DYNAMIC REGRESSIONS MODELS

Autoregressive Disturbance Processes

Economic variables often follow slowly-evolving trends and they tend to be strongly correlated with each other. If the disturbance term is compounded from such variables, then it should have similar characteristics.

In the classical model, the disturbances $\varepsilon(t) = \{\varepsilon_t; t = 0, \pm 1, \pm 2, \ldots\}$ are independently and identically distributed such that

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

A sequence of 50 observations generated by such a process is plotted in Figure 1. The sequence is of a highly volatile nature; and its past values are of no use in predicting its future values.

Our task is to find models for the disturbance process that are more in accordance with economic circumstances.

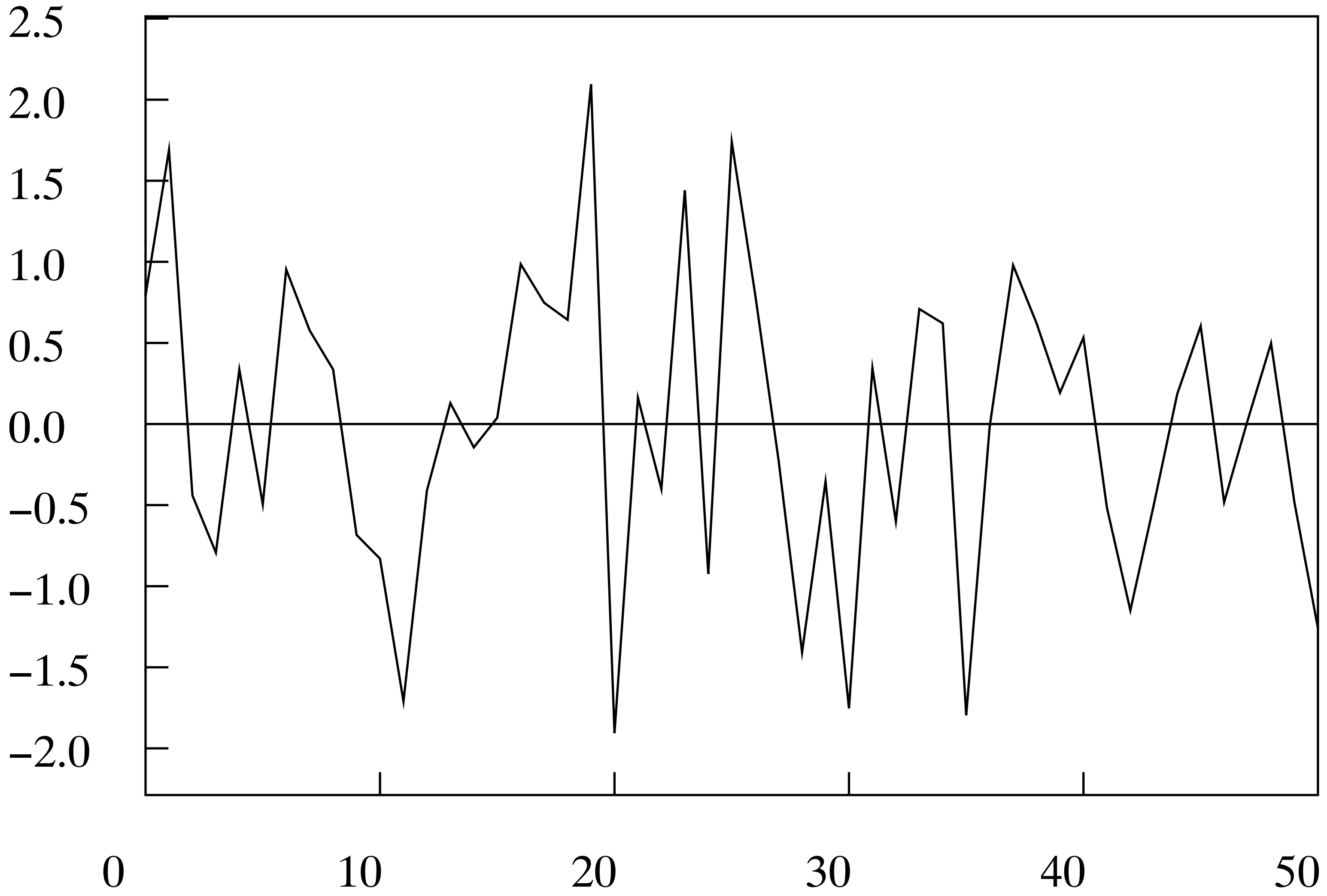

Figure 1. 50 observations on a white-noise process $\varepsilon(t)$ of unit variance.

The traditional means of representing the inertial properties of the disturbance process has been to adopt a simple first-order autoregressive model, or AR(1) model, whose equation takes the form of

$$(3) \qquad \eta_t = \phi\eta_{t-1} + \varepsilon_t, \qquad \text{where} \qquad \phi \in (-1, 1).$$

In many econometric applications, the value of ϕ falls in the more restricted interval $[0, 1)$.

The conditional expectation of η_t given η_{t-1} is $E(\eta_t|\eta_{t-1}) = \phi\eta_{t-1}$. If ϕ is close to unity, there will be a high degree of correlation amongst successive elements of $\eta(t) = \{\eta_t; t = 0, \pm 1, \pm 2, \ldots\}$. Figure 2 shows 50 observations on an AR(1) process with $\phi = 0.9$

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

$$(4) \qquad C(\eta_{t-\tau}, \eta_t) = \gamma_\tau = \sigma^2 \frac{\phi^\tau}{1 - \phi^2}.$$

Their correlation declines as their temporal separation increases.

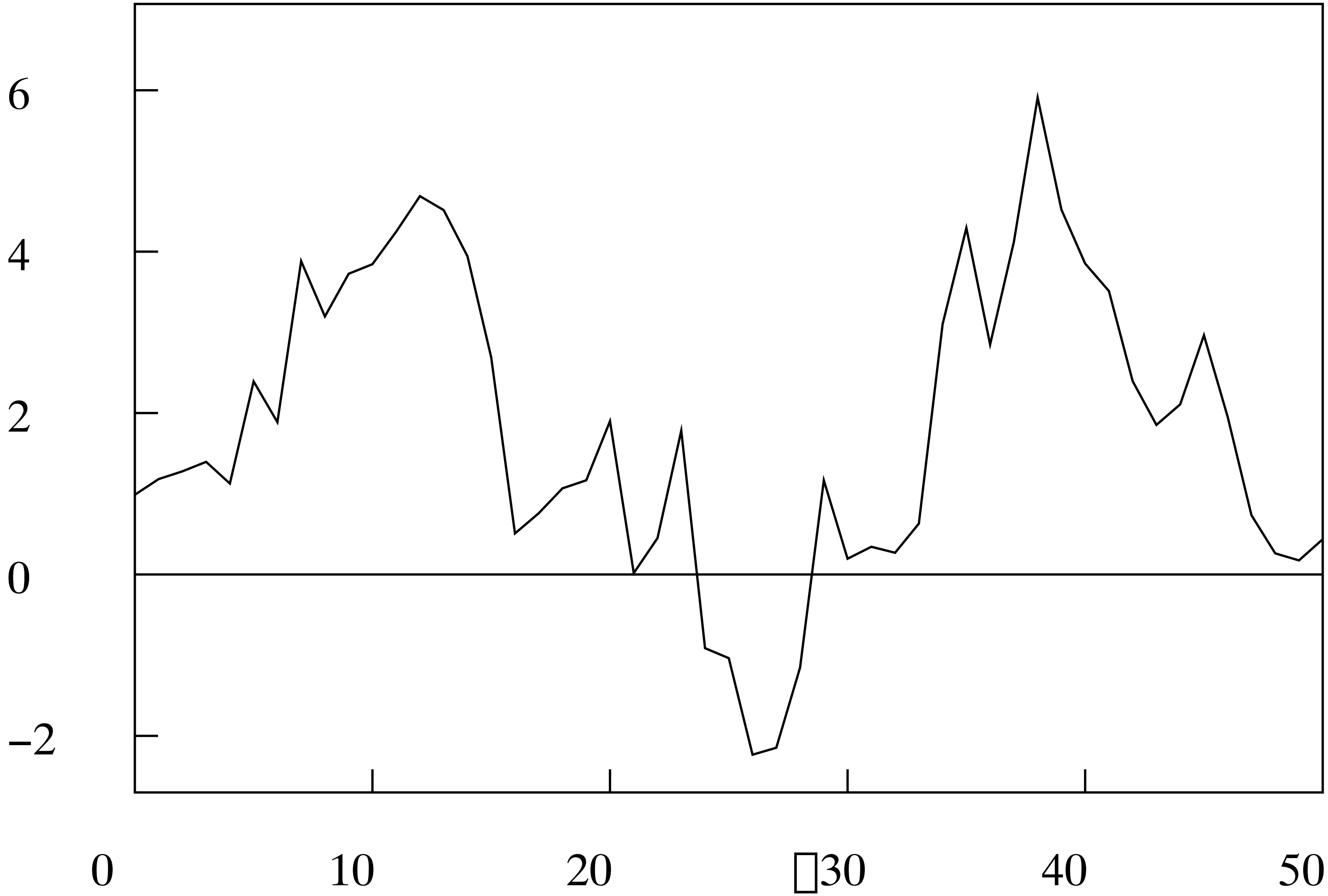

Figure 2. 50 observations on an AR(1) process $\eta(t) = 0.9\eta(t-1) + \varepsilon(t)$.

Detecting Serial Correlation in the Regression Disturbances

Imagine that the sequence $\eta(t) = \{\eta_t; t = 0, \pm 1, \pm 2, \ldots\}$ of the regression disturbances follows an AR(1) process such that

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t), \quad \text{with} \quad \rho \in [0,1). \tag{12}$$

If this could be observed directly, then the serial correlation could be detected by testing the significance of the estimate of ρ, which is

$$h = \frac{\sum_{t=2}^{T} \eta_{t-1}\eta_t}{\sum_{t=1}^{T} \eta_t^2}. \tag{13}$$

When residuals are used in the place of disturbances, this will be affected, in finite samples, by the values of the explanatory variables.

The traditional approach to testing for serial correlation is due to Durbin and Watson. They have attempted to make an explicit allowance for the uncertainties that arise from not knowing the precise distribution of the test statistic in any particular instance.

The test statistic of Durbin and Watson, which is based upon the sequence $\{e_t; t = 1, 2, \ldots, T\}$ of ordinary least-square residuals, is defined by

$$(14) \qquad d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}.$$

This statistic may be used for detecting any problem of misspecification indicated by seeming violations of the assumption of i.i.d disturbances.

Expanding the numerator of the D–W statistic, we find that

$$(15) \qquad d = \frac{1}{\sum_{t=1}^{T} e_t^2}\left\{\sum_{t=2}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1} + \sum_{t=2}^{T} e_{t-1}^2\right\} \simeq 2 - 2r,$$

where

$$(16) \qquad r = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$$

is an estimate of the coefficient ρ of serial correlation based on the ordinary least-squares residuals.

If ρ and likewise r are close to 1, then d will be close to zero; and there is a strong indication of serial correlation. If ρ is close to zero, so that the i.i.d assumption is more or less valid, then d will be close to 2.

Given that their statistic is based on residuals rather than on disturbances, Durbin and Watson provided a table of critical values that included a region of indecision. Their decision rules are as follows:

$$
\begin{aligned}
&\text{if} \quad d < d_L, \quad \text{then acknowledge the presence of serial correlation,}\\
&\text{if} \quad d_L \leq d \leq d_U, \quad \text{then remain undecided,}\\
&\text{if} \quad d_U < d, \quad \text{then deny the presence of serial correlation.}
\end{aligned}
$$

The values of d_L and d_U in the table depend on sample size and the number of variables included in the regression, not counting the intercept term.

As the number of degrees of freedom increases, the region of indecision lying between d_L and d_U becomes smaller, until a point is reached were it is no longer necessary to make any allowance for it.

Estimating a Regression Model with AR(1) Disturbances

Assume that the regression equation takes the form of

$$(18)\qquad y(t)=\alpha+\beta x(t)+\eta(t),\quad \text{with}\quad \eta(t)=\rho\eta(t-1)+\varepsilon(t).$$

Subtracting $\rho y(t-1)=\rho\alpha+\rho\beta x(t-1)+\rho\varepsilon(t-1)$, and using $Lx(t)=x(t-1)$, $Ly(t)=y(t-1)$ and $L\varepsilon(t)=\varepsilon(t-1)$, gives

$$(21)\qquad \begin{aligned}(1-\rho L)y(t)&=(1-\rho L)\alpha+(1-\rho L)\beta x(t)+\varepsilon(t)\\&=\mu+(1-\rho L)\beta x(t)+\varepsilon(t),\end{aligned}$$

where $\mu=(1-\rho)\alpha$. On defining the transformed variables,

$$(23)\qquad q(t)=(1-\rho L)y(t)\quad\text{and}\quad w(t)=(1-\rho L)x(t).$$

this can be written as

$$(22)\qquad q(t)=\mu+\beta w(t)+\varepsilon(t),$$

If a value is given to ρ, then we can form

$$
(25)\qquad
\begin{aligned}
q_2 &= y_2 - \rho y_1, & w_2 &= x_2 - \rho x_1,\\
q_3 &= y_3 - \rho y_2, & w_3 &= x_3 - \rho x_2,\\
&\vdots & &\vdots\\
q_T &= y_T - \rho y_{T-1}, & w_T &= x_T - \rho x_{T-1},
\end{aligned}
$$

and the equations

$$
(26)\qquad q_t = \mu + \beta w_t + u_t; \qquad t = 2, \ldots, T
$$

can be subjected to an ordinary least-squares regression. The regression can be repeated for various values of ρ; and the definitive estimates of ρ, $\alpha = \mu/(1-\alpha)$ and β are those corresponding to the minimum of the residual sum of squares.

The procedure of searching for the optimal value of ρ may be conducted in a systematic manner using a line-search algorithm such as the method of Fibonacci Search or the method of Golden-Section Search, which are described in textbooks of numerical optimisation.

In the Cochrane–Orcutt method, ρ is estimated via a subsidiary regression. The method is an iterative one in which each stage comprises two ordinary least-squares regressions.

Given an initial value for ρ, the parameters μ and β are determined from

$$
\begin{aligned}
y_t - \rho y_{t-1} &= \mu + \beta(x_t - \rho x_{t-1}) + \varepsilon_t \quad \text{or, equivalently,} \\
q_t &= \mu + \beta w_t + \varepsilon_t.
\end{aligned} \tag{27}
$$

Given values for β and $\alpha = \mu/(1-\rho)$, a revised value for ρ can be determined via a second regression applied to

$$
\begin{aligned}
(y_t - \alpha - \beta x_t) &= \rho(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t \quad \text{or, equivalently,} \\
\eta_t &= \rho\eta_{t-1} + \varepsilon_t.
\end{aligned} \tag{28}
$$

The revised value of ρ can fed back into equation (27), from which revised values of α and β can be obtained. The procedure can be pursued through successive iterations, until it converges.

The Feasible Generalised Least-Squares Estimator

The search procedure and the Cochrane–Orcutt procedure can be viewed within the context of the generalised least-squares (GLS) estimator that takes account of the dispersion matrix of the vector of disturbances.

The efficient GLS estimator of β in the model $(y; X\beta, \sigma^2 Q)$ is

$$\beta^* = (X'Q^{-1}X)^{-1}X'Q^{-1}y. \tag{31}$$

The dispersion matrix of the vector $\eta = [\eta_1, \eta_2, \eta_3, \ldots \eta_T]'$, generated by an AR(1) process is $[\gamma_{|i-j|}] = \sigma_\varepsilon^2 Q$, where

$$Q = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \ldots & \phi^{T-1} \\ \phi & 1 & \phi & \ldots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \ldots & \phi^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \ldots & 1 \end{bmatrix}. \tag{29}$$

It can be confirmed directly that

$$(30) \qquad Q^{-1} = \begin{bmatrix} 1 & -\phi & 0 & \ldots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & \ldots & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & \ldots & -\phi & 1 \end{bmatrix}.$$

Given its sparsity, the matrix Q^{-1} could be used directly in implementing the GLS estimator for which the formula is

$$(31) \qquad \beta^* = (X'Q^{-1}X)^{-1}X'Q^{-1}y.$$

By exploiting the factorisation $Q^{-1} = T'T$, the estimates can be obtained by applying OLS to the transformed data $W = TX$ and $g = Ty$. Thus, it can be seen that

$$(32) \qquad \begin{aligned} \beta^* = (W'W)^{-1}W'g &= (X'T'TX)^{-1}X'T'Ty \\ &= (X'Q^{-1}X)^{-1}X'Q^{-1}y. \end{aligned}$$

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

$$T = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \ldots & 0 \\ -\phi & 1 & 0 & \ldots & 0 \\ 0 & -\phi & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}. \tag{33}$$

This effects a simple transformation of the data. The element y_1 within $y = [y_1, y_2, y_3, \ldots, y_T]'$ is replaced $y_1\sqrt{1-\phi^2}$, whereas y_t is replaced by $y_t - \phi y_{t-1}$, for all $t > 1$.

Since ρ must be estimated by one or other of the methods that we have outlined above, the resulting estimator of β is apt to be described as a feasible GLS estimator. The true GLS estimator, which would require a precise knowledge of ρ, is an infeasible estimator.

Distributed Lags

In the early days of econometrics, attempts were made to model the dynamic responses primarily by including lagged values of x on the RHS of the regression equation; and the so-called distributed-lag model was commonly adopted, which takes the form of

$$y(t) = \beta_0 x(t) + \beta_1 x(t-1) + \cdots + \beta_k x(t-k) + \varepsilon(t). \tag{34}$$

Here, the sequence of coefficients $\{\beta_0, \beta_1, \ldots, \beta_k\}$ constitutes the impulse-response function of the mapping from $x(t)$ to $y(t)$. That is to say, if we imagine that, on the input side, the signal $x(t)$ is a unit impulse of the form

$$x(t) = \{\ldots, 0, 1, 0, \ldots, 0, 0 \ldots\} \tag{35}$$

which has zero values at all but one instant, then the output of the transfer function would be

$$r(t) = \{\ldots, 0, \beta_0, \beta_1, \ldots, \beta_k, 0, \ldots\}. \tag{36}$$

Another concept that helps in understanding the dynamic response is the step-response of the transfer function. Imagine that the input sequence is zero-valued up to a point in time when it assumes a constant unit value:

$$x(t) = \{\ldots, 0, 1, 1, \ldots, 1, 1 \ldots\}. \tag{37}$$

The output of the transfer function would be the sequence

$$s(t) = \{\ldots, 0, s_0, s_1, \ldots, s_k, s_k, \ldots\}, \tag{38}$$

where

$$\begin{aligned} s_0 &= \beta_0, \\ s_1 &= \beta_0 + \beta_1, \\ &\vdots \\ s_k &= \beta_0 + \beta_1 + \cdots + \beta_k. \end{aligned} \tag{39}$$

Here, the value s_k, which is attained by the sequence when the full adjustment has been accomplished after k periods, is called the (steady-state) gain of the transfer function; and it is denoted by $\gamma = s_k$.

The distributed-lag formulation of equation (34) is profligate in its use of parameters; and, given that the sequence $x(t)$ is likely to show strong serial correlation, we may expect to encounter problems of multicollinearity.

The Geometric Lag Structure

Another early approach to the problem of defining a lag structure, which depends on two parameters β and ϕ, is the geometric lag scheme of Koyk:

$$y(t) = \beta\Big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\Big\} + \varepsilon(t). \tag{40}$$

The impulse-response function of the Koyk model is a geometrically declining sequence $\{\beta, \phi\beta, \phi^2\beta, \ldots\}$

The gain of the transfer function, which is obtained by summing the geometric series, has the value of

$$\gamma = \frac{\beta}{1-\phi}. \tag{41}$$

By lagging the equation by one period and multiplying by ϕ, we get

$$\text{(42)}\quad \phi y(t-1) = \beta\big\{\phi x(t-1) + \phi^2 x(t-2) + \phi^3 x(t-3) + \cdots\big\} + \phi\varepsilon(t-1).$$

Taking the latter from (40) gives

$$y(t) - \phi y(t-1) = \beta x(t) + \big\{\varepsilon(t) - \phi\varepsilon(t-1)\big\} \quad \text{or}$$

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

of which the rational form is

$$\text{(45)}\qquad y(t) = \frac{\beta}{1-\phi L}x(t) + \varepsilon(t).$$

The following expansion can be used to recover equation (40):

$$\text{(46)}\qquad \begin{aligned} \frac{\beta}{1-\phi L}x(t) &= \beta\big\{1 + \phi L + \phi^2 L^2 + \cdots\big\}x(t) \\ &= \beta\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} \end{aligned}$$

Equation (43) cannot be estimated consistently by OLS regression because the composite disturbance term $\{\varepsilon(t) - \phi\varepsilon(t-1)\}$ is correlated with the lagged dependent variable $y(t-1)$.

A simple consistent estimation procedure is based on the equation under (40). The elements of $y(t)$ within the sample may be expressed as

$$
\begin{aligned}
y_t = \beta \sum_{i=0}^{\infty} \phi^i x_{t-i} + \varepsilon_t &= \theta\phi^t + \beta \sum_{i=0}^{t-1} \phi^i x_{t-i} + \varepsilon_t \\
&= \theta\phi^t + \beta z_t + \varepsilon_t.
\end{aligned} \tag{47}
$$

Here

$$
\theta = \beta\big\{x_0 + \phi x_{-1} + \phi^2 x_{-2} + \cdots\big\} \tag{48}
$$

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

$$
z_t = x_t + \phi x_{t-1} + \cdots + \phi^{t-1} x_1 \tag{49}
$$

is a synthetic variable based on the observations $x_t, x_{t-1}, \ldots, x_1$ and on the value attributed to ϕ.

The procedure for estimating ϕ and β that is based on equation (47) 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 that minimise the residual sum of squares.

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

$$(50) \qquad y(t) = \frac{\beta}{1 - \phi L} x(t) + \frac{1}{1 - \rho L} \varepsilon(t),$$

which has a first-order autoregressive disturbance scheme in place of the white-noise disturbance to be found in equation (45).

The parameters ϕ and ρ may be estimated by searching within the square defined by $-1 < \rho, \phi < 1$. The search might be confined to the quadrant defined by $0 \leq \rho, \phi < 1$.

One can afford, to ignore autoregressive nature of the disturbance process while searching for an optimum value for ϕ. When this has been found, the residuals will constitute estimates of the autoregressive disturbances, to which an AR(1) model can be fitted by OLS regression.

Lagged Dependent Variables

A regression equation can also be set in motion by including lagged values of the dependent variable on the RHS. With one lagged value, we get

$$y(t) = \phi y(t-1) + \beta x(t) + \varepsilon(t). \tag{51}$$

In terms of the lag operator, this is

$$(1-\phi L)y(t) = \beta x(t) + \varepsilon(t), \tag{52}$$

of which the rational form is

$$y(t) = \frac{\beta}{1-\phi L}x(t) + \frac{1}{1-\phi L}\varepsilon(t). \tag{53}$$

The advantage of equation (51) is that it is amenable to estimation by ordinary least-squares regression. Although the estimates will be biased in finite samples, they will be consistent, if the model is correctly specified.

The disadvantage is the restrictive assumption that the systematic and disturbance parts have the same dynamics.

Partial Adjustment and Adaptive Expectations

A simple partial-adjustment model has the form

$$y(t) = \lambda\big\{\gamma x(t)\big\} + (1-\lambda)y(t-1) + \varepsilon(t), \tag{55}$$

If $y(t)$ is current consumption, $x(t)$ is disposable income, then $\gamma x(t) = y^*(t)$ is "desired" consumption.

If habits of consumption persist, then current consumption will be 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 $\lambda\gamma = (1-\phi)\gamma = \beta$ and $(1-\lambda) = \phi$ is adopted, equation (55) becomes identical to equation (51), which is the regression model with a lagged dependent variable.

According to Friedman's Permanent Income Hypothesis, the consumption function is specified as

$$y(t) = \gamma x^*(t) + \varepsilon(t), \quad \text{where}$$

$$x^*(t) = (1-\phi)\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} = \frac{1-\phi}{1-\phi L}x(t)$$

is the value of permanent or expected income, which is formed as a geometrically weighted sum of all past values of income. Observe that this is a case of adaptive expectations:

$$x^*(t) = \phi x^*(t-1) + (1-\phi)x(t).$$

On substituting the expression for permanent income into the equation of the consumption function, we get

$$y(t) = \gamma\frac{(1-\phi)}{1-\phi L}x(t) + \varepsilon(t). \tag{59}$$

When the notation $\gamma(1-\phi) = \beta$ is adopted, equation (59) becomes identical to the equation (45) of the Koyk model.

Error-Correction Forms, and Nonstationary Signals

The usual linear regression procedures presuppose that the relevant moment matrices will converge asymptotically to fixed limits as the sample size increases. This cannot happen if the data are trended, in which case, the standard techniques of statistical inference will not be applicable.

A common approach is to subject the data to as many differencing operations as may be required to achieve stationarity. However, differencing tends to remove some of the essential information regarding the behaviour of economic agents. Moreover, it is often discovered that the regression model looses much of its explanatory power when the differences of the data are used instead.

In such circumstances, one might use the so-called error-correction model. The model depicts a mechanism whereby two trended economic variables maintain an enduring long-term proportionality with each other.

The data sequences comprised by the model are stationary, either individually or in an appropriate combination; and this enables us apply the standard procedures of statistical inference that are appropriate to models comprising data from stationary processes.

Consider taking $y(t-1)$ from both sides of the equation of (51) which represents the first-order dynamic model. This gives

$$
\begin{aligned}
\nabla y(t) = y(t) - y(t-1) &= (\phi - 1)y(t-1) + \beta x(t) + \varepsilon(t) \\
&= (1-\phi)\left\{ \frac{\beta}{1-\phi} x(t) - y(t-1) \right\} + \varepsilon(t) \\
&= \lambda\big\{\gamma x(t) - y(t-1)\big\} + \varepsilon(t),
\end{aligned}
\tag{60}
$$

where $\lambda = 1-\phi$ and where γ is the gain of the transfer function as defined under (41). 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 that maintain a long-run proportionality are said to be cointegrated. It is easy to obtain an accurate estimate of γ, which is the coefficient of proportionality, simply by running a regression of $y(t-1)$ on $x(t)$.

To see how to derive an error-correction form for a more general autoregressive distributed-lag model, consider the second-order model:

$$(61) \qquad y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \beta_0 x(t) + \beta_1 x(t-1) + \varepsilon(t).$$

The part $\phi_1 y(t-1) + \phi_2 y(t-2)$ comprising the lagged dependent variables can be reparameterised as follows:

$$\left\{ [\,\phi_1 \quad \phi_2\,] \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\} \left\{ \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y(t-1) \\ y(t-2)) \end{bmatrix} \right\} = [\,\theta \quad \rho\,] \begin{bmatrix} y(t-1) \\ \nabla y(t-1)) \end{bmatrix}.$$

Here, the matrix that postmultiplies the row vector of the parameters is the inverse of the matrix that premultiplies the column vector of the variables. The sum $\beta_0 x(t) + \beta_1 x(t-1)$ can be reparametrised, similarly, to become

$$\left\{ [\,\beta_0 \quad \beta_1\,] \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\} \left\{ \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1)) \end{bmatrix} \right\} = [\,\kappa \quad \delta\,] \begin{bmatrix} x(t-1) \\ \nabla x(t) \end{bmatrix}.$$

If follows that equation (61) can be recast in the form of

$$(62) \qquad y(t) = \theta y(t-1) + \rho \nabla y(t-1) + \kappa x(t-1) + \delta \nabla x(t) + \varepsilon(t).$$

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

$$y(t) = \theta y(t-1) + \rho\nabla y(t-1) + \kappa x(t-1) + \delta\nabla x(t) + \varepsilon(t). \tag{63}$$

and rearranging it gives

$$\begin{aligned}\nabla y(t) &= (1-\theta)\left\{\frac{\kappa}{1-\theta}x(t-1) - y(t-1)\right\} + \rho\nabla y(t-1) + \delta\nabla x(t) + \varepsilon(t)\\ &= \lambda\left\{\gamma x(t-1) - y(t-1)\right\} + \rho\nabla y(t-1) + \delta\nabla x(t) + \varepsilon(t).\end{aligned}$$

This is an elaboration of equation (51); and it includes the differenced sequences $\nabla y(t-1)$ and $\nabla x(t)$. These are deemed to be stationary, as is the composite error sequence $\gamma x(t-1) - y(t-1)$.

Additional lagged differences can be added to the equation (63); and this is tantamount to increasing the number of lags of the dependent variable $y(t)$ and the number of lags of the input variable $x(t)$ within equation (61).

Lagged Dependent Variables and Autoregressive Residuals

A common approach to building a dynamic econometric model is to begin with a model with a single lagged dependent variable and, if this proves inadequate on account serial correlation in the residuals, to enlarge the model to include an AR(1) disturbance process.

The two equations

$$y(t) = \phi y(t-1) + \beta x(t) + \eta(t) \tag{64}$$

and

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t) \tag{65}$$

of the resulting model may be combined to form an equation which may be expressed in the form

$$(1 - \phi L)y(t) = \beta x(t) + \frac{1}{1 - \rho L}\varepsilon(t) \tag{66}$$

or in the form

$$(1 - \phi L)(1 - \rho L)y(t) = \beta(1 - \rho L)x(t) + \varepsilon(t) \tag{67}$$

or in the rational from

$$(68) \qquad y(t) = \frac{\beta}{1-\phi L}x(t) + \frac{1}{(1-\phi L)(1-\rho L)}\varepsilon(t).$$

Equation (67) can be envisaged as a restricted version of the equation

$$(69) \qquad (1-\phi_1 L-\phi_2 L^2)y(t) = (\beta_0+\beta_1 L)x(t) + \varepsilon(t)$$

wherein the lag-operator polynomials

$$(70) \qquad \begin{aligned} &1-\phi_1 L-\phi_2 L^2 = (1-\phi L)(1-\rho L) \qquad \text{and} \\ &\beta_0+\beta_1 L = \beta(1-\rho L) \end{aligned}$$

have a common factor of $1-\rho L$.

Some authorities maintain that one should begin by estimating equation (69) as it stands. Then, one should use tests to ascertain whether the common-factor restriction is justifiable.

Only if the restriction is acceptable, should one then proceed to estimate the model with a single lagged dependent variable and with autoregressive residuals. This strategy of model building proceeds from a general model to a particular model.