15 : CHAPTER

Temporal Regression Models in Econometrics

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

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

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:

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

We rewrite this equation as

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

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 $\alpha(z) = 1-\phi_1 z-\phi_2 z^2 = 0$ 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.

Models with Lagged-Dependent Variables

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

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

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

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

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

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

and the second is the equation

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

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

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

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

Combining the latter gives

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Error-Correction Forms, and Nonstationary Signals

Many econometric data sequences are nonstationary, with evident trends that persist for long periods. However, the usual linear regression procedures, which might be used in estimating the relationships existing amongst the data, presuppose that the relevant moment matrices will converge asymptotically to fixed limits as the sample size increases. This condition will not be fulfilled if the data are strongly trended, in which case, the standard techniques of statistical inference will not be applicable.

In order to apply the regression procedures successfully, it is necessary to find some means of reducing the data to stationarity. A common approach is to subject the data to as many differencing operations as may be required to achieve stationarity. Often, only a single differencing is required.

An objection that can be raised against the recourse to differencing is that it tends to remove, or at least to attenuate severely, some of the essential information regarding the behaviour of economic agents. There are processes of equilibration, which are evident in the original data, by which the relative proportions of econometric variables are maintained over long periods of time. The evidence will be lost in the process of differencing the data.

When the original undifferenced data sequences share a common trend, the coefficient of determination in a fitted regression is liable to be high; but it is often discovered that the regression model looses much of its explanatory power when the differences of the data are used instead.

A formulation which can sometimes be used to good effect in such circumstances is the so-called error-correction model. The form of error correction model that we shall examine is a re-parametrised version of an ordinary autoregressive–distributed lag model. One of the features of the model is that it depicts a mechanism whereby two trended economic variables maintain an enduring long-term proportionality with each other. Moreover, the data sequences comprised by the model are stationary, either individually or in an appropriate combination; and this feature enables us apply the standard procedures of statistical inference that are available to models comprising data from stationary processes.

In this section, we shall begin by deriving the error-correction formulation which corresponds to a simple first-order dynamic model. Thereafter, we shall consider models of higher orders.

Consider taking $y(t-1)$ from both sides of the equation under (1) which represents the first-order dynamic model. This gives

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

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

The error-correction form provides the basis for estimating the parameters of the model when the signal series $x(t)$ is trended or nonstationary. A pair of nonstationary series which maintain a long-run proportionality are said to be cointegrated.

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 $\{\gamma x(t) - y(t-1)\}$. However, if the error-correction model is an unrestricted reparametrisation of an original model in levels, as it will be in the majority of the cases that we shall be considering here, then its parameters can be estimated by ordinary least-squares regression. The same estimates can also be inferred from the least-squares estimates of the parameters of the original model in levels.

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

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

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

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

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

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

The first two terms on the RHS combine to give

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

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

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

In the case of a nonstationary signal $x(t)$, this is amenable to precisely the same principle of estimation as was the simpler first-order equation under (27). That is to say, we can begin by estimating the gain γ by a simple regression of $y(t-1)$ on $x(t)$. Then, when a value for γ is available, we can proceed to find the remaining parameters of the model via a second regression. Alternatively, the parameters of the error-correction model can be estimated directly by ordinary least-squares regression, or they can be inferred from the estimated parameters of a model in levels that has been fitted by least-squares regression.

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

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

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

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

and

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

The two matrices which affect the transformation upon the variables and upon their associated parameters stand in an inverse relationship to one another.

Example. The reparameterisation of a dynamic system can be achieved in a variety of ways which lead to alternative expressions. A reparametrisation which is common, and which is somewhat less straighforward than that of the previous example, can be illustrated in the context of the equation

$$\sum_{j=0}^{3}\alpha_j y(t-j) = \sum_{j=0}^{3}\beta_j y(t-j) + \varepsilon(t). \tag{36}$$

On the RHS, the variables are transformed as follows:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1) \\ x(t-2) \\ x(t-3) \end{bmatrix} = \begin{bmatrix} x(t-1) \\ -\nabla x(t) \\ -\nabla x(t-1) \\ -\nabla x(t-2) \end{bmatrix}. \tag{37}$$

The corresponding transformation of the parameters is given by

$$[\,\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3\,] \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix} = [\,\kappa \quad -\delta_0 \quad -\delta_1 \quad -\delta_2\,]. \tag{38}$$

On the LHS, the transformed variables are given by

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} y(t) \\ y(t-1) \\ y(t-2) \\ y(t-3) \end{bmatrix} = \begin{bmatrix} y(t) \\ y(t-1) \\ -\nabla y(t-1) \\ -\nabla y(t-2) \end{bmatrix}, \tag{39}$$

and the corresponding parameters by

$$[\,1 \quad \alpha_1 \quad \alpha_2 \quad \alpha_3\,] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} = [\,1 \quad -\theta \quad \rho_1 \quad \rho_2\,]. \tag{40}$$

Thus the equation becomes

$$\begin{aligned} y(t) = \theta y(t-1) + \kappa x(t-1) + \rho_1\nabla y(t-1) + \rho_2\nabla y(t-2) \\ + \delta_0\nabla x(t) + \delta_1\nabla x(t-1) + \delta_2\nabla x(t-2) + \varepsilon(t). \end{aligned} \tag{41}$$

Taking $y(t-1)$ from both sides of the latter gives an equation with $\nabla y(t)$ on the LHS and an error-correction term in the form of $\lambda\{\gamma x(t-1) - y(t-1)\}$ on the RHS, where $\lambda = 1-\theta$ and $\gamma = \kappa/(1-\theta)$ as before.

Example. In many models, the natural logarithms of the economic variables are used in the regression equation. Consider the case where the dependent variable Y maintains a long-term proportionality with the explanatory variable X. For example, Y and X might be consumption and income respectively. In that case, the equilibrium condition which characterises a stationary state or an equilibrium growth path is $Y = KX$. Letting $x = \ln X$, $y = \ln Y$ and $k = \ln K$ gives $y = k + x$; and, on an equilibrium path, the proportional rate of growth of the two variables will be $\nabla y = \nabla x = r$.

Our purpose is to reconcile these equilibrium conditions with a dynamic regression equation in the form of

$$y(t) = \mu + \phi_1 y(t-1) + \beta_0 x(t) + \beta_1 x(t-1) + \varepsilon(t). \tag{42}$$

This is rendered in error-correction form by taking $y(t-1)$ from both sides of the equation and supplementing the RHS by $\pm\beta_0 x(t-1)$. The result is

$$\begin{aligned} \nabla y(t) &= \mu + (\phi_1 - 1)y(t-1) + \beta_0 \nabla x(t) + (\beta_1 + \beta_0)x(t-1) + \varepsilon(t) \\ &= \mu + (1-\phi_1)\left\{\frac{\beta_1+\beta_0}{1-\phi_1}x(t-1) - y(t-1)\right\} + \beta_0\nabla x(t) + \varepsilon(t). \end{aligned} \tag{43}$$

Let the rate of growth be r so that $Y_t = Y_{t-1}e^r$ and $X_t = X_{t-1}e^r$, which give $y_t = y_{t-1} + r$ and $x_t = x_{t-1} + r$, respectively. Putting these conditions in (42), eliminating the disturbance term and suppressing the temporal indices gives

$$y + r = \mu + \phi_1 y + \beta_0(x + r) + \beta_1 x, \tag{44}$$

from which

$$y = \frac{\mu + (\beta_0 - 1)r}{1-\phi_1} + \frac{\beta_0 + \beta_1}{1-\phi_1}x. \tag{45}$$

To reconcile this with the equation $y = k + x$ which characterises the growth path, we must impose the condition that $(\beta_0+\beta_1)/(1-\phi_1) = 1$ or, equivalently, that $\beta_0 + \beta_1 + \phi_1 = 1$. Notice that $k = \{\mu + (\beta_0 - 1)r\}/(1-\phi_1)$ is dependent on the growth rate r.

One might doubt whether it is reasonable to postulate an equilibrium growth rate that prevails over the entire sample period. However, if such a postulate is accepted, then it becomes appropriate to fit an equation the form of

$$\nabla y(t) = \mu + \lambda\{x(t-1) - y(t-1)\} + \beta_0 \nabla x(t) + \varepsilon(t), \tag{46}$$

which is the resulting specialisation of equation (43). This equation and the foregoing analysis were the basis of an influential article on consumer's expenditure in the U.K. by Davidson *et al.*

To understand the dynamic implications of the equation, let us set $\mu = 0$. Then, in a steady state, with a growth rate of r and with $\varepsilon(t) = 0$ for all t, we should have $\nabla y(t) = r$, $\beta_0 \nabla x(t) = \beta_0 r$ and $\lambda \{x(t-1) - y(t-1)\} = -\lambda\kappa = (1-\beta_0)r$. It follows that, the faster the growth rate, the wider is the gap between income and consumption. In the absense of an intercept term μ, the gap would disappear altogether at a zero rate of growth. It seems that, in order to avoid this implication, an intercept term should be present in the equation.

Cointegration in a Multivariate System

In the previous section, we have imagined that the sequences $x(t)$ and $y(t)$ are both generated by nonstationary stochastic processes, and we have supposed that these processes are coupled or "cointegrated" in the sense that their values maintain a long-run proportionality. In particular, we have implied that the processes are to be found within a system of the form

$$\begin{bmatrix} \alpha(L) & -\beta(L) \\ 0 & \omega(L) \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}, \tag{47}$$

where $\omega(L)$ is a nonstationary autoregressive operator with one or more roots of unit value. If the disturbance sequences $\varepsilon(t)$ and $\eta(t)$ are statistically independent, then we will be justified in treating $x(t)$ as a variable which is exogenous with respect to the processes generating $y(t)$.

However, it may be more meaningful, sometimes, to regard $x(t)$ and $y(t)$ as pair of mutually dependent variables which enter into several statistical relationships which might comprise other variables as well. Let us therefore imagine that $x(t)$ and $y(t)$ are elements of a vector sequence $z(t)$, which is not constrained to be stationary, and let us consider an equation in the form of

$$\Pi(L)z(t) = z(t) + \Pi_1 z(t-1) + \cdots + \Pi_n z(t-n) = \zeta(t), \tag{48}$$

which purports to describe how $z(t)$ is generated. In this equation, the individual processes within the disturbance vector $\zeta(t)$ on the RHS *are* presumed to be stationary and, therefore, the combination $\Pi(L)z(t)$ of the LHS must be be stationary likewise.

There are a variety of ways in which the stationarity of the LHS can arise. It may indeed be attributable to the stationary of each of the elements of $z(t)$. Alternatively, it may be that the operator $\Pi(L)$ is effective in taking differences of the nonstationary elements of $z(t)$. However, it is also possible for the stationarity to result from the combination of cointegrated nonstationary processes which follow common trends.

In order to demonstrate this third possibility, let us, for convenience, make the assumption that $n = 2$. Then equation (48) can be written as

$$z(t) + \Pi_1 z(t-1) + \Pi_2 z(t-2) = \zeta(t); \tag{49}$$

and, on applying the transformation described in the previous section, this becomes

$$
\begin{aligned}
&[\,I \quad \Pi_1 \quad \Pi_2\,]\begin{bmatrix} I & 0 & 0 \\ I & I & 0 \\ I & I & I \end{bmatrix}\begin{bmatrix} I & 0 & 0 \\ -I & I & 0 \\ 0 & -I & I \end{bmatrix}\begin{bmatrix} z(t) \\ z(t-1) \\ z(t-2) \end{bmatrix} \\
&= [\,\Gamma \quad B_1 \quad B_2\,]\begin{bmatrix} z(t) \\ \nabla z(t) \\ \nabla z(t-1) \end{bmatrix} = \zeta(t),
\end{aligned} \tag{50}
$$

where

$$
\Gamma = I + \Pi_1 + \Pi_2, \quad -B_1 = \Pi_1 + \Pi_2, \quad \text{and} \quad -B_2 = \Pi_2. \tag{51}
$$

Thus, in place of (49), we have an equivalent equation

$$
\Gamma z(t) + B_1 \nabla z(t) + B_2 \nabla z(t-1) = \zeta(t). \tag{52}
$$

Equation (52) contains a mixture of differenced and undifferenced variables. We imagine that the differencing is sufficient to reduce the variables to stationarity. Therefore, if the model is to be consistent, the term $\Gamma z(t)$ must also be stationary. This will be impossible $z(t)$ is nonstationary and if Γ has full rank. It will only be possible if there are one or more cointegrating relationships between the variables such that there can be found linear combinations, embedded within $\Gamma z(t)$, which render the variables stationary.

A cointegrating relationship represents a restriction on the variables of the sytem which asserts that, in the long run, they will tend to maintain a certain proportionality. The greater the number of cointergrating relationships, the more closely are these prorportions governed. In the limiting, case where the number of relationship is one less that the number of variables, every ratio amongst the variables is governed.

The number of cointegrating relationships is equal to the rank of Γ. If the matrix Γ is of full rank, then every arbitrary combination of the sequences must be stationary; which means that each of the sequences must be stationary. Then there will be no call for differencing. On the other hand, if Γ is null, with a rank of zero, then there will be no cointegrating relationships, and each sequence will be following its own independent random walk, which will be present in the equation only in its stationary differenced form.

Example. For the simplest example of these relationships, we may consider the equations

$$
\begin{aligned}
&y(t) = \beta x(t) + \phi y(t-1) + \varepsilon(t) \quad \text{and} \\
&x(t) = x(t-1) + \eta(t),
\end{aligned} \tag{53}
$$

which can be written together as

$$(54)\qquad \begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} - \begin{bmatrix} \phi & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y(t-1) \\ x(t-1) \end{bmatrix} = \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix},$$

or as

$$(55)\qquad \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} - \begin{bmatrix} \phi & \beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y(t-1) \\ x(t-1) \end{bmatrix} = \begin{bmatrix} \varepsilon(t) + \beta\eta(t) \\ \eta(t) \end{bmatrix},$$

which comers from setting $x(t) = x(t-1) + \eta(t)$ in the first equation. Then

$$(56)\qquad \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \phi & \beta \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-\phi & -\beta \\ 0 & 0 \end{bmatrix}$$

is clearly a matrix of rank one as it is required to be. Also, it is clear that

$$(57)\qquad \begin{bmatrix} 1-\phi & -\beta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} (1-\phi)x(y) - \beta(y(t) \\ 0 \end{bmatrix}$$

comprises the disequilibrium error from equation (27), which is a stationary random variable.