SOME STATISTICAL PROPERTIES OF THE OLS ESTIMATOR

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

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

The expectation is

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

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

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

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

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

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

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

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

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

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

ORTHOGONALITY AND OMITTED-VARIABLES BIAS

Let us now investigate the effect that a condition of orthogonality amongst the regressors might have upon the ordinary least-squares estimates of the regression parameters. Let us take the partitioned regression model of equation (109) which was written as

$$y = [\, X_1, X_2 \,] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{166}$$

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

$$X'X = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} [\, X_1 \quad X_2 \,] = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}, \tag{167}$$

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

$$(X'X)^{-1} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix}. \tag{168}$$

We also have

$$X'y = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}. \tag{169}$$

On combining these elements, we find that

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y \\ (X_2'X_2)^{-1}X_2'y \end{bmatrix}. \tag{170}$$

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 (112) and (117):

$$\begin{aligned}\hat{\beta}_1 &= (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2),\\ \hat{\beta}_2 &= \left\{X_2'(I - P_1)X_2\right\}^{-1}X_2'(I - P_1)y, \quad P_1 = X_1(X_1'X_1)^{-1}X_1'.\end{aligned} \tag{171}$$

We can easily confirm that these formulae do specialise to those under (170) 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

$$\begin{aligned}\tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y\\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon)\\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon.\end{aligned} \tag{172}$$

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

$$E(\tilde{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2, \tag{173}$$

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 of the regression parameters does not arise.

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

The standard test of an hypothesis regarding the vector β in the model $N(y;X\beta,\sigma^2I)$ 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

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

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

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

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

CHARACTERISTIC ROOTS AND VECTORS OF A SYMMETRIC MATRIX

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

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

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

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

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

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

The Diagonalisation of a Symmetric Matrix

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

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

which have the property that

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

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

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

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

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

SEEMINGLY-UNRELATED REGRESSION EQUATIONS

The Algebra of the Kronecker Product. Consider the matrix equation $Y = AXB'$ where

$$
\begin{aligned}
Y &= [y_{kl}]; k := 1, \ldots, r, l = 1, \ldots, s, \\
X &= [x_{ij}]; i := 1, \ldots, m, j = 1, \ldots, n, \\
A &= [a_{ki}]; k := 1, \ldots, r, i = 1, \ldots, m, \\
B &= [b_{lj}]; l := 1, \ldots, s, j = 1, \ldots, n.
\end{aligned} \tag{1}
$$

The object is to reformulate this matrix equation so that it can be treated as an ordinary vector equation. Amongst the advantages which this will entail is the possibility of solving the equation by the methods which are commonly applied in finding the solutions to vector equations.

Therefore consider writing $Y = AXB'$ more explicitly as

$$
\begin{aligned}
[y_{.1}, y_{.2}, \ldots, y_{.s}] &= A[x_{.1}, x_{.2}, \ldots, x_{.n}][b'_{1.}, b'_{2.}, \ldots, b'_{s.}] \\
&= [Ax_{.1}, Ax_{.2}, \ldots, Ax_{.n}][b'_{1.}, b'_{2.}, \ldots, b_{s.}]
\end{aligned}. \tag{2}
$$

In this notation, the expression $x_{.j}$ stands for the jth column of the matrix X whilst the notation $b_{l.}$ stands for the lth *row* of B. Therefore the transposed vector $b'_{l.} = [b_{l1}, b_{l2}, \ldots, b_{ln}]'$ is a *column* vector of n elements—as it must be if the multiplication of the two expressions on the RHS of (2) is to be properly defined. By performing that multiplication, we find that

$$
\begin{aligned}
[y_{.1}, y_{.2}, \ldots, y_{.s}] = \Big[&\{b_{11}Ax_{.1} + b_{12}Ax_{.2} + \cdots + b_{1n}Ax_{.n}\}, \\
&\{b_{21}Ax_{.1} + b_{22}Ax_{.2} + \cdots + b_{2n}Ax_{.n}\}, \ldots, \\
&\{b_{s1}Ax_{.1} + b_{s2}Ax_{.2} + \cdots + b_{sn}Ax_{.n}\}\Big]
\end{aligned} \tag{3}
$$

Here, each of the expressions on the the RHS within braces $\{,\}$ stands for one of the vectors $y_{.1}, y_{.2}, \ldots, y_{.s}$ on the LHS. These LHS vectors may be stacked vertically one below the other to form long vectors. When the RHS of the equation is rearranged likewise, a system is derived which takes the form of

$$
\begin{bmatrix} y_{.1} \\ y_{.2} \\ \vdots \\ y_{.s} \end{bmatrix} = \begin{bmatrix} b_{11}A & b_{12}A & \ldots & b_{1n}A \\ b_{21}A & b_{22}A & \ldots & b_{2n}A \\ \vdots & & & \vdots \\ b_{s1}A & b_{s2}A & \ldots & b_{sn}A \end{bmatrix} \begin{bmatrix} x_{.1} \\ x_{.2} \\ \vdots \\ x_{.n} \end{bmatrix}. \tag{4}
$$

The system can be written is a summary notation as

$$
Y^c = (AXB')^c = (B \otimes A)X^c. \tag{5}
$$

Here the long vectors Y^c and X^c are derived simply by slicing the matrices and rearranging the columns in the manner which we have described. The matrix $B \otimes A = [b_{lj}A]$, whose (lj)th partition contains the matrix $b_{lj}A$, is described as the Kronecker product of B and A.

The following rules govern the use of the Kronecker product:

$$
\begin{aligned}
&\text{(i)} \quad (A \otimes B)(C \otimes D) = AC \otimes BD, \\
&\text{(ii)} \quad A \otimes (B + C) = (A \otimes B) + (A \otimes B), \\
&\text{(iii)} \quad \lambda(A \otimes B) = \lambda A \otimes B = A \otimes \lambda B, \\
&\text{(iv)} \quad (A \otimes B)^{-1} = (A^{-1} \otimes B^{-1}).
\end{aligned} \tag{6}
$$

The Kronecker product is non-commutative, which is to say that $A \otimes B \neq B \otimes A$. However, observe that

$$
A \otimes B = (A \otimes I)(I \otimes B) = (I \otimes B)(A \otimes I). \tag{7}
$$

Systems with Multiple Outputs. The typical regression equation describes a system which transforms k observable inputs and a stochastic disturbance into a single output. We now wish to consider a system which produces M outputs. Consider, therefore, the equations

$$(8)\qquad \begin{aligned}[y_{t1}, y_{t2}, \cdots, y_{tM}] &= [x_{t.}\beta_{.1}, x_{t.}\beta_{.2}, \cdots, x_{t.}\beta_{.M}] + [\varepsilon_{t1}, \varepsilon_{t2}, \cdots, \varepsilon_{tM}]\\ &= x_{t.}[\beta_{.1}, \beta_{.2}, \cdots, \beta_{.M}] + [\varepsilon_{t1}, \varepsilon_{t2}, \cdots, \varepsilon_{tM}].\end{aligned}$$

Here the generic equation is

$$(9)\qquad y_{tm} = x_{t.}\beta_{.m} + \varepsilon_{tm};$$

and this has the form of a single regression equation. In a notation which mixes matrices and vectors, the system under (8) may be written as

$$(10)\qquad y_{t.} = x_{t.}B + \varepsilon_{tm},$$

where $B = [\beta_{.1}, \beta_{.2}, \cdots, \beta_{.M}]$, and T realisations of the latter may be compiled to give the equation

$$(11)\qquad Y = XB + \mathcal{E},$$

or

$$(12)\qquad [y_{.1}, y_{.2}, \cdots, y_{.M}] = [x_{.1}, x_{.2}, \cdots, x_{.k}]B + [\varepsilon_{.1}, \varepsilon_{.2}, \cdots, \varepsilon_{.M}].$$

When the latter equation is vectorised, we have

$$(13)\qquad Y^c = (XBI)^c + \mathcal{E}^c = (I \otimes X)B^c + \mathcal{E}^c,$$

which can be written more explicitly as

$$(14)\qquad \begin{bmatrix} y_{.1}\\ y_{.2}\\ \vdots\\ y_{.s}\end{bmatrix} = \begin{bmatrix} X & 0 & \dots & 0\\ 0 & X & \dots & 0\\ \vdots & & & \vdots\\ 0 & 0 & \dots & X\end{bmatrix}\begin{bmatrix}\beta_{.1}\\ \beta_{.2}\\ \vdots\\ \beta_{.M}\end{bmatrix} + \begin{bmatrix}\varepsilon_{.1}\\ \varepsilon_{.2}\\ \vdots\\ \varepsilon_{.M}\end{bmatrix} = \begin{bmatrix} X\beta_{.1}\\ X\beta_{.2}\\ \vdots\\ X\beta_{.M}\end{bmatrix} + \begin{bmatrix}\varepsilon_{.1}\\ \varepsilon_{.2}\\ \vdots\\ \varepsilon_{.M}\end{bmatrix}.$$

Some assumptions must now be made regarding the disturbance terms of the model. We shall assume that, taken separately, the M equations $y_{tm}x_{t.}\beta_{.m} + \varepsilon_{tm}; t = 1, \ldots T, m = 1, \ldots, M$ have the stochastic structure of the classical linear model; which is to say that the disturbances are independently and identically distributed with an expected value of zero and a common variance. However, we shall assume that the M contemporaneous disturbances in the vector $\varepsilon_{t.} = [\varepsilon_{t1}, \ldots, \varepsilon_{tM}]$ have nonzero covariances such that

$$(15)\qquad D(\varepsilon_{t.}) = E(\varepsilon'_{t.}\varepsilon_{t.}) = \Sigma = [\sigma_{ml}] \quad \text{for all} \quad t.$$

Thus, if $\varepsilon.m$ and $\varepsilon.l$ are vectors of T disturbances from the equations $y_{.m} = X\beta_{.m} + \varepsilon_{.m}$ and $y_{.l} = X\beta_{.l} + \varepsilon_{.l}$ respectively, then we should have

$$\begin{aligned} &E(\varepsilon_{.m}) = E(\varepsilon_{.l}) = 0 \qquad \text{and} \\ &D(\varepsilon_{.m}) = \sigma_{mm} I_T, \quad D(\varepsilon_{.l}) = \sigma_{ll} I_T, \\ &C(\varepsilon_{.m}, \varepsilon_{.l}) = \sigma_{ml} I_T, \end{aligned} \tag{16}$$

where $C(\varepsilon_{.m}, \varepsilon_{.l}) = E(\varepsilon_{.m}\varepsilon'_{.l})$ is the covariance matrix of the two vectors. Putting these assumptions together, we get

$$E(\mathcal{E}^c) = 0 \qquad \text{and} \qquad D(\mathcal{E}^c) = E(\mathcal{E}^c \mathcal{E}^{c\prime}) = \Sigma \otimes I_T. \tag{17}$$

It may be appropriate to write these in a manner which makes them more explicit. First there is the assumption concerning the expected value of the long vector of disturbances. Writing this vector in transposed form gives

$$E(\mathcal{E}^{c\prime}) = E[\varepsilon'_{.1}, \varepsilon'_{.2}, \ldots, \varepsilon'_{.M}] = [\ 0,\ 0, \ldots, 0\ ]. \tag{18}$$

The assumptions concerning the dispersion matrix of this vector can be written as

$$\begin{aligned} D\begin{bmatrix} \varepsilon_{.1} \\ \varepsilon_{.2} \\ \vdots \\ \varepsilon_{.M} \end{bmatrix} &= E\begin{bmatrix} \varepsilon_{.1}\varepsilon'_{.1} & \varepsilon_{.1}\varepsilon'_{.2} & \ldots & \varepsilon_{.1}\varepsilon'_{.M} \\ \varepsilon_{.2}\varepsilon'_{.1} & \varepsilon_{.2}\varepsilon'_{.2} & \ldots & \varepsilon_{.2}\varepsilon'_{.M} \\ \vdots & & & \vdots \\ \varepsilon_{.M}\varepsilon'_{.1} & \varepsilon_{.M}\varepsilon'_{.2} & \ldots & \varepsilon_{.M}\varepsilon'_{.M} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \ldots & \sigma_{1M}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \ldots & \sigma_{2M}I_T \\ \vdots & & & \vdots \\ \sigma_{M1}I_T & \sigma_{M2}I_T & \ldots & \sigma_{MM}I_T \end{bmatrix}. \end{aligned} \tag{19}$$

It is common to denote the regression model $y = X\beta + \varepsilon$ in which $E(\varepsilon) = 0$ and $E(\varepsilon) = \sigma^2 Q$ by the triplet $(y; X\beta, \sigma^2 Q)$. Using the same notation, we may now denote the vectorised version of the model with M outputs as $(Y^c, (I \otimes X)B^c, \Sigma \otimes I)$. It is apparent that the two models are isomorphic, which is to say that they share the same structure. Therefore it is possible to estimate the parameters of the M-output model, once it has been cast in the appropriate form, by using methods which have been developed in the context of a single-equation model.

The approrpriate method is generalised least-squares regression. When it is applied to the model $(y; X\beta, \sigma^2 Q)$, this method delivers the estimate $\hat{\beta} =$

$(X'Q^{-1}X)^{-1}X'Q^{-1}y$. When it is applied to the M-equation model the method delivers the estimate

$$\hat{B} = \Big\{(I \otimes X)'(\Sigma \otimes I)^{-1}(I \otimes X)\Big\}^{-1}(I \otimes X)'(\Sigma \otimes I)^{-1}Y^c. \tag{20}$$

The algebraic rules under (6) can now be invoked to simplify this result. It can be see that

$$\begin{aligned}\hat{B} &= (\Sigma^{-1} \otimes X'X)^{-1}(\Sigma \otimes X')^{-1}Y^c\\ &= \big\{I \otimes (X'X)^{-1}X'\big\}Y^c\\ &= \big\{(X'X)^{-1}X'Y\big\}^c.\end{aligned} \tag{21}$$

Thus it transpires that the efficient system-wide estimator amounts to nothing more than the repeated application of the ordinary least-squares procedure to generate the regression estimates $\hat{\beta}_{.m} = (X'X)^{-1}X'Y_{.m}; m = 1, \ldots, M$.

We can use the residual vectors $e_{.m} = y_{.m} - X\hat{\beta}_{.m}$ from these M estimations to derive estimates of the elements of $\Sigma = [\sigma_{ml}]$. Thus an unbiased estimator of σ_{ml} is

$$\begin{aligned}\hat{\sigma}_{ml} &= \frac{e'_{.m}e_{.l}}{T-k} = \frac{(y_{.m} - X\hat{\beta}_{.m})'(y_{.l} - X\hat{\beta}_{.l})}{T-k}\\ &= \frac{y'_{.m}\{I - X(X'X)^{-1}X'\}y_{.l}}{T-k}.\end{aligned} \tag{22}$$

The reduction of the system-wide estimator to an M-fold application of ordinary least-square regression occurs only when all the variables in X are present in each of the M equations and when no other variables are present in any of them. If some of the variables are missing, or if we have *a priori* information relating to the parameter vectors $\beta_{.m}; m = 1, \ldots, M$, then, to obtain efficient estimates, we must use the available information on Σ. For example, let X_m be the submatrix containing only those variables which are present in the mth equation. Then the system of equations assumes the following form:

$$\begin{bmatrix} y_{.1}\\ y_{.2}\\ \vdots\\ y_{.s}\end{bmatrix} = \begin{bmatrix} X_1 & 0 & \ldots & 0\\ 0 & X_2 & \ldots & 0\\ \vdots & & & \vdots\\ 0 & 0 & \ldots & X_M\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_M\end{bmatrix} + \begin{bmatrix}\varepsilon_{.1}\\ \varepsilon_{.2}\\ \vdots\\ \varepsilon_{.M}\end{bmatrix}. \tag{23}$$

This can be written in summary notation as

$$Y^c = W\delta + \mathcal{E}^c. \tag{24}$$

Now the block-diagonal or "staircase" matrix X no longer has the structure of a Kronecker product. Nor can the subvectors of $\delta' = [\beta_1', \beta_2, \ldots, \beta_M']$ be stacked together in a matrix B, for the reason that they are liable to be of different lengths. The efficient generalised least-squares estimator of the parameters now takes the form of

$$\hat{\delta} = \left\{W'(\Sigma^{-1} \otimes I)\right\}^{-1} W'(\Sigma^{-1} \otimes I)Y^c; \tag{25}$$

and there in no longer any possibility of simplifying or reducing the expression.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

We rewrite this equation as

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

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

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

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

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

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

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

DYNAMIC REGRESSION MODELS: THE GEOMETRIC LAG SCHEME

Anr 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:

$$y(t) = \beta\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} + \varepsilon(t). \tag{256}$$

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

$$\gamma = \frac{\beta}{1-\phi}. \tag{257}$$

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

$$\phi y(t-1) = \beta\big\{\phi x(t-1) + \phi^2 x(t-2) + \phi^3 x(t-3) + \cdots\big\} + \phi\varepsilon(t-1). \tag{258}$$

Taking the latter from (256) gives

$$y(t) - \phi y(t-1) = \beta x(t) + \big\{\varepsilon(t) - \phi\varepsilon(t-1)\big\}. \tag{259}$$

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

$$(1-\phi L)y(t) = \beta x(t) + (1-\phi L)\varepsilon(t), \tag{260}$$

of which the rational form is

$$y(t) = \frac{\beta}{1 - \phi L} x(t) + \varepsilon(t). \tag{261}$$

In fact, by using the expansion

$$\begin{aligned} \frac{\beta}{1 - \phi L} x(t) &= \beta\big\{1 + \phi L + \phi L^2 + \cdots\big\} x(t) \\ &= \beta\big\{x(t) + \phi x(t-1) + \phi x(t-2) + \cdots\big\} \end{aligned} \tag{262}$$

within equation (261), we can recover the original form under (256).

Equation (259) 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 (256). In view of that equation, we may express the elements of $y(t)$ which fall within the sample as

$$\begin{aligned} y_t &= \beta \sum_{i=0}^{\infty} \phi^i x_{t-i} + \varepsilon_t \\ &= \theta\phi^t + \beta \sum_{i=0}^{t-1} \phi^i x_{t-i} + \varepsilon_t \\ &= \theta\phi^t + \beta z_t + \varepsilon_t. \end{aligned} \tag{263}$$

Here

$$\theta = \beta\big\{x_0 + \phi x_{-1} + \phi^2 x_{-2} + \cdots\big\} \tag{264}$$

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

$$z_t = x_t + \phi x_{t-1} + \cdots + \phi^{t-1} x_1 \tag{265}$$

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 (263) 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

$$y(t) = \frac{\beta}{1 - \phi L} x(t) + \frac{1}{1 - \rho L} \varepsilon(t), \tag{266}$$

which has a first-order autoregressive disturbance scheme in place of the white-noise disturbance to be found in equation (261). 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

$$y(t) = \phi y(t-1) + \beta x(t) + \varepsilon(t). \tag{267}$$

In terms of the lag operator, this is

$$(1 - \phi L) y(t) = \beta x(t) + \varepsilon(t), \tag{268}$$

of which the rational form is

$$y(t) = \frac{\beta}{1 - \phi L} x(t) + \frac{1}{1 - \phi L} \varepsilon(t). \tag{269}$$

The advantage of equation (267) 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 (262) to the rational form of the present model given in equation (269) and by comparing the result with (256). The comparison of equation (269) with the corresponding rational equation (261) for the Koyk model shows that we now have an AR(1) disturbance process described by the equation

$$\eta(t) = \phi\eta(t-1) + \varepsilon(t) \tag{270}$$

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 (266) with ρ and ϕ taking very different values. Imagine that, nevertheless, it is decided to fit the equation under (269). 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 (266). 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

$$y(t) = \lambda\{\gamma x(t)\} + (1-\lambda)y(t-1) + \varepsilon(t), \tag{271}$$

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 (271) becomes identical to equation (267) 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

$$y(t) = \delta x^*(t) + \varepsilon(t), \tag{272}$$

where

$$\begin{aligned} x^*(t) &= (1-\phi)\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} \\ &= \frac{1-\phi}{1-\phi L} x(t) \end{aligned} \tag{273}$$

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 (273) by $1 - \phi L$ and rearranging the result. Thus

$$x^*(t) - x^*(t-1) = (1-\phi)\big\{x(t) - x^*(t-1)\big\}, \tag{274}$$

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 (273) into the equation (272) of the consumption function, we get

$$y(t) = \delta \frac{(1-\phi)}{1-\phi L} x(t) + \varepsilon(t). \tag{275}$$

When the notation $\delta(1 - \phi) = \beta$ is adopted, equation (275) becomes identical to the equation (261) of the Koyk model.