LECTURE 8

Dynamic Regressions

Autoregressive Disturbance Processes

The interpretation that is given to the disturbance term of a regression model depends upon the context in which the analysis is conducted. Imagine that we are are fitting the regression equation

$$y_t = \beta_0 + x_{t1}\beta_1 + \cdots + x_{tk}\beta_k + \varepsilon_t \tag{1}$$

to a set of economic variables observed through time. Then, it is usual to assume that the disturbance ε_t represents the net effect upon the dependent variable y_t of a large number of subsidiary variables, which, individually, are of insufficient importance to be included in the systematic part of the model.

It may be reasonable to imagine that, instead of accumulating over time, the effects of the subsidiary variables will tend to cancel each other in any period. Then, their overall effect might have a small constant expected value. The inclusion of the intercept term in the regression equation allows us to assume that $E(\varepsilon_t) = 0$; for any nonzero net effect of the subsidiary variables will be absorbed by β_0.

Economic variables are liable to follow slowly-evolving trends and they are also liable to be strongly correlated with each other. If the disturbance term is indeed compounded from such variables, then we should expect that it too will follow a slowly-evolving trend.

The assumptions of the classical regression model regarding the disturbance term are at variance with these expectations. In the classical model, it is assumed that the disturbances constitute a sequence $\varepsilon(t) = \{\varepsilon_t; t = 0, \pm 1, \pm 2, \ldots\}$ of independently and identically distributed random variables such that

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

The process that generates such disturbances is often called a white-noise process. A sequence of 50 observations generated by a white-noise 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. Given the paucity of econometric data, we are unlikely to be able to estimate the parameters of complicated models with any degree of precision; and, in econometrics, the traditional means of representing the inertial properties of the disturbance process has been to adopt a simple first-order autoregressive model, or AR(1) model, whose equation takes the form of

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

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

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

We have shown, in a previous lecture, that the covariance of two elements of the sequence $\eta(t)$ that are separated by τ time periods is given by

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

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

$$V(\eta_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi^2}. \tag{5}$$

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

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

$$\text{Corr}(\eta_{t-\tau}, \eta_t) = \frac{C(\eta_{t-\tau}, \eta_t)}{\sqrt{V(\eta_{t-\tau})V(\eta_t)}} = \frac{C(\eta_{t-\tau}, \eta_t)}{V(\eta_t)} = \frac{\gamma_\tau}{\gamma_0}. \tag{6}$$

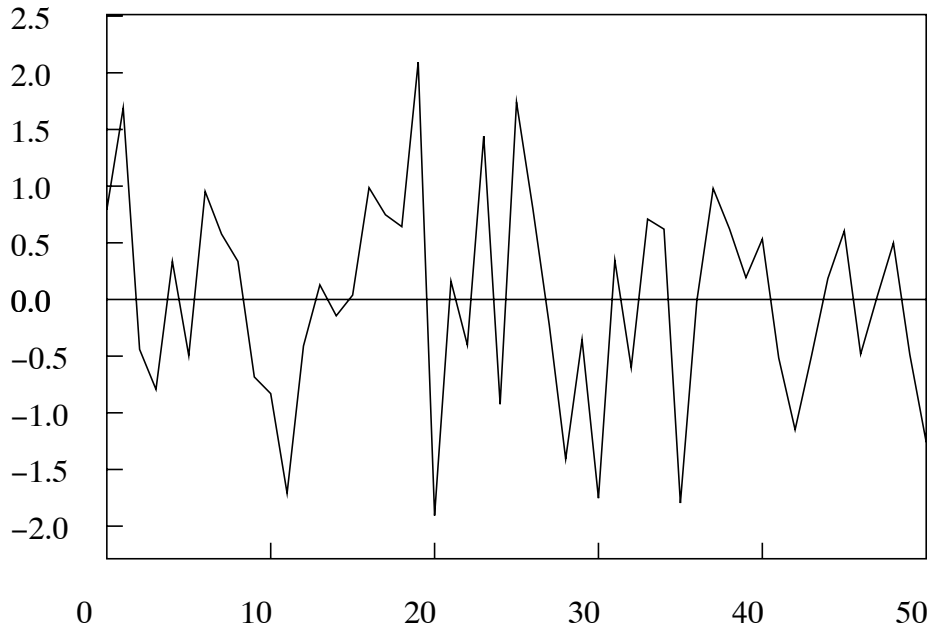

Figure 1. 50 observations on a white-noise process $\varepsilon(t)$ of unit variance.

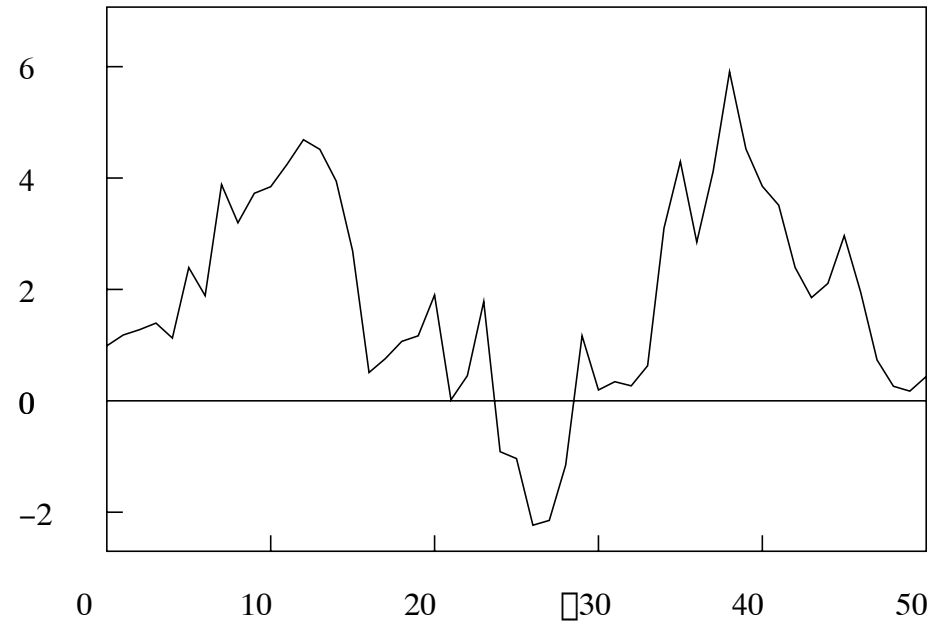

Figure 2. 50 observations on an AR(1) process $\eta(t) = 0.9\eta(t-1) + \varepsilon(t)$.

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

Detecting Serial Correlation in the Regression Disturbances

If all of the other assumptions of the Classical Model are granted, then the presence of serial correlation amongst the disturbances will affect neither the unbiasedness of the ordinary least-squares estimates nor their consistency. However, the standard errors and confidence intervals that are calculated under the i.i.d assumption will be incorrect, which will affect any inferences that we might make.

To overcome the problems of serial correlation, we should have to build a model for the disturbance process as an integral part of our regression strategy. First, however, we have to detect the presence of serial correlation.

The residuals of an ordinary least-squares regression represent estimates of the disturbances, from which they may differ significantly; and the misrepresentation of the disturbances can affect our ability to detect serial correlation.

To demonstrate the differences between the disturbances and the residuals, we may resort to matrix algebra. The equation that is fitted to the classical linear model $(y; X\beta, \sigma^2 I)$ by ordinary least-squares regression is represented by

$$y = X\hat{\beta} + e, \quad \text{where} \quad \hat{\beta} = (X'X)^{-1}X'y. \tag{7}$$

The residual vector is

$$e = y - X\hat{\beta} = y - X(X'X)^{-1}X'y; \tag{8}$$

and, on defining $P = X(X'X)^{-1}X'$, we can write this as

$$e = (I - P)y = (I - P)(X\beta + \varepsilon). \tag{9}$$

Now, observe that $(I - P)X = 0$. It follows that

$$e = (I - P)\varepsilon. \tag{10}$$

This shows that the residuals represent a transformation of the disturbances; and, since the matrix $I - P$ of the transformation is non-invertible, it is impossible to recover the disturbances. Next, observe that $PP' = P$ and that $(I - P)(I - P)' = (I - P)^2 = (I - P)$. It follows that the variance–covariance or dispersion matrix of the residuals is given by

$$\begin{aligned} D(e) &= E\big\{(I - P)\varepsilon\varepsilon'(I - P)'\big\} \\ &= \sigma^2(I - P) \end{aligned} \tag{11}$$

This is to be compared with the variance–covariance matrix of the disturbances which is $D(\varepsilon) = \sigma^2 I$. Thus, we see that the residuals will be serially correlated even when the disturbances are independently and identically distributed. Nevertheless, if the remainder of regression model, which is to say its systematic part, is correctly specified, then we shall find that $e \to \varepsilon$ as $T \to \infty$. Likewise, we shall find that, as the sample size increases, $D(e) = \sigma^2(I - P)$ will tend to $D(\varepsilon) = \sigma^2 I$.

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

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

Then, if we could observe the sequence directly, the best way of detecting the serial correlation would be to estimate the value of ρ and to test the significance of the estimate. The appropriate estimate would be

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

When residuals are used in the place of disturbances, we face a variety of options.

The first of these options, which we may be constrained to follow in some circumstances, is to treat the residuals as if they were equal to the disturbances. This approach is acceptable when the sample contains sufficient information to determine estimates of the systematic parameters which are close to the true values. Thus, it may be said that the procedure has asymptotic validity in the sense that it becomes increasingly acceptable as the sample size increases.

A second approach is to derive a sequence of revised residuals which have the same statistical properties as the disturbances, such that they are independently and identically distributed whenever the disturbances are. Then we should be able to make exact inferences based on the usual statistical tables. The problem with this approach is that the test statistics that are based upon the revised residuals tend to lack power in small samples.

A third approach, which has become practicable only in recent years, is to devise tests that take full account of whatever statistical properties the ordinary least-squares residuals would possess if the null hypothesis asserting the i.i.d nature of the disturbances were true. This means that we must calculate the exact sampling distribution of the test statistic under the null hypothesis in view of the precise value of the moment matrix $X'X/T$ that characterises the regression. Such an approach requires considerable computing resources.

The traditional approach to the problem of testing for the presence of serial correlation in a regression model 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. This has led them to acknowledge and to identify the circumstances where their test statistic delivers inconclusive results.

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

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

In fact, this statistic may be used quite generally for detecting any problem of misspecification whose symptoms are seeming violations of the i.i.d assumption concerning the disturbances.

In attempting to understand the nature of the statistic, attention should be focussed on its numerator. The numerator is a sum of squares of differences between adjacent residuals. Clearly, if the disturbance sequence is highly correlated, with adjacent values lying close to each other, then we should expect the numerator to have a low value. Conversely, if adjacent values of the disturbance sequence are uncorrelated, then we should expect adjacent residuals, likewise, to show little correlation, and we should expect the numerator to have a high value.

On expanding the numerator of the Durbin–Watson statistic, we find that

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

where

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

is an estimate of the coefficient ρ of serial correlation, which is based on the ordinary least-squares residuals. If ρ is close to 1, and r likewise, then d will be close to zero; and we shall have a strong indication of the presence 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.

The response of Durbin and Watson to the fact that their statistic is based on residuals rather than on disturbances was to provide a table of critical values, which included a region of indecision. The resulting decision rules can be expressed as follows

if $d < d_L$, then acknowledge the presence of serial correlation,

if $d_L \leq d \leq d_U$, then remain undecided,

if $d_U < d$, then deny the presence of serial correlation.

The values of d_L and d_U to be found in the table depend upon two parameters. The first is the number of degrees of freedom available to the regression and the second is the number of variables included in the regression. The latter is reckoned without including the intercept term. It is evident from the table that, as the number of degrees of freedom increase, 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

Now let us consider various ways of estimating a regression model with a first-order autoregressive disturbance process. Assume that the regression equation takes the form of

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

where

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t) \quad \text{with} \quad \rho \in [0, 1). \tag{18}$$

With the use of the lag operator, the latter equation can be written as

$$\eta(t) = \frac{1}{1-\rho L}\varepsilon(t) \tag{19}$$

which is substituted into equation (17) to give

$$y(t) = \alpha + \beta x(t) + \frac{1}{1-\rho L}\varepsilon(t). \tag{20}$$

Multiplying this equation throughout by $1 - \rho L$ gives

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

where $\mu = (1-\rho)\alpha$. This can be written as

$$q(t) = \mu + \beta w(t) + \varepsilon(t), \tag{22}$$

where

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

If the value of ρ were known, then the sequences $q(t)$ and $w(t)$ could be formed and the parameters μ and β could be estimated by applying ordinary least-squares regression to equation (22). An estimate for $\alpha = \mu/(1-\rho)$ would then be recoverable from the estimate of μ.

There are various various ways in which the estimation of the equation (20) can be approached when ρ is unknown.

A simple approach to the estimation of the equation, which requires only a single application of ordinary least-squares regression, depends upon rewriting equation (21) as

$$y(t) = \mu + \rho y(t-1) + \beta_0 x(t) + \beta_1 x(t-1) + \varepsilon(t), \tag{24}$$

where $\beta_0 = \beta$ and $\beta_1 = -\rho\beta$. The estimation of this equation by ordinary least-squares regression takes no account of the fact that the parameters are bound by the restriction that $\beta_1 = -\rho\beta_0$; and, therefore, the number of parameters to be estimated becomes four instead of three. Some statistical efficiency will be lost by ignoring the parametric restriction.

The second approach to the estimation of equation (17) is based on equation (22), and it involves searching for the optimal value of ρ by running a number of trial regressions. For a given value of ρ, the elements of $q(t) = (1-\rho L)y(t)$ and $w(t) = (1-\rho L)x(t)$ for $t = 2, \ldots, T$ are constructed according to the following scheme:

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

Then the corresponding equation

$$q_t = \mu + \beta w_t + u_t \tag{26}$$

is subjected to an ordinary least-squares regression; and the value of the residual sum of squares is recorded. The procedure is repeated for various values of ρ; and the definitive estimates of ρ, α and β are those that correspond to the least value of the residual sum of squares.

The procedure of searching for the optimal value of ρ may be conducted in a systematic and efficient manner by using a line-search algorithm such as the method of Fibonacci Search or the method of Golden-Section Search. These algorithms are described in textbooks on unconstrained numerical optimisation.

The third method for estimating the parameters of equation (17) is the well-known Cochrane–Orcutt procedure. This is an iterative method in which each stage comprises two ordinary least-squares regressions.

The first of these regressions takes a value for ρ and estimates the parameters of the equation

$$y_t - \rho y_{t-1} = \mu + \beta(x_t - \rho x_{t-1}) + \varepsilon_t$$

or, equivalently, (27)

$$q_t = \mu + \beta w_t + \varepsilon_t$$

in the manner which we have just described. This gives rise to the conditional estimates $\beta = \beta(\rho)$ and $\mu = \mu(\rho)$. From the latter, we obtain the estimate $\alpha = \alpha(\rho)$.

The second of the regressions is applied to the equation

$$(y_t - \alpha - \beta x_t) = \rho(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

(28) or, equivalently,

$$\eta_t = \rho\eta_{t-1} + \varepsilon_t,$$

which incorporates the previously-determined values of α and β. This gives rise to the conditional estimate $\rho = \rho(\alpha, \beta)$. The later is then fed back into the previous equation (27); at which point another stage of the iterative process may ensue. The process is continued until the estimates generated by successive stages are virtually identical. The final estimates correspond to the unconditional minimum of the residual sum of squares.

The Feasible Generalised Least-Squares Estimator

The search procedure and the Cochrane–Orcutt procedure may be refined by placing them in the context of the generalised least-squares estimator that takes account of the dispersion matrix of the vector of disturbances.

Consider writing the regression equation as as $y = X\beta + \eta$, where y and η are vectors of T elements and X is a matrix or order $T \times k$. The variance–covariance or dispersion matrix of the vector $\eta = [\eta_1, \eta_2, \eta_3, \ldots \eta_T]'$, which is generated by the first-order autoregressive process, takes the form of $[\gamma_{|i-j|}] = \sigma_\varepsilon^2 Q$, where

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

and it can be confirmed directly that

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

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

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

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

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

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

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

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

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

Since the parameter ρ requires to 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 generalised least-squares estimator. The true generalised least-squares estimator, which would require a knowledge of the precise value of the parameter ρ is considered to be an infeasible estimator.

Distributed Lags

In an experimental situation, where we might be investigating the effects of an input variable x on a mechanism or on an organism, we can set the value of x and then wait until the system has achieved an equilibrium before recording the corresponding value of the output variable y. In economics, we are often interested in the dynamic response of y to changes in x; and, given that x is continuously changing, the system might never reach an equilibrium. Moreover,

it is in the nature of economic relationships that the adjustment of y to changes in x is distributed widely through time.

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

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

It is difficult to specify *a priori* what the form of a lag response will be in any particular econometric context. Nevertheless, there is a common presumption that the coefficients will all be of the same sign, and that, if this sign is positive, their values will rise rapidly to a peak before declining gently to zero. In that case, the sequence of coefficients bears a useful analogy to the ordinates of a discrete probability distribution; and one may speak of such measures as the mean lag and the median lag.

Whilst this may seem to be a reasonable presumption, it ignores the possibility of overshooting. Imagine, for example, that the impulse represents a windfall increase in income which is never repeated. A consumer may respond rapidly by increasing his expenditure; and, if he does so in the expectation of a permanently increased income, he will soon find himself in debt. His response, when he recognises that his good fortune has been temporary, should be to save; and, if he has overspent, then his retrenchment will lead him temporarily to a lower level of consumption than he was accustomed to before the increase.

Another concept which helps us to understand the nature of a dynamic response is the step-response of the transfer function. We may imagine that the input sequence is zero-valued up to a point in time when it assumes a constant unit value:

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

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 gain represents the amount by which y would increase, in the long run, if x, which has been constant hitherto, were to increase in value by one unit and to remain constant thereafter.

A problem with the distributed-lag formulation of equation (34) is that it is profligate in its use of parameters; and given that, in a dynamic econometric context, the sequence $x(t)$ is likely to show strong serial correlation, we may expect to encounter problems of multicollinearity—which is to say that the estimates of the parameters will be ill-determined with large standard errors.

There are several ways of constructing a lag scheme which has a parsimonious parametrisation. One of them is to make the parameters $\beta_0, \ldots, \beta_k$ functionally dependent upon a smaller number of latent parameters $\theta_0, \ldots, \theta_g$, where $g < k$. For example, in the Almon lag scheme, the parameters $\beta_0, \ldots, \beta_k$ are the ordinates of a polynomial of degree g.

The Geometric Lag Structure

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

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

Here, although there is an infinite set of lagged values of $x(t)$, there are 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 that 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

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

To make the Koyk model amenable to estimation, the equation must be transformed. By lagging the equation by one period and multiplying the result by ϕ, we get

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

Taking the latter from (40) gives

$$y(t) - \phi y(t-1) = \beta x(t) + \big\{\varepsilon(t) - \phi\varepsilon(t-1)\big\}. \tag{43}$$

With the use of the lag operator, this can be written as

$$(1-\phi L)y(t) = \beta x(t) + (1-\phi L)\varepsilon(t), \tag{44}$$

of which the rational form is

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

In fact, by using the expansion

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

within equation (45), the original form under (40) can be recovered.

Equation (43) 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 (40). In view of that equation, we may express the elements of $y(t)$ which fall within the sample as

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

Here

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

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

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

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 (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 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

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

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

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

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 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 (46) to the rational form of the present model given in equation (53) and by comparing the result with (40). The comparison of equation (53) with the corresponding rational equation (45) for the Koyk model shows that we now have an AR(1) disturbance process described by the equation

$$\eta(t) = \phi\eta(t-1) + \varepsilon(t) \tag{54}$$

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

under (53). 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 (50). If this estimate of ϕ is taken to represent the systematic dynamics of the model, then our inferences about such matters as the speed of convergence of the impulse response and the value of the steady-state gain are liable to be misleading.

Partial Adjustment and Adaptive Expectations

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

Consider a partial-adjustment model of the form

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

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 $\lambda\gamma = (1-\phi)\gamma = \beta$ and $(1-\lambda) = \phi$ is adopted, equation (55) becomes identical to equation (51) 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

$$y(t) = \gamma x^*(t) + \varepsilon(t), \tag{56}$$

where

$$\begin{aligned} x^*(t) &= (1-\phi)\big\{x(t) + \phi x(t-1) + \phi^2 x(t-2) + \cdots\big\} \\ &= \frac{1-\phi}{1-\phi L}x(t) \end{aligned} \tag{57}$$

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

$$x^*(t) - x^*(t-1) = (1-\phi)\big\{x(t) - x^*(t-1)\big\}, \tag{58}$$

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 (57) into the equation (56) 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 Non-stationary Signals

Many econometric data sequences are non-stationary, with trends that persist for long periods. However, 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.

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. In that case, the orignal data sequence is said to be integrated of order one, denoted by $I(1)$, and its differenced version, which is stationary, may be described as an $I(0)$ sequence.

A problem with 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 may be processes of equilibration by which the relative proportions of econometric variables are maintained over long periods of time. The evidence of this 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.

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. Moreover, 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 available 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 non-stationary. A pair of non-stationary 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 $x(t)$.

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, 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 straightforward to derive an error-correction form for the more general autoregressive distributed-lag model. The technique can be illustrated with the following second-order model:

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

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

$$\left\{[\,\phi_1 \quad \phi_2\,]\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}\right\}\left\{\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\begin{bmatrix}y(t-1)\\ y(t-2)\end{bmatrix}\right\} = [\,\theta \quad \rho\,]\begin{bmatrix}y(t-1)\\ \nabla y(t-1))\end{bmatrix}. \tag{62}$$

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 to become

$$\left\{[\,\beta_0 \quad \beta_1\,]\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}\right\}\left\{\begin{bmatrix}0 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}x(t)\\ x(t-1)\end{bmatrix}\right\} = [\,\kappa \quad \delta\,]\begin{bmatrix}x(t-1)\\ \nabla x(t)\end{bmatrix}. \tag{63}$$

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

$$y(t) = \theta y(t-1) + \rho\nabla y(t-1) + \kappa x(t-1) + \delta\nabla x(t) + \varepsilon(t). \tag{64}$$

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

$$\begin{aligned}\nabla y(t) &= (1-\theta)\left\{\frac{\kappa}{1-\theta}x(t-1) - y(t-1)\right\} + \rho\nabla y(t-1) + \delta\nabla x(t) + \varepsilon(t)\\ &= \lambda\left\{\gamma x(t-1) - y(t-1)\right\} + \rho\nabla y(t-1) + \delta\nabla x(t) + \varepsilon(t).\end{aligned} \tag{65}$$

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 $\gamma x(t-1) - y(t-1)$.

Observe that, in contrast to equation (60), the error-correction term of (65) comprises the lagged value $x(t-1)$ in place of $x(t)$. Had the reparametrising transformation that has been employed in equation (62) also been used in (63), then the consequence would have been to generate an error-correction term of the form $\gamma x(t) - y(t-1)$. It should also be observed that the parameter associated with $x(t)$ in (65), which is

$$\gamma = \frac{\kappa}{1-\theta} = \frac{\beta_0 + \beta_1}{1 - \phi_1 - \phi_2}, \tag{66}$$

is the steady state gain of the transfer function from $x(t)$ to $y(t)$.

Additional lagged differences can be added to the equation (65); 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). The general form of the autoregressive distributed-lag model may be expressed as

$$y(t) = \phi_1 y(t-1) + \cdots + \phi_p y(t-p) + \beta_0 x(t) + \cdots + \beta_k x(t-k) + \varepsilon(t), \tag{67}$$

which can be written is summary notation as

$$\alpha(L)y(t) = \beta(L)x(t) + \varepsilon(t), \tag{68}$$

with $\alpha(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p = 1 - \phi_1 L - \cdots - \phi_p L^p$. The usual form of the corresponding error-correction model, which is a reparametrised version of the autoregressive distributed-lag model, is

$$\nabla y(t) = \lambda\{\gamma x(t-1) - y(t-1)\} + \delta(L)\nabla x(t) + \rho(L)\nabla y(t) + \varepsilon(t), \tag{69}$$

where $\delta(L) = \delta_0 + \delta_1 L + \cdots + \delta_{k-1}L^{k-1}$ and $\rho(L) = \rho_1 L + \rho_2 L^2 + \cdots + \rho_{p-1}L^{p-1}$.

There are various alternative reparametrisations of the autoregressive distributed-lag model that will serve the purpose of creating an error-correction model that is appropriate to data that require to be differenced to achieve stationarity. Their algebraic details are given in the following chapter, which is devoted to transfer functions.

Detecting Non-stationarity via the Dickey–Fuller Test

Dickey and Fuller (1979, 1981) devised a test for the presence of a unit root in the autoregressive operator that would signify that the process in question follows a random walk. Although these tests relate to a very specific cause of non-stationarity, they have come to be regarded, by many econometricians, as the essential test of non-stationarity.

The purpose is to discover the presence of the factor $1 - z$ within the polynomial $\alpha(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_p z^p$, which corresponds to the autoregressive operator within the equation $\alpha(L)y(t) = \varepsilon(t)$.

The initial task is to find a means of revealing the presence of such a factor. Thereafter, the test may be pursued within the context of a variety of models. These include the ordinary pth-order autoregressive model, of which the first-order model is special case, and the autoregressive model with an incorporated drift or a trend.

The polynomial $\alpha(z)$ may be divided by $1 - z$ to yield a quotient $\beta(z)$, which is a polynomial of degree $p - 1$, and a remainder γ, which is a constant:

$$\alpha(z) = (1 - z)\beta(z) + \gamma. \tag{70}$$

Setting $z = 1$ in this equation gives the following value for the remainder:

$$\gamma = \alpha(1) = \alpha_0 + \alpha_1 + \cdots + \alpha_p. \tag{71}$$

But, if the polynomial contains a unit root, such that $1 - z$ is one of its factors, then the remainder will be zero, i.e. $\gamma = 0$; and this is the condition that is the subject of the tests.

The Auxiliary Regressions

In practice, the test of the unit root hypothesis is usually conducted using an auxiliary regression. Consider the case of a first-order AR(1) model

$$y(t) = \mu + \phi y(t - 1) + \varepsilon(t). \tag{72}$$

Subtracting $y(t - 1)$ from both sides gives

$$\begin{aligned} y(t) - y(t - 1) &= \mu + (\phi - 1)y(t - 1) + \varepsilon(t) \quad \text{or} \\ \nabla y(t) &= \mu - \gamma y(t) + \varepsilon(t). \end{aligned} \tag{73}$$

When $\gamma = 0$, this becomes a model of a random walk with a drift, which is an upward drift if $\mu > 0$. By accumulating the differences, it can be seen that, in this case, the process is the sum of a random walk and a linear time trend. If $\mu = 0$, then there is an ordinary random walk.

Within the context of equation (4), the hypothesis that $\gamma = 0$ is tested using the standard t-statistic of an ordinary least-squares (OLS) regression. Thus, to test the null hypothesis $H_0 : \phi = 1$ that the process is a random walk—either an ordinary random walk or one with drift—against the alternative hypothesis $H_1 : \phi < 1$ that the process is stationary, the appropriate Dickey–Fuller test statistic is

$$DF = \frac{-\gamma^*}{SE(\gamma^*)}, \tag{74}$$

where $SE(\gamma^*)$ is the OLS standard error of the estimate of γ. In fact, this is identical to the standard error that would arise from applying and OLS regression to equation (73).

The test statistic does not have a t-distribution under the the null hypothesis of a non-stationary process. In particular, the distribution is skewed to the left, with a long lower tail, which implies that the critical values are more distant from the centre than those of the t-distribution or its normal approximation. The various levels of significance are tabulated below:

Asymptotic Critical Values for the Unit Root Test

Significance Level	1%	2.5%	5%	10%
Critical Value	−3.43	−3.12	−2.86	−2.57

The alternative hypothesis that the process is stationary about a constant mean value might not be an appropriate one to consider. It might be more appropriate to envisage a process that is stationary about some polynomial function of time.

In that case, the appropriate context within which to conduct the Dickey–Fuller test might be a model of the form

$$\nabla y(t) = \mu_0 + \mu_1 t - \gamma y(t) + \varepsilon(t). \tag{75}$$

If $\gamma = 0$ then, by accumulating the differences, we can see that the process is the sum of a random walk and a quadratic time trend. However, the model does allow for the possibility that $\mu_1 = 0$, which would reduce equation (75) to equation (73).

When a time trend is included in the regression, new critical values must be found for the test. This is because the time trend offers a competing explanation for the non-stationarity of the data. Therefore, more extreme values of the t-statistic are required in order to reject the null hypothesis. The modified critical values that are appropriate to equation (75) are tabulated below:

Asymptotic Critical Values for the Unit Root Test in the Presence of a Time Trend

Significance Level	1%	2.5%	5%	10%
Critical Value	−3.96	−3.66	−3.41	−3.12

A test for the presence of a single unit root in pth-order AR(p) autoregressive model may be conducted via a straightforward extension of the procedure that is appropriate to an AR(1) model, by including lagged differences $\nabla y(t)$ on the RHS. The equation that can be used for testing the condition of (71) takes the form of

$$\nabla y(t) = \mu - \gamma y(t) + \delta_1 \nabla y(t-1) + \cdots + \delta_{p-1}\nabla y(t-p+1) + \varepsilon(t), \tag{76}$$

and the resulting test is described as the augmented Dickey–Fuller test. The critical values with respect to the estimate of $-\gamma$ are the same as before, but the estimated coefficients associated with the lagged values of $\nabla y(t)$, found on the RHS, have approximately the t-distribution.

It is commonly asserted that the reason for the inclusion of the lagged differences on the RHS of the equation is to ensure that the error term of the regression will tend asymptotically to a white-noise process. This is a necessary condition for the asymptotic validity of the critical values of the tests.

However, it needs to be demonstrated that the parameter γ in this equation is the remainder term from the division of the polynomial $\alpha(z)$ by $\nabla(z) = 1 - z$. The requirement is to show that equation (76) is just a reparametrised version of the p-th order autoregressive equation. In demonstrating this, one can afford to ignore the intercept term μ, which simplifies the algebra.

The Reparametrisation of the Autoregressive Model

In order to demonstrate the reparametrisation of the autoregressive model, it is necessary to derive expressions for the coefficients of the quotient polynomial $\beta(z)$ of equation (70). There is a wide variety of ways in which these coefficients may be derived, including the familiar method of long division. Probably, the easiest way is via the method of synthetic division which may be illustrated by an example.

Consider the case where $p = 3$. Then

$$\alpha_0 + \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 = (\beta_0 + \beta_1 z + \beta_2 z^2)(1 - z) + \gamma. \tag{77}$$

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

$$\begin{aligned} \alpha_3 &= -\beta_2 \\ \alpha_2 &= -\beta_1 + \beta_2 \\ \alpha_1 &= -\beta_0 + \beta_1 \\ \alpha_0 &= \gamma + \beta_0. \end{aligned} \tag{78}$$

These can be rearranged to give

$$\begin{aligned}
\beta_2 &= -\alpha_3 \\
\beta_1 &= -\alpha_2 + \beta_2 &&= -(\alpha_2 + \alpha_3) \\
\beta_0 &= -\alpha_1 + \beta_1 &&= -(\alpha_1 + \alpha_2 + \alpha_3) \\
\gamma &= \alpha_0 - \beta_0 &&= \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3.
\end{aligned} \tag{79}$$

An alternative expression for the RHS of (70), is obtained using the identity

$$\gamma = \alpha(1) = z\alpha(1) + (1 - z)\alpha(1). \tag{80}$$

Substituting this into equation (1) gives

$$\alpha(z) = z\alpha(1) + \nabla(z)\{\alpha(1) + \beta(z)\}. \tag{81}$$

Now observe that, in terms of the example and more generally, there is

$$\begin{aligned}
\alpha(1) + \beta_0 &= (\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3) - (\alpha_1 + \alpha_2 + \alpha_3) \\
&= \alpha_0.
\end{aligned} \tag{82}$$

It can be assumed that $\alpha_0 = 1$, which is the usual normalisation for the autoregressive polynomial. Therefore, one can write

$$\alpha(1) + \beta(z) = 1 - \delta_1 z - \cdots - \delta_{p-1} z^{p-1}. \tag{83}$$

Now let $\alpha(L)$ be the lag operator polynomial within the p-th order autoregressive model $\alpha(L)y(t) = \varepsilon(t)$. Then, setting $z = L$ in equations (71) and (73) gives

$$\alpha(L) = \gamma L + \nabla - (\delta_1 z + \cdots + \delta_{p-1})\nabla. \tag{84}$$

With this, the model can be written in the form of

$$\nabla y(t) = -\gamma y(t-1) + \delta_1 \nabla y(t-1) + \cdots + \delta_{p-1} \nabla y(t-p+1), \tag{85}$$

which is the basis of equation (76).

The Dickey–Fuller Test in gretl

The augmented Dickey–Fuller test that is implemented in *gretl* is based of the following regression equation:

$$\nabla y(t) = \mu(t) - \gamma y(t-1) + \sum_{j=1}^{p} \delta_j \nabla y(t-j). \tag{86}$$

Here, $\mu(t)$ may be a zero, a constant, a linear function or a quadratic function, depending on various command options that are detailed in the manual. These options are not mutually exclusive, and, when they are used together, the statistic will be reported separately for each case.

For each case, approximate probability values are calculated. The number p of the lagged values of $\nabla y(t)$ should be chosen to ensure that the parametrisation is flexible enough to represent adequately the dynamics of $\nabla y(t)$.

Setting p too low results in a mispecification of the model, which will invalidate the critical values of the test, whereas setting it too high will prejudice the power of the test. The parameter p can be determined automatically. Setting p to a negative number triggers a sequential procedure that starts with a high value of p and then reduces the value until the t-statistic for the parameter δ_p exceeds 1.645 in absolute value, which is when it is judged to be significantly different from zero.

Testing for Cointegration

If the two sequences $x(t)$ and $y(t)$ are cointegrated, then the error correction term $y(t) - \gamma x(t)$ should constitute a stationary process. To test whether or not that is the case, one might employ the Dickey–Fuller test or the augmented Dickey–Fuller test. However, the parameter γ requires to be estimated. This can be achieved by regressing $y(t)$ on $x(t)$.

If these two variables are generated by random walk processes and if there is a genuine cointegrating relationship, then it follows that the estimated coefficient $\hat{\gamma}$ will converge rapidly to the true value of γ. The rate of convergence is such that a non-degenerate distribution can be attributed to $T\hat{\gamma}$ as opposed to $\sqrt{T}\hat{\gamma}$, which would be the case if the two data sequences were stationary. Nevertheless, the fact that the coefficient requires to be estimated does have an effect on the distribution of the elements $u_t = y_t - \hat{\gamma}x_t$ that are subject to the test, of which the asymptotic critical values are given in the table below:

Asymptotic Critical Values for the Cointegration Test

Significance Level	1%	2.5%	5%	10%
Critical Value	−3.90	−3.59	−3.34	−3.04

Recall that the null hypothesis of the the Dickey–Fuller test is that the data sequence in question is from a non-stationary $I(1)$ process. In the present context, the null hypothesis of the test is that $u(t) = y(t) - \gamma x(t)$ is non-stationary and, therefore, that the two processes $x(t)$ and $y(t)$ are not cointegrated.

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

$$y(t) = \phi y(t-1) + \beta x(t) + \eta(t) \tag{87}$$

and

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t) \tag{88}$$

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

$$(1 - \phi L)y(t) = \beta x(t) + \frac{1}{1-\rho L}\varepsilon(t) \tag{89}$$

or in the form

$$(1 - \phi L)(1 - \rho L)y(t) = \beta(1 - \rho L)x(t) + \varepsilon(t) \tag{90}$$

or in the rational from

$$y(t) = \frac{\beta}{1-\phi L}x(t) + \frac{1}{(1-\phi L)(1-\rho L)}\varepsilon(t). \tag{91}$$

Equation (90) can be envisaged as a restricted version of the equation

$$(1 - \phi_1 L - \phi_2 L^2)y(t) = (\beta_0 + \beta_1 L)x(t) + \varepsilon(t) \tag{92}$$

wherein the lag-operator polynomials

$$\begin{aligned} &1 - \phi_1 L - \phi_2 L^2 = (1 - \phi L)(1 - \rho L) \qquad \text{and} \\ &\beta_0 + \beta_1 L = \beta(1 - \rho L) \end{aligned} \tag{93}$$

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

Recognition of this fact has led to a certain model-building prescription. It is maintained, by some authorities, that one should begin the model-building procedure by estimating equation (92) as it stands. Then, one should apply tests to the estimated model 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 is one which proceeds from a general model to a particular model.

Notwithstanding such prescriptions, many practitioners continue to build their models by proceeding in the reverse direction. That is to say, they begin by estimating the equation under (52) which is a simple regression equation with a single lagged dependent variable. Then they proceed to examine the results of tests of misspecification which might induce them to supplement their model with a disturbance scheme. Therefore, their strategy of model building is to proceed from a particular model to a more general model in the event of a misspecification.

In is important to recognise that, when the regression model contains a lagged dependent variable, it is no longer valid to use the Durbin–Watson statistic to test for the presence of serial correlation amongst the disturbances. The problem is that, if the disturbances are serially correlated, then the application of ordinary least-squares regression to the equation no longer results in consistent estimates of β and ϕ. Therefore it is erroneous to imagine that the regression residuals will provide adequate substitutes for the unobservable disturbances if one is intent on determining the character of the latter.

The inconsistency of the ordinary least-squares estimates of the parameters of equation (87) in attributable to the correlation of the disturbances of $\eta(t)$ with the elements of $y(t-1)$ which assume the role of regressors. Thus, if $\eta(t)$ and $y(t)$ are serially correlated sequences—which they clearly are in view of equations (88) and (87) respectivey—and if the elements of $\eta(t)$ form part of the contemporaneous elements of $y(t)$, then contemporaneous elements of $y(t-1)$ and $\eta(t)$ must be serially correlated.

There are ways of testing for the presence of serial correlation in the disturbances of the regression model containing a lagged dependent variable which are valid in large samples. Thus Durbin has suggested using the statistic

$$h = r\sqrt{\frac{T}{1 - TV(\phi)}} \tag{94}$$

wherein T is the sample size, r is the autocorrelation of the residuals defined under (3) and $V(\phi)$ is the estimated variance of the coefficient associated with the lagged dependent variable in the fitted equation. Notice that, in view of (15), we may put $(1 - d/2) \simeq r$ in place of r. Under the null hypothesis that there is no serial correlation amongst the disturbances, the distribution of statistic h tends to the standard normal distribution. The statistic is undefined if the quantity under the square-root sign is negative.

The h statistic is applicable only to cases where the regression model contains the dependent variable lagged by one period. A statistic which serves the same purpose as the h statistic and which is also available in a wider range

of circumstances is the Lagrange-multiplier test-statistic which generally distributed as a χ^2 variate.