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

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

The expectation is

(2)
$$\begin{aligned}E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon) \\ &= \beta.\end{aligned}$$

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

(3)
$$\begin{aligned}D(\hat{\beta}) &= E\Big[\big\{\hat{\beta} - E(\hat{\beta})\big\}\big\{\hat{\beta} - E(\hat{\beta})\big\}'\Big] \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

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

(4) 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

(5)
$$\begin{aligned}D(\beta^*) &= AD(y)A' \\ &= \sigma^2\big\{(X'X)^{-1}X' + G\big\}\big\{X(X'X)^{-1} + G'\big\} \\ &= \sigma^2(X'X)^{-1} + \sigma^2 GG' \\ &= D(\hat{\beta}) + \sigma^2 GG'.\end{aligned}$$

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

$$\begin{aligned} V(q'\beta^*) &= q'D(\hat{\beta})q + \sigma^2 q'GG'q \\ &\geq q'D(\hat{\beta})q = V(q'\hat{\beta}); \end{aligned} \tag{6}$$

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

2. THE PARTITIONED REGRESSION MODEL AND OMITTED VARIABLES BIAS

Consider taking a regression equation in the form of

$$y = [\,X_1 \quad X_2\,]\begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{1}$$

Here $[X_1, X_2] = X$ and $[\beta_1', \beta_2']' = \beta$ are obtained by partitioning the matrix X and vector β of the equation $y = X\beta + \varepsilon$ in a conformable manner. The normal equations $X'X\beta = X'y$ can be partitioned likewise. Writing the equations without the surrounding matrix braces gives

$$X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y, \tag{2}$$

$$X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y. \tag{3}$$

From (2), we get the equation $X_1'X_1\beta_1 = X_1'(y - X_2\beta_2)$ which gives an expression for the leading subvector of $\hat{\beta}$:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2). \tag{4}$$

To obtain an expression for $\hat{\beta}_2$, we must eliminate β_1 from equation (3). For this purpose, we multiply equation (2) by $X_2'X_1(X_1'X_1)^{-1}$ to give

$$X_2'X_1\beta_1 + X_2'X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 = X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{5}$$

When the latter is taken from equation (3), we get

$$\Big\{X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\Big\}\beta_2 = X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{6}$$

On defining

$$P_1 = X_1(X_1'X_1)^{-1}X_1', \tag{7}$$

can we rewrite (6) as

$$\Big\{X_2'(I - P_1)X_2\Big\}\beta_2 = X_2'(I - P_1)y, \tag{8}$$

whence

$$\hat{\beta}_2 = \left\{X_2'(I - P_1)X_2\right\}^{-1} X_2'(I - P_1)y. \tag{9}$$

Let us now by investige the effect that a condition of orthogonality amongst the regressors might have upon the ordinary least-squares estimates of the regression parameters. Let us consider a partitioned regression model which can be written as

$$y = [\,X_1, X_2\,] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{10}$$

We may assume that the variables in this equation are in deviation form. Let us imagine that the columns of X_1 are orthogonal to the columns of X_2 such that $X_1'X_2 = 0$. This is the same as imagining that the empirical correlation between variables in X_1 and variables in X_2 is zero.

To see the effect upon the ordinary least-squares estimator, we may examine the partitioned form of the formula $\hat{\beta} = (X'X)^{-1}X'y$. Here we have

$$X'X = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} [\,X_1 \quad X_2\,] = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}, \tag{11}$$

where the final equality follows from the condition of orthogonality. The inverse of the partitioned form of $X'X$ in the case of $X_1'X_2 = 0$ is

$$(X'X)^{-1} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix}. \tag{12}$$

We also have

$$X'y = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}. \tag{13}$$

On combining these elements, we find that

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y \\ (X_2'X_2)^{-1}X_2'y \end{bmatrix}. \tag{14}$$

In this special case, the coefficients of the regression of y on $X = [X_1, X_2]$ can be obtained from the separate regressions of y on X_1 and y on X_2.

We should make it clear that this result does not hold true in general. The general formulae for $\hat{\beta}_1$ and $\hat{\beta}_2$ are those which we have given already under (4) and (9):

$$\begin{aligned} &\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2), \\ &\hat{\beta}_2 = \left\{X_2'(I - P_1)X_2\right\}^{-1} X_2'(I - P_1)y, \quad P_1 = X_1(X_1'X_1)^{-1}X_1'. \end{aligned} \tag{15}$$

We can easily confirm that these formulae do specialise to those under (14) in the case of $X_1'X_2 = 0$.

The purpose of including X_2 in the regression equation when, in fact, our interest is confined to the parameters of β_1 is to avoid falsely attributing the explanatory power of the variables of X_2 to those of X_1.

Let us investigate the effects of erroneously excluding X_2 from the regression. In that case, our estimate will be

$$\begin{aligned}\tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon.\end{aligned} \tag{16}$$

On applying the expectations operator to these equations, we find that

$$E(\tilde{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2, \tag{17}$$

since $E\{(X_1'X_1)^{-1}X_1'\varepsilon\} = (X_1'X_1)^{-1}X_1'E(\varepsilon) = 0$. Thus, in general, we have $E(\tilde{\beta}_1) \neq \beta_1$, which is to say that $\tilde{\beta}_1$ is a biased estimator. The only circumstances in which the estimator will be unbiased are when either $X_1'X_2 = 0$ or $\beta_2 = 0$. In other circumstances, the estimator will suffer from a problem which is commonly described as *omitted-variables bias*.

We need to ask whether it matters that the estimated regression parameters are biased. The answer depends upon the use to which we wish to put the estimated regression equation. The issue is whether the equation is to be used simply for predicting the values of the dependent variable y or whether it is to be used for some kind of structural analysis.

If the regression equation purports to describe a structural or a behavioral relationship within the economy, and if some of the explanatory variables on the RHS are destined to become the instruments of an economic policy, then it is important to have unbiased estimators of the associated parameters. For these parameters indicate the leverage of the policy instruments. Examples of such instruments are provided by interest rates, tax rates, exchange rates and the like.

On the other hand, if the estimated regression equation is to be viewed solely as a predictive device—that it to say, if it is simply an estimate of the function $E(y|x_1,\ldots,x_k)$ which specifies the conditional expectation of y given the values of $x_1,\ldots,x_n$—then, provided that the underlying statistical mechanism which has generated these variables is preserved, the question of the unbiasedness the regression parameters does not arise.

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.

(1) 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 $x_2' A x_1 = \lambda_1 x_2' x_1$ and $x_1' A x_2 = \lambda_2 x_1' x_2$. But $A = A'$ implies that $x_2' A x_1 = x_1' A x_2$, whence $\lambda_1 x_2' x_1 = \lambda_2 x_1' x_2$. Since $\lambda_1 \neq \lambda_2$, it must be that $x_1' x_2 = 0$.

In x_1 and x_2 are two linearly independent vectors corresponding to the same root λ, such that $Ax_1 = \lambda x_1$ and $Ax_2 = \lambda x_2$, then they span a two-dimensional characteristic subspace. Within this subspace, it is possible to find two mutually orthogonal vectors which can be added to the set of the characteristic vectors of A.

More generally, if there exists a set $x_1, \ldots, x_p$ of p linearly independent vectors such that $Ax_i = \lambda x_i; i = 1, \ldots, p$, then λ is described as a characteristic root of multiplicity p; and a set of p mutually orthogonal characteristic vectors can be found which correspond to λ.

A characteristic vector corresponding to a particular root is defined only up to a factor of proportionality. For let x be a vector 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.

(2) 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

$$(3) \qquad c_1 = \frac{x_1}{\sqrt{x_1' x_1}}, \; \ldots, \; c_n = \frac{x_n}{\sqrt{x_n' x_n}},$$

which have the property that

$$c_i'c_j = \begin{cases} 0, & \text{if } i \neq j; \\ 1, & \text{if } i = j. \end{cases} \tag{4}$$

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

(5) 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 $\text{Rank}(X) = k$. 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

$$\begin{aligned} \frac{\varepsilon' C_1 C_1' \varepsilon}{\sigma^2} &= \frac{\varepsilon' P \varepsilon}{\sigma^2} \sim \chi^2(k) \quad \text{and} \\ \frac{\varepsilon' C_2 C_2' \varepsilon}{\sigma^2} &= \frac{\varepsilon'(I-P)\varepsilon}{\sigma^2} \sim \chi^2(T-k) \end{aligned} \tag{9}$$

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

$$\frac{\varepsilon' C_1 C_1' \varepsilon}{\sigma^2} + \frac{\varepsilon' C_2 C_2' \varepsilon}{\sigma^2} = \frac{\varepsilon'\varepsilon}{\sigma^2} \sim \chi^2(T); \tag{10}$$

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:

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

We rewrite this equation as

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

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.

FIRST-ORDER AUTOREGRESSIVE DISTURBANCES IN THE CLASSICAL LINEAR REGRESSION MODEL

In the classical linear regression model, it is assumed that the disturbances constitute a sequence $\varepsilon(t) = \{\varepsilon_t; t = 0, \pm 1, \pm 2, \ldots\}$ of independently and identically distributed random variables such that

$$(1) \qquad E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2, & \text{if } t = s; \\ 0, & \text{if } t \neq s. \end{cases}$$

The process which generates such disturbances is often called a white-noise process.

Our task is to find models for the disturbance process which are more in accordance with the circumstances of economics where the variables tend to show a high degree of inertia. In econometrics the traditional means of representing the inertial properties of the disturbance process has been to adopt a simple first-order autoregressive model, or AR(1) model, whose equation takes the form of

$$(2) \qquad \eta_t = \phi\eta_{t-1} + \varepsilon_t, \qquad \text{where} \qquad \phi \in (-1, 1).$$

Here it continues to be assumed that ε_t is generated by a white-noise process with $E(\varepsilon_t) = 0$. In many econometric applications, the value of ϕ falls in the more restricted interval $[0, 1)$.

According to this model, the conditional expectation of η_t given η_{t-1} is $E(\eta_t|\eta_{t-1}) = \phi\eta_{t-1}$. That is to say, the expectation of the current disturbance is ϕ times the value of the previous disturbance. This implies that, for a value of ϕ which is closer to unity that to zero, there will be a high degree of correlation amongst successive elements of the sequence $\eta(t) = \{\eta_t; t = 0, \pm1, \pm2, \ldots\}$. This result is illustrated in figure 2 which gives a sequence of 50 observation on an AR(1) process with $\phi = 0.9$

We can show that the covariance of two elements of the sequence $\eta(t)$ which are separated by τ time periods is given by

$$C(\eta_{t-\tau}, \eta_t) = \gamma_\tau = \sigma^2 \frac{\phi^\tau}{1-\phi^2}. \tag{3}$$

It follows that variance of the process, which is formally the autocovariance of lag $\tau = 0$, is given by

$$V(\eta_t) = \gamma_0 = \frac{\sigma^2}{1-\phi^2}. \tag{4}$$

As ϕ tends to unity, the variance increases without bound. In fact, the sequences in figures 1 and 2 share the same underlying white noise-process which has a unit variance; and it is evident that the autocorrelated sequence of figure 2 has the wider dispersion.

To find the correlation of two elements from the autoregressive sequence, we note that

$$\mathrm{Corr}(\eta_{t-\tau}, \eta_t) = \frac{C(\eta_{t-\tau}, \eta_t)}{\sqrt{V(\eta_{t-\tau})V(\eta_t)}} = \frac{C(\eta_{t-\tau}, \eta_t)}{V(\eta_t)} = \frac{\gamma_\tau}{\gamma_0}. \tag{5}$$

This implies that the correlation of the two elements separated by τ periods is just ϕ^τ; and thus, as the temporal separation increases, the correlation tends to zero in the manner of a convergent geometric progression.

To demonstrate these results, let us consider substituting for $\eta_{t-1} = \phi\eta_{t-2} + \varepsilon_{t-1}$ in the equation under (2) and then substituting for $\eta_{t-2} = \phi\eta_{t-3} + \varepsilon_{t-2}$, and so on indefinitely. By this process, we find that

$$\begin{aligned} \eta_t &= \phi\eta_{t-1} + \varepsilon_t \\ &= \phi^2\eta_{t-2} + \varepsilon_t + \phi\varepsilon_{t-1} \\ &\;\vdots \\ &= \left\{\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-1} + \cdots\right\} \\ &= \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}. \end{aligned} \tag{6}$$

Here the final expression is justified by the fact that $\phi^n \to 0$ as $n \to \infty$ in consequence of the restriction that $|\phi| < 1$. Thus we see that η_t is formed as a geometrically declining weighted average of all past values of the sequence $\varepsilon(t)$.

Using this result, we can now write

$$\begin{aligned}\gamma_\tau &= C(\eta_{t-\tau}, \eta_t) = E(\eta_{t-\tau}\eta_t)\\ &= E\left(\Big\{\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-\tau-i}\Big\}\Big\{\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\Big\}\right)\\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \phi^i \phi^j E(\varepsilon_{t-\tau-i}\varepsilon_{t-j}).\end{aligned} \tag{7}$$

But the assumption that $\varepsilon(t)$ is a white-noise process with zero-valued autocovariances at all nonzero lags implies that

$$E(\varepsilon_{t-\tau-i}\varepsilon_{t-j}) = \begin{cases} \sigma^2, & \text{if } j = \tau + i;\\ 0, & \text{if } j \neq \tau + i.\end{cases} \tag{8}$$

Therefore, on using the above conditions in (7) and on setting $j = \tau + i$, we find that

$$\begin{aligned}\gamma_\tau &= \sigma^2 \sum_{i}^{\infty} \phi^i \phi^{i+\tau} = \sigma^2 \phi^\tau \sum_{i}^{\infty} \phi^{2i}\\ &= \sigma^2 \phi^\tau \{1 + \phi^2 + \phi^4 + \phi^6 + \cdots\}\\ &= \sigma^2 \frac{\phi^\tau}{1 - \phi^2}.\end{aligned} \tag{9}$$

This establishes the result under (4).

Now let us imagine a linear regression model in the form of

$$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + \cdots + x_{tk}\beta_k + \eta_t, \tag{10}$$

where η_t follows a first-order autoregressive process. A set of T instances of the relationship would be written as $y = X\beta + \eta$, where y and η are vectors of T elements and X is a matrix or order $T \times k$. The variance–covariance or dispersion matrix of the vector $\eta = [\eta_1, \eta_2, \eta_3, \ldots \eta_T]'$ takes the form of $[\gamma_{|i-j|}] = \sigma_\varepsilon^2 Q$, where

$$Q = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \ldots & \phi^3\\ \phi & 1 & \phi & \ldots & \phi^2\\ \phi^2 & \phi & 1 & \ldots & \phi\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \phi^3 & \phi^2 & \phi & \ldots & 1\end{bmatrix}; \tag{11}$$

and it can be confirmed directly that

$$(12)\qquad Q^{-1}=\begin{bmatrix} 1 & -\phi & 0 & \ldots & 0\\ -\phi & 1+\phi^2 & -\phi & \ldots & 0\\ 0 & -\phi & 1+\phi^2 & \ldots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & -\phi & 1 \end{bmatrix}.$$

This is a matrix of three nonzero diagonal bands. The elements of principal diagonal, apart from the first and the last, have the value of $1+\phi^2$. The first and last elements are units. The elements of the supradiagonal band and of the subdiagonal band have the value of ϕ.

Given its sparcity, the matrix Q^{-1} could be used directly in implementing the generalised least-squares estimator for which the formula is

$$(13)\qquad \beta^*=(XQ^{-1}X)^{-1}XQ^{-1}y.$$

However, by exploiting the factorisation $Q^{-1}=T'T$, we are able to to implement the estimator by applying an ordinary least-squares procedure to the transformed data $W=TX$ and $g=Ty$. The following equation demonstrates the equivalence of the procedures:

$$(14)\qquad \begin{aligned}\beta^*&=(W'W)^{-1}W'g\\ &=(XT'TX)^{-1}XT'Ty\\ &=(XQ^{-1}X)^{-1}XQ^{-1}y\end{aligned}$$

The factor T of the matrix $Q^{-1}=T'T$ takes the form of

$$(15)\qquad T=\begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \ldots & 0\\ -\phi & 1 & 0 & \ldots & 0\\ 0 & -\phi & 1 & \ldots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

this effect a very simple transformation the data. Thus, for example, the elementy_1 within the vector $y=[y_1,y_2,Y_3,\ldots,y_T]'$ is replaced $y_1\sqrt{1-\phi^2}$ whilst y_t is replaced by $y_t-\phi y_{t-1}$, for all $t>1$.

Consider, for example, the simple regression model

$$(16)\qquad y_t=x_t\beta+\eta_t \qquad \text{with} \quad \eta_t=\phi\eta_{t-1}+\varepsilon_t.$$

The transformation gives the equation

$$(17)\qquad y_t-\phi y_{t-1}=(x_t-\phi x_{t-1}\beta+\varepsilon_t,$$

which represents a model that fulfils the classical assumptions and for which ordinary least squares regression is the appropriate method of estimation.

THE CLASSICAL SIMULTANEOUS-EQUATION MODEL AND THE 2SLS ESTIMATION OF A STRUCTURAL EQUATION

The classical simultaneous-equation model of econometrics is a system of M structural equations which can may be compiled to give the following equation:

$$(1)\qquad \begin{aligned}&[y_{t1}, y_{t2}, \ldots, y_{tM}][\gamma_{.1}, \gamma_{.2}, \ldots, \gamma_{.M}] + x_{t.}[\beta_{.1}, \beta_{.2}, \ldots, \beta_{.M}]\\ &\qquad + [\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tM}] = [0, 0, \ldots, 0],\end{aligned}$$

This can be written in summary notation as

$$(2)\qquad y_{t.}\Gamma + x_{t.}B + \varepsilon_{t.} = 0,$$

where $\Gamma = [\gamma_{.1}, \gamma_{.2}, \ldots, \gamma_{.j}]$. The elements of the vector $\varepsilon_{t.} = [\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tM}]$, of the M structural disturbances are assumed to be distributed independently of time such that, for every t, there are

$$(3)\qquad E(\varepsilon_{t.}) = 0 \quad \text{and} \quad D(\varepsilon_{t.}) = E(\varepsilon_{t.}'\varepsilon_{t.}) = \Sigma_{\varepsilon\varepsilon}.$$

It is also assumed that the structural disturbances are distributed independently of the exogenous variables so that $C(\varepsilon_{t.}, x_{s.}) = 0$ for all t and s.

The reduced form of the system is obtained from equation (2) by postmultiplying it by the inverse of the matrix Γ. This gives

$$(6)\qquad y_{t.} = x_{t.}\Pi + \eta_{t.} \quad \text{with} \quad \Pi = -B\Gamma^{-1} \quad \text{and} \quad \eta_{t.} = -\varepsilon_{t.}\Gamma^{-1}.$$

It follows that the vector $\eta_{t.} = -\varepsilon_{t.}\Gamma^{-1}$ of reduced-form disturbances has

$$(7)\qquad E(\eta_{t.}) = 0 \quad \text{and} \quad D(\eta_{t.}) = \Gamma'^{-1}D(\varepsilon_{t.})\Gamma^{-1} = \Gamma'^{-1}\Sigma_{\varepsilon\varepsilon}\Gamma^{-1} = \Omega.$$

It is assumed that the statistical properties of the data can be described completely in terms of its first and second moments. The dispersion matrices of $x_{t.}$ and $y_{t.}$can be denoted by $D(x_{t.}) = \Sigma_{xx}$ and $D(y_{t.}) = \Sigma_{yy}$ and their covariance matrix by $C(x_{t.}, y_{t.}) = \Sigma_{xy}$. By combining the reduced-form regression relationship of (6) with a trivial identity in $x_{t.}$, we get the following system:

$$(8)\qquad [\,y_{t.} \quad x_{t.}\,]\begin{bmatrix} I & 0 \\ -\Pi & I \end{bmatrix} = [\,\eta_{t.} \quad x_{t.}\,].$$

Given the assumptions that $D(\eta_{t.}) = \Omega$ and that $C(\eta_{t.}, x_{t.}) = 0$, it follows that

$$(9)\qquad \begin{bmatrix} I & -\Pi' \\ 0 & I \end{bmatrix}\begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}\begin{bmatrix} I & 0 \\ -\Pi & I \end{bmatrix} = \begin{bmatrix} \Omega & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}.$$

Premultiplying this system by the inverse of the leading matrix gives an equivalent equation in the form of

$$\begin{aligned}\begin{bmatrix}\Sigma_{yy} & \Sigma_{yx}\\ \Sigma_{xy} & \Sigma_{xx}\end{bmatrix}\begin{bmatrix}I & 0\\ -\Pi & I\end{bmatrix} &= \begin{bmatrix}I & \Pi'\\ 0 & I\end{bmatrix}\begin{bmatrix}\Omega & 0\\ 0 & \Sigma_{xx}\end{bmatrix}\\ &= \begin{bmatrix}\Omega & \Pi'\Sigma_{xx}\\ 0 & \Sigma_{xx}\end{bmatrix}.\end{aligned} \tag{10}$$

From this system, the equations $\Sigma_{yy} - \Sigma_{yx}\Pi = \Omega$ and $\Sigma_{xy} - \Sigma_{xx}\Pi = 0$ may be extracted, from which are obtained the parameters that characterise the reduced-from relationship:

$$\Pi = \Sigma_{xx}^{-1}\Sigma_{xy} \qquad \text{and} \qquad \Omega = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}. \tag{11}$$

These parameters can be estimated provided that the empirical counterparts of the moment matrices Σ_{xx}, Σ_{yy} and Σ_{xy} are available in the form of $M_{xx} = T^{-1}\sum_t x'_{t.}x_{t.}$, $M_{yy} = T^{-1}\sum_t y'_{t.}y_{t.}$ and $M_{xy} = T^{-1}\sum_t x'_{t.}y_{t.}$.

Now consider combining the structural equation of (2) with a trivial identity to form the counterpart of equation (8). This is the equation

$$[\,y_{t.} \quad x_{t.}\,]\begin{bmatrix}\Gamma & 0\\ B & I\end{bmatrix} = [\,\varepsilon_{t.} \quad x_{t.}\,]. \tag{12}$$

Given that $D(\varepsilon) = \Sigma_{\varepsilon\varepsilon}$ and that $C(\varepsilon, x) = 0$, it follow that

$$\begin{bmatrix}\Gamma' & B'\\ 0 & I\end{bmatrix}\begin{bmatrix}\Sigma_{yy} & \Sigma_{yx}\\ \Sigma_{xy} & \Sigma_{xx}\end{bmatrix}\begin{bmatrix}\Gamma & 0\\ B & I\end{bmatrix} = \begin{bmatrix}\Sigma_{\varepsilon\varepsilon} & 0\\ 0 & \Sigma_{xx}\end{bmatrix}, \tag{13}$$

and, from this, an equivalent expression can be obtained the form of

$$\begin{aligned}\begin{bmatrix}\Sigma_{yy} & \Sigma_{yx}\\ \Sigma_{xy} & \Sigma_{xx}\end{bmatrix}\begin{bmatrix}\Gamma & 0\\ B & I\end{bmatrix} &= \begin{bmatrix}\Gamma'^{-1} & \Pi'\\ 0 & I\end{bmatrix}\begin{bmatrix}\Sigma_{\varepsilon\varepsilon} & 0\\ 0 & \Sigma_{xx}\end{bmatrix}\\ &= \begin{bmatrix}\Omega\Gamma & \Pi'\Sigma_{xx}\\ 0 & \Sigma_{xx}\end{bmatrix}.\end{aligned} \tag{14}$$

This identity provides the fundamental equations that relate the structural parameters Γ, B to the moment matrices of the data variables:

$$\begin{bmatrix}0\\ 0\end{bmatrix} = \begin{bmatrix}\Pi'\Sigma_{xy} & \Pi'\Sigma_{xx}\\ \Sigma_{xy} & \Sigma_{xx}\end{bmatrix}\begin{bmatrix}\Gamma\\ B\end{bmatrix}. \tag{15}$$

By setting $\Pi' = \Sigma_{yx}\Sigma_{xx}^{-1}$, we can express the matrix of the second equation in terms of the data moments alone.

Equation (15) is the basis from which the values the structural parameters Γ and B must be inferred. As it stands, the system contains insufficient information for the purpose. In particular, the constituent equation $\Pi'\Sigma_{xx}\Gamma + \Pi'\Sigma_{yx}B = 0$ is a transformation of its companion $\Sigma_{xx}\Gamma + \Sigma_{yx}B = 0$; and, therefore, it contains no additional information. In order for the parameters of the structrural equations to be identifiable, sufficient prior information regarding the structure must be available.

In practice, the prior information commonly takes the form of normalisation rules that set the diagonal elements of Γ to -1 and the exclusion restrictions that set certain of the elements of Γ and B to zeros. If none of restrictions affect more than one equation, then it is possible to treat each equation in isolation.

If the restrictions on the parameters of the jth equation are in the form of exclusion restrictions and a normalisation rule, then they can be represented by the equation

$$(17)\qquad \begin{bmatrix} R'_\diamond & 0 \\ 0 & R'_* \end{bmatrix} \begin{bmatrix} \gamma_{.j} \\ \beta_{.j} \end{bmatrix} = \begin{bmatrix} r_j \\ 0 \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} R'_\diamond & 0 \\ 0 & R'_* \end{bmatrix} \begin{bmatrix} \gamma_{.j} + e_j \\ \beta_{.j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where R_* comprises a selection of columns from the identity matrix I_K of order K, $R_\diamond$ comprises, likewise, a set of columns from the identity matrix I_M of order M, and r_j is a vector containing zeros and an element of minus one corresponding to the normalisation rule. The vector e_j is the jth column of I_M whose unit cancels with the normalised element of $\gamma_{.j}$.

The general solution to these restrictions is

$$(18)\qquad \begin{bmatrix} \gamma_{.j} \\ \beta_{.j} \end{bmatrix} = \begin{bmatrix} S_\diamond & 0 \\ 0 & S_* \end{bmatrix} \begin{bmatrix} \gamma_{\diamond j} \\ \beta_{*j} \end{bmatrix} - \begin{bmatrix} e_j \\ 0 \end{bmatrix},$$

where $\gamma_{\diamond j}$ and β_{*j} are composed of the M_j and K_j unrestricted elements of $\gamma_{.j}$ and $\beta_{.j}$ respectively, and where $S_\diamond$ and S_* are the complements of $R_\diamond$ and R_* within I_M and I_K respectively.

On substituting the solution of (18) into the equation $\Sigma_{xy}\gamma_{.j} + \Sigma_{xx}\beta_{.j} = 0$, which is from the jth equation of (16), we get

$$(19)\qquad \Sigma_{xy}S_\diamond\gamma_{\diamond j} + \Sigma_{xx}S_*\beta_{*j} = \Sigma_{xy}e_j.$$

This is a set of K equations in $M_j + K_j$ unknowns; and, given that the matrix $[\Sigma_{xy}\ ,\ \Sigma_{xx}]$ is of full rank, it follows that the necessary and sufficient condition for the identifiability of the parameters of the jth equation is that $K \geq M_j + K_j$.

If this condition is fulfilled, then any subset of $M_j + K_j$ of the equations of (19) will serve to determine $\gamma_{\diamond j}$ and β_{*j}. However, we shall be particularly interested in a set of $M_j + K_j$ independent equations in the form of

$$(20)\qquad \begin{bmatrix} P'_\diamond\Pi'\Sigma_{xy}S_\diamond & S'_\diamond\Pi'\Sigma_{xx}S_* \\ S'_*\Sigma_{xy}S_\diamond & S'_*\Sigma_{xx}S_* \end{bmatrix} \begin{bmatrix} \gamma_{\diamond j} \\ \beta_{*j} \end{bmatrix} = \begin{bmatrix} S'_\diamond\Pi'\Sigma_{xy}e_j \\ S'_*\Sigma_{xy}e_j \end{bmatrix},$$

which are derived by premultiplying equation (19) by the matrix $[\Pi S_\diamond, S_*]'$.

The so-called two stage least-square estimates are derived from these equation by substituting the empirical moments M_{xx}, M_{xy} and the estimate $\hat{\Pi} = M_{xx}^{-1}M_{xy}$ in place of S_{xx}, S_{xy} and $\Pi = S_{xx}^{-1}S_{xy}$ repectively and solving the resulting equations for $\gamma_{\diamond j}$ and β_{*j}.

Two-Stage Least Squares and Instrumental Variables Estimation

The 2SLS estimating equations were derived independently by Theil and by Basmann, who followed a different line of reasoning from the one which we have pursued above. Their approach was to highlight the reason for the failure of ordinary least-squares regression to deliver consistent estimates of the parameters of a structural equation.

The failure is due to the violation of an essential condition of regression analysis which is that the disturbances must be uncorrelated with the explanatory variables on the RHS of the equation. Within the equation $y_j = Y_\diamond\gamma_{\diamond j} + X_*\beta_{*j} + \varepsilon_j$, there is a direct dependence of $Y_\diamond$ on the structural disturbances of ε. However, the disturbances are independent of the exogenous variables in X_*.

The original derivations of the 2SLS estimator were inspired by the idea that, if it were possible to purge the variables of $Y_\diamond$ of their dependence on ε, then ordinary least-squares regression would become the appropriate method of estimation. Thus, if $X\Pi_{X\diamond}$ were available, then this could be put in place of $Y_\diamond$; and the problem of dependence would be overcome.

Although $X\Pi_{X\diamond}$ is an unknown quantity, a consistent estimate of it is available in the form of $\hat{Y}_\diamond = X\hat{\Pi}_{X\diamond}$. Finding the estimate $\hat{\Pi}_{X\diamond}$ represents the first stage of the 2SLS procedure. Applying ordinary least-squares regression to the equation $y_j = \hat{Y}_\diamond\gamma_{\diamond j} + X_*\beta_{*j} + e$ is the second stage.

An alternative approach which leads to the same 2SLS estimator is via the method of instrumental-variables estimation. The method depends upon finding a set of variables which are correlated with the regressors yet uncorrelated with the disturbances.

In the case of the structural equation, the appropriate instrumental variables are the exogenous variables of the system as a whole which are contained in the matrix X. Premultiplying the structural equation by X' gives

$$X'y_j = X'Y_\diamond\gamma_{\diamond j} + X'X_*\beta_{*j} + X'\varepsilon. \tag{21}$$

Within this system, the cross products correspond to a set of moment matrices which have the following limiting values:

$$\begin{aligned}
&\text{plim}(T^{-1}X'y_j) = \Sigma_{xy}e_j,\\
&\text{plim}(T^{-1}X'Y_\diamond) = \Sigma_{xy}S_\diamond,\\
&\text{plim}(T^{-1}X'X_*) = \Sigma_{xx}S_*,\\
&\text{plim}(T^{-1}X'\varepsilon) = 0.
\end{aligned} \tag{22}$$

When the moment matrices are replaced by their limiting values, we obtain the equation

$$\Sigma_{xy}e_j = \Sigma_{xy}S_\diamond\gamma_{\diamond j} + \Sigma_{xx}S_*\beta_{*j}, \tag{23}$$

which has been presented already as equation (19). In this system, there are K equations in $M_\diamond + K_*$ parameters. We may assume that $[\Sigma_{xy}, \Sigma_{xy}]$ is of full rank. In that case, the necessary condition for the indentifiablity of the parameters $\gamma_{\diamond j}$ and β_{*j} is that $K \geq M_\diamond + K_*$, which is to say that the number of exogenous variables in the system as a whole must be no less that the number of structural parameters that need to be estimated.

The empircal counterpart of (23) is the equation

$$X'y_j = X'Y_\diamond\gamma_{\diamond j} + X'X_*\beta_{*j}. \tag{24}$$

If $K = M_\diamond + K_*$, then this equation can be solved directly to provide the estimates. However, if $K > M_\diamond + K_*$, then the equation is bound to be algebraically inconsistent and the parameters are said to be overidentified. To resolve the inconsistency, we may apply to (21) the method of generalised least-squares regression. The disturbance term in (21), which is $X'\varepsilon$, had a dispersion matrix $D(X'\varepsilon) = \sigma^2 X'X$. When this is used in the context of the generalised least-squares estimator, we obtain, once again, the 2SLS estimates.

LEONTIEFF'S INPUT–OUTPUT ANALYSIS.

According to the postulate of Leontieff, the value x_{ij} of goods shipped from the ith sector of the economy to the jth sector is proportional to the activity level x_j of the latter: $x_{ij} = a_{ij}x_j$. Also, the activity level of the ith sector is reckoned as the sum of (the values of) the output, x_{ii}, consumed within that sector, the goods, $x_{ij}; j = 1, \ldots, n$, shipped to other sectors, and the goods, y_i, consumed in final demand.

Imagine a closed economy of three sectors which is characterised by the following activity levels and trade flows:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 100 \\ 200 \\ 100 \end{bmatrix}, \quad \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} 10 & 30 & 10 \\ 30 & 50 & 20 \\ 10 & 20 & 20 \end{bmatrix}, \quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 50 \\ 100 \\ 50 \end{bmatrix}.$$

Construct the complete input–output table including a row for the value added to each sector by factor services, and confirm that the various accounting identities have been observed in the construction of the table.

Calculate the matrix $A = [a_{ij}]$ of input–output coefficients. Use the method of Gaussian elimination and the method of back-substitution to solve the equation $(I - A)x = y$ to find the vector $x = [x_1, x_2, x_3]'$ of the activity levels in the three sectors when the levels of final demand are given by $y = [y_1, y_2, y_3]' = [60, 120, 60]'$.

Answer. The trade flows, the activity levels and the final demands are displayed in the following input–output table:

	Sector 1	*Sector 2*	*Sector 3*	*Final Demand*	*Total Demand*
Sector 1	10	30	10	50	100
Sector 2	30	50	20	100	200
Sector 3	10	20	20	50	100
Factors	50	100	50	200	
Activity Level	100	200	100		

The matrix A of input–output coefficients and the Leontieff matrix $I - A$ are

$$A = \begin{bmatrix} 0.1 & 0.15 & 0.1 \\ 0.3 & 0.25 & 0.2 \\ 0.1 & 0.1 & 0.2 \end{bmatrix}, \qquad I - A = \begin{bmatrix} 0.9 & -0.15 & -0.1 \\ -0.3 & 0.75 & -0.2 \\ -0.1 & -0.1 & 0.8 \end{bmatrix}.$$

Imagine that the vector of final demands becomes $y = [y_1, y_2, y_3]' = [60, 120, 60]'$. Then, to find the corresponding activity levels in $x = [x_1, x_2, x_3]'$, we must solve the system $(I - A)x = y$. We have

$$\begin{bmatrix} 0.9 & -0.15 & -0.1 \\ -0.3 & 0.75 & -0.2 \\ -0.1 & -0.1 & 0.8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 60 \\ 120 \\ 60 \end{bmatrix} \iff \begin{bmatrix} 0.9 & -0.15 & -0.1 \\ -0.9 & 2.25 & -0.6 \\ -0.9 & -0.9 & 7.2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 60 \\ 360 \\ 540 \end{bmatrix}.$$

Adding the first row to the second row and to the third gives

$$\begin{bmatrix} 0.9 & -0.15 & -0.1 \\ 0.0 & 2.1 & -0.7 \\ 0.0 & -1.05 & 7.1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 60 \\ 420 \\ 600 \end{bmatrix} \iff \begin{bmatrix} 0.9 & -0.15 & -0.1 \\ 0.0 & 2.1 & -0.7 \\ 0.0 & -2.1 & 14.2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 60 \\ 420 \\ 1200 \end{bmatrix}.$$

Adding the second row of the final expression to the third row gives the following triangular system:

$$\begin{bmatrix} 0.9 & -0.15 & -0.1 \\ 0.0 & 2.1 & -0.7 \\ 0.0 & 0.0 & 13.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 60 \\ 420 \\ 1620 \end{bmatrix}.$$

The solution of this system is

$$x_3 = 120, \qquad x_2 = 240, \qquad x_1 = 120.$$

Queen Mary & Westfield College

UNIVERSITY OF LONDON
BSc (ECONOMICS)
EXAMINATION OF ASSOCIATE STUDENTS

Econometric Theory

December, 1999

Answer THREE questions in TWO HOURS

1. What do you understand by the omitted-variables bias?

Derive explicit expressions for the ordinary least-squares estimators of β_1 and β_2 in the partitioned model $(y; X_1\beta_1 + X_2\beta_2, \sigma^2 I)$ and compare these with the estimators of β_1 and β_2 which are obtained by regressing y on X_1 and X_2 separately. Under what conditions would the different estimators of β_1 and β_2 be equal?

2. 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.

3. 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?

4. 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.

5. Imagine that the temporal regression equation

$$y(t) = x(t)\beta + \eta(t)$$

embodies a disturbance sequence $\eta(t)$ which follows a first-order autoregressive process such that

$$\eta(t) = \phi\eta(t-1) + \varepsilon(t),$$

where $\varepsilon(t)$ is a white-noise sequence of independently and identically distributed random variables. Show that the covariance of any two elements of the disturbance process which are separated by τ time periods is given by

$$C(\eta_t, \eta_{t-\tau}) = \sigma_\varepsilon^2 \frac{\phi^\tau}{1-\phi^2}.$$

What is the variances covariance matrix of a vector $\eta = [\eta_0, \ldots, \eta_{T-1}]'$ of T elements of the process and what is the form of the inverse of this matrix. How should one attempt to estimate the parameters ϕ and β.

6. Demonstrate that a dynamic econometric equation in the form of

$$y(t) = \phi_1 y(t-1) + \cdots + \phi_p y(t-p) + \beta_0 x(t) + \cdots + \beta_k x(t-k) + \varepsilon(t)$$

can be rewritten as

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

where $\gamma = (\beta_0 + \cdots + \beta_k)/(1 - \phi_1 - \cdots - \phi_p)$ is the steady state gain of the transfer function from $x(t)$ to $y(t)$. What is the customary interpretation of the second equation?

Describe how you would estimate the dynamic equation (a) in the case where $x(t)$ is generated by a stationary stochastic process and (b) in the case where $x(t)$ is generated by a nonstationary unit-root process.

7. 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?

8. 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

