THE STATISTICAL PROPERTIES OF THE OLS ESIMATOR: UNBIASEDNESS AND EFFICIENCY

Some Statistical Properties of the Estimator

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

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

The expectation is

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

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

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

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

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

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

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

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

Estimating the Variance of the Disturbance

The principle of least squares does not, of its own, suggest a means of estimating the disturbance variance $\sigma^2 = V(\varepsilon_t)$. However it is natural to estimate the moments of a probability distribution by their empirical counterparts. Given that $e_t = y_t - x_{t.}\hat{\beta}$ is an estimate of ε_t, it follows that $T^{-1}\sum_t e_t^2$ may be used to estimate σ^2. However, it transpires that this is biased. An unbiased estimate is provided by

$$(7)\qquad \begin{aligned} \hat{\sigma}^2 &= \frac{1}{T-k}\sum_{t=1}^{T} e_t^2 \\ &= \frac{1}{T-k}(y - X\hat{\beta})'(y - X\hat{\beta}). \end{aligned}$$

The unbiasedness of this estimate may be demonstrated by finding the expected value of $(y - X\hat{\beta})'(y - X\hat{\beta}) = y'(I-P)y$. Given that $(I-P)y = (I-P)(X\beta + \varepsilon) = (I-P)\varepsilon$ in consequence of the condition $(I-P)X = 0$, it follows that

$$(8)\qquad E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = E(\varepsilon'\varepsilon) - E(\varepsilon'P\varepsilon).$$

The value of the first term on the RHS is given by

$$(9)\qquad E(\varepsilon'\varepsilon) = \sum_{t=1}^{T} E(e_t^2) = T\sigma^2.$$

The value of the second term on the RHS is given by

$$(10)\qquad \begin{aligned} E(\varepsilon'P\varepsilon) &= \text{Trace}\big\{E(\varepsilon'P\varepsilon)\big\} = E\big\{\text{Trace}(\varepsilon'P\varepsilon)\big\} = E\big\{\text{Trace}(\varepsilon\varepsilon'P)\big\} \\ &= \text{Trace}\big\{E(\varepsilon\varepsilon')P\big\} = \text{Trace}\big\{\sigma^2 P\big\} = \sigma^2\text{Trace}(P) \\ &= \sigma^2 k. \end{aligned}$$

The final equality follows from the fact that $\text{Trace}(P) = \text{Trace}(I_k) = k$. Putting the results of (9) and (10) into (8), gives

$$(11)\qquad E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = \sigma^2(T-k);$$

and, from this, the unbiasedness of the estimator in (7) follows directly.

THE PARTITIONED REGRESSSION MODEL

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

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

Now let us investigate the effect that conditions of orthogonality amongst the regressors have upon the ordinary least-squares estimates of the regression parameters. Consider a partitioned regression model, which can be written as

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

It can be assumed that the variables in this equation are in deviation form. 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 assuming that the empirical correlation between variables in X_1 and variables in X_2 is zero.

The effect upon the ordinary least-squares estimator can be seen by examining 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.

It should be understood 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):

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

It can be confirmed easily 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, 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, the 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 estimates does not arise.

DIAGONALISATION OF A SYMMETRIC MATRIX

Characteristic Roots and Characteristic Vectors. 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 $x_2'Ax_1 = \lambda_1 x_2'x_1$ 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.

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

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. Post-multiplying 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

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

The Geometry of Quadratic Forms

The Circle Let the coordinates of the points in the Cartersian plane be the denoted by (z_1, z_2). Then the equation of a circle of radius r centred on the origin is just

$$z_1^2 + z_2^2 = r^2. \tag{1}$$

This follows immediately from Pythagorus. The so-called parametic equations for the coordiantes of the circle are

$$z_1 = r\cos(\omega), \quad \text{and} \quad z_2 = r\sin(\omega). \tag{2}$$

The Ellipse The equation of an ellipse whose principal axes are aligned with those of the coordinate system in the (y_1, y_2) plane is

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = r^2, \tag{2}$$

On setting $\lambda_1 y_1^2 = z_1^2$ and $\lambda_2 y_2^2 = z_2^2$, we can see that

$$y_1 = \frac{z_1}{\sqrt{\lambda_1}} = \frac{r}{\sqrt{\lambda_1}}\cos(\omega), \quad y_2 = \frac{z_2}{\sqrt{\lambda_2}} = \frac{r}{\sqrt{\lambda_2}}\sin(\omega). \tag{3}$$

We can write equation (2) in matrix notation as

$$r^2 = [\, y_1 \quad y_2 \,] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = z_1^2 + z_2^2. \tag{4}$$

This implies

$$(5)\qquad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and

$$(6)\qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{\lambda_1} & 0 \\ 0 & 1/\sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

The Oblique Ellipse An oblique ellipse is one whose principal axes are not aligned with those of the coordinate system. Its general equation is

$$(6)\qquad a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = r^2;$$

which is subject to the ccondition that $a_{11}a_{22} - 2a_{12} > 0$. We can write this in matrix notation:

$$(7)\qquad \begin{aligned} r^2 &= [\,x_1 \quad x_2\,] \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= [\,y_1 \quad y_2\,] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = z_1^2 + z_2^2 \end{aligned}$$

The coefficients of the equation (6) are the elements of the matrix

$$(8)\qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \lambda_1\cos^2\theta + \lambda_2\sin^2\theta & (\lambda_2 - \lambda_1)\cos\theta\sin\theta \\ (\lambda_2 - \lambda_1)\cos\theta\sin\theta & \lambda_1\sin^2\theta + \lambda_2\cos^2\theta \end{bmatrix}.$$

Notice that if $\lambda_1 = \lambda_2$, which is to say that both axes are are rescaled by the same factor, then the equation is that of a circle of radius λ_1, and the rotation of the circle has no effect.

The mapping from the ellipse to the circle is

$$(9)\qquad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1}(x_1\cos\theta - x_2\sin\theta) \\ \sqrt{\lambda_2}(x_1\sin\theta + x_2\cos\theta) \end{bmatrix}.$$

and the inverse mapping, from the circle to the ellipse, is

$$(10)\qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1/\sqrt{\lambda_1} & 0 \\ 0 & 1/\sqrt{\lambda_2} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

We see from the latter that the circle is converted to an oblique ellipse via two operations. The first is an operation of scaling which produces an ellipse whose principal axes are aligned with those of the coordinate system. The second operation is a rotation which tilts the ellipse.

The vectors of the matrix which affects the rotation define the axes of the ellipse. They have the property that, when they are mapped through the matrix A, their orientation is preserved and only their length is altered. Thus

$$
\begin{aligned}
&\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}. \\
&= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_1 \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}
\end{aligned} \tag{11}
$$

Such vectors are described as the characteristic vectors of the matrix, and the factors λ_1 and λ_2 by which their lengths are altered under the transformation are described as the corresponding characteristic roots.

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 equations in this notation is that it enables us to envisage more clearly the Pythagorean relationship between the vectors. Thus, from the condition that $P = P' = P^2$, which is equivalent to the condition that $P'(I - P) = 0$, it can be established that

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

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 is given by $\varepsilon' P\varepsilon = (X\hat{\beta} - X\beta_0)'(X\hat{\beta} - X\beta_0)$. 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

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

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

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

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

$$\begin{bmatrix} C_1'\varepsilon \\ C_2'\varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 I_k & 0 \\ 0 & \sigma^2 I_{T-k} \end{bmatrix} \right), \tag{8}$$

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 expressed 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-square 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.

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

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

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:

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

We rewrite this equation as

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

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.

THE GEOMETRIC LAG SCHEME

An early approach to the problem of defining a lag structure which depends on a limited number of parameters was that of Koyk who proposed the following geometric lag scheme:

$$(1) \qquad y(t) = \beta\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} + \varepsilon(t).$$

Here, although we have an infinite set of lagged values of $x(t)$, we have only two parameters which are β and ϕ.

It can be seen that the impulse-response function of the Koyk model takes a very restricted form. It begins with an immediate response to the impulse. Thereafter, the response dies away in the manner of a convergent geometric series, or of a decaying exponential function of the sort which also characterises processes of radioactive decay.

The values of the coefficients in the Koyk distributed-lag scheme tend asymptotically to zero; and so it can said that the full response is never accomplished in a finite time. To characterise the speed of response, we may calculate the median lag which is analogous to the half-life of a process of radioactive decay. The gain of the transfer function, which is obtained by summing the geometric series $\{\beta, \phi\beta, \phi^2\beta, \ldots\}$, has the value of

$$(2) \qquad \gamma = \frac{\beta}{1-\phi}.$$

To make the Koyk model amenable to estimation, we might first transform the equation. By lagging the equation by one period and multiplying the result by ϕ, we get

$$(3) \qquad \phi y(t-1) = \beta\big\{\phi x(t-1) + \phi^2 x(t-2) + \phi^3 x(t-3) + \cdots\big\} + \phi\varepsilon(t-1).$$

Taking the latter from (1) gives

$$(4) \qquad y(t) - \phi y(t-1) = \beta x(t) + \big\{\varepsilon(t) - \phi\varepsilon(t-1)\big\}.$$

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

$$(5) \qquad (1-\phi L)y(t) = \beta x(t) + (1-\phi L)\varepsilon(t),$$

of which the rational form is

$$y(t) = \frac{\beta}{1 - \phi L} x(t) + \varepsilon(t). \tag{6}$$

In fact, by using the expansion

$$\begin{aligned} \frac{\beta}{1 - \phi L} x(t) &= \beta\big\{1 + \phi L + \phi L^2 + \cdots\big\} x(t) \\ &= \beta\big\{x(t) + \phi x(t-1) + \phi x(t-2) + \cdots\big\} \end{aligned} \tag{7}$$

within equation (6), we can recover the original form under (1).

Equation (4) is not amenable to consistent estimation by ordinary least squares regression. The reason is that the composite disturbance term $\{\varepsilon(t) - \phi\varepsilon(t-1)\}$ is correlated with the lagged dependent variable $y(t-1)$—since the elements of $\varepsilon(t-1)$ form part of the contemporaneous elements of $y(t-1)$. This conflicts with one of the basic conditions for the consistency of ordinary least-squares estimation which is that the disturbances must be uncorrelated with the regressors. Nevertheless, there is available a wide variety of simple procedures for estimating the parameters of the Koyk model consistently.

One of the simplest procedures for estimating the geometric-lag scheme is based on the original form of the equation under (1). In view of that equation, we may express the elements of $y(t)$ which fall within the sample as

$$\begin{aligned} y_t &= \beta \sum_{i=0}^{\infty} \phi^i x_{t-i} + \varepsilon_t \\ &= \theta\phi^t + \beta \sum_{i=0}^{t-1} \phi^i x_{t-i} + \varepsilon_t \\ &= \theta\phi^t + \beta z_t + \varepsilon_t. \end{aligned} \tag{8}$$

Here

$$\theta = \beta\big\{x_0 + \phi x_{-1} + \phi^2 x_{-2} + \cdots\big\} \tag{9}$$

is a nuisance parameter which embodies the presample elements of the sequence $x(t)$, whilst

$$z_t = x_t + \phi x_{t-1} + \cdots + \phi^{t-1} x_1 \tag{10}$$

is an explanatory variable compounded from the observations $x_t, x_{t-1}, \ldots, x_1$ and from the value attributed to ϕ.

The procedure for estimating ϕ and β which is based on equation (8) involves running a number of trial regressions with differing values of ϕ and therefore of the regressors ϕ^t and z_t; $t = 1, \ldots, T$. The definitive estimates are those which correspond to the least value of the residual sum of squares.

It is possible to elaborate this procedure so as to obtain the estimates of the parameters of the equation

$$(11) \qquad y(t) = \frac{\beta}{1 - \phi L} x(t) + \frac{1}{1 - \rho L} \varepsilon(t),$$

which has a first-order autoregressive disturbance scheme in place of the white-noise disturbance to be found in equation (6). An estimation procedure may be devised which entails searching for the optimal values of ϕ and ρ within the square defined by $-1 < \rho, \phi < 1$. There may even be good reason to suspect that these values will be found within the quadrant defined by $0 \leq \rho, \phi < 1$.

The task of finding estimates of ϕ and ρ is assisted by the fact that we can afford, at first, to ignore autoregressive nature of the disturbance process while searching for the optimum value of the systematic parameter ϕ.

When a value has been found for ϕ, we shall have residuals which are consistent estimates of the corresponding disturbances. Therefore, we can proceed to fit the AR(1) model to the residuals in the knowledge that we will then be generating a consistent estimate of the parameter ρ; and, indeed, we can might use ordinary least-squares regression for this purpose. Having found the estimate for ρ, we should wish to revise our estimate of ϕ.

Lagged Dependent Variables

In spite of the relative ease with which one may estimate the Koyk model, it has been common throughout the history of econometrics to adopt an even simpler approach in the attempt to model the systematic dynamics.

Perhaps the easiest way of setting a regression equation in motion is to include a lagged value of the dependent variable on the RHS in the company of the explanatory variable x. The resulting equation has the form of

$$(12) \qquad y(t) = \phi y(t-1) + \beta x(t) + \varepsilon(t).$$

In terms of the lag operator, this is

$$(13) \qquad (1 - \phi L) y(t) = \beta x(t) + \varepsilon(t),$$

of which the rational form is

$$(14) \qquad y(t) = \frac{\beta}{1 - \phi L} x(t) + \frac{1}{1 - \phi L} \varepsilon(t).$$

The advantage of equation (12) is that it is amenable to estimation by ordinary least-squares regression. Although the estimates will be biased in finite samples, they are, nevertheless, consistent in the sense that they will tend to converge upon the true values as the sample size increases—provided, of course, that the model corresponds to the processes underlying the data.

The model with a lagged dependent variable generates precisely the same geometric distributed-lag schemes as does the Koyk model. This can be confirmed by applying the expansion given under (7) to the rational form of the present model given in equation (14) and by comparing the result with (1). The comparison of equation (14) with the corresponding rational equation (6) for the Koyk model shows that we now have an AR(1) disturbance process described by the equation

$$\eta(t) = \phi\eta(t-1) + \varepsilon(t) \tag{15}$$

in place of a white-noise disturbance $\varepsilon(t)$.

This might be viewed as an enhancement of the model were it not for the constraint that the parameter ϕ in the systematic transfer function is the same as the parameter ϕ in the disturbance transfer function. For such a constraint is appropriate only if it can be argued that the disturbance dynamics are the same as the systematic dynamics—and they need not be.

To understand the detriment of imposing the constraint, let us imagine that the true model is of the form given under (11) with ρ and ϕ taking very different values. Imagine that, nevertheless, it is decided to fit the equation under (14). Then the estimate of ϕ will be a biased and an inconsistent one whose value falls somewhere between the true values of ρ and ϕ in equation (11). If this estimate of ϕ is taken to represent the systematic dynamics of the model, then our inferences about such matters as the speed of convergence of the impulse response and the value of the steady-state gain are liable to be misleading.

Partial Adjustment and Adaptive Expectations

There are some tenuous justifications both for the Koyk model and for the model with a lagged dependent variable which arise from economic theory.

Consider a partial-adjustment model of the form

$$y(t) = \lambda\big\{\gamma x(t)\big\} + (1-\lambda)y(t-1) + \varepsilon(t), \tag{16}$$

where, for the sake of a concrete example, $y(t)$ is current consumption, $x(t)$ is disposable income and $\gamma x(t) = y^*(t)$ is "desired"consumption. Here we are supposing that habits of consumption persist, so that what is consumed in the current period is a weighted combination of the previous consumption and present desired consumption. The weights of the combination depend on the partial-adjustment parameter $\lambda \in (0, 1]$. If $\lambda = 1$, then the consumers adjust their consumption instantaneously to the desired value. As $\lambda \to 0$, their consumption habits become increasingly persistent. When the notation $\lambda\gamma = \beta$ and $(1 - \lambda) = \phi$ is adopted, equation (16) becomes identical to equation (12) which relates to a simple regression model with a lagged dependent variable.

An alternative model of consumers' behaviour derives from Friedman's Permanent Income Hypothesis. In this case, the consumption function is specified as

$$y(t) = \delta x^*(t) + \varepsilon(t), \tag{17}$$

where

$$\begin{aligned} x^*(t) &= (1-\phi)\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} \\ &= \frac{1-\phi}{1-\phi L}x(t) \end{aligned} \tag{18}$$

is the value of permanent or expected income which is formed as a geometrically weighted sum of all past values of income. Here it is asserted that a consumer plans his expenditures in view of his customary income, which he assesses by taking a long view over all of his past income receipts.

An alternative expression for the sequence of permanent income is obtained by multiplying both sides of (18) by $1 - \phi L$ and rearranging the result. Thus

$$x^*(t) - x^*(t-1) = (1-\phi)\big\{x(t) - x^*(t-1)\big\}, \tag{19}$$

which depicts the change of permanent income as a fraction of the prediction error $x(t) - x^*(t-1)$. The equation depicts a so-called adaptive-expectations mechanism.

On substituting the expression for permanent income under (18) into the equation (17) of the consumption function, we get

$$y(t) = \delta\frac{(1-\phi)}{1-\phi L}x(t) + \varepsilon(t). \tag{20}$$

When the notation $\delta(1 - \phi) = \beta$ is adopted, equation (20) becomes identical to the equation (6) of the Koyk model.

ERROR CORRECTION, NONSTATIONARITY AND COINTERGRATED VARIABLES

In this note, we shall consider some of the methods that are available for modelling tranfer function relationships which map non-stationary input sequences into nonstationary outputs.

We can begin by considering a first-order simple dynamic model of the form

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

Taking $y(t-1)$ from both sides of this equation gives

$$\begin{aligned} \nabla y(t) = y(t) - y(t-1) &= (\phi - 1)y(t-1) + \beta x(t) + \varepsilon(t) \\ &= (1-\phi)\left\{\frac{\beta}{1-\phi}x(t) - y(t-1)\right\} + \varepsilon(t) \\ &= \lambda\big\{\gamma x(t) - y(t-1)\big\} + \varepsilon(t), \end{aligned} \tag{2}$$

where $\lambda = 1 - \phi$ and where γ is the gain of the transfer function which maps from $x(t)$ to $y(t)$. 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 a more general model denoted by

$$y(t) = \phi_1 y(t-1) + \cdots + \phi_p y(t-p) + \beta_0 x(t) + \cdots + \beta_k x(t-k) + \varepsilon(t). \tag{3}$$

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

$$\begin{aligned} y(t) = \theta y(t-1) &+ \rho_1 \nabla y(t-1) + \cdots + \rho_p \nabla y(t-p+1) \\ &+ \kappa x(t) + \delta_0 \nabla x(t) + \cdots + \delta_k \nabla x(t-k+1) + \varepsilon(t), \end{aligned} \tag{4}$$

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 (4). This gives

$$\begin{aligned} \nabla y(t) =& (\theta - 1)y(t-1) + \kappa x(t) \\ &+ \rho_1 \nabla y(t-1) + \cdots + \rho_p \nabla y(t-p+1) \\ &+ \delta_0 \nabla x(t) + \cdots + \delta_k \nabla x(t-k+1) + \varepsilon(t). \end{aligned} \tag{5}$$

The first two terms on the RHS combine to give

$$(6)\qquad \begin{aligned}(\theta-1)y(t-1)+\kappa x(t) &= (1-\theta)\left\{\frac{\kappa}{1-\theta}x(t)-y(t-1)\right\}\\ &= \lambda\big\{\gamma x(t)-y(t-1)\big\}\end{aligned}$$

which is an error-correction term in which γ is the value of the gain of the tranfer function. It follows that the error-correction form of equation (3) is

$$(7)\quad \nabla y(t)=\lambda\big\{\gamma x(t)-y(t-1)\big\}+\sum_{i=1}^{p-1}\rho_i\nabla y(t-i)+\sum_{i=0}^{k-1}\delta_i\nabla x(t-i)+\varepsilon(t).$$

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 (2). 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 (3) into equation (4), let us consider the following example:

$$(8)\qquad \begin{aligned}&\beta_0x(t)+\beta_1x(t-1)+\beta_2x(t-2)+\beta_3x(t-3)\\ &\quad=\big\{\beta_0+\beta_1+\beta_2+\beta_3\big\}x(t)-\big\{\beta_1+\beta_2+\beta_3\big\}\big\{x(t)-x(t-1)\big\}\\ &\qquad\qquad-\big\{\beta_2+\beta_3\big\}\big\{x(t-1)-x(t-2)\big\}\\ &\qquad\qquad-\beta_3\big\{x(t-2)-x(t-3)\big\}\\ &\quad=\kappa x(t)+\delta_0\nabla x(t)+\delta_1\nabla x(t-1)+\delta_2\nabla x(t-2).\end{aligned}$$

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:

$$(9)\qquad \begin{bmatrix}1&0&0&0\\-1&1&0&0\\0&-1&1&0\\0&0&-1&1\end{bmatrix}\begin{bmatrix}x(t)\\x(t-1)\\x(t-2)\\x(t-3)\end{bmatrix}=\begin{bmatrix}x(t)\\-\nabla x(t)\\-\nabla x(t-1)\\-\nabla x(t-2)\end{bmatrix}$$

and

$$(10)\qquad [\,\beta_0\quad\beta_1\quad\beta_2\quad\beta_3\,]\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&1&1&0\\1&1&1&1\end{bmatrix}=[\,\kappa\quad-\delta_0\quad-\delta_1\quad-\delta_2\,].$$

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 (3) into equation (4).

Reparametrisation of Dynamic Models

In this note, we demonstrate a simple identity affecting polynomials in the lag operator and we show how this can be used in describing relationships between cointegrated time series. The same identity has been used to express an ARIMA process as the sum of a stationary stochastic process and an ordinary random walk. This expression is commonly know the Beveridge–Nelson decomposition after its original proponents.

A Polynomial Identity

Let $\beta(z) = \beta_0 + \beta_1 z + \cdots + \beta_k z^k = \sum_{j=0}^{k} \beta_j z^j$ be a polynomial of degree k in the argument z. We wish to show that this can be written in the following forms:

$$\begin{aligned} \beta(z) &= \beta(1) + \nabla(z)\gamma(z) \\ &= z^n\beta(1) + \nabla(z)\delta_n(z), \end{aligned} \tag{1}$$

where $\nabla(z) = 1 - z$, where $0 \leq n \leq k$ and where $\gamma(z)$ and $\delta_n(z)$ are both polynomials of degree $k-1$. Also, $\beta(1)$ is the constant which is obtained by setting $z = 1$ in the polynomial $\beta(z)$.

To obtain the first expression on the RHS of (1), we divide $\beta(z)$ by $\nabla(z) = 1 - z$ to obtain a quotient of $\gamma(z)$ and a remainder of δ:

$$\beta(z) = \gamma(z)(1-z) + \delta. \tag{2}$$

Setting $z = 1$ in this equation gives

$$\delta = \beta(1) = \beta_0 + \beta_1 + \cdots + \beta_k. \tag{3}$$

This is an instance of the well-known remainder theorem of polynomial division. The coefficients of the quotient polynomial $\gamma(z)$ are given by

$$\gamma_p = -\sum_{j=p+1}^{k} \beta_j, \quad \text{where} \quad p = 0, \ldots, k-1. \tag{4}$$

There is a wide variety of ways in which these coefficients may de derived, including the familiar method of long division. Probably, the easiest way is via the method of synthetic division which may be illustrated by an example.

Example. Consider the case where $k = 3$. Then

$$\beta_0 + \beta_1 z + \beta_2 z^2 + \beta_3 z^3 = (\gamma_0 + \gamma_1 z + \gamma_2 z^2)(1 - z) + \delta. \tag{5}$$

By equating the coefficients associated with the same powers of z on either side of the equation, we obtain the following identities:

$$\begin{aligned} \beta_3 &= -\gamma_2 \\ \beta_2 &= -\gamma_1 + \gamma_2 \\ \beta_1 &= -\gamma_0 + \gamma_1 \\ \beta_0 &= \delta + \gamma_0. \end{aligned} \tag{6}$$

These can be rearranged to give

$$\begin{aligned} \gamma_2 &= -\beta_3 \\ \gamma_1 &= -\beta_2 + \gamma_2 &&= -(\beta_2 + \beta_3) \\ \gamma_0 &= -\beta_1 + \gamma_1 &&= -(\beta_1 + \beta_2 + \beta_3) \\ \delta &= \beta_0 - \gamma_0 &&= \beta_0 + \beta_1 + \beta_2 + \beta_3. \end{aligned} \tag{7}$$

To obtain the second expression on the RHS of (1), consider the identity

$$1 = z^n + \nabla(z)(1 + z + \cdots + z^{n-1}), \tag{8}$$

where $1 < n \leq k$. Multiplying both sides by $\beta(1)$ gives

$$\beta(1) = z^n \beta(1) + \nabla(z)\{1 + z + \cdots + z^{n-1}\}\beta(1), \tag{9}$$

On substituting this expression into the first equation of (1) and on defining

$$\delta_n(z) = \gamma(z) + \{1 + z + \cdots + z^{n-1}\}\beta(1), \tag{10}$$

we can write

$$\beta(z) = z^n \beta(1) + \nabla(z)\delta_n(z). \tag{11}$$

This is a general expression which covers both equations of (1), since setting $n = 0$ reduces it to the first equation.

A leading instance is obtained by setting $n = 1$. In that case, equation (9) gives $\beta(1) = z\beta(1) + \nabla(z)\beta(1)$. Substituting this into $\beta(z) = \beta(1) + \nabla(z)\gamma(z)$ gives the following specialisation of equation (11):

$$\begin{aligned} \beta(z) &= z\beta(1) + \nabla(z)\{\gamma(z) + \beta(1)\} \\ &= z\beta(1) + \nabla(z)\delta_1(z). \end{aligned} \tag{12}$$

By comparing coefficients associated with the same powers of z on both sides of the equation, it can be seen that the constant term of the polynomial $\delta_1(z)$ is just β_0. Reference to (3) and (7) confirms this result.

Reparametrisation of a Distributed Lag Model

Consider a distributed-lag model of the form

$$y(t) = \beta_0 x(t) + \beta_1 x(t-1) + \cdots + \beta_k x(t-k) + \varepsilon(t). \tag{13}$$

This can be written is summary notation as

$$y(t) = \beta(L)x(t) + \varepsilon(t), \tag{14}$$

where $\beta(L) = \beta_0 + \beta_1 L + \cdots + \beta_k L^k$ is a polynomial in the lag operator L.

Using the basic identity from (1), we can set $\beta(L) = \beta(1) + \nabla\delta(L)$. Then taking $y(t-1)$ from both sides of the resulting equation gives

$$\nabla y(t) = \{\beta(1)x(t) - y(t-1)\} + \nabla\delta(L)x(t) + \varepsilon(t), \tag{15}$$

which is an error-correction formulation of equation (14).

Reparametrisation of an Autoregressive Distributed Lag Model

Now consider an equation in the form of

$$y(t) = \phi_1 y(t-1) + \cdots + \phi_p y(t-p) + \beta_0 x(t) + \cdots + \beta_k x(t-k) + \varepsilon(t), \tag{16}$$

which can be written is summary notation as

$$\alpha(L)y(t) = \beta(L)x(t) + \varepsilon(t), \tag{17}$$

with $\alpha(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p = 1 - \phi_1 L - \cdots - \phi_p L^p$. On setting $\alpha(L) = \alpha(1)L + \theta_1(L)\nabla$ and $\beta(L) = \beta(1)L + \delta_1(L)\nabla$, this can be rewritten as

$$\{\alpha(1)L + \theta_1(L)\nabla\}y(t) = \{\beta(1)L + \delta_1(L)\nabla\}x(t) + \varepsilon(t), \tag{18}$$

where the leading element of $\theta_1(L)$ is $\alpha_0 = 1$. Define

$$\rho(L) = \rho_1 L + \rho_2 L^2 + \cdots + \rho_p L^p = 1 - \theta_1(L). \tag{19}$$

Then equation (18) can be rearranged to give

$$\begin{aligned}\nabla y(t) &= \{\beta(1)Lx(t) - \alpha(1)Ly(t)\} + \delta_1(L)\nabla x(t) + \rho(L)\nabla y(t) + \varepsilon(t)\\ &= \lambda\{\gamma x(t-1) - y(t-1)\} + \delta_1(L)\nabla x(t) + \rho(L)\nabla y(t) + \varepsilon(t),\end{aligned} \tag{20}$$

where $\lambda = \alpha(1)$ is the so-called adjustment parameter and where $\gamma = \beta(1)/\alpha(1)$ is the steady-state gain of the rational transfer function $\beta(L)/\alpha(L)$. The term $\gamma x(t-1) - y(t-1)$ is described as the equilibrium error; and the value of the error will tend to zero if a steady state is maintained by $x(t)$ and if there are no disturbances. Equation (20) is the classical form of the error-correction equation.

The Beveridge–Nelson Decomposition

Consider an ARIMA model which is represented by the equation

$$\alpha(L)\nabla y(t) = \mu(L)\varepsilon(t) \tag{21}$$

Dividing both sides by $\alpha(L)$ gives

$$\begin{aligned}\nabla y(t) &= \frac{\mu(L)}{\alpha(L)}\varepsilon(t)\\ &= \psi(L)\varepsilon(t),\end{aligned} \tag{22}$$

where $\psi(L)$ stands for the power series expansion of the rational function. If the coefficients of this expansion form an absolutely summable sequence, then $\psi(L)$ will be subject to a decomposition in the form of equation (1). Then

$$\begin{aligned}\nabla y(t) &= \psi(1)\varepsilon(t) + \lambda(L)\nabla\varepsilon(t)\\ &= \nabla v(t) + \nabla w(t).\end{aligned} \tag{23}$$

The first term on the RHS is $\nabla v(t) = \psi(1)\varepsilon(t)$. Here $v(t)$ stands for a random walk. The difference operator can be elimiated from the second term to give $w(t) = \lambda(L)\varepsilon(t)$, which stands for a stationary stochastic process.

For another perspective on this decomposition, we may consider the following partial-fraction decompostion:

$$\frac{\mu(z)}{\alpha(z)\nabla(z)} = \frac{\gamma(z)}{\alpha(z)} + \frac{\delta}{\nabla(z)}. \tag{24}$$

Multiplying both sides by $\nabla(z)$ gives

$$\frac{\mu(z)}{\alpha(z)} = \frac{\nabla(z)\gamma(z)}{\alpha(z)} + \delta, \tag{25}$$

whence setting $z = 1$ gives $\delta = \mu(1)/\alpha(1)$. Thereafter, we can find $\gamma(z) = \{\mu(z) - \delta\alpha(z)\}/\nabla(z)$.

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

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

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

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

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

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