

Introduction and Overview

Since the mid-1980's applied economists attempting to estimate time series econometric models have been aware of certain difficulties that arise when unit roots are present in the data. To ignore this fact and to proceed to estimate a regression model containing non-stationary variables at best ignores important information about the underlying (statistical and economic) processes generating the data, and at worst leads to nonsensical (or spurious) results. For this reason, it is incumbent on the applied researcher to test for the presence of unit roots and if they are present (and the evidence suggests that they generally are) to use appropriate modelling procedures. De-trending is not appropriate (Chapter 2) and simply differencing the data¹ to remove the non-stationary (stochastic) trend is only part of the answer. While the use of differenced variables will avoid the spurious regression problem, it will also remove any long-run information. In modelling time series data we need to retain this long-run information, but to ensure that it reflects the co-movement of variables due to the underlying equilibrating tendencies of economic forces, rather than those due to common, but unrelated, time trends in the data.

Modelling the long run when the variables are non-stationary is an expanding area of econometrics (both theoretical and applied). It is still fairly new in that while it is possible to find antecedents in the literature dating back to, for example, the seminal work of Sargan (1964) on early forms of the error-correction model, it was really only in 1986 (following the March special issue of the *Oxford Bulletin of Economics and Statistics*) that cointegration became a familiar term in the literature.² It is also a continually expanding area, as witnessed by the number of articles that have been published since the mid-1980s. There have been and continue to be major new developments.

¹ That is, converting x_t to Δx_t , where $\Delta x_t = x_t - x_{t-1}$, will remove the non-stationary trend from the variable (and if it does not, because the trend is increasing over time, then x_t will need to be differenced twice, etc.).

² Work on testing for unit roots developed a little earlier (e.g., the PhD work of Dickey, 1976 and Fuller, 1976).

The purpose of this book is to present to the reader those techniques that have generally gained most acceptance (including the latest developments surrounding such techniques) and to present them in as non-technical a way as possible while still retaining an understanding of what they are designed to do. Those who want a more rigorous treatment to supplement the current text are referred to Banerjee, Dolado, Galbraith and Hendry (1993) and Johansen (1995a) in the first instance and then of course to the appropriate journals. It is useful to begin by covering some introductory concepts, leaving a full treatment of the standard econometric techniques relating to time series data to other texts (see, for example, Hendry, 1995). This is followed by an overview of the remainder of the book, providing a route map through the topics covered starting with a simple discussion of long-run and short-run models (Chapter 2) and then proceeding through to estimating these models using multivariate techniques (Chapters 5 and 6). We then cover panel data tests for unit roots and cointegration (Chapter 7) before concluding with an in-depth look at modelling and forecasting financial time series (Chapter 8).

SOME INITIAL CONCEPTS

This section will review some of the most important concepts and ideas in time series modelling, providing a reference point for later on in the book. A fuller treatment is available in a standard text such as Harvey (1990). We begin with the idea of a data-generating process (hereafter d.g.p.), in terms of autoregressive and moving-average representations of dynamic processes. This will also necessitate some discussion of the properties of the error term in a regression model and statistical inferences based on the assumption that such residuals are 'white noise'.

Data-generating Processes

As economists, we only have limited knowledge about the economic processes that determine the observed data. Thus, while models involving such data are formulated by economic theory and then tested using econometric techniques, it has to be recognized that theory in itself is not enough. For instance, theory may provide little evidence about the processes of adjustment, which variables are exogenous and indeed which are irrelevant or constant for the particular model under investigation (Hendry, Pagan and Sargan, 1984). A contrasting approach is based on statistical theory, which involves trying to characterize the statistical processes whereby the data were generated.

We begin with a very simple stationary univariate model observed over the sequence of time $t = 1, \dots, T$:

$$\text{or } \left. \begin{array}{l} y_t = \rho y_{t-1} + u_t \\ (1 - \rho L)y_t = u_t \end{array} \right\} \quad |\rho| < 1 \quad (1.1)$$

where L is the lag operator such that $Ly_t = y_{t-1}$. This statistical model states that the variable y_t is generated by its own past together with a disturbance (or residual) term u_t . The latter represents the influence of all other variables excluded from the model, which are presumed to be random (or unpredictable) such that u_t has the following statistical properties: its expected value (or mean) is zero [$E(u_t) = 0$] fluctuations around this mean value are not growing or declining over time (i.e., it has constant variance denoted $E(u_t^2) = \sigma^2$); and it is uncorrelated with its own past [$E(u_t u_{t-i}) = 0$]. Having u_t in (1.1) allows y_t to also be treated as a random (stochastic) variable.³

This model can be described as a d.g.p., if the observed realization of y_t over time is simply one of an infinite number of possible outcomes, each dependent on drawing a sequence of random numbers u_t from an appropriate (e.g., standard normal) distribution.⁴ Despite the fact that in practice only a single sequence of y_t is observed, in theory any number of realizations is possible over the same time period. Statistical inferences with respect to this model are now possible based on its underlying probability distribution.

The model given by equation (1.1) is described as a first-order autoregressive (AR) model or more simply an AR(1) model. It is straightforward to derive the statistical properties of a series generated by this model. First, note that (1.1) can be rearranged as:

$$y_t = [1/(1 - \rho L)]u_t \quad (1.2)$$

It can be shown that $1/(1 - \rho L) = (1 + \rho L + \rho^2 L^2 + \rho^3 L^3 \dots)$, and therefore the AR(1) model (1.1) can be converted to an infinite order moving average of the lagged disturbance terms:⁵

$$y_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots \quad (1.3)$$

Taking expectations gives $E(y_t) = 0$ (since $E(u_t) = 0$ for all t), thus the mean of y_t , when the d.g.p. is (1.1), is zero. The formula for the variance of y_t is $\text{var}(y_t) = E[\{y_t - E(y_t)\}]^2$. Since in this case the mean of y_t is zero, the formula for the variance simplifies to $E(y_t^2)$. Using this gives:

$$\begin{aligned} E(y_t^2) &= E(\rho y_{t-1} + u_t)^2 \\ &= E(\rho^2 y_{t-1}^2) + E(u_t^2) + 2\rho E(y_{t-1} u_t) \\ &= \rho^2 E(y_{t-1}^2) + \sigma^2 \end{aligned} \quad (1.4)$$

³ In contrast, y_t would be a deterministic (or fixed) process if it were characterized as $y_t = \rho y_{t-1}$, which, given an initial starting value of y_0 , results in y_t being known with complete certainty each time period. Note also that deterministic variables (such as an intercept or time trend) can also be introduced into (1.1).

⁴ The standard normal distribution is of course appropriate in the sense that it has a zero mean and constant variance and each observation is uncorrelated with any other.

⁵ This property is known as *invertibility*.

Repeatedly substituting for $E(y_{t-1}^2)$ on the right-hand side of (1.4) leads to a geometric series that converges to $E(y_t^2) = \sigma^2/(1 - \rho^2)$.

The autocovariance of a time series is a measure of dependence between observations. It is straightforward to derive the autocovariance for an AR(1) process. Generally, the autocovariance is $\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)]$ for $k \neq 0$, where μ represents the mean of y_t . When y_t is generated by (1.1), since $E(y_t) = 0$, the autocovariance formula simplifies to $E(y_t y_{t-k})$. Using this formula, it can be shown that the k th autocovariance is given by:

$$\gamma_k = \rho^k \gamma_0 \quad k = 1, 2, \dots \quad (1.5)$$

The autocorrelation coefficient for a time series is a standardized measure of the autocovariance restricted to lie between -1 and 1 . The k th autocorrelation is given by:

$$\frac{E[(y_t - \mu)(y_{t-k} - \mu)]}{E[(y_t - \mu)^2]} = \frac{\gamma_k}{\gamma_0} \quad (1.6)$$

Thus the k th autocorrelation when y_t is generated by (1.1) is given by ρ^k . Note that the autocovariances and autocorrelation coefficients discussed above are population parameters. In practice, the sample equivalents of these amounts are employed. In particular they are used when specifying time series models for a particular data set and evaluating how appropriate those models are, as in the Box–Jenkins procedure for time series analysis (Box and Jenkins, 1970). These authors were the first to develop a structured approach to time series modelling and forecasting. The Box–Jenkins approach recognizes the importance of using information on the autocovariances and autocorrelations of the series to help identify the correct time series model to estimate, and when evaluating the fitted disturbances from this model.

Another simple model that is popular in time series econometrics is the AR(1) model with a constant:

$$y_t = \delta + \rho y_{t-1} + u_t \quad |\rho| < 1 \quad (1.7)$$

Adding a constant to (1.1) allows y_t to have a non-zero mean. Specifically, the mean of y_t when (1.7) is the d.g.p. is given by $E(y_t) = \delta/(1 - \rho)$. To see this note that (1.7) can be written as:

$$(1 - \rho L)y_t = \delta + u_t \quad (1.8)$$

so that

$$\begin{aligned} y_t &= [1/(1 - \rho L)](\delta + u_t) \\ &= (1 + \rho + \rho^2 + \dots)\delta + (u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots) \end{aligned} \quad (1.9)$$

Since we are assuming that $E(u_t) = 0$, the expected value of (1.9) simplifies to:

$$E(y_t) = (1 + \rho + \rho^2 + \dots)\delta \quad (1.10)$$

which is a geometric series that converges to $E(y_t) = \delta/(1 - \rho)$. To calculate the variance of y_t when the d.g.p. is (1.7), it is easiest to work with the demeaned series $x_t = y_t - \mu$. We can then rewrite (1.7) as:

$$x_t = \rho x_{t-1} + u_t \quad (1.11)$$

It follows that $\text{var}(y_t) = E(x_t^2)$, and that $E(x_t^2) = \sigma^2/(1 - \rho^2)$. Therefore y_t generated by the AR(1) model with a constant has a mean of $E(y_t) = \delta/(1 - \rho)$ and a variance of $\text{var}(y_t) = \sigma^2/(1 - \rho^2)$.

The simple time series model (1.1) can be extended to let y_t depend on past values up to a lag length of p :

$$\left. \begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_p y_{t-p} + u_t \\ \text{or } A(L)y_t &= u_t \end{aligned} \right\} \quad (1.12)$$

where $A(L)$ is the polynomial lag operator $1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^p$. The d.g.p. in (1.12) is described as a p th-order AR model.⁶ The mean, variance and covariance of AR(p) processes when $p > 1$ can also be computed algebraically. For example, for the AR(2) model with a constant:

$$y_t = \delta + \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t \quad (1.13)$$

assuming $\rho_1 + \rho_2 < 1$ and that u_t is defined as before, the mean of y_t is $E(y_t) = \delta/(1 - \rho_1 - \rho_2)$ and the variance of y_t is:⁷

$$\text{var}(y_t) = \frac{(1 - \rho_2)\sigma^2}{(1 + \rho_2)(1 - \rho_1 - \rho_2)(1 + \rho_1 - \rho_2)} \quad (1.14)$$

An alternative to the AR model is to specify the dependence of y_t on its own past as a moving average (MA) process, such as the following first-order MA model:

$$y_t = u_t + \theta u_{t-1} \quad |\theta| < 1 \quad (1.15)$$

or a model with past values up to a lag length of q :

$$\left. \begin{aligned} y_t &= u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q} \\ \text{or } y_t &= B(L)u_t \end{aligned} \right\} \quad (1.16)$$

where $B(L)$ is the polynomial lag operator $1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$. In practice, lower order MA models have been found to be more useful in econometrics than higher order MA models, and it is straightforward to derive the statistical properties of such models. For example, for the first-order MA model (the MA(1) model) given by (1.15), the mean of y_t is simply $E(y_t) = 0$, while the variance of y_t is $\text{var}(y_t) = (1 + \theta^2)\sigma^2$. It turns out that,

⁶ Hence, (1.1) was a first-order AR process.

⁷ The importance of the assumption $\rho_1 + \rho_2 < 1$ will become clear in the next chapter.

for the MA(1) model, the first autocovariance is $\gamma_1 = \theta\sigma^2$, but that higher autocovariances are all equal to zero. Similarly, the first autocorrelation coefficient is $\rho_1 = \theta/(1 + \theta^2)$, but higher autocorrelation coefficients are all equal to zero.

Finally, it is possible to specify a mixed autoregressive moving average (ARMA) model:

$$A(L)y_t = B(L)u_t \quad (1.17)$$

which is the most flexible d.g.p. for a univariate series. Consider, for example, the ARMA(1, 1) model:

$$y_t = \rho_1 y_{t-1} + u_t + \theta_1 u_{t-1} \quad |\rho_1| < 1, |\theta_1| < 1 \quad (1.18)$$

As with the AR(1) model, note that the ARMA(1, 1) model can be rewritten as an infinite order MA process:

$$\begin{aligned} y_t &= (1 + \theta_1 L)(1 - \rho_1 L)^{-1} u_t \\ &= \sum_{j=0}^{\infty} \omega_j u_{t-j} \end{aligned} \quad (1.19)$$

Since we are assuming that $E(u_t) = 0$, it follows that $E(y_t) = 0$. The variance of y_t is given by:

$$\begin{aligned} E(y_t^2) &= E[(\rho_1 y_{t-1} + u_t + \theta_1 u_{t-1})^2] \\ &= E(\rho_1^2 y_{t-1}^2 + 2\rho_1 \theta_1 y_{t-1} u_{t-1} + u_t^2 + \theta_1^2 u_{t-1}^2) \end{aligned} \quad (1.20)$$

Using the autocovariance notation, the variance of y_t can be written:

$$\gamma_0 = \rho_1^2 \gamma_0 + 2\rho_1 \theta_1 \sigma^2 + \sigma^2 + \theta_1^2 \sigma^2 \quad (1.21)$$

which can be rearranged as:

$$\gamma_0 = \left(\frac{1 + \theta_1^2 + 2\rho_1 \theta_1}{1 - \rho_1^2} \right) \sigma^2 \quad (1.22)$$

The higher autocovariances can be obtained in a similar way, and it can be shown that:

$$\begin{aligned} \gamma_1 &= \rho_1 \gamma_0 + \theta_1 \sigma^2 \\ &= \left(\frac{(1 + \rho_1 \theta_1)(\rho_1 + \theta_1)}{(1 - \rho_1^2)} \right) \sigma^2 \end{aligned} \quad (1.23)$$

$$\gamma_2 = \rho_1 \gamma_1 \quad (1.24)$$

and $\gamma_k = \rho_1 \gamma_{k-1}$ for $k \geq 2$. The autocorrelation coefficients are given by:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 + \rho_1 \theta_1)(\rho_1 + \theta_1)}{(1 + \theta_1^2 + 2\rho_1 \theta_1)} \quad (1.25)$$

and $\rho_k = \rho_1 \rho_{k-1}$ for $k \geq 2$.

So far the d.g.p. underlying the univariate time series y_t contains no economic information. That is, while it is valid to model y_t as a statistical process (cf. the Box–Jenkins approach), this is of little use if we are looking to establish (causal) linkages between variables. Thus, (1.1) can be generalized to include other variables (both stochastic, such as x_t , and deterministic, such as an intercept), for example:

$$y_t = \alpha_0 + \gamma_0 x_t + \alpha_1 y_{t-1} + u_t \quad (1.26)$$

Since x_t is stochastic, let its underlying d.g.p. be given by:

$$x_t = \xi x_{t-1} + \varepsilon_t \quad |\xi| < 1 \quad \text{and} \quad \varepsilon_t \sim \text{IN}(0, \sigma_\varepsilon^2) \quad (1.27)^8$$

If u_t and ε_t are not correlated, we can state that $E(u_t \varepsilon_s) = 0$ for all t and s , and then it is possible to treat x_t as if it were fixed for the purposes of estimating (1.26). That is, x_t is independent of u_t (denoted $E(x_t u_t) = 0$) and we can treat it as (strongly) exogenous in terms of (1.26) with x_t being said to Granger-cause y_t . Equation (1.26) is called a conditional model in that y_t is conditional on x_t (with x_t determined by the marginal model given in (1.27)). Therefore, for strong exogeneity to exist x_t must not be Granger-caused by y_t , and this leads on to the concept of weak exogeneity.

Note, if (1.27) is reformulated as:

$$x_t = \xi_1 x_{t-1} + \xi_2 y_{t-1} + \varepsilon_t \quad (1.28)$$

then $E(x_t u_t) = 0$ is retained, but since past values of y_t now determine x_t the latter can only be considered *weakly* exogenous in the conditional model (1.26).⁹

Lastly, weak exogeneity is a necessary condition for super-exogeneity, but the latter also requires that the conditional model is structurally invariant; that is, changes in the distribution of the marginal model for x_t (equation (1.27) or (1.28)) do not affect the parameters in (1.26). In particular, if there are regime shifts in x_t then these must be invariant to (α_t, γ_0) in (1.26).

All three concepts of exogeneity will be tested later, but it is useful at this point to provide a brief example of testing for super-exogeneity in order to make the concept clearer.¹⁰ Assuming that known institutional (e.g., policy)

⁸ Note that $\varepsilon_t \sim \text{IN}(0, \sigma_\varepsilon^2)$ states that the residual term is independently and normally distributed with zero mean and constant variance σ_ε^2 . The fact that σ_ε^2 is multiplied by a (not shown) value of 1 means that ε_t is not autocorrelated with its own past.

⁹ That is, x_t still causes y_t , but not in the Granger sense, because of the lagged values of y_t determining x_t . For a review of these concepts of weak and strong exogeneity, together with their full properties, see Engle, Hendry and Richard (1983).

¹⁰ This example is based on Hendry (1995, p. 537). Further discussion of super-exogeneity can be found in Engle and Hendry (1993), Hendry (1995, p. 172) and Favero (2001, p. 146).

and historical shifts (shocks) can be identified that affected x_t , it should be possible to construct a dummy variable (e.g., POL_t) that augments (1.28):

$$x_t = \xi_1 x_{t-1} + \xi_2 y_{t-1} + \xi_3 POL_t + \varepsilon_t \quad (1.28^*)$$

Assuming that the estimate of $\hat{\xi}_3$ is (highly) significant in determining x_t , then super-exogeneity can be tested by including POL_t in the conditional model (1.26), and if this dummy is significant then super-exogeneity is rejected.¹¹

The importance of these three concepts of exogeneity are discussed in Favero (2001, p. 146): (i) if we are primarily interested in inference on the (α_i, γ_0) parameters in (1.26), then if x_t is weakly exogenous we only need to estimate (1.26) and not also (1.28); (ii) if we wish to dynamically simulate y_t and x_t is strongly exogenous, again we only need to estimate (1.26) and not also (1.28); and (iii) if the objective of modelling y_t is for econometric policy evaluation, we only need to estimate the conditional model (1.26) if x_t has the property of being super-exogenous. The latter is a necessary condition to avoid the Lucas Critique (see Lucas, 1976). For example, suppose y_t is a policy variable of government (e.g., the money supply) and x_t is the instrument used to set its outcome (e.g., the interest rate), then x_t must be super-exogenous to avoid the Lucas Critique. Otherwise, setting x_t would change the policy model (the parameters of 1.26), and the policy outcome would not be what the model (1.26) had predicted.¹²

As with the univariate case, the d.g.p. denoted by (1.26) can be generalized to obtain what is known as an autoregressive distributed lag (ADL) model:

$$A(L)y_t = B(L)x_t + u_t \quad (1.29)$$

where the polynomial lag operators $A(L)$ and $B(L)$ have already been defined.¹³ Extending to the multivariate case is straightforward, replacing y_t and x_t by vectors of variables, \mathbf{y}_t and \mathbf{x}_t .

The great strength of using an equation like (1.29) as the basis for econometric modelling is that it provides a good first approximation to the (unknown) d.g.p. Recall the above arguments that theory usually has little

¹¹That is, its exclusion from (1.26) would alter the estimates of (α_i, γ_0) . Note also that the residuals $\hat{\varepsilon}_t$ from (1.28*) should not be a significant determinant of y_t in equation (1.26).

¹²For example, suppose the government uses the immediate history of y_t to determine what it wishes current y_t to be; hence, it alters x_t to achieve this policy outcome. However, economic agents also 'know' the model (the policy rule) underlying (1.26) and (1.28*). Thus when POL_t changes, agents alter their behaviour (the parameters of 1.26 change) since they have anticipated the intended impact of government policy. Econometric models that fail to separate out the expectations formulation by economic agents from the behavioural relationships in the model itself will be subject to Lucas's critique.

¹³While we could further extend this to allow for an MA error process, it can be shown that a relatively simple form of the MA error process can be approximated by sufficiently large values of p and q in (1.29).

to say about the form of the (dynamic) adjustment process (which (1.29) is flexible enough to capture), nor about which variables are exogenous (this model can also be used as a basis for testing for exogeneity). In fact, Hendry et al. (1984) argue that the process of econometric modelling is an attempt to match the unknown d.g.p. with a validly specified econometric model, and thus ‘... economic theory restrictions on the analysis are essential; and while the data are the result of economic behaviour, the actual statistical properties of the observables corresponding to y and z are also obviously relevant to correctly analysing their empirical relationship. In a nutshell, measurement without theory is as valueless as the converse is non-operational.’ In practical terms, and according to the Hendry-type approach, the test of model adequacy is whether the model is *congruent* with the data evidence, which in a single equation model is defined in terms of the statistical properties of the model (e.g., a ‘white noise’ error term and parameters that are constant over time) and whether the model is consistent with the theory from which it is derived and with the data it admits. Finally, congruency requires the model to encompass rival models.¹⁴

Role of the Error Term u_t and Statistical Inference

As stated above, the error term u_t represents the influence of all other variables excluded from the model that are presumed to be random (or unpredictable) such that u_t has the following statistical properties: its mean is zero [$E(u_t) = 0$]; it has constant variance [$E(u_t^2) = \sigma^2$]; and it is uncorrelated with its own past [$E(u_t u_{t-i}) = 0$]. To this we can add that the determining variable(s) in the model, assuming they are stochastic, must be independent of the error term [$E(x_t u_t) = 0$].¹⁵ If these assumptions hold, then it is shown in standard texts like Johnston (1984) that estimators like the ordinary least squares (OLS) estimator will lead to unbiased estimates of the parameter coefficients of the model (indeed, OLS is the best linear unbiased estimator). If it is further assumed that u_t is drawn from the (multivariate) normal distribution, then this suffices to establish inference procedures for testing hypotheses involving the parameters of the model, based on χ^2 , t - and F -tests and their associated probability distributions.

Thus, testing to ensure that $u_t \sim \text{IN}(0, \sigma_u^2)$ (i.e., an independently distributed random ‘white noise’ process drawn from the normal distribution) is an essential part of the modelling process. Its failure leads to invalid inference

¹⁴ A good discussion of congruency and modelling procedures is given in Doornik and Hendry (2001).

¹⁵ Although not considered above, clearly this condition is not met in (1.1) and similar dynamic models, where y_{t-1} is a predetermined explanatory variable, since $E(y_t u_{t-i}) \neq 0$ for $i \geq 1$. However, it is possible to show by applying the Mann–Wald theorem (Johnston, 1984, p. 362) that with a sufficiently large sample size this will not lead to bias when estimating the parameter coefficients of the regression model.

procedures unless alternative estimators (e.g., generalized least squares—GLS—or systems estimators) and/or alternative probability distributions (such as the Dickey–Fuller distribution) are invoked.

FORECASTING

In applied economics, particularly applied macroeconomics and financial econometrics, often the main reason for estimating an econometric model is so that the estimated model can be used to compute forecasts of the series. While any type of econometric model can be used to compute forecasts (e.g., multivariate regression model, ADL model), it is univariate time series models such as the AR and ARMA models that have proved to be the most popular. The forecasting theory for univariate time series models has long been established (see in particular the work of Box and Jenkins, 1970) and univariate Box–Jenkins methods have continued to be popular with econometricians. Granger and Newbold (1986) set out a number of reasons why univariate forecasting methods in particular deserve consideration. Perhaps the most pertinent of these is the first reason they give:

They are quick and inexpensive to apply, and may well produce forecasts of sufficient quality for the purposes at hand. The cost of making particular forecasting errors should always be balanced against the cost of *producing* forecasts, for it is hardly worth expanding large resources to obtain a relatively small increase in forecast accuracy if the payoff, in terms of improved decision making is likely to be only marginally beneficial (p. 151).

This is an important point, not just for forecasting but for econometrics as a whole. There are usually a number of alternative models or techniques in econometrics that could be employed to undertake any one task, ranging from the simple to the very complex—and the complex techniques are typically more costly to use than the simple. Granger and Newbold (1986) sensibly argue that only when the benefits of the complex techniques outweigh the additional costs of using them should they be the preferred choice. It is often the case that forecasts made from simple linear univariate models such as AR models are more accurate, or are only marginally less accurate than forecasts from more complex alternatives.

In this section we will briefly review how to compute optimal forecasts from some of the models discussed so far, beginning with the most simple univariate time series model—the AR(1) model. Let the h -steps ahead forecast of a time series y_{T+h} be represented by \hat{y}_{T+h} , where T is the sample size (thus we assume forecasts of the series are from the end of the sample onward). The forecast error $e_{T+h} = y_{T+h} - \hat{y}_{T+h}$ plays a vital role in the literature on forecasting. Note in particular that the optimal forecast of y_{T+h} is the forecast that

minimizes the expected value of the squared forecast error $E[e_{T+h}^2]$. It can be proved that the h -steps ahead forecast that minimizes the expected value of the squared forecast error is simply the conditional expectation of y_{T+h} :¹⁶

$$\hat{y}_{T+h} = E(y_{T+h} | y_T, y_{T-1}, \dots, y_1) \quad (1.30)$$

which can be written more concisely as:

$$\hat{y}_{T+h} = E(y_{T+h} | \Omega_T) \quad (1.31)$$

where Ω_T represents the information set at time T . The forecast function will include unknown population parameters, and in practice these parameters are replaced with their estimated values.

If we assume that the d.g.p. is the AR(1) model given by equation (1.1), the optimal h -steps ahead forecast is given by:

$$\begin{aligned} \hat{y}_{T+h} &= E(\rho y_{T+h-1} + u_{T+h} | \Omega_T) \\ &= \rho y_{T+h-1} \end{aligned} \quad (1.32)$$

where Ω_T is the relevant information set at time T . So in the case of a 1-step ahead forecast ($h = 1$), the optimal forecast is simply ρy_T . Forecasts greater than 1-step ahead are computed recursively. So, for example, in the case of 2-steps ahead forecasts, the optimal forecast is:

$$\begin{aligned} \hat{y}_{T+2} &= E(\rho y_{T+1} + u_{T+2} | \Omega_T) \\ &= E(\rho \rho y_T + u_{T+2} | \Omega_T) \\ &= \rho^2 y_T \end{aligned} \quad (1.33)$$

and for 3-steps ahead forecasts $\hat{y}_{T+3} = \rho^3 y_T$. It follows that the forecasting function for the optimal h -steps ahead forecast (1.32) can be rewritten:

$$\hat{y}_{T+h} = \rho^h y_T \quad (1.34)$$

Clearly, assuming $|\rho| < 1$, as $h \rightarrow \infty$ the forecast of y_{T+h} converges to zero. In fact, for this d.g.p., as $h \rightarrow \infty$ the h -steps ahead forecast of y_{T+h} converges to the mean of y_t , which in this case is zero. If the d.g.p. is the AR(1) model with a constant:

$$y_t = \delta + \rho y_{t-1} + u_t \quad (1.35)$$

where $|\rho| < 1$ and $u_t \sim \text{IID}(0, 1)$, then y_t has a non-zero mean equal to $\delta/(1 - \rho)$. The 1-step ahead forecast is given by:

$$\begin{aligned} \hat{y}_{T+1} &= E(\delta + \rho y_T + u_{T+1} | \Omega_T) \\ &= \delta + \rho y_T \end{aligned} \quad (1.36)$$

¹⁶ The conditional expectation of y_{T+h} is the expected value of y_{T+h} conditioning on all information known about y at time T .

and the h -step ahead forecast is given by:

$$\hat{y}_{T+h} = (\rho^{h-1} + \rho^{h-2} + \cdots + \rho + 1)\delta + \rho^h y_T \quad (1.37)$$

Again, for this d.g.p., as $h \rightarrow \infty$ the forecast converges to the mean of y_t . This can be seen more clearly by noting that (1.37) is a geometric series, and as $h \rightarrow \infty$ it converges to:

$$\hat{y}_{T+h} = \frac{\delta}{(1 - \rho)} \quad (1.38)$$

Both the forecast functions and properties of the forecasts depend on the exact d.g.p. assumed for y_t . For example, if the d.g.p. is the first-order MA model given by equation (1.15) the 1-step ahead forecast is given by:

$$\begin{aligned} \hat{y}_{T+1} &= E(u_{T+1} + \theta u_T | \Omega_T) \\ &= \theta u_T \end{aligned} \quad (1.39)$$

The optimal forecast of the first-order MA process when the forecast horizon is greater than 1-step ahead is just the mean of the series, which in this case is zero:

$$\begin{aligned} \hat{y}_{T+h} &= E(u_{T+h} + \theta u_{T+h-1} | \Omega_T) \quad h > 1 \\ &= 0 \end{aligned} \quad (1.40)$$

For the ARMA(1, 1) model (1.18) the 1-step ahead forecast is given by:

$$\begin{aligned} \hat{y}_{T+1} &= E(\rho_1 y_T + u_{T+1} + \theta_1 u_T | \Omega_T) \\ &= \rho_1 y_T + \theta_1 u_T \end{aligned} \quad (1.41)$$

As with the AR model, when the forecast horizon is greater than 1-step ahead, forecasts from an ARMA model are computed recursively. For example, the 2-steps ahead forecast from an ARMA(1, 1) model is given by:

$$\begin{aligned} \hat{y}_{T+2} &= E(\rho_1 y_{T+1} + u_{T+2} + \theta_1 u_{T+1} | \Omega_T) \\ &= E[\rho_1(\rho_1 y_T + \theta_1 u_T) + u_{T+2} + \theta_1 u_{T+1} | \Omega_T] \\ &= \rho_1^2 y_T + \rho_1 \theta_1 u_T \end{aligned} \quad (1.42)$$

(note that $E(u_{T+2} | \Omega_T) = 0$ and $E(\theta_1 u_{T+1} | \Omega_T) = 0$), and the h -steps ahead forecast is given by:

$$\hat{y}_{T+h} = \rho_1^h y_T + \rho_1^{h-1} \theta_1 u_T \quad (1.43)$$

Assuming $|\rho_1| < 1$, as $h \rightarrow \infty$ again the forecast converges to the mean of the series—zero.

The forecasts referred to above are *point forecasts*; however, since the forecast itself (and forecast error) is a random variable, it is often helpful to compute *interval forecasts*. For the AR(1) model given by equation (1.1) the h -steps ahead forecast error is:

$$\begin{aligned} e_{T+h} &= y_{T+h} - \hat{y}_{T+h} \\ &= \rho y_{T+h-1} + u_{T+h} - \rho^h y_T \end{aligned} \quad (1.44)$$

which after repeated substitution can be rewritten as:

$$e_{T+h} = \rho^h y_T + u_{T+h} + \rho u_{T+h-1} + \cdots + \rho^{h-1} u_{T+1} - \rho^h y_T \quad (1.45)$$

Since it has a mean of zero, the variance of the h -steps ahead forecast error is:

$$E(e_{T+h}^2) = (1 + \rho^2 + \rho^4 + \cdots + \rho^{2h-2})\sigma^2 \quad (1.46)$$

and thus, for a 1-step ahead forecast from the first-order AR model, the variance of the forecast error is simply equal to the variance of the disturbance term u_t . Assuming that the forecast errors are normally distributed we can obtain a 95% confidence interval for the 1-step ahead forecast by computing $\hat{y}_{T+1} \pm 1.96\hat{\sigma}$, where $\hat{\sigma}$ is the estimated value of σ obtained when estimating the parameters of the fitted model. For the h -steps ahead forecast the 95% confidence interval is:

$$\hat{y}_{T+h} \pm 1.96\hat{\sigma}\sqrt{(1 + \rho^2 + \rho^4 + \cdots + \rho^{2h-2})} \quad (1.47)$$

If y_t is generated by the ARMA model (1.17) then computing forecast confidence intervals is more involved. The relevant theory is given in Box 1.1.

As already mentioned, we can compute forecasts from any type of econometric model. Consider a multivariate regression model:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_t \quad (1.48)$$

where $\mathbf{x}'_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ are explanatory variables and $\boldsymbol{\beta}$ is a vector of parameters of dimension $(k \times 1)$. Subject to the standard assumptions of the classical linear regression model, it can be shown the optimal forecast of y_{T+h} is given by the conditional expectation of y_{T+h} .¹⁷

$$\hat{y}_{T+h} = \mathbf{x}'_{T+h} \boldsymbol{\beta} \quad (1.49)$$

In practice $\boldsymbol{\beta}$ is replaced by the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ where $\mathbf{y} = (y_1, y_2, \dots, y_T)'$, $\mathbf{X}' = (\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_T)$. Obviously, to compute this forecast requires knowledge of the values of the explanatory variables at time $T+h$. Assuming these values are known, then (1.49) is the appropriate forecasting function. However, if these values are not known, then the appropriate forecasting function is:

$$\hat{y}_{T+h} = \hat{\mathbf{x}}'_{T+h} \hat{\boldsymbol{\beta}} \quad (1.50)$$

¹⁷See Granger and Newbold (1986, ch. 6, sect. 6.2).

Box 1.1 Interval forecasts for ARMA model

First, note that the ARMA model:

$$A(L)y_t = B(L)u_t \quad (1.1.1)$$

can be rewritten as an infinite-order MA model:

$$\begin{aligned} y_t &= A^{-1}(L)B(L)u_t \\ &= \omega(L)u_t \\ &= \sum_{j=0}^{\infty} \omega_j u_{t-j} \end{aligned} \quad (1.1.2)$$

It follows that y_{T+h} is given by:

$$y_{T+h} = \omega_0 u_{T+h} + \omega_1 u_{T+h-1} + \cdots + \omega_{h-1} u_{T+1} + \sum_{j=0}^{\infty} \omega_{h+j} u_{T-j} \quad (1.1.3)$$

A candidate forecast of y_{T+h} can be written:

$$\hat{y}_{T+h} = \sum_{j=0}^{\infty} \omega_{h+j}^* u_{T-j} \quad (1.1.4)$$

in which case the forecast error is:

$$\begin{aligned} e_{T+h} &= y_{T+h} - \hat{y}_{T+h} \\ &= \omega_0 u_{T+h} + \omega_1 u_{T+h-1} + \cdots + \omega_{h-1} u_{T+1} + \sum_{j=0}^{\infty} (\omega_{h+j} - \omega_{h+j}^*) u_{T-j} \end{aligned} \quad (1.1.5)$$

It follows that the expected value of the squared forecast error is:

$$E(e_{T+h}^2) = (\omega_0^2 + \omega_1^2 + \cdots + \omega_{h-1}^2) \sigma^2 + \sum_{j=0}^{\infty} (\omega_{h+j} - \omega_{h+j}^*)^2 \sigma^2 \quad (1.1.6)$$

which is minimized if $\omega_{h+j}^* = \omega_{h+j}$. Thus the optimal forecast of y_{T+h} is given by (1.1.4) where $\omega_{h+j}^* = \omega_{h+j}$. The variance of the forecast error associated with the optimal forecast is:

$$E(e_{T+h}^2) = (\omega_0^2 + \omega_1^2 + \cdots + \omega_{h-1}^2) \sigma^2 \quad (1.1.7)$$

and so the 95% forecast confidence interval is computed using the formula:

$$\hat{y}_{T+h} \pm 1.96 \hat{\sigma} \sqrt{\sum_{j=0}^{h-1} \omega_j^2} \quad (1.1.8)$$

where $\hat{\mathbf{x}}'_{T+h}$ are h -step ahead forecasts of the explanatory variables, which could be obtained via univariate methods.

OUTLINE OF THE BOOK

The next chapter deals with short- and long-run models. Inherent in the distinction is the notion of equilibrium; that is, the long run is a state of equilibrium where there is no inherent tendency to change since economic forces are in balance, while the short run depicts the disequilibrium state. Long-run models are often termed ‘static models’, but there is no necessity to actually achieve equilibrium at any point in time, even as $t \rightarrow \infty$. All that is required is that economic forces move the system toward the equilibrium defined by the long-run relationship posited. Put another way, the static equilibrium needs to be reinterpreted empirically since most economic variables grow over time. Thus, what matters is the idea of a steady-state relationship between variables that are evolving over time. This is the way the term ‘equilibrium’ is used in this book.

When considering long-run relationships, it becomes necessary to consider the underlying properties of the processes that generate time series variables. That is, we must distinguish between stationary and non-stationary variables, since failure to do so can lead to a problem of spurious regression whereby the results suggest that there are statistically significant long-run relationships between the variables in the regression model—when in fact all that is being obtained is evidence of contemporaneous correlations rather than meaningful causal relations. Simple examples of stationary and non-stationary processes are provided, and it is shown that whether a variable is stationary depends on whether it has a unit root. Comparing stationary and non-stationary variables is also related to the different types of time trends that can be found in variables. Non-stationary variables are shown to contain stochastic (i.e., random) trends, while stationary variables contain deterministic (i.e., fixed) trends. Since random trends in the data can lead to spurious correlations, an example of a spurious regression is given together with some explanations of why this occurs.

This leads naturally to the question of when it is possible to infer a causal long-run relationship(s) between non-stationary time series. The simple answer is: when the variables are cointegrated. The Engle and Granger (1987) definition of cointegration is explained, alongside the economic interpretation of cointegration that states that if two (or more) series are linked to form an equilibrium relationship spanning the long run, then even though the series themselves may contain stochastic trends (i.e., be non-stationary), they will nevertheless move closely together over time and the difference between them is constant (i.e., stationary). Thus the concept of cointegration mimics the existence of a long-run equilibrium to which an economic system converges over time. The absence of cointegration leads back to the problem of spurious regression.

Finally, Chapter 2 discusses short-run (dynamic) models. Simple examples of dynamic models are presented and linked to their long-run steady-state (equilibrium) solutions. It is pointed out that estimating a dynamic equation in the levels of the variables is problematic and differencing the variables is not a solution, since this then removes any information about the long run. The more suitable approach is to convert the dynamic model into an error correction (sometimes called an equilibrium correction) model (ECM), and it is shown that this contains information on both the short-run and long-run properties of the model, with disequilibrium as a process of adjustment to the long-run model. The relationship between ECMs and the concept of cointegration is also explored, to show that if two variables y_t and x_t are cointegrated, then there must exist an ECM (and, conversely, that an ECM generates cointegrated series).

Having discussed the importance of unit roots, the next task (Chapter 3) is to test for their presence in time series data. This begins with a discussion of the Dickey–Fuller (DF) test for a unit root, showing that a t -test of the null hypothesis of non-stationarity is not based on the standard t -distribution, but the non-standard DF distribution. Assumptions about what is the most appropriate d.g.p. for the variable being tested are found to be important when performing the test; that is, should an intercept and trend (i.e., deterministic components) be included in the test equation? Not only does inclusion and exclusion lead to different critical values for the DF test, but they are also important to ensure that the test for a unit root nests both the null hypothesis and the alternative hypothesis. To do this it is necessary to have as many deterministic regressors in the equation used for testing as there are deterministic components in the assumed underlying d.g.p. In order to test what will probably be in practice the most common form of the null hypothesis (that the d.g.p. contains a stochastic trend against the alternative of being trend-stationary), it is necessary to allow both an intercept and a time trend t to enter the regression model used to test for a unit root.

To overcome the problems associated with which (if any) deterministic components should enter the DF test (including problems associated with test power), the sequential testing procedure put forward by Perron (1988) is discussed. Then the DF test is extended to allow for situations when more complicated time series processes underlie the d.g.p. This results in the augmented Dickey–Fuller (ADF) test, which entails adding lagged terms of the dependent variable to the test equation. A question that often arises in applied work is how many extra lagged terms should be added, and there is some discussion of this problem. This in turn leads to a consideration of the power and size properties of the ADF test (i.e., the tendency to under-reject the null when it is false and over-reject the null when it is true, respectively). In finite samples it can be shown that any trend-stationary process can be approximated arbitrarily well by a unit root process and, similarly, any unit root process can be approximated by a trend-stationary process, especially for smaller sample sizes. That is, some unit root processes display finite sample

behaviour closer to (stationary) ‘white noise’ than to a (non-stationary) random walk (while some trend-stationary processes behave more like random walks in finite samples). This implies that a unit root test ‘... with high power against *any* stationary alternative *necessarily* will have correspondingly high probability of false rejection of the unit root null when applied to near stationary processes’ (Blough, 1992, p. 298). This follows from the closeness of the finite sample distribution of any statistic under a particular trend-stationary process and the finite sample distribution of the statistic under a difference-stationary process that approximates the trend-stationary process. Thus, Blough (1992, p. 299) states that there is a trade-off between size and power in that unit root tests must have either high probability of falsely rejecting the null of non-stationarity when the true d.g.p. is a nearly stationary process (poor size properties) or low power against any stationary alternative. This problem of the size and power properties of unit root tests means that any results obtained must be treated with some caution, although we consider some recent improvements that in principle have good size and power properties (cf. Ng and Perron, 2002). We also cover recent developments such as asymmetric tests for unit roots (panel tests are covered in Chapter 7).

There are further ‘problems’ associated with testing for non-stationarity. A structural break in a series will have serious consequences for the power of the test, if it is ignored. Taking into account the possibility that the intercept and/or slope of the underlying d.g.p. has changed (at an unknown date or dates) can be handled using the testing methods outlined in Perron (1994). Examples are provided and discussed. Finally Chapter 3 discusses testing for seasonal unit roots (including when there are structural breaks) and periodic integration. First of all, it is suggested that where possible seasonally *unadjusted* data should be used when testing for unit roots, since the filters used to adjust for seasonal patterns often distort the underlying properties of the data. In particular, there is a tendency of the DF test to be biased toward rejecting the null hypothesis of non-stationarity substantially less often than it should when seasonally adjusted series are tested. However, using unadjusted data that exhibit strong seasonal patterns opens up the possibility that these series may contain seasonal unit roots (i.e., the seasonal processes themselves are non-stationary). Tests for seasonal unit roots are discussed based on the Hylleberg, Engle, Granger and Yoo (1990) approach, and an example is presented using UK data on consumption, income and wealth. Structural breaks and their impact on seasonal unit root-testing is covered next, and the chapter concludes with a discussion of the situation where observations on a variable y_t can be described by a different model for each quarter, with the result being a periodic autoregressive model.

After testing for unit roots in the data and assuming they are present, the next task is to estimate the long-run relationship(s). Chapter 4 deals with cointegration in single equations, while Chapter 5 considers the possibility of more than one cointegration relationship. The most common single equation approach to testing for cointegration is the Engle–Granger (EG) approach.

This amounts to estimating the static OLS regression model in order to obtain an estimate of the cointegration vector (i.e., the estimate of β that establishes a long-run *stationary* relationship between the non-stationary variables in the model). Such a simple and popular approach, which of course ignores any short-run dynamic effects and the issue of endogeneity, is justified on the grounds of the ‘superconsistency’ of the OLS estimator. The latter states that the OLS estimator of β with non-stationary $I(1)$ variables converges to its true value at a much faster rate than the usual OLS estimator with stationary $I(0)$ variables, assuming cointegration (Stock, 1987). The most common form of testing for cointegration is based on an ADF unit root test of the residuals from the OLS regression. The need to use the correct critical values for testing the null hypothesis of no cointegration is discussed along with its dependence on the presence or otherwise of $I(2)$ variables in the regression. A first potential problem with the test procedure is also discussed (namely, the common factor restriction imposed on the long-run model by the ADF test for cointegration). We also consider testing for cointegration using the EG approach with a structural break, using the procedure developed in Gregory and Hansen (1996).

Despite the popularity of the EG approach, there are other serious problems such as small sample bias and the inability to test statistical hypotheses; hence, the advent of alternative testing procedures. Testing whether the speed-of-adjustment coefficient is significant in an error correction model is one alternative, and this is comparable to estimating a dynamic ADL model and testing whether the model converges to a steady-state solution. The major advantage of the ADL approach is that it generally provides unbiased estimates of the long-run model and valid t -statistics (even, on the basis of Monte Carlo evidence, when some of the regressors in the model are endogenous). The fully modified estimator is also discussed, but yields few advantages over the standard OLS estimator.

However, there still remain several disadvantages with a single equation approach. The major problem is that when there are more than two variables in the model, there can be more than one cointegration relationship among these variables. If there is, then adopting a single equation approach is inefficient in the sense that we can only obtain a linear combination of these vectors. However, the drawbacks of the single equation model extend beyond its inability to validly estimate all the long-run relationships between the variables; even if there is only one cointegration relationship, estimating a single equation is potentially inefficient (i.e., it does not lead to the smallest variance against alternative approaches). It is shown that this results from the fact that, unless all the right-hand-side variables in the cointegration vector are weakly exogenous, information is lost by not estimating a system that allows each endogenous variable to appear on the left-hand side of the estimated equations in the multivariate model. Thus, it is only really applicable to use the single equation approach when there is a single unique cointegration vector and when all the right-hand-side variables are weakly exogenous. Before proceeding to

the multivariate approach, Chapter 4 considers the short-run (EG) model based on a single equation and in particular gives an example of Hendry's general-to-specific modelling approach using the *PcGive* software package. The chapter then considers testing for seasonal cointegration and periodic cointegration, using single equation techniques and concludes with asymmetric testing for cointegration.

Chapter 5 is given over entirely to the Johansen procedure. Starting with a *vector* error correction model (VECM), it is shown that this contains information on both the short- and long-run adjustment to changes in the variables in the model. In particular, the problem faced is to decompose the long-run relationships into those that are stationary (and thus comprise the cointegration vectors) and those that are non-stationary (and thus comprise the 'common trends'). To do this, Johansen specifies a method based on reduced rank regressions, which is discussed in Box 5.1 used throughout the book to present the more difficult material. Before using Johansen's approach, it is important to consider whether the multivariate model contains $I(0)$ and $I(1)$ variables alone, in which case the modelling procedure is much simpler, or whether $I(2)$ variables are also present (i.e., variables that need to be differenced twice to achieve stationarity). If the latter, then the situation becomes far more complicated and Johansen has developed a procedure to handle the $I(2)$ model, although (at the time of writing) this is not fully available in *PcGive*. Instead, current practice is to test for the presence of $I(2)$ variables, and if they are present to seek to replace them through some form of differencing (e.g., if money supply and prices are $I(2)$, we could reformulate the model to consider real money $m_t - p_t$).

Since the Johansen approach requires a correctly specified VECM, it is necessary to ensure that the residuals in the model have the appropriate, standard Gaussian properties of being independently drawn from a normal distribution. This, *inter alia*, involves setting the appropriate lag length in the model and including (usually dummy) variables that only affect the short-run behaviour of the model. It is pointed out that residual mis-specification can arise as a consequence of omitting these important conditioning variables, and increasing the lag-length is often not the solution (as it usually is, for example, when autocorrelation is present). The procedures for testing the properties of the residuals are discussed and illustrated through examples. We then consider the method of testing for 'reduced rank' (i.e., testing how many cointegration vectors are present in the model). This involves a discussion of Johansen's trace and maximal eigenvalue tests and consideration of the small sample reliability of these statistics (at the same time an example of a likely $I(2)$ system is considered and the testing procedure for $I(2)$ variables is discussed). At this stage a major issue is confronted that presents considerable difficulty in applied work (namely, that the reduced rank regression procedure provides information on how many unique cointegration vectors *span* the cointegration space, while any linear combination of the stationary vectors is itself also a stationary vector and thus the estimates produced for any particular vector in β are not

necessarily unique). To overcome this ‘problem’ will involve testing the validity of linear restrictions on β . Before this, it is necessary to turn to the question of whether an intercept and trend should enter the short- and/or long-run model. Various models are presented and discussed along with the testing procedure for deciding which should be used in empirical work. An example of the use of the so-called Pantula principle is provided.

Weak exogeneity is considered next. This amounts to testing whether rows of the speed-of-adjustment matrix α are zero, and if such hypotheses are accepted the VECM can be respecified by conditioning on the weakly exogenous variables. The reasons for doing this, as well as a discussion of conditional and marginal models, are presented, while the concept of ‘weak exogeneity’ and how it is defined in various contexts is also discussed. The actual procedures that are used to perform tests of the null hypothesis that elements of α are zero are discussed together with examples that use *PcGive*. This then leads on to testing hypotheses about the cointegration relations involving β , which involves imposing restrictions motivated by economic arguments (e.g., that some of the β_{ij} are zero or that homogeneity restrictions are needed such as $\beta_{1j} = -\beta_{2j}$) and then testing whether the columns of β are identified. The form of the linear restrictions is discussed in some detail, along with various examples.

Lastly, the discussion moves on to testing for unique cointegration vectors (and hence structural long-run relationships). This involves testing that the restrictions placed on each of the cointegration vectors (the columns of β) in fact lead to an identified system (i.e., a model where any one cointegration vector cannot be represented by a linear combination of the other vectors). Johansen’s method for identification is carefully discussed and illustrated by several examples. The importance of this approach is stressed, since the unrestricted estimates of β are often hard to interpret in terms of their economic information.

One point that is worth emphasizing on testing for cointegration, and which should be fairly obvious from the above overview of the book thus far, is that an applied economist should really begin his or her analysis by using a multivariate framework and not by using a single equation approach. The exception will obviously be when only two variables are involved. The main reason for taking a systems approach from the outset is that to do otherwise restricts the practitioner to considering only one cointegration relationship when there may in fact be more, and even if he or she is only interested in one vector, it is probable that he or she will not get consistent and efficient estimates without allowing for the possibility of other cointegration vectors. Of course, where tests for weak exogeneity permit, moving down to the single equation approach can be justified after using the Johansen procedure.

Chapter 6 considers modelling the short-run multivariate system and concludes with a short discussion on structural macroeconomic modelling. First of all, it is stressed that obtaining long-run estimates of the cointegration relationships is only a first step to estimating the complete model. The short-run structure of the model is also important in terms of the information it

conveys on the short-run adjustment behaviour of economic variables, and this is likely to be at least as interesting from a policy viewpoint as estimates of the long run. Another important aspect of modelling both the short- and long-run structures of the system is that we can attempt to model the contemporaneous interactions between variables (i.e., we can estimate a simultaneous system, and this then provides an additional layer of valuable information). Based on the example of a small monetary model for the UK developed in Hendry and Mizon (1993) and Hendry and Doornik (1994), the following steps are illustrated: (i) use the Johansen approach to obtain the long-run cointegration relationships between the variables in the system; (ii) estimate the short-run vector autoregression (VAR) in error correction form (hence VECM) with the cointegration relationships explicitly included and obtain a parsimonious representation of the system; (iii) condition on any (weakly) exogenous variables thus obtaining a conditional parsimonious VAR (PVAR) model; and (iv) model any simultaneous effects between the variables in the (conditional) model, and test to ensure that the resulting restricted model parsimoniously encompasses the PVAR.

Chapter 7 considers testing for unit roots and cointegration with panel data (i.e., cross sectional time series data with $i = 1, \dots, N$ 'individuals' in each time period and with $t = 1, \dots, T$ observations for each individual over time). This offers the potential to increase the power of tests for integration and cointegration, since adding the cross section dimension to the time series dimension means that non-stationarity from the time series can be dealt with *and* combined with the increased data and power that the cross section brings. The latter acts as repeated draws from the same distribution, and thus while it is known that the standard DF-type tests lack power in distinguishing the unit root null from stationary alternatives, using the cross sectional dimension of panel data increases the power of unit root (and cointegration) tests that are based on a single draw from the population under consideration. The chapter considers in detail the various panel unit root tests that have been developed by, *inter alia*, Levin and Lin (1992, 1993); Im, Pesaran and Shin (1995, 1997); Harris and Tzavalis (1999); Maddala and Wu (1999); and Breitung (2000). All of these take non-stationarity as the null hypothesis and involve differing alternatives (depending on differing assumptions about the homogeneity of the cross sections in the panel) that all involve stationarity. The size and power of these tests is discussed and examples are given from estimating a well-known data set. Similarly, we consider the tests for cointegration and methods for estimation of the cointegration vector that have been developed in the literature. Cointegration tests using a single equation approach developed by Pedroni (1995, 1999) and Kao (1999) are discussed, where the null hypothesis is that there is no cointegration, while we also consider the approach taken by McKoskey and Kao (1998), who developed a residual-based test for the null of cointegration rather than the null of no cointegration in panels. The Larsson, Lyhagen and Lothgren (2001) use of a multi-equation framework to construct a panel test for cointegration rank in heterogeneous

panels is considered, which is based on the average of the individual rank trace statistics developed by Johansen (1995a).

In terms of estimating cointegration vectors using panel data sets, we look at the various estimators available which include within- and between-group fully modified (FMOLS) and dynamic (DOLS) estimators. In particular the estimators devised by Kao and Chiang (2000) and Pedroni (2000, 2001) are presented. In addition, some progress has recently been made toward developing a multivariate approach to panel cointegration estimation, with Breitung (2002) having developed a two-step procedure that is based on estimating a VECM. All of these estimators are compared using appropriate empirical examples.

Chapter 8 focuses on conditional heteroscedasticity models and forecasting. It can be viewed as a stand-alone chapter of particular relevance to those studying on courses in financial econometrics, although the reader will benefit from having read previous chapters, as when we mention unit roots and cointegration in Chapter 8 we do so assuming a good knowledge of the concepts. Conditional heteroscedasticity models such as the autoregressive conditional heteroscedastic (ARCH) model introduced by Engle (1982) and the generalized version of this model (GARCH), introduced by Bollerslev (1986), have become extremely popular in financial econometrics. For economists studying at final year undergraduate or postgraduate level, hoping to pursue careers in financial economics, an understanding of ARCH and GARCH models is important given their widespread use. We begin the chapter assuming no previous knowledge of conditional heteroscedasticity models and spend some time introducing concepts. We work through the standard ARCH and GARCH models and go on to discuss multivariate versions of these models. The estimation of ARCH and GARCH models is then considered before moving on to demonstrate the models with an empirical application to US stock market data. Beginning with conventional ARCH and GARCH models we then continue with the same data set to illustrate the main extensions of these models, including the ARCH-M model in which the conditional variance appears as a regressor in the conditional mean. A common feature of financial time series is that negative shocks tend to increase volatility by more than positive shocks of the same absolute magnitude—this characteristic has been labelled the ‘asymmetry effect’ or ‘leverage effect’.¹⁸ A number of GARCH specifications have been proposed to capture this effect, and we consider the most popular. After briefly introducing integrated and fractionally integrated GARCH models, we move on to discuss the impact of conditional heteroscedasticity on conventional unit root and cointegration tests. In empirical analyses it is common practice to apply conventional unit root and cointegration tests ignoring the presence of conditional heteroscedasticity,

¹⁸ The title ‘leverage effect’ is used because it is thought that the operating leverage of companies is responsible for the asymmetric behaviour of their share prices in response to ‘good’ and ‘bad’ news. See Nelson (1991), fn. 3.

since conventional unit root tests have been shown to be asymptotically robust to its presence. However, research has indicated that the finite sample properties of unit root tests can be adversely affected if ARCH or GARCH is present and is ignored. Furthermore, recently it has been shown that if the model used to compute a unit root test takes account of the ARCH effect and the parameters of this model are estimated simultaneously by maximum likelihood, the unit root test statistic does not have its conventional distribution. While research on these issues is still at a relatively early stage, we feel that they are important issues and are likely to be the subject of considerably more research in the future, hence we introduce the literature here.

The final part of this chapter considers some forecasting issues. We begin by discussing forecasting from ARCH or GARCH models and illustrate with a simple empirical application using the US stock market data previously employed. Forecasts are computed from ARCH and GARCH models and are evaluated using conventional measures of forecast accuracy such as mean squared error and graphical techniques. There have been a number of important developments in forecast evaluation, primarily published in the specialist econometrics and forecasting literature. In particular the development of tests of equal forecasting accuracy by Diebold and Mariano (1995), Harvey, Leybourne and Newbold (1997) and tests of forecast-encompassing of Chong and Hendry (1986), Harvey, Leybourne and Newbold (1998) and Clark and McCracken (2000, 2001). These tests allow the practitioner to test whether apparent differences in forecast accuracy are statistically significant and whether forecasts from one model contain information that is not present in the forecasts from a competing model. We illustrate the application of some of these tests using US stock market data (although they are applicable to forecasts from any kind of econometric model).

Important Terms and Concepts

Differencing and levels	d.g.p.	Congruency
ADL models	Strong, weak and super-exogeneity	'White noise' error term
Random (stochastic) variables	AR processes	MA processes
Polynomial lag operator	Forecasting	

