LECTURE 6

Dynamic Regressions

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

$$y(t) = \beta_0 x(t) + \beta_1 x(t-1) + \cdots + \beta_k x(t-k) + \varepsilon(t). \tag{250}$$

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

$$x(t) = \{\ldots, 0, 1, 0, \ldots, 0, 0\ldots\} \tag{251}$$

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

$$r(t) = \{\ldots, 0, \beta_0, \beta_1, \ldots, \beta_k, 0, \ldots\}. \tag{252}$$

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:

$$x(t) = \{\ldots, 0, 1, 1, \ldots, 1, 1 \ldots\}. \tag{253}$$

The output of the transfer function would be the sequence

$$s(t) = \{\ldots, 0, s_0, s_1, \ldots, s_k, s_k, \ldots\}, \tag{254}$$

where

$$\begin{aligned} s_0 &= \beta_0, \\ s_1 &= \beta_0 + \beta_1, \\ &\vdots \\ s_k &= \beta_0 + \beta_1 + \cdots + \beta_k. \end{aligned} \tag{255}$$

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 (250) 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 which depends on a limited number of parameters was that of Koyk who proposed the following geometric lag scheme:

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

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

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

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

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

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

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

Taking the latter from (256) gives

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

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

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

of which the rational form is

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

In fact, by using the expansion

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

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

Equation (259) is not amenable to consistent estimation by ordinary least squares regression. The reason is that the composite disturbance term $\{\varepsilon(t) - \phi\varepsilon(t-1)\}$ is correlated with the lagged dependent variable $y(t-1)$—since the elements of $\varepsilon(t-1)$ form part of the contemporaneous elements of $y(t-1)$. This conflicts with one of the basic conditions for the consistency of ordinary least-squares estimation which is that the disturbances must be uncorrelated with the regressors. Nevertheless, there is available a wide variety of simple procedures for estimating the parameters of the Koyk model consistently.

One of the simplest procedures for estimating the geometric-lag scheme is based on the original form of the equation under (256). In view of that equation, we may express the elements of $y(t)$ which fall within the sample as

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

Here

$$\text{(264)}\qquad \theta=\beta\{x_0+\phi x_{-1}+\phi^2x_{-2}+\cdots\}$$

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

$$\text{(265)}\qquad z_t=x_t+\phi x_{t-1}+\cdots+\phi^{t-1}x_1$$

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

The procedure for estimating ϕ and β which is based on equation (263) involves running a number of trial regressions with differing values of ϕ and therefore of the regressors ϕ^t and z_t; $t = 1, \ldots, T$. The definitive estimates are those which correspond to the least value of the residual sum of squares.

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

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

which has a first-order autoregressive disturbance scheme in place of the white-noise disturbance to be found in equation (261). An estimation procedure may be devised which entails searching for the optimal values of ϕ and ρ within the square defined by $-1 < \rho, \phi < 1$. There may even be good reason to suspect that these values will be found within the quadrant defined by $0 \leq \rho, \phi < 1$.

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

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

Lagged Dependent Variables

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

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

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

In terms of the lag operator, this is

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

of which the rational form is

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

The advantage of equation (267) is that it is amenable to estimation by ordinary least-squares regression. Although the estimates will be biased in finite samples, they are, nevertheless, consistent in the sense that they will tend to converge upon the true values as the sample size increases—provided, of course, that the model corresponds to the processes underlying the data.

The model with a lagged dependent variable generates precisely the same geometric distributed-lag schemes as does the Koyk model. This can be confirmed by applying the expansion given under (262) to the rational form of the present model given in equation (269) and by comparing the result with (256). The comparison of equation (269) with the corresponding rational equation (261) for the Koyk model shows that we now have an AR(1) disturbance process described by the equation

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

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

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

To understand the detriment of imposing the constraint, let us imagine that the true model is of the form given under (266) with ρ and ϕ taking very different values. Imagine that, nevertheless, it is decided to fit the equation under (269). Then the estimate of ϕ will be a biased and an inconsistent one whose value falls somewhere between the true values of ρ and ϕ in equation (266). If this estimate of ϕ is taken to represent the systematic dynamics of the model, then our inferences about such matters as the speed of convergence of the impulse response and the value of the steady-state gain are liable to be misleading.

Partial Adjustment and Adaptive Expectations

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

Consider a partial-adjustment model of the form

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

where, for the sake of a concrete example, $y(t)$ is current consumption, $x(t)$ is disposable income and $\gamma x(t) = y^*(t)$ is "desired" consumption. Here we are supposing that habits of consumption persist, so that what is consumed in the current period is a weighted combination of the previous consumption and present desired consumption. The weights of the combination depend on the partial-adjustment parameter $\lambda \in (0, 1]$. If $\lambda = 1$, then the consumers adjust their consumption instantaneously to the desired value. As $\lambda \to 0$, their consumption habits become increasingly persistent. When the notation $\lambda\gamma = (1-\phi)\gamma = \beta$ and $(1-\lambda) = \phi$ is adopted, equation (271) becomes identical to equation (267) which relates to a simple regression model with a lagged dependent variable.

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

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

where

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

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

An alternative expression for the sequence of permanent income is obtained by multiplying both sides of (273) by $1-\phi L$ and rearranging the result. Thus

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

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

On substituting the expression for permanent income under (273) into the equation (272) of the consumption function, we get

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

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

Error-Correction Forms, and Nonstationary Signals

Many econometric data sequences are nonstationary, 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.

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 (267) which represents the first-order dynamic model. This gives

$$\begin{aligned}\nabla y(t) = y(t) - y(t-1) &= (\phi - 1)y(t-1) + \beta x(t) + \varepsilon(t)\\ &= (1-\phi)\left\{\frac{\beta}{1-\phi}x(t) - y(t-1)\right\} + \varepsilon(t)\\ &= \lambda\big\{\gamma x(t) - y(t-1)\big\} + \varepsilon(t),\end{aligned} \tag{276}$$

where $\lambda = 1 - \phi$ and where γ is the gain of the transfer function as defined under (257). 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 $x(t)$.

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:

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \beta_0 x(t) + \beta_1 x(t-1) + \varepsilon(t). \tag{277}$$

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

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

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, likewise, to become

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

If follows that equation (277) can be recast in the form of

$$y(t) = \theta y(t-1) + \rho \nabla y(t-1) + \kappa x(t) + \delta \nabla x(t) + \varepsilon(t). \tag{278}$$

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

$$\begin{aligned} \nabla y(t) &= (1-\theta) \left\{ \frac{\kappa}{1-\theta} x(t) - y(t-1) \right\} + \rho \nabla y(t-1) + \delta \nabla x(t) + \varepsilon(t) \\ &= \lambda \left\{ \gamma x(t) - y(t-1) \right\} + \rho \nabla y(t-1) + \delta \nabla x(t) + \varepsilon(t). \end{aligned} \tag{279}$$

This is an elaboration of equation (267); 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 $\gamma x(t) - y(t-1)$.

Additional lagged differences can be added to the equation (279); 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 (277).

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

$$y(t) = \phi y(t-1) + \beta x(t) + \eta(t) \tag{280}$$

and

$$\eta(t) = \rho\eta(t-1) + \varepsilon(t) \tag{281}$$

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

$$(1 - \phi L)y(t) = \beta x(t) + \frac{1}{1-\rho L}\varepsilon(t) \tag{282}$$

or in the form

$$(1 - \phi L)(1 - \rho L)y(t) = \beta(1 - \rho L)x(t) + \varepsilon(t) \tag{283}$$

or in the rational from

$$y(t) = \frac{\beta}{1-\phi L}x(t) + \frac{1}{(1-\phi L)(1-\rho L)}\varepsilon(t). \tag{284}$$

Equation (283) can be envisaged as a restricted version of the equation

$$(1 - \phi_1 L - \phi_2 L^2)y(t) = (\beta_0 + \beta_1 L)x(t) + \varepsilon(t) \tag{285}$$

wherein the lag-operator polynomials

$$\begin{aligned} &1 - \phi_1 L - \phi_2 L^2 = (1 - \phi L)(1 - \rho L) \qquad \text{and} \\ &\beta_0 + \beta_1 L = \beta(1 - \rho L) \end{aligned} \tag{286}$$

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 (285) 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 (268) 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 (280) 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 (281) and (280) 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

$$h = r\sqrt{\frac{T}{1 - TV(\phi)}} \tag{287}$$

wherein T is the sample size, r is the autocorrelation of the residuals defined under (226) 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 (225), 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.