimate of γ simply by running a regression of $y(t-1)$ on $x(t)$; for all that is required of the regression is that it should determine the fundamental coefficient of proportionality which, in the long term, dominates the relationship which exists between the two series. Once a value for γ is available, the remaining parameter λ may be estimated by regressing $\nabla y(t)$ upon the composite variable $\{\gamma x(t) - y(t-1)\}$.

It is possible to derive an error-correction form for the more general model to be found under (10). We may begin by writing the model in the form of

$$y(t) = \phi_1 y(t-1) + \cdots + \phi_p y(t-p) + \beta_0 x(t) + \cdots + \beta_k x(t-k) + \varepsilon(t). \tag{28}$$

We can proceed to reparametrise this model so that it assumes the equivalent form of

$$\begin{aligned}y(t) = \theta y(t-1) &+ \rho_1 \nabla y(t-1) + \cdots + \rho_p \nabla y(t-p+1)\\ &+\kappa x(t) + \delta_0 \nabla x(t) + \cdots + \delta_k \nabla x(t-k+1) + \varepsilon(t),\end{aligned} \tag{29}$$

where $\theta = \phi_1 + \cdots + \phi_p$ and $\kappa = \beta_0 + \cdots + \beta_k$. Now let us subtract $y(t-1)$ from both sides of equation (29). This gives

$$\begin{aligned}\nabla y(t) =&(\theta-1)y(t-1) + \kappa x(t)\\ &+ \rho_1 \nabla y(t-1) + \cdots + \rho_p \nabla y(t-p+1)\\ &+ \delta_0 \nabla x(t) + \cdots + \delta_k \nabla x(t-k+1) + \varepsilon(t).\end{aligned} \tag{30}$$

The first two terms on the RHS combine to give

$$\begin{aligned}(\theta-1)y(t-1) + \kappa x(t) &= (1-\theta)\left\{\frac{\kappa}{1-\theta}x(t) - y(t-1)\right\}\\ &= \lambda\big\{\gamma x(t) - y(t-1)\big\}\end{aligned} \tag{31}$$

which is an error-correction term in which γ is the value of the gain defined in (14) above. It follows that the error-correction form of equation (28) is

$$\nabla y(t) = \lambda\big\{\gamma x(t) - y(t-1)\big\} + \sum_{i=1}^{p-1} \rho_i \nabla y(t-i) + \sum_{i=0}^{k-1} \delta_i \nabla x(t-i) + \varepsilon(t). \tag{32}$$

In the case of a nonstationary signal $x(t)$, this is amenable to precisely the same principle of estimation as was the simpler first-order equation equation under (27). That is to say, we can begin by estimating the gain γ by a simple regression of $y(t-1)$ on $x(t)$. Then, when a value for γ is available, we can proceed to find the remaining parameters of the model via a second regression.

Example. To reveal the nature of the reparameterisation which transforms equation (28) into equation (29), let us consider the following example:

$$\begin{aligned}
&\beta_0 x(t) + \beta_1 x(t-1) + \beta_2 x(t-2) + \beta_3 x(t-3)\\
&\quad = \big\{\beta_0 + \beta_1 + \beta_2 + \beta_3\big\}x(t) - \big\{\beta_1 + \beta_2 + \beta_3\big\}\big\{x(t) - x(t-1)\big\}\\
&\qquad\qquad - \big\{\beta_2 + \beta_3\big\}\big\{x(t-1) - x(t-2)\big\}\\
&\qquad\qquad - \beta_3\big\{x(t-2) - x(t-3)\big\}\\
&\quad = \kappa x(t) + \delta_0 \nabla x(t) + \delta_1 \nabla x(t-1) + \delta_2 \nabla x(t-2).
\end{aligned} \tag{33}$$

The example may be systematised. Consider the product $\beta' x$ wherein $x = [x(t), x(t-1), x(t-2), x(t-3)]'$ and $\beta' = [\beta_0, \beta_1, \beta_2, \beta_3]$. Let Λ be an arbitrary nonsingular, i.e. invertible, matrix of order 4×4. Then $\beta' x = \{\beta'\Lambda^{-1}\}\{\Lambda x\} = \delta' z$ where $z = \Lambda x$ and $\delta' = \beta'\Lambda^{-1}$. That is to say, the expression in terms of z and δ is equivalent to the original expression in terms of x and β. With these results in mind, let us consider the following transformations:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1) \\ x(t-2) \\ x(t-3) \end{bmatrix} = \begin{bmatrix} x(t) \\ -\nabla x(t) \\ -\nabla x(t-1) \\ -\nabla x(t-2) \end{bmatrix} \tag{34}$$

and

$$\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \kappa & -\delta_0 & -\delta_1 & -\delta_2 \end{bmatrix}. \tag{35}$$

Here the two matrices which affect the transformation upon the variables and upon their associated parameters stand in an inverse relationship to one another. They are, in fact, the matrix analogues, repsctively, of the operators $1-L$ and $(1-L)^{-1} = \{1+L+L^2+\cdots\}$. The transformations provide a simple example of what is entailed in converting equation (28) into equation (29).

Queen Mary & Westfield College

UNIVERSITY OF LONDON
BSc (ECONOMICS)
EXAMINATION OF ASSOCIATE STUDENTS

Econometric Theory

December, 1997

Answer THREE questions in TWO HOURS

1. Prove that the characteristic vectors $x_1, \ldots, x_n$ of an $n \times n$ symmetric matrix corresponding to n distinct roots $\lambda_1, \ldots, \lambda_n$ are mutually orthogonal.

Show how we can reduce a symmetric matrix to a diagonal matrix using an orthonormal matrix, and thence prove that, for any symmetric matrix Q, there exists a matrix T such that $TQT' = I$ and $T'T = Q^{-1}$. Using the latter result, show how the generalised least-squares estimator of β in the model $(y; X\beta, \sigma^2 Q)$ may be obtained as the ordinary least-squares estimator of the regression parameters of a transformed model.

2. Describe the essential differences between the permanent-income and the partial-adjustment models of consumption behaviour, both in respect of their mathematical formulations and their behavioral implications.

Describe and account for the distortions which might affect the estimates if a parsimonious model of the form

$$\alpha(L)y(t) = \beta(L)x(t) + \varepsilon(t)$$

were fitted to data generated by an equation in the form of

$$y(t) = \frac{\delta(L)}{\gamma(L)}x(t) + \frac{\theta(L)}{\phi(L)}\varepsilon(t).$$

3. Demonstrate the unbiasedness of the estimators $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(T - k)$ in the classical regression model $(y; X\beta, \sigma^2 I)$.

Prove that $V(q'\hat{\beta}) \leq V(q'\beta^*)$ where β^* is any other linear unbiased estimator of β and q is an arbitrary nonstochastic vector.

4. Explain what is mean by the phase, the amplitude, the damping and the period of a complex impulse response generated by a rational transfer function in the form of

$$\omega(z) = \frac{\beta_0 + \beta_1 z}{1 - \phi_1 z - \phi_2 z^2}.$$

Explain why all four coefficients β_0, β_1, ϕ_1 and ϕ_2 must be present if the transfer function is to provide a sufficiently flexible means of representing complex dynamic behaviour.

Show how the coefficients of the series expansion of $\omega(z)$ may be obtained, and find the first four coefficients in the case where $\beta_0 = 1$, $\beta_2 = 3$, $\phi_1 = -0.5$ and $\phi_2 = 0.9$. Is this response complex or not and is it damped or explosive?

5. Consider the model

$$y(t) = y(t-1)\beta + \eta(t)$$

wherein

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t)$$

is a sequence of disturbances generated by a first-order autoregressive process which is driven by a white-noise sequence $\varepsilon(t)$ of independently and identically distributed random variables. Show that the estimate of β, obtained by applying the ordinary least-squares procedure, would tend to the value of $(\beta + \rho)/(1 + \rho\beta)$ as the size of the sample increases.

Imagine that the model

$$y(t) = \beta_1 y(t-1) + \beta_2 y(t-2) + \varepsilon(t)$$

is fitted to the data via the ordinary least-squares procedure. What values would you expect to obtain for β_1 and β_2 in the limit, as the size of the sample increases indefinitely?

6. Let $P = X(X'X)^{-1}X'$, where X is the matrix of explanatory variables within the regression equation $y = X\beta + \varepsilon$ which comprises a vector $\varepsilon \sim N(0, \sigma^2 I)$ of normally distributed disturbances. Demonstrate that

$$\frac{\varepsilon'\varepsilon}{\sigma^2} = \frac{\varepsilon' P\varepsilon}{\sigma^2} + \frac{\varepsilon'(I-P)\varepsilon}{\sigma^2}$$

represents the decomposition of a chi-square variate $\varepsilon'\varepsilon/\sigma^2 \sim \chi^2(T)$ into a pair of independent chi-square variates $\varepsilon'P\varepsilon/\sigma^2 \sim \chi^2(k)$ and $\varepsilon'(I-P)\varepsilon/\sigma^2 \sim \chi^2(T-k)$. What is the practical use of this result?

