TEMA 4: FORECASTING WITH DYNAMIC REGRESSION MODELS

Antoni Espasa y Ana Pérez Espartero.

25 noviembre 2008

TEMA 4: FORECASTING WITH DYNAMIC REGRESSION MODELS

4.1. RELATING VARIABLES

In the previous two chapters forecasts have been made by relating a variable to its own past, so that if a forecast is required of y_{n+h}, the information set considered is I_n: y_{n-j}, $j \geq 0$. It is now time to move to wider information sets, such that the forecast of y_{n+h} will be based not only on its own past but also on the past and present values of other time series related to it. For example, a forecast of future unemployment could be based on past and present unemployment and also on past and present production figures; a forecast of inflation could include information on variables such us unitary labour costs, gross domestic product (GDP), quantity of money or interest rate differentials.

In the examples above, the enlargement of the information set involves considering data on other variables rather than y_t. This implies moving from a univariate information set, I_n: y_{n-j}, $j \geq 0$, to a multivariate data set that includes information on several variables.

Figure 4.1 displays a scheme of different directions in which the univariate information set can be enlarged. One of them consists of increasing the frequency of observation of the phenomenon of interest as, for example, passing to consider GDP from annual to quarterly level, or the monetary aggregate M3 from monthly to weekly level or the electricity consumption from monthly to daily, hourly or minute-by-minute levels. In this case, labelled as 1 in Figure 4.1, the resulting data set keeps, in principle, the initial period of time but with more data points. It is still univariate and the techniques explained in previous chapters still apply.

Another way of enlarging the information set is by including more variables, so that we move to a multivariate set. These new variables to be included can be of very different nature. For instance, we can apply a certain functional disaggregation (case 2 in Figure 4.1) so that components in which the series to be forecast can be broken down are included in the information set. As an example, employment can be disaggregated by population groups according to age, sex and education. Another possible disaggregation (case 3 in Figure 4.1) is by geographic areas: the regions in a country, or the states in a federal state or in a monetary union. In cases 2 and 3, if the weights connecting the components with the aggregate change through time, it can be interesting to include also the aggregate y_t in the information set.

In other cases, the enlarged information set includes data on other external variables. The relationship between these new variables and the phenomenon of interest may be merely empirical, like that between an industrial production index and a corresponding confidence indicator, or a consumer price index and a production price index. In this case, labelled as 4 in Figure 4.1, these additional variables are often called *leading indicators*. These will be studied in Chapter ??.

It could also be the case (case 5 in Figure 4.1) that new variables are added to the information set according to economic theory which postulates a relationship between the new variables and the one to be forecast. The examples of unemployment and inflation mentioned above fit in this framework. In this book models including explanatory variables will be denoted as *econometric models*. The main difficulty in building such models, and then using them to forecast, is to decide which variables are

going to be included as the information set. Thus, questions such as the dynamic structure involving each variable and what economic time series are appropriate are very important and can make a considerable difference to the forecasting performance of the model.

INSERTAR FIGURE 4.1

The data sets represented in cases 2 to 5 in Figure 4.1 contain time series on k+1 different variables with some interrelationships among them. Thus, we will presumably need a model with k+1 equations to explain these situations. However, in some especial cases to be discussed in next section, where there is only one variable of interest and for certain types of relationships connecting the variables, it will be possible to use single equation models. In this chapter we will focus on these models and defer the more sophisticated multivariate econometric models with several equations to Chapter 5.

In general, one can expect that if the quality of the information on the additional variables is good, the model relating the variables is well specified and correctly estimated, and the extra variables in the information set can be accurately forecast, the econometric models to be studied in this chapter and Chapter 5, will generate better forecasts than the univariate ARIMA models presented in Chapter 3. Moreover, an econometric model may provide not only a reliable forecast of an economic phenomenon, but also a better understanding of its underlying structure and an explanation of the factors which determine the forecast. Needless to say that the construction of the econometric model is more complex and costly than the ARIMA model, as it will become clear in next sections.

4.2. SINGLE-EQUATION MODELS

Suppose that one is interested in forecasting the dividend that some corporation is going to declare next month. If the sequence of previous annual dividends is denoted by y_t, the methods of earlier chapters would involve building a model from this sequence and then using it to forecast. However, anyone at all versed in economics may well think that it would be worthwhile expanding the information set by trying to explain dividends in terms of company earnings, otherwise known as after-tax profits. We then have a multivariate information set consisting of present and past values of both variables. Let us denote y_t the variable to be forecast and x_t the related variable added to the information set in order to improve the forecast of y_t. This kind of additional variables are called *explanatory variables*. In the example above, we have only one explanatory variable x_t but, in general, there will be k of these variables denoted by $x_{1t},..., x_{kt}$.

In this multivariate framework, we are faced with a forecasting problem which involves k+1 economic variables - the variable of interest and k explanatory variables - and consequently, we will presumably need to construct a *multi-equation* or *vector* model with k+1 equations, one for each variable, in order to describe the interrelationships among them. For instance, in studying the relationship between the quantity demanded of a product in terms of the price of that product, it should be clear that, in general, these two variables are interrelated, so that prices have an influence on demand, but demand also affects prices, and the causality can go in both directions. Similarly, it is known that production of a commodity will depend on consumption and consumption on employment and employment on production, for example. Therefore, feedback between

these variables is present and a model comprising one equation for each variable will be needed, even if our interest is only on variable y_t. In other words, if y_t depends on k explanatory variables, $x_{1t},..., x_{kt}$, but these variables also depend on y_t, a model to capture the relationship of our variable of interest y_t with $x_{1t},..., x_{kt}$ will need, in general, to include also equations explaining the dependency between the variables x_{it} and y_t.

However, in some cases, it could happen that the relationship between the variable y_t and the explanatory variables $x_{1t},..., x_{kt}$, is such that all the causality goes from $x_{1t},..., x_{kt}$ to y_t, so that the former influence the latter, but not on the other way round. That is, the explanatory variables are not subject to any feedback from the relevant phenomenon y_t. For example, the amount of electricity used by a household may depend both on the income of the household and the price of electricity[1], but it is not expected that an increase in the consumption of electricity will change the income of the household. As another example, the exports of a country will depend on a measure of the world production, but the latter is not expected to depend on the former. In these cases, under certain conditions that will be further studied in the next chapter, the econometric model for the variable of interest, y_t, can be formulated in only one equation, even though the information set is multivariate, and this can be done without loosing any information about y_t contained in the vector model. The explanatory variables are then called *exogenous* and the variable y_t is called *dependent* or *endogenous.*

The concept of exogeneity will be developed more precisely in the next chapter, where the distinction between causality and exogeneity will be done. Herein we just point out some remarks on the topic. When we have a multivariate information set, the corresponding distribution function of the data is a joint multivariate distribution, say $F(y, x_1, ..., x_k)$. The property of exogeneity for the explanatory variables implies a restriction on this joint distribution such that the analysis to be done on y_t can be based on the conditional distribution, $F(y / x_1, ..., x_k)$, without loosing any relevant information as compared to what should be achieved using a vector model from the multivariate distribution. In such framework, the required analysis on y_t can be carried out conditional on the explanatory variables and a single equation from the conditional distribution $F(y / x_1, ..., x_k)$ will be enough to represent the relationship between the endogenous and the exogenous variables. This single equation is one that explains the value of the variable of interest, y_t, in terms of present values of one or more related variables, $x_{1t},..., x_{kt}$, and possibly also the past values of these variables and the own past of the series being explained. This type of single-equation models with explanatory variables will be referred to as *regression models* and since they usually will have lagged values of the variables involved they will be more precisely referred to as *dynamic regression models.* In these models, exogeneity implies that the explanatory variables are independent of the error term in the regression.[2]

As a simple example of a regression model, let us suppose that we want to forecast the industrial production index of a certain sector of the U.S. economy, denoted by y_t, and that we have a qualitative indicator for such production, denoted by x_t. For simplicity of exposition, assume also that there is a contemporaneous relationship between these two variables of the form:

$$y_t = a + b\, x_t + \eta_t \qquad (4.1)$$

[1] For the purpose of this example, this price is considered as an administrative price.
[2] Unless otherwise stated, we only consider here linear models, as in previous chapters.

where in this and other equations to follow η_t just represents some residual or "error" series with indeterminate stationary properties. The coefficients *a* and *b* are chosen from the available data to give the best "fit" possible, using a least-squares criteria, as explained briefly in Chapter 2. Suppose that *a* and *b* have been estimated and so can be taken as being known and that a forecast of y_{n+1} at time *n* is required. Equation (4.1) then gives:

$$y_{n+1} = a + b\, x_{n+1} + \eta_{n+1} \tag{4.2}$$

and so it is immediately seen that the future value of y_{n+1} depends on the future value of x_{n+1}. In this framework, forecasting can be performed in two different ways that will be further explained in section 4.79. The first one is to use the model to provide *conditional* or *ex post* forecasts, that is, to forecast the value of y_{n+1} given some specific value for x_{n+1}. For example, if the indicator has been published just at the end of the month of reference, as it is often the case, and the production index is published some weeks later, then equation (4.2) may be used to forecast production index conditional on the given value of the indicator using:

$$\text{forecast of } y_{n+1} = a + b\, x_{n+1}.$$

However, as will be explained later, this may not be the optimal forecasting use of the model, as the forecastability of the error term is being ignored.

Equation (4.2) can also be used to provide forecasts of the variable being explained by linking it with forecasts of the explanatory variables. This will be called an *unconditional* or *ex-ante* forecast. For example, the US Production Price Index for the Industrial Sector (PPIND) can be regarded as a leading indicator for the Consumer Price Index for Industrial Goods (CPIND). Let us denote by PPI_j the production price index of a specific good *j* included in PPIND and CPI_j the corresponding consumer price, included in CPIND. Assuming a relationship of the form (4.2) for these two variables, being y_t the CPI_j and x_t the PPI_j, a forecast for x_t could be converted into a forecast for y_t by using (4.2) in the form

$$\text{forecast of } y_{n+1} = a + b\ (\text{forecast of } x_{n+1}). \tag{4.3}$$

Any forecast for production prices can be used here, but obviously the better this forecast the better the forecast for consumption prices. A very simple but very often useful model for industrial prices, if these do not show seasonal oscillations, could be

$$x_t = m + x_{t-1} + \varepsilon_t, \tag{4.4}$$

where ε_t is white noise. As we saw in previous chapters, a variable generated by such a model is said to be a random walk with drift. Accepting it as true, the corresponding forecast of industrial prices for $t=n+1$ is,

$$\text{forecast of } x_{n+1} = m + x_n,$$

and substituting into (4.3) gives

$$\text{forecast of } y_{n+1} = a + b\,(m + x_n).$$

As mentioned above, this forecast is not necessarily optimal. Note also that this forecast is strictly derived from the pair of relationships (4.2) and (4.4), but it could equally well have been reached by simply regressing y_{n+1} on x_n, that is, by considering a model of the form

$$y_t = \alpha + \beta x_{t-1} + \eta_t',$$

where $\eta_t' = \eta_t + b\varepsilon_t$. However, using (4.2) and (4.4) might be thought preferable since more use is then made of an economic theory, however naïve. This will be the case when one is interested in the forecast of y_{n+1} as well as in the parameter b itself, which will be the contemporaneous elasticity of y_t with respect to x_t if both variables are measured in logs.

This very simple-minded example discussed above illustrates a number of important points concerning forecasting methods based on regression models. First, that once the necessary model is obtained, it is obviously quite easy to construct forecasts; and second, and most important, that the main problem is to achieve the correct regression or forecasting model in the first place. As it will be seen through next sections, this problem is not an easy one, and much care should be taken in building up such models.

In general, we could say that the usual stages in forming a dynamic regression model, which will be further discussed in next sections, are as follows:

(i) For the variable to be forecast, list explanatory or causal variables. This list should be based on some specific theory, if possible, but introspection is also an acceptable source.

(ii) Gather relevant data to be used to estimate and evaluate the model. Time series will be required for all variables involved, and the series should be as long as possible in time span, although very ancient data may not be relevant. Exactly what data to use has to be a judgement call, both for the time span and the time interval between observations, if data are available daily, weekly, monthly, etc. In practice, there is less choice and often only a few years' monthly data, say, will be available.

(iii) Propose a model for the relationship between the variable of interest and its explanatory variables and estimate it from the data, as it will be explained below. The question of how many lags of the variables to use is a very difficult one, and it is doubtful if a satisfactory but simple method exists capable of answering it. If sufficient data are available, choose a relatively high number of lags that allows for a sufficiently large dynamic structure in the model (e.g., with quarterly data, it is usually recommended to include four or five lags) but ensures, at the same time, that the total number of parameters is at most around 1/3 of the number of observations. Moreover, if the data contain a seasonal component, this should also be taken into account. In practice, various alternative lagged models should be fitted to the data and the most satisfactory one would be chosen, that is, the model which leads to the smallest value of some information criterion. A more complex way to selecting a model will be discussed below when talking about automatic econometric modelling.

(iv) Carry out careful diagnostic checking on any fitted model. One essential aspect of the model eventually chosen is that its error residual series should be white noise. If this is not the case, the model is not well specified and the addition of more dynamics to the model or the inclusion of omitted variables should be considered. The reason for this requirement is that if the errors are not white noise the estimation procedure can lead to unsatisfactory estimates, as will be further explained below, and also because the full forecasting possibilities are not being realized if the errors are themselves somewhat forecastable.

The extent to which economic theory should be used in the procedure sketched above is still rather controversial. Some modellers will believe in a theory so strongly that they will insist on constraining parameters to agree with this theory. Others have so little confidence in theories that these are totally ignored and their specification comes just from analysis of the data. A correct theory can be very helpful in pointing towards a satisfactory model specification, limiting the class of models that need to be considered, or at the very least suggesting explanatory variables which should be included. Very rarely will a theory completely specify the model, including exactly what lags to use.

In practice, the analyst faces a complex task in building empirical models for forecasting. Following Granger (1999), who describes in detail these problems, we can say that the modern analyst starts the process of building a model by assuming that the specification of such model is unknown and finding out the best one is one of his main goals. Specifying an empirical economic model will first require knowledge of the economic theories that may help to select and structure the variables. It will also require the analyst to know the different available measures and data sources for those variables and the institutional characteristics and particular facts that may concern the phenomenon to be modelled.

4.3. DYNAMIC REGRESSION MODELS

As we said before, a dynamic regression model is a single equation model in which a variable is explained in terms of its own past and the present and/or past of other variables related to it. Consequently, these models involve different relationships between variables and so may represent many real situations in economics, business and many other fields dealing with time series data. For example, if we think of how production affects employment, it seems clear that an increase in production will have an effect on the current employment, but since the immediate adjustment of employment to production can be very costly or simply impossible, that increment on production will also affect the employment in the future. Therefore, the number of employees working next month will depend on the production in that month but also on the production in the previous months. Therefore, the effect of a change in the explanatory variable (production) is not only instantaneous but takes additional time to completely work through to the response variable (employment). Sometimes it can also be the case that only a delayed response exists. This can happen, for example, in the monthly relationship between a consumer price index and a relevant measure of interest rates. In this case, it can be expected that an increment of the interest rate at month t will only have a significant impact on prices after t.

In the examples above, a dynamic regression model is required. However, when building macro-econometric models, the data are very often available only at annual or

quarterly frequency, and therefore some relationships are necessarily specified as contemporaneous. But as the time disaggregation increases, such contemporaneous relationship may perhaps disappear and more complicated dynamic structures emerge.

In this section we will assume that all the variables involved in the econometric model are stationary, otherwise, they would have been made stationary by a suitable transformation. The study of relationships between nonstationary time series is postponed until section 4.7.

4.3.1. Autoregressive Distributed Lag models

There are two main ways of formulating a dynamic regression model: either as a usual regression model with lagged variables, in which case it is called an *Autoregressive Distributed Lags* (ADL) model, or with a structure of rational lag polynomials, in which case it is referred to as a *Transfer Function* (TF) model. In this section we discuss ADL models and the next is devoted to TF models.

A simple example of an ADL model may include one lagged value of the dependent variable, y_{t-1}, and one explanatory variable, x_t, and its first lag, x_{t-1}, giving rise to the following equation:

$$y_t = c + a_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + \varepsilon_t \tag{4.5}$$

where c, a_1, b_0 and b_1 are parameters and ε_t is a disturbance term assumed to be white noise. If quarterly data are available, this model states that the value of y in the present quarter depends on its value in the previous quarter and also on the current and previous values of an explanatory variable. For example, the value of the exports of one country could depend on an index of the world trade this quarter and the previous one (although this simple example should obviously include more variables like real exchange rates).

Despite its simplicity, model (4.5) is widely encountered in empirical time series econometrics. Actually, it encompasses as special cases several distinct types of dynamic models which have turned out relevant in describing some important economic relationships. See Hendry, Pagan and Sargan (1984) and Hendry (1995).

The simple model (4.5) may be generalized to include more explanatory variables and more lagged values of the dependent and the explanatory variables. For example, including more lags of both y_t and x_t in (4.5) leads to the following general formulation for the case of a single explanatory variable:

$$y_t = c + a_1 y_{t-1} + \ldots + a_r y_{t-r} + b_0 x_t + b_1 x_{t-1} \ldots + b_s x_{t-s} + \varepsilon_t. \tag{4.6}$$

This equation can also be written in terms of lag polynomials as:

$$a(L)\, y_t = c + b(L)\, x_t + \varepsilon_t \tag{4.7}$$

where $a(L)=(1-\sum_{j=1}^{r} a_j L^j)$, $b(L)=(\sum_{j=0}^{s} b_j L^j)$ and ε_t is white noise. Equation (4.7) is usually referred to as an ADL model, and because there are r lags on y_t and s lags on x_t,

it is sometimes denoted as ADL(r,s). In particular, model (4.5) is an ADL(1,1) model.

The obvious generalization of (4.7) to include more explanatory variables, say x_{1t},..., x_{kt}, gives the following equation:

$$a(L)\, y_t = c + \sum_{i=1}^{k} b_i(L)\, x_{it} + \varepsilon_t \tag{4.8}$$

where $a(L)$, $b_1(L)$,…, $b_k(L)$ are finite order lag polynomials of degrees r, s_1,..., s_k, respectively, and ε_t is assumed to be white noise. To ensure that the model is not explosive, all the roots of $a(L)$ should be outside the unit circle. Model (4.8) is usually denoted as ADL(r,s_1,...,s_k). Obviously, (4.5) and (4.7) are particular cases of it.

It should be emphasized that equation (4.8), where $a(L)$, $b_1(L)$,…, $b_k(L)$ are polynomials of appropriate finite orders r, s_1,..., s_k, is a very general representation of a dynamic regression model that enables modelling many kinds of relationships between economic variables. As an illustration, let us think about the following simple example. Consider a regression model with only one explanatory variable lagged once and where, for simplicity of exposition, it is assumed that there is not intercept and the errors are not independent but show some correlation between them which can be represented as an AR(1) process; that is, suppose that we have the following model

$$y_t = b_1\, x_{t-1} + \eta_t \tag{4.9a}$$
$$\eta_t = a_1\, \eta_{t-1} + \varepsilon_t \tag{4.9b}$$

where ε_t is white noise. From (4.9a) it turns out that

$$\eta_{t-1} = y_{t-1} - b_1\, x_{t-2},$$

and inserting this expression into equation (4.9b) this can be written as

$$\eta_t = a_1\, y_{t-1} + \beta\, x_{t-2} + \varepsilon_t$$

where $\beta = -a_1 b_1$. Now, putting back this expression in equation (4.9a) yields

$$y_t = a_1\, y_{t-1} + b_1\, x_{t-1} + \beta\, x_{t-2} + \varepsilon_t, \tag{4.10}$$

and this can be alternatively written in terms of lag polynomials as

$$(1 - a_1 L)\, y_t = b_1\, (1 - a_1 L)\, x_{t-1} + \varepsilon_t. \tag{4.11}$$

This last representation is an ADL(1,2) model with one explanatory variable including lagged values of both the endogenous and the exogenous variables and a residual term, ε_t, which is white noise. Moreover, in this case, the polynomials $a(L)=(1-a_1L)$ and $b(L)=b_1(1-a_1L)$ are subject to some constraints as they share a common factor $(1-a_1L)$. Therefore, the model (4.9), which has two structural parameters a_1 and b_1, can be thought of as a particular specification of the more general ADL model (4.10) – which has three parameters $\{a_1, b_1, \beta\}$ - with an imposed constraint on the lag coefficients of the polynomials, namely $\beta = -a_1 b_1$.

On the other way round, this example shows that an ADL model like (4.10), with a restriction on the parameters leading to a common factor between the dynamics of the variables, can be simplified as in (4.11) by making explicit the common factor. Then by passing that factor to the residual term, (4.11) can be formulated by a two-equation model like (4.9a)-(4.9b) in which the orders of the dynamic polynomials of the variables are reduced at the expense of including an error term with serial dependence. Hence, it seems clear that a dynamic regression model like (4.9a) is incomplete without specifying the structure of its error term η_t. Furthermore, as it has just been shown, assuming η_t to be white noise when it is not, the model will be wrong since the dynamic residual properties will have been ignored. In general, the dynamic structure of η_t will be unknown. But, as the above example illustrates, such information is not required to specify an ADL model because the model itself can account for the residual dynamics by enlarging the dynamics of both the endogenous and exogenous variables, i.e. by including a sufficiently large number of lags of these variables in the model. In doing so, we end up with a model like (4.10) in our example, where the residual ε_t is white noise and lagged values of both y_t and x_t are included.

It should also be remarked that specifications such as (4.5) are not usually given at the beginning of the building process of the empirical model. This is something that must be explored by the analyst, who should start from a general formulation and then proceed to test down for zero restrictions and common factors in the corresponding polynomials in order to get a simpler model. For instance, in the above example, the starting point could be a model like (4.6) that, for values of *r* and *s* sufficient large, can be taken as a general unrestricted model for testing the relevant hypothesis on its parameters. In particular, testing first for zero restrictions on the parameters of $a(L)$ and $b(L)$, we could find that the hypothesis H_0: $c=a_2=...= a_r= b_0= b_3=...= b_s =0$ is not rejected and so specify the model as

$$y_t = a_1 y_{t-1} + b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t$$

or equivalently as

$$(1-a_1L)\, y_t = (b_1 + b_2L)\, x_{t-1} + \varepsilon_t.$$

Now, this model provides a framework to test the hypothesis $a_1 = -b_2/b_1$. If this hypothesis is not rejected, this will imply that the dynamic filters of y_t and x_t has a common factor, $(1- a_1 L)$, and then the model will collapse to (4.11).

As we said before, a necessary condition for stability in the ADL model (4.8) is that the roots of $a(L)$ should be outside the unit circle. If this is the case, polynomial $a(L)$ can be inverted and, ignoring the intercept to keep notation simple, the ADL model can be written as

$$y_t = \sum_{i=1}^{k} \beta_i(L)\, x_{it} + \frac{1}{a(L)}\, \varepsilon_t. \qquad (4.12)$$

where $\beta_i(L)=b_i(L)/a(L)$. This equation decompose the value of y_t in two terms: the first term captures the systematic dynamics due to the influence of the explanatory variables

on the actual value of y_t while the second term reflects residual dynamics, which is the dynamics in y_t not explained by the k explanatory variables in the model. This representation is a special case of a *Transfer Function* model. We will go further into this kind of models in the next section. For the moment, it is enough to note that the residual term in (4.12), $\varepsilon_t / a(L)$, has its own lag structure and therefore, is somehow forecastable, while this is not the case in the ADL model in equation (4.8), where the disturbance term, ε_t, is already white noise.

Example 1: Dividends and Earnings

In this first example, the objective is to forecast an aggregate, economy-wide measure of dividend yield that is the dividend divided by share price. An obvious explanatory variable is aggregate earnings (profits) divided by share prices. The data used are

D_t – Dividend yield, aggregate dividend/price ratio
E_t – Aggregate company earnings/price ratio

The series are observed quarterly, starting in 1978:1 and ending in 2006:4 (UPDATE). The sample size is n = 116 and the source is Standard and Poors. This example can be representative for the dividend yield of a particular stock and the earnings of the corresponding company. For an individual investor it could be very useful to have an accurate forecast of the income which he could expect from this stock for, say, next quarter or next year. This could be forecast by an ARIMA model, using only past data on dividends, but since dividends depend on the earnings of the company, enlarging the univariate information set with earnings one can built a model which should provide more accurate forecasts. Besides, since in a given quarter, earnings are announced before dividends, that information of earnings could provide an additional improvement of the forecast of current quarter dividends, as we will see below.

The data are plotted in Figure 4.2. As you can see, these series seem to evolve rather smoothly with possible upward and downward local trends along time. Unlike, their first differences, plotted in Figure 4.3, move up and down in a sharply way around a constant zero mean, a typical pattern of the stationary series. In fact, both variables D_t and E_t are nonstationary and, for reasons to be explained in section 4.7, their relationship will be formulated on their stationary transformations, ΔD_t and ΔE_t.

INSERTAR FIGURE 4.2. + FIGURE 4.3

Since the variable to be forecast is D_t and we consider E_t as exogenous (we will formally test this hypothesis in next chapter), a single-equation model of the ADL type is proposed. We keep some observations left for forecasting evaluation (2005:1 – 2006:4(UPDATE)) and so the estimation is carried out only with data up to 2004:4. The general ADL formulation used to start the modelling process has been an ADL(5,5). Testing for zero restrictions in the coefficients of that model, it has been found that the null hypothesis that the coefficients of lags 1, 2, 3 and 5 of ΔD_t and of lags 2, 3 and 5 of ΔE_t are equal zero, is not rejected. This leads to the following model

$$\Delta D_t = 0.18\ \Delta D_{t-4} + 0.13\ \Delta E_t + 0.09\ \Delta E_{t-1} - 0.06\ \Delta E_{t-4} + \varepsilon_t. \qquad (4.13)$$

This model states that the increment in dividends in a given quarter is being explained by the contemporaneous increments in earnings and increments in earnings in the previous quarter and the previous year, and also by the increment in dividends in the same quarter of the year before. The model can also be written as:

$$\Delta D_t = 0.12\ \Delta D_{t-4} + 0.13\ \Delta E_t + 0.09\ \Delta E_{t-1} + 0.06\ (\Delta D_{t-4} - \Delta E_{t-4}) + \varepsilon_t.$$

This representation indicates that not only increments in earnings matter in explaining increments in dividends, but also the past differences between both increments, in particular the difference in the previous year.

In terms of lag-polynomials, model (4.13) can be written as

$$(1 - 0.18L^4)\ \Delta D_t = (0.13 + 0.09L - 0.06L^4)\ \Delta E_t + \varepsilon_t. \qquad (4.14)$$

Since both polynomials in (4.14) have not common roots, model (4.13) can not be simplified further and can be taken as a final model to forecast ΔD_t or D_t. In this last case, the model for D_t is derived from (4.13) by passing D_{t-1} to the right hand side as follows:

$$D_t = D_{t-1} + 0.18\ \Delta D_{t-4} + 0.13\ \Delta E_t + 0.09 \Delta E_{t-1} - 0.06\ \Delta E_{t-4} + \varepsilon_t.$$

We will further develop this example in next sections.

4.3.2. Transfer Function models

An alternative choice to ADL models for modelling dynamic relations are the *transfer functions* models mentioned above. These models can be represented in a more general form than that in equation (4.12). In fact, these models make use of rational distributed lag structures in modelling both the cross-variable dynamics and the residual dynamics. A general representation of this type would be given by the following equation:

$$y_t = \sum_{i=1}^{k} \frac{\omega_i(L)}{\delta_i(L)} x_{it} + \eta_t \qquad (4.15)$$

where $\omega_i(L)$ and $\delta_i(L)$ are finite order lag polynomials, $\omega_i(L) = \omega_{i0} + \omega_{i1} L + \ldots + \omega_{is_i} L^{s_i}$ and $\delta_i(L) = \delta_{i0} + \delta_{i1} L + \ldots + \delta_{ir_i} L^{r_i}$, and the residual term, η_t, follows an ARMA process given by:

$$\eta_t = \frac{\theta(L)}{\phi(L)} \varepsilon_t, \qquad (4.16)$$

where $\theta(L)$ and $\phi(L)$ are finite order lag polynomials and ε_t is white noise. Since we are assuming that the variables are stationary all the polynomials $\delta_i(L)$ and $\phi(L)$ must have all their roots outside the unit circle. We also assume invertibility and then all the roots of $\theta(L)$ are also outside the unit circle.

Equation (4.15) shows that y_t can be split up in two terms. The first one, $\sum_{i=1}^{k} \frac{\omega_i(L)}{\delta_i(L)} x_{it}$, represents the dynamic contribution of the explanatory variables to the contemporaneous value of the dependent variable. For each variable x_{it}, the filter $\omega_i(L)/\delta_i(L)$ embodies the dynamic relationship between this variable and y_t and its coefficients are called the *impulse response function* from x_{it} to y_t. This function, which will be further explained in section 4.4, represents how a transitory unit shock to x_{it} is transferred to the dependent variable in current and future periods.

The second term in the right hand side of equation (4.15), η_t, gathers the dynamics of y_t which are not explained by the explanatory variables. This term depends on its own past values, which can be derived from the values of x_t and y_t using (4.15), and therefore is somehow forecastable. Hence, the forecastability of y_t is not completely accounted for by the explanatory variables, but also by the forecastable part of the residual term η_t. This can be better illustrated by noting that the invertible and stationary ARMA structure of η_t can be approximated by a high order AR(p) polynomial, say

$$\frac{\theta(L)}{\phi(L)} \cong \frac{1}{a(L)}, \tag{4.17}$$

with $a(L)= (1–a_1L–\ldots– a_pL^p)$. Then the residual term η_t can be written down as

$$a(L)\eta_t = \varepsilon_t,$$

or equivalently as

$$\eta_t = (a_1L+\ldots+ a_pL^p)\eta_t + \varepsilon_t.$$

The first term in the right hand side of the above equation, say $\hat{\eta}_t = (a_1L+\ldots+ a_pL^p)\eta_t$, captures the part of η_t which is explained by its own past. Therefore, it is known given the model and having information on all the variables up to time t-1. Unlike, the second term, ε_t, is the random element (innovation) which appears at time t. Using this decomposition of η_t, model (4.15) admits an equivalent representation as

$$y_t = \sum_{i=1}^{k} \frac{\omega_i(L)}{\delta_i(L)} x_{it} + \hat{\eta}_t + \varepsilon_t.$$

In this equation the first two terms represent the forecastable part of the model, i.e. the conditional expectation of y_t given the information of all the explanatory variables up to time t and the information on y up to time t-1. The last term, ε_t, is the innovation.

The dynamics behind model (4.15) can be summarized in the scheme in Figure 4.4. In this figure, we first find the dynamic filter determining the way in which each explanatory variable x_{it} affects the dependent variable y_t. Secondly, we have the residual structure, given by an ARMA process on the innovations ε_t. The joint effect of both elements makes up the observed value of the dependent variable y_t. The equality in equation (4.15) implies that all the properties of y_t must also held in the right hand side of such equation. In Figure 4.4 we see that these properties come from the inputs,

explanatory variables and innovations (circles in the figure), and from the filters acting on them (rectangles in the figure). If the model is well specified –all the relevant explanatory variables are included and the filters in the model are correct–, the residual term η_t can be represented in terms of a white noise error, like in (4.16), and we say that the model is balanced. If the model is misspecified because it does not include the proper explanatory variables and/or filters, the only way to balance the equation is through an error term that is not white noise. Thus, we see once more that the analysis of the errors is crucial in model building. In section 4.7 we will come back on this figure to discuss the sources of the non-stationarity of y_t, when this is the case.

INSERTAR FIGURE 4.4

4.3.3. Further discussion

It was mentioned in section 4.3.1 that ADL models can be represented as transfer function models with some restrictions on the polynomials. In fact, model (4.12) is a particular case of (4.15) with $\delta_i(L)=a(L)$, $\omega_i(L)=b_i(L)$ for any i, $\phi(L)=a(L)$ and $\theta(L)=1$. On the other way round, a transfer function model could also be regarded as an ADL model under stationarity conditions on the dynamic filters. Under such conditions, the ARMA structure of the error term in (4.16) can be well approximated by a high order autoregressive structure, as in (4.17), and each rational polynomial associated with the explanatory variables can also be approximated as

$$\frac{\omega_i(L)}{\delta_i(L)} \cong \beta_i(L).$$

Thus, the transfer function model (4.15) can be alternatively written as:

$$y_t = \sum_{i=1}^{k} \beta_i(L)\, x_{it} + \frac{1}{a(L)}\, \varepsilon_t.$$

Note that this is a generalization of our example (4.9) that includes more explanatory variables and polynomials of any finite order. Multiplying now both sides of the equation above by $a(L)$, the transformation $a(L)$ will apply to each explanatory variable of the first term in the right hand side of the equation and the second component will reduce to the innovation ε_t. Hence, we will end up with the following representation:

$$a(L)\, y_t = \sum_{i=1}^{k} \beta_i(L) a(L)\, x_{it} + \varepsilon_t, \qquad (4.18)$$

which is an ADL model like that in (4.8), with $c=0$, $b_i(L)=\beta_i(L)a(L)$. Note that in this ADL model, $a(L)$ is a common factor in all the polynomials.

Therefore, we have seen that to get rid of the dynamic effect gathered in the residuals of the transfer function model (4.15), we can incorporate additional dynamic structure to y_t and x_{it} to end up with an ADL model like (4.18), where the error term ε_t is white noise.

Consequently, once we know the dynamic model, it does not matter in which form this is given, either ADL or TF model, because one formulation can approximate very well the other. Whether to use one or another representation is up to the user, the question arises when the model is not known and must be estimated from the data. How to proceed in that situation is an open debate and only some remarks about it will be briefly sketched below.

First, it must be noted that transfer function models are more complex than ADL models, in the sense that its specification requires not only the economic theory to establish the relevant explanatory variables to be included, but also some knowledge about the dynamic interrelations between the endogenous and exogenous variables, which is rarely available. This knowledge is important to establish the structure of the rational lag polynomials in (4.15) which represent the response of *y* to changes in the exogenous variables. Thus, for each explanatory variable *x*, the information required will be at least referred to: (i) whether there is a delayed or an immediate response of *y* to *x* and (ii) which kind of response is expected, a short or a large one over time.

Meanwhile, the ADL model in (4.8) only needs to call on economic theory to select the explanatory variables. Then, the number of lagged values of each of the variables in the model will be detected empirically in the stages of estimating and diagnostic checking. Consequently, the transfer function model becomes useful when there is some prior information about the response of each explanatory variable to the dependent variable. If such information is not available, the formulation of an ADL model with sufficiently large orders in all the polynomials to ensure a white noise residual is a good choice. Anyhow, the decision as to which formulation is appropriate is sometimes based on pragmatic issues, such as ease of estimation and model specification. In general, we could say that econometricians have tended to favour ADL models, while transfer function models are more widely used among engineers and other scientists working with experimental data. Herein, we will focus on ADL models.

4.4. THE IMPULSE RESPONSE FUNCTION

An appealing feature of the *TF* model (4.15) and, hence, of the transfer function representation of the ADL model in (4.12), is that it provides an interpretation in economic terms of the dynamics of the system. Let us denote:

$$\upsilon_i(L) = \omega_i(L)/\delta_i(L)$$

so that equation (4.15) turns out to be:

$$y_t = \sum_{i=1}^{k} \upsilon_i(L) x_{it} + \eta_t$$

where η_t is given in (4.16) and the lag polynomial $\upsilon_i(L)$ can be expanded as

$$\upsilon_i(L) = \upsilon_{i0} + \upsilon_{i1} L + \upsilon_{i2} L^2 + \ldots$$

The coefficients of this lag polynomial, $\{\upsilon_{i0}, \upsilon_{i1}, \ldots\}$, are called the *impulse response function (IRF)* and track the complete dynamic response of *y* to the explanatory variable

x_i. In particular, υ_{ij} represents the effect on the dependent variable of a transitory unit shock in the variable x_i occurred j periods before. Furthermore, the sum of all these coefficients, $\{\upsilon_{i0}+ \upsilon_{i1}+\ldots\}$, is usually called the *gain* and obviously represents the accumulation of all the impacts on y from a transitory unit shocks in x_i. This value is obtained by taking L=1 in the corresponding lag polynomials, i.e.,

$$gain = \upsilon_i(1) = \omega_i(1)/\delta_i(1) = (\omega_{i0}+\ldots+\omega_{is})/(\delta_{i0}+\ldots+\delta_{ir}).$$

In many cases the impulse response function can be characterized as having three phases. These are shown in Figure 4.5, which displays the effect on y_{t+j}, $j \geq 0$, of a transitory change in an explanatory variable x_t.

INSERTAR FIGURE 4.5

First, there can be a delaying phase (Phase I) in which a transitory change in x_t does not affect the variable y. This phase lasts s periods if the reaction of y does not occur up to s periods after the change in x. For example, let us consider a very simple model like

$$y_t = c + 0.8\, x_{t-2} + 2\, x_{t-3} + \varepsilon_t, \qquad (4.19)$$

where y_t and x_t could be the weekly sales of a certain product and the expenses on advertising it, respectively. A transitory change in the advertising expenditure made at time t* does not affect weekly sales till two weeks later.

In order to better understand the meaning of the IRF in this example we will assume that the system is in equilibrium, i.e. the error term is zero, and x_t is taking an equilibrium value, say x^e. Then we have

$$y_t = y^e = c + 0.8\, x^e + 2\, x^e$$

However, if a unit transitory shock in x occurs at time t*, so that x_{t*} becomes x^e+1, and afterwards comes back to x^e, we can write down the model equation for the periods at and after that change as follows

$$y_{t*} = c + 0.8\, x^e + 2\, x^e = y^e$$
$$y_{t*+1} = c + 0.8\, x^e + 2\, x^e = y^e$$
$$y_{t*+2} = c + 0.8\, (x^e + 1) + 2\, x^e = (c + 0.8\, x^e + 2\, x^e) + \mathbf{0.8} = y^e + \mathbf{0.8}$$
$$y_{t*+3} = c + 0.8\, x^e + 2\, (x^e + 1) = (c + 0.8\, x^e + 2\, x^e) + \mathbf{2} = y^e + \mathbf{2}$$
$$y_{t*+4} = c + 0.8\, x^e + 2\, x^e = y^e$$

and so on. In these equations above, we note that the values of y_{t*} and y_{t*+1} are unaffected by the change in x_{t*} and only at time t*+2 changes start in y. Thus, we can say that in this relationship there is a delaying phase of two periods, t* and t*+1, in which there is no impact on sales due to the change in the advertising expenditures at t*. Obviously, in a contemporaneous relationship between x_t and y_t, as in the dividends and earnings example (4.13), this delaying phase does not exist.

After the delaying phase, if it exists, or otherwise from the very beginning, y_t can show a free response phase (Phase II), i.e. a response with unconstrained parameters to the transitory change in x_{t^*}. In the sales-advertising example above, the response of y starts at t^*+2 and is given by the extra terms in the third and fourth equations above, namely 0.8 and 2, respectively. These terms represent the change in the value of y at time t^*+2 and t^*+3 due to the transitory unit impulse in x_{t^*}. In fact, if y_t and x_t are measured in logs, this model tells us that an increment of 1% in advertising at time t^* will increase sales by 0.8% at time t^*+2 and by 2% at time t^*+3. From t^*+4 onwards the effect of that change becomes again unnoticed. Thus, in this example, after two periods of no response (t^* and t^*+1), the subsequent free response phase lasts two periods (t^*+2 and t^*+3) where the effects on the endogenous variable are given by free specific parameters not related to each other. The same conclusion would have been reached by noting that the impulse response filter in model (4.19) is $\upsilon(L)= 0.8L^2 + 2L^3$. The interpretation of its coefficients is the same as stated above, with only two values different from zero, namely $\upsilon_2= 0.8$ and $\upsilon_3= 2$.

Finally, in some cases, there is a third phase, which we call the convergence phase, in which the effect on y of a transitory change in x declines smoothly to zero. Obviously, this phase does not exist in the example above, where immediately after period t^*+3, the change in x has no longer effects on y and the impact on y_{t^*+3+h}, $h>0$, becomes abruptly zero without a smooth transition. However, the convergence phase does appear in models with lagged endogenous variables. For instance, in a model like

$$y_t = c + 0.8\, x_{t-2} + 2\, x_{t-3} + 0.5\, y_{t-1} + \varepsilon_t, \tag{4.20}$$

the presence of the first lag of the endogenous variable with a coefficient 0.5 keeps 50% of the value of y_t in next period. In general, if this coefficient is α, 100α % of the value of y_t is always kept in next period. In order to illustrate this property, let us write down model (4.20) as

$$(1\text{-}0.5L)\, y_t = c + (0.8\, L^2 + 2L^3)\, x_t + \varepsilon_t, \tag{4.21}$$

and pass the autoregressive polynomial multiplying y_t to the right hand side. Then the model becomes:

$$y_t = c' + \frac{0.8L^2 + 2L^3}{1 - 0.5L}\, x_t + \frac{\varepsilon_t}{1 - 0.5L},$$

and the impulse response filter, $\upsilon(L)$, takes the form:

$$\upsilon(L)=b(L)/a(L)$$

where $a(L)=1\text{-}0.5L$ and $b(L)= 0.8L^2 + 2L^3$. Now the filter has a rational form and $a(L)$ is the polynomial that generates the convergence phase in the impulse response function. In this case, the coefficients of $\upsilon(L)$ are obtained by considering that

$$(1\text{-}0.5L)\ \upsilon(L)= (0.8L^2 + 2L^3).$$

Expanding $\upsilon(L)$ as $\upsilon(L)=\upsilon_0+\upsilon_1 L+\upsilon_2 L^2+\ldots$ and working out the product of polynomials on the left hand side of the equation above, the coefficients $\{\upsilon_0, \upsilon_1, \upsilon_2, ..\}$ are derived by equating powers of L on both sides of the equation. In particular, it turns out that:

$$\upsilon_0 = 0;$$

$$\upsilon_1 = 0;$$

$$\upsilon_2 = 0.8;$$

$$\upsilon_3 = 2 + \mathbf{0.5} * 0.8 = 2.4;$$

$$\upsilon_4 = \mathbf{0.5}\ (2 + 0.5 * 0.8) = 1.2;$$

$$\upsilon_5 = \mathbf{0.5}\ [0.5\ (2 + 0.5 * 0.8)] = 0.5^2\ (2 + 0.5 * 0.8) = 0.6;$$

and so on. Thus, in this model, a change in x at time t^* has no effect on y at times t^* and t^*+1 ($\upsilon_0=\upsilon_1=0$) but afterwards the sequence of responses of y at times t^*+2, t^*+3, … are given by the coefficients υ_2, υ_3, …, respectively. The first non-zero coefficient, υ_2, tells us that an increase of 1% in advertising expenses in a given week will generate an increase in sales of 0.8% two weeks later. The subsequent coefficient tell us that there will be an additional effect of 2.4% three weeks later and from there onwards the subsequent effects will be just the 50% of the previous one, i.e. 1.2% four weeks later, and so on. Summing up all these coefficients we get the gain or cumulative effect of the transitory unit shocks along all future time, which is given, in this case, by $\upsilon(1)=b(1)/a(1)=(0.8+2)/(1-0.5) = 5.6$. That is, an increase of 1% in current advertising expenses will amount to a global increase of 5.6% on future sales.

Note that in this example, from t^*+4 onwards, the response of y to a transitory change in x_{t^*} is given by some coefficients which are constrained by the expression

$$\upsilon_j = 0.5\ \upsilon_{j-1} = 0.5^{j-3}\upsilon_3, \quad \text{for } j \geq 4, \tag{4.22}$$

with $\upsilon_3=2.4$. Note also that as the periods go further in the future, the impact decreases and it becomes zero in the very far apart lags ($\upsilon_j \to 0$ as $j \to \infty$). This is so because the $a(L)$ polynomial in (4.21) has its roots outside the unit circle, in fact, it has only one root with a value of two. When the IRF converges to zero, either in a smooth way, if phase III exists, or in an abrupt one if it does not, we see that the relationship between these two variables is such that an impulse –a transitory shock- in one variable (exogenous) has not a permanent effect on the other (endogenous).

To summarize, in this second example, which is actually the one represented in Figure 4.5, there is a delaying phase of no response which lasts two periods (t^* and t^*+1), following by a second phase of free response in periods t^*+2 and t^*+3, and finally a convergence phase starting at t^*+4. Thus, this example shows that the presence of the first lag of the endogenous variable (this can be generalized to the presence of more lags or lags of any order) extends the dynamic relationship between y_t and x_t beyond the horizon given by the largest lag of x_t included in the model. Furthermore, in this case, this extension is done according to the converge restriction (4.22), which ensures that the effect on y of changes in x tends exponentially to zero because the coefficient of y_{t-1} in (4.20), say $a_1=0.5$, is such that $|a_1|<1$, or equivalently the root of the lag-polynomial $a(L)$ in (4.21), namely 2, is outside the unit circle.

It becomes clearer now that dynamic relationships given by filters that only include lagged values of the explanatory variables, as that in (4.19), only contain phases I and II of the prototype described in Figure 4.5. This kind of filters, where $\upsilon(L)=b(L)$, are sometimes denoted as moving average filters, and they are appropriate if a change in x is transferred to y for only a few periods later. For the presence of phase III we need a model with a rational filter, $\upsilon(L)=b(L)/a(L)$, i.e. a filter including lagged values of the endogenous variable, as that in (4.20). In this case, $a(L)$ is called the autoregressive polynomial of the filter and its presence is suitable when the effect of a change in the explanatory variable takes a long time to completely work through to the response variable.

The points of interest about a IRF are: a) the presence or not of a lag response, b) the presence of endogenous lags, indicating that the dynamic relationship between y_t and x_t is longer than what the largest lag of x_t in the model indicates and c) that there are short-run effects, represented by coefficients, $\{\upsilon_0, \upsilon_1, \ldots\}$, and a long-run effect given by the sum of all υ_j coefficients. The specific form of the IRF, i.e. the values of the coefficients υ_j, can not be taken too seriously, because minor changes in the coefficients of $b(L)$ and $a(L)$ –which are not significantly different from the estimated values– could imply substantial changes in the IRF.

In general, the polynomial $a(L)$ will be of any finite order r and could be factorized like

$$a(L) = (1-G_1L) \ldots (1-G_rL) \tag{4.23}$$

where G_i, i=1,..,r, are the inverses of the r roots of the polynomial. For a convergent phase III in the impulse response function all the coefficients G_i's in (4.23) must be in absolute value less than one. Thus if at least one root, say $1/G_r$, is inside the unit circle, then $|G_r|>1$ and the coefficients of the filter polynomial, $\upsilon(L)=b(L)/\{(1-G_1L) \ldots (1-G_rL)\}$, will make up an increasing sequence. In this case, the model becomes explosive, a type of behaviour that can be excluded in the relationships between economic variables. In the case of a unit root in (4.23), say $G_r=1$, the series y_t is non-stationary and the coefficients of $\upsilon(L)$ do not converge to zero but tend to a constant value. In this case, the general ADL model can be written as

$$(1-G_1L) \ldots (1-G_{r-1}L)\, \Delta y_t = c + b(L)\, x_t + \varepsilon_t$$

with $\Delta y_t = (1-L)\, y_t$ as the dependent variable. Thus, an ADL(r,s) model with a unit root for a non-stationary variable y_t can be converted into an ADL(r-1,s) for the stationary variable Δy_t. Hence, we can see that if we relate a I(1) variable, y_t, with a I(0) variable, x_t, the only way to explain the non-stationarity of y_t is by having a unit root in the $a(L)$ polynomial in the model (4.7) and this implies that what we really have is a model relating Δy_t with x_t, in which both variables are I(0). This would happen if we were trying to relate a I(2) price index with a I(1) interest rate, indicating that the relationship is between the increments of prices (inflation) and interest rates. Of course this relationship will in general be much more complex with more explanatory variables, but the point of interest now is to realise that a relationship between a variable I(d) an another I(d-1) must be formulated on the first differences of the first one.

To summarize, if $|G_i|>1$ for some i, then y_t responds in a explosive way to an impulse change of x_t and, as already mentioned, this is an unrealistic model for economic data. Cases with $|G_i|<1$ for all i are stationary models, provided that the explanatory variables and the residual term are also stationary, and may be appropriate for economic series which are stationary in levels or are stationary after differencing. Finally, cases with $G_i=1$ implies that the model should be specified with Δy_t as the dependent variable.

4.5. INFERENCE AND DIAGNOSTIC CHECKING IN *ADL* MODELS

Once a certain approach for modelling has been chosen, the starting point should be a very general model that may be reduced by a sequence of inference procedures until a more specific, reasonably parsimonious and readily interpretable formulation is obtained. This reduction process includes testing on the significance of the parameters of the model, selecting among competing models and carrying out careful diagnostic to ensure that the error term is white noise. This is the *general-to-specific* methodology, sometimes also referred to as the *London School of Economics* methodology. This approach has played a major role in recent empirical econometric modelling as it has several advantages over the simple-to-general approach; see *i.e.* Hendry (1995).[3] Moreover, automatic model selection from a general-to-specific methodology can be implemented through a computer program called *PcGets* developed by Hendry, Doornik and Krolzig (??)[4]. We will further describe this methodology and illustrate the use of this package throughout next sections.

In this section, we will discuss briefly some of the general issues that arise when one attempts to specify and estimate dynamic regression models. Under the assumption that the explanatory variables are exogenous, ADL models can be estimated by Ordinary Least Squares (OLS). For instance, in the ADL(1,1) case, this entails considering (4.5) as a regression equation of y_t on the vector of regressors $\boldsymbol{x}_t=(1,\ y_{t-1},\ x_t,\ x_{t-1})'$ with parameter vector $\boldsymbol{\beta} = (c,\ a_1,\ b_0,\ b_1)'$ and ε_t white noise. Then, the OLS estimator is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

where $\mathbf{X}$ is the matrix with the observations of $\boldsymbol{x}_t$ and $\mathbf{Y}$ is the column vector with the observed values of y_t, t=2,…,n, with n being the sample size. Note that dividing and multiplying the right hand side of the above equation by n, $\hat{\boldsymbol{\beta}}$ becomes the sample covariance vector between y_t and $\boldsymbol{x}_t$ premultiplied by the inverse of the sample covariance matrix of $\boldsymbol{x}_t$. Thus, $\hat{\boldsymbol{\beta}}$ is a function of the data and therefore is a random variable and the estimator of its variance-covariance matrix is $\hat{\sigma}_e^2(\mathbf{X}'\mathbf{X})^{-1}$, where $\hat{\sigma}_e^2$ is an unbiased estimator of the residual variance.

The first problem that arises here is that $\mathbf{X}$ is stochastic because x_t can be stochastic and certainly the lagged dependent variables are and some of its elements are correlated with past disturbances, rendering OLS estimators biased in small samples. There may also exist small sample biases in estimating the variances of the coefficient estimates, and this could lead to wrong decisions when using the *t*-values to test the significance of

[3] It must be noted that abandoning the class of linear models, the general-to-specific methodology could not be a good one in non-linear models.
[4] See Appendix A for a brief review of what PcGets can do.

such coefficients. However, OLS estimators still retain desirable asymptotic properties, provided that ε_t is white noise. In particular, the OLS estimator for $\boldsymbol{\beta}$ and the variance estimators are consistent and asymptotically unbiased and $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed. This implies that the usual tests on the parameters of the model are asymptotically justified in this case. Moreover, if ε_t is normally distributed, the OLS estimator is essentially the Maximum Likelihood estimator and therefore is asymptotically efficient.

The fulfilment of the asymptotic properties of the OLS estimators relies entirely on the assumption that ε_t is white noise. If this is not the case and the disturbances are serially correlated, OLS may not be even consistent. Therefore, checking that ε_t is white noise is essential for the asymptotic properties to hold. Other diagnostic checks should be applied to the estimated model and methods for comparing alternatives are available to help make a choice.

Four types of statistics are provided by most computer regression packages to help evaluate the quality of the estimated model. First, those concerning the estimation results, such as the estimates of the parameters themselves, their standard deviations and the *t*-values associated with them. Second, some summary statistics for comparison between alternative fitted models, such as the R^2 coefficient, the residual variance and some information criteria like the AIC and the SIC. Third, some diagnostic tests for possible mis-specification, such as those concerning the hypothesis of white noise, homoscedasticity and normality of the residuals. Finally, some statistics to test if the model is stable along time, although these are not always included in the computer packages. All of these statistics will be explained below. In all cases, it will be assumed that the sample size is large enough for the asymptotic results to be a reasonably valid approximation to the distributional properties of the statistics computed with the finite sample available.

Estimation results

- Attached to every estimated parameter in the model most packages display a standard error, a *t*-value and a probability called *p*-value or *t*-prob. As it was explained in Chapter ??, a *t*-value is the ratio of the estimated parameter value divided by its standard error. Under the null hypothesis that the parameter is zero, this ratio has a Student-*t* distribution. Hence, testing that hypothesis entails comparing the *t*-value with the corresponding percentile of the Student distribution. Furthermore, a small *p*-value for this parameter means that is very unlikely that a model with this parameter being zero has generated our data. Hence, a small *p*-value implies that the parameter is significantly different from zero and the variable with this coefficient should be kept in the model. The usual rather inexact convention is to say that the parameter is significantly different from zero if the *t*-value is at least 2 in magnitude or if the estimated parameter is at least twice the standard error in magnitude. Alternatively, it is said that the parameter is significant if its associated *p*-value is less than 0.05. This interpretation corresponds to assuming a 5% risk of wrongly rejecting the null and is appropriate provided certain assumptions about the model hold. Unfortunately, these assumptions are sometimes not correct, as it will happen in regression models with nonstationary variables or models with heteroscedastic errors, which makes interpretation of *t*-values difficult. Below in this

section we discuss the tests for heteroscedasticity and section 4.7 is devoted to models with non-stationary variables.

Model-selection criteria

- The residual variance or, more commonly, the standard deviation of the error term (usually called the standard error of the regression), is almost always provided. In this case, a model with a smaller value is preferred to one with a larger value.
- A value of R^2 or R_c^2, which is R^2 corrected for degrees of freedom so that the number of regressors used is accounted for, is also provided. These are interpreted as measures of the *goodness of fit* of the model and are defined by

$$R^2 = 1 - \frac{\text{(variance of error terms in model)}}{\text{(variance of the dependent variable explained by model)}}, \quad R_c^2 = 1 - \frac{n}{n-K}(1-R^2)$$

where n is the sample size and K is the number of regressors used.

The general interpretation of these quantities is that a model with a larger value is preferred to one with a smaller value, provided that the endogenous variable is the *same*. However, the R^2 or R_c^2 values cannot be usefully interpreted in absolute terms. It is not correct, for instance, to say that a model with R^2 greater than 0.8 is satisfactory but that one with R^2 less than 0.3 is unsatisfactory. The value achieved for R^2 depends solely on how explainable one variable is in terms of its dynamic relationships with other variables. Thus, R^2 can only be used to *rank* alternative models that are trying to explain the *same* dependent variable. For example, if one fits dynamic models for either y_t or its first difference, Δy_t, as dependent variables, one can not use R^2 values to compare both specifications. Then the comparison should be done in terms of the residual variance which is, in both cases, a measure of the predictability of both dependent variables one period ahead. Actually, for forecasting purposes the criterion that one is most interested in when comparing different models is the residual variance (or the standard deviation) of the error term, as this will be a guide to the variance of the forecast errors when the model is used to form forecasts. The standard deviation has the advantage of being in the same units as the dependent variable and if this one is in logs, the standard deviation multiplied by 100 can be interpreted as percentage points.

For example, if one has the pair of models

$$y_t = 18 + 0.64\, x_{t-1} + \varepsilon_{1t}, \qquad R_c^2 = 0.86,$$
$$\Delta y_t = 0.23\, \Delta x_{t-1} + 0.13\, \Delta z_{t-1} + \varepsilon_{2t}, \qquad R_c^2 = 0.22,$$

and the residual variances are $\text{var}(\varepsilon_{1t})=3.6$ and $\text{var}(\varepsilon_{2t})=1.3$, then the second model is more useful for forecasting, as its variance of error is smaller. This is so because although the second model has $\Delta y_t = y_t - y_{t-1}$ as the dependent variable, it could also be regarded as a model for y_t by just passing the term y_{t-1} to the right hand side. Therefore, comparing in both models the unpredictability of y_t one period ahead amounts to compare the unpredictability of their error terms, and this is measured by

their variances. However, as we stated above, the R^2 values are not useful here as the form of the dependent variable, and therefore its variance, is not the same for the two models. In any case, other diagnostic statistics for both models, as those explained below, should also be considered.

- The Akaike (AIC), Schwarz (SIC) and Hannan-Quinn (HQ) information criteria (see Chapter ???, page ?? for their exact definitions) are based on the minus maximised value of the respective likelihood function and include a penalty for the number of estimated parameters, so that the more parsimonious models are less penalized. The decision rule is to select the model for which such criterion is minimum, that is, to select the fitted model which makes more likely our observed data. The SIC criterion's penalty is harsher than AIC and therefore it tends to select a more parsimonious model.

Model diagnostic tests

- If an ADL model without an intercept has been estimated, a test for the null hypothesis that the mean of the errors is zero should also be carried out to ensure that the residuals are zero-mean white noise. In fact, in the estimation process models should always include an intercept and if this is not significantly different from zero at the final stage, it will be dropped from the model, mainly when the model is going to be used for forecasting.
- The Durbin-Watson *d* statistic is a classical statistical tool originally designed to deal with annual data and to check only for first-order uncorrelatedness. To a close approximation, the *d* statistic is given by $d \approx 2(1-r_1)$, where r_1 is the estimated first order autocorrelation for the error of the model. Thus, if errors are white noise, *d* will approximately equal 2, but it will be less than 2 if the errors have a positive first autocorrelation. In ADL models the *d*-statistic is biased toward 2 and it is more appropriate to use the *h*-statistic proposed by Durbin (1970), whose asymptotic distribution, under the null hypothesis of ε_t being white noise, is standard Normal. Thus, the null is rejected against positive first-order autocorrelation if *h* is larger than the corresponding critical value in the Normal distribution. Nowadays, these tests have been outperformed by other statistics which consider error autocorrelations of order greater than one, as those mentioned below.
- Tests on the correlogram of the residuals are the main tools to check that these are white noise. If this is so, the asymptotic distribution of a particular order sample autocorrelation of the residuals should be N(0,1/*n*). Therefore, the plot of such correlations against the lag, i.e. the correlogram that was introduced in Chapter 3, together with the 95% confidence bands, approximated by $\pm 1.96/\sqrt{n}$, may help to decide on rejecting the hypothesis of uncorrelated disturbances. Some computer programs use better approximations than 1/*n* for the sample variance of the autocorrelations and then the corresponding confidence bands are not just $\pm 1.96/\sqrt{n}$.
- The above statistics are designed to test the zero null hypotheses for each term of the error autocorrelation function individually. Additionally, a portmanteau test for the null hypothesis that *P* successive autocorrelations are jointly equal to zero may be carried out with the Box-Ljung *Q*(*P*) statistic defined in Chapter 3. High values of this statistic lead to a rejection of the null. However, it is not generally appropriate

to apply these tests to the residuals from dynamic regression models that include exogenous variables as well as lagged values of the dependent variable.

- Alternatively, the null hypothesis of white noise residuals against the alternative of an autoregressive structure can be performed using the Breusch-Godfrey statistic. This test is appropriate to detect higher-order autocorrelations and is based on the Lagrange Multiplier (LM) principle. For instance, to test against the alternative of AR(P) residuals with autoregressive parameters $\phi_1,\ldots,\phi_P$, a test for the null $H_0:\phi_1=\ldots=\phi_P=0$ (white noise) is performed by regressing the residuals on all the regressors of the original model and the P lagged residuals and testing the joint significance of the latter. High values of the LM test compared to the critical values of the corresponding asymptotic Chi-squared distribution rejects the null.

- The hypothesis of white noise for the residuals also requires them to have constant variance, i.e. to be homoscedastic. Thus, tests for unconditional heteroscedasticity are usually provided by most econometric computer packages. Heteroscedasticity is a particularly strange word meaning unequal variance, and if a model is estimated for a situation in which the errors have this property, somewhat unsatisfactory estimates of the parameters and, particularly, of the associated t statistics, will usually occur. A frequently used strategy for at least reducing this problem is to build models on logarithms of the data rather than on the raw data. However, taking logs does not necessary alleviate the problem and so a formal test for heteroscedasticity should be performed. The most commonly used one is the White test where the null of homoscedasticity is tested against the alternative that the residual variance depends on the regressors and its squares. High values of the test statistic indicate rejection of the null. Other tests for heteroscedasticity, such as Breusch-Pagan-Godfrey can also be performed.

- The ARCH test for AutoRegressive Conditional Heteroscedasticity is sometimes also displayed by some packages. This test is based on analysing whether the squared residuals are uncorrelated. If the residuals were white noise, so there should be their squares. Thus, if this is not the case, the model is regarded as misspecified.

- The most widely used test for the hypothesis of normality of the residual term is the Jarque-Bera test, which is based on the asymmetry and kurtosis of the residuals. This statistic measures to what extent the values of the sample asymmetry and kurtosis differ, in squared terms, from those of a normal distribution, namely 0 and 3, respectively. The test statistic has an asymptotic Chi-squared distribution and so large positive values will reject the null. The results of the normality tests are usually accompanied with some plots such as the histogram and the estimated density of the residuals, which is compared to the Normal density, and the residual QQ plot. The latter graphs the quantiles in the cumulative distribution function of the residuals versus those of the corresponding Normal distribution. That is, it represents points $(Q_R(p),Q_N(p))$, for several values of p $(0<p<1)$, where $Q_R(p)$ is the p-quantile of the residual sample cumulative distribution, i.e. the value which cumulates a $100p$% frequency so that $100p$% of the residuals are not larger than it, and $Q_N(p)$ is the corresponding quantile of the Normal distribution, so that there is a probability p for the Normal density to be less than $Q_N(p)$. If residuals are Gaussian, these two values should be quite similar and the points should be scattered around the diagonal straight line. Departures from this line indicate departures from Normality.

- PcGets also performs some checks for parameter constancy along the sample period. This includes computing the Chow test by splitting the sample in two subsamples and cross-validating the estimated model in both of them. To test for the null of parameter constancy, the Chow procedure consists of comparing the Residuals Sum of Squares from the two subsamples through a ratio statistic with an F distribution. High values of this statistic indicate rejection of the null.
- The task can be completed with recursive procedures to analyze the behaviour of the parameter estimates, the standardized residuals and the Chow test itself along the sample. This entails estimating the model and computing the statistics for a first sample size and then repeats the estimation procedure for enlarged samples obtained by adding one observation at each stage. That is, if we start estimating the model based on the first M observations with $M<n$, being n the full sample size, we will have parameter estimates and test statistics values at each t, for $t = M, M+1,\ldots,n$. This allows the analyst to see how the estimated model evolves over time and discover possible changes in the model parameters along the sample period.

Some very relevant diagnostic statistics described above are those related to the absence of autocorrelation in the residuals. Many unsophisticated users of regression techniques could be content to get a high R^2 value and ignore an unsatisfactory value of $Q(P)$, for instance, but this can lead to very poor models. As it has already been remarked, the assumption that the errors are white noise is essential for the asymptotic properties of the estimators of the ADL model to hold and, therefore, for the reliability of the other statistics explained above. Indeed, it can be shown that if OLS estimators of a regression model are formed when the errors are not white noise, then the t-values are overestimated in magnitude, so that parameters that are actually insignificant from zero may appear to be significant. In the extreme case, when the first-order correlation in the estimated residuals is near 1, so that d is very small, both theory and statistical experimentation have shown that spurious results not only can, but are very likely to occur. This is particularly likely if the levels of economic variables are used in the model. In this section we are restricted to regression models with stationary variables, but if the model is specified using variables in levels, the stationary condition could not be fulfilled. In fact, as it was discussed in Chapter 3 such variables often need to be differenced to achieve stationarity. Using them in an undifferenced form can easily lead to spurious regressions, which can be distinguished by very low d values even though the R^2 values may seem to be satisfactory or even rather high (see section 4.7.1 for a further discussion on this issue). If differenced data are used, the problem of non-white-noise errors will not totally disappear, but the effect is likely to be very much less important. For forecasting purposes, there is no preference for a model on levels rather than on changes, as forecasts are easily produced from either model, but there should be a strong preference for a model in whose structure and estimated parameters one can have some confidence.

It is hoped that this section will at least have given an indication of the very real and important practical problems that arise when using dynamic regression models. These problems are often ignored by unsophisticated users of such models and often result in unsatisfactory or suboptimal forecasts. A complete description of all aspects of these problems and the strategies to overcome them is well beyond the scope of this text. Descriptions can be found in the more advanced text in econometrics and forecasting.

4.6. REGRESSION MODELLING: AN EMPIRICAL EXAMPLE

The examples shown here are designed to illustrate the approach and issues discussed in the previous sections. The models are chosen to be useful for forecasting.

Example 1: Dividends and Earnings (continued)

In the example on Earnings (E_t) and Dividends (D_t) introduced in section 4.3.1, a model for the dynamic relationship between the variables in first differences, ΔD_t and ΔE_t, was established based on quarterly data for the period 1978:1 - 2004:4; see equation (4.13). Now we will describe in more detail the estimation process and diagnostic checking that led to such model, including the estimation results, statistics for model selection and some diagnostic tests on residuals.

Table 4.1 displays the estimation results from the fitted model with ΔD_t as dependent variable and ΔE_t as explanatory variable. The model selected includes as regressors the endogenous variable lagged four periods and the contemporaneous explanatory variable as well as its first and fourth lags. The *p*-values associated with the *t*-statistics are very low, especially for the contemporaneous and lagged Earnings, indicating that these variables are highly significant.

TABLE 4.1

The Durbin-Watson statistic, reported in Table 4.1, and the Box-Ljung statistics for residual correlation at different lags, displayed in Table 4.2, indicate that the residuals seem to be white noise. This result is confirmed by the residual correlogram, whose values are also given in table 4.2 and displayed in panel (b) of Figure 4.6. In this Figure, it can be observed that all the sample correlations fall inside the 95% confidence bands, indicating that the null hypothesis of uncorrelated residuals is not rejected. Also the LM test for uncorrelated residuals fails to reject the null against the alternative AR(4) residuals and the same happens with the Heteroscedasticity test, which does not reject (at 5% significance level) the hypothesis of homoscedastic residuals; see Table 4.1. The plots shown in the bottom panels of Figure 4.6 indicate that the hypothesis of Normality is not unrealistic. The Q-Q plot moves around the diagonal straight line and the histogram and estimated density of the residuals seem to fit quite well with the Normal density. The test of Normality in Table 4.1 confirms this impression with a large *p*-value that fails to reject the Normality assumption. Finally, the two Chow tests for parameter constancy along two subsamples, reported in Table 4.1, neither reject, so we can conclude that the estimated model is rather stable along the period considered.

TABLE 4.2 + FIGURE 4.6

Let us now consider the Impulse Response Function which in this case is given by

$$\upsilon(L)= \frac{b(L)}{a(L)} = \frac{0.13+0.09L-0.06L^4}{1-0.18L^4}.$$

Working out the product of the following polynomials

$$(1 - 0.18L^4)\,(\upsilon_0 + \upsilon_1 L + \upsilon_2 L^2 + \ldots) = (0.13 + 0.09L - 0.06L^4)$$

and equating powers of L on both sides of the equation, it turns out that:

$$\upsilon_0 = b_0 = 0.13;$$
$$\upsilon_1 = b_1 = 0.09;$$
$$\upsilon_{4k} = (a_4)^{k-1}\,(a_4\, b_0 + b_4) = 0.18^{k-1} \text{ x } (-0.04);$$
$$\upsilon_{4k+1} = (a_4)^{k}\, b_1 = 0.18^{k} \text{ x } 0.09$$
$$\upsilon_j = 0, \text{ otherwise.}$$

Therefore, in this example, the Impulse Response Function, displayed in Figure 4.7, has no delaying phase since the presence of ΔE_t as a regressor indicates an immediate response of ΔD_t to the exogenous variable. Also, there is a phase III with a declining non-free response subject to constraints imposed by the presence of the lagged endogenous variable.

INCLUDE ANOTHER EXAMPLE?

4.7. REGRESSION WITH INTEGRATED VARIABLES

In the previous sections, we have assumed that the variables involved in the dynamic regression model are all stationary. Herein we will allow for nonstationarity and explain the implications of this generalization on the specification, estimation and validation of the model. In particular, we will explain how to deal with ADL models that involve integrated variables.

First of all, let us recall a concept that was introduced in Chapter 3. A series is said to be integrated of order one, and denoted by $I(1)$, if it requires to be differentiated once to become stationary. Thus a stationary series is said to be $I(0)$. More generally, a series is integrated of order d, and denoted $I(d)$, if it is necessary to apply d differences to make it stationary. In this section we will concentrate on models with variables being $I(1)$.[5]

As a starting point, let us think of the general dynamic regression model defined in section 4.3.1, where the variable of interest, y_t, is split in two terms like:

$$y_t = \sum_{i=1}^{k} \beta_i(L)\, x_{it} + \frac{1}{a(L)}\, \varepsilon_t, \qquad (4.24)$$

where the first term captures the dynamics due to the explanatory variables and the second term gathers the residual dynamics. For simplicity of the exposition the constant term has been omitted in equation (4.24). As mentioned above, this model is said to be balanced if its disturbance, ε_t, is actually an innovation, i.e. it is white noise and is unrelated with all the explanatory variables and with the past of the endogenous variable. Thus, recalling Figure 4.4, it can be observed that in models like (4.24), if one

[5] Recall that with the notation $I(d,m^s)$ introduced in Chapter 3, an $I(1)$ variable can just have local oscillations of level if the mean of the differenced series is zero, which we denote as $I(1,0)$, and will have a systematic growth if that mean is non-zero, in which case the series is denoted as $I(1,1)$.

input (innovation) is white noise and the filters of the explanatory variables are stationary, as it is always assume with economic data, the nonstationarity of y_t can only be explained either by the nature of the explanatory variables, being themselves nonstationary, or by nonstationarity of the dynamic residual filter. The latter implies the presence of unit roots in the autoregressive polynomial $a(L)$ of the residual term in (4.24).

In the first case, where the endogenous variable y_t is integrated and the explanatory variables x_{1t}, ..., x_{kt} completely explain its nonstationarity, the residual term, which accounts for the possible effect on y_t of the omitted variables, will be stationary by definition and the model can be specified for the variables in levels. This type of models reveals that, although the variables are individually nonstationary, there exists a restriction on their non-stationarity, in the sense that being y_t and at least some x_{it} non-stationary, the linear combination $[y_t - \sum_{i=1}^{k} \beta_i(L)\, x_{it}\,]$ is stationary. In this case, we will say that the variables are cointegrated and the model will admit an alternative formulation in terms of what is called the equilibrium correction. This formulation will be more appropriate for interpretation as it will be written in terms of the parameters which are usually of economic interest. All these concepts will be further explained in subsections 4.7.2 and 4.7.3.

Unlike, if the endogenous variable y_t in (4.24) has a unit root and its nonstationarity is not completely accounted for by the possible nonstationarity of the exogenous variables, then the residuals of the model in levels are also not stationary, i.e. there is a unit root in its autoregressive polynomial $a(L)$, and we could end up with a case of spurious regression (see section 4.7.1). In these cases, the model will be better specified and estimated in first differences for the endogenous and the exogenous variables, so that with this transformation all the variables involved are stationary. In particular, if the autoregressive polynomial $a(L)$ in (4.24) has a unit root and (r-1) stationary roots, this can be written as

$$a(L)=(1- G_1L) \ldots (1- G_{r-1}L)\,(1-L) = a_{r-1}(L)\,\Delta$$

where $a_{r-1}(L)= (1- G_1L)\ldots(1- G_{r-1}L)$ and $\Delta=1-L$. Then, premultiplying both sides of equation (4.24) by Δ yields the following representation:

$$\Delta y_t = \sum_{\mathrm{i}=1}^{\mathrm{k}} \beta_{\mathrm{i}}(L)\ \Delta x_{it} + \frac{1}{a_{r-1}(L)}\,\varepsilon_t, \qquad (4.25)$$

where all the variables, endogenous and exogenous, appear in first differences and are in turn stationary and the residual term now is also stationary. Then, it is clear that a unit root in the polynomial $a(L)$ contributes to the nonstationarity of y_t.

The model (4.25) above can be alternatively written as

$$a_{r-1}(L)\ \Delta y_t = \sum_{\mathrm{i}=1}^{\mathrm{k}} b_{\mathrm{i}}^{*}(L)\ \Delta x_{it} + \varepsilon_t, \qquad (4.26)$$

where $b_i^*(L) = \beta_i(L)\, a_{r-1}(L)$. This equation allows for a new interpretation of the model as an ADL(r-1,s_1,...,s_k) model for the variables in first differences, with Δy_t being explained in terms of Δx_{1t}, …, Δx_{kt} and their lags and the lags of the endogenous variable. Once this model has been estimated, it is equivalent to interpret it as a model in levels with a non-stationary residual, as in (4.24), or as a model in first differences with stationary errors, like in (4.26). The first formulation could be more appropriate for forecasting, but the latter should be used for inference on the model parameters.

The discussion above shows up that, when relating economic variables in levels which have unit roots, it is very important to test whether the residual term, $\varepsilon_t/a(L)$, is non-stationary with an autoregressive unit root because, depending on the result of this test, a model in levels could be spurious. We will deal with this issue below when discussing cointegration in more detail.

4.7.1. Spurious regression

Many time series in business and economics are integrated and, as we discussed in previous chapters, show an upward trend over time. This type of series displays positive sample autocorrelations, indicating that a value above the sample mean tends to be followed by other values above this mean. Now, let us consider two of these variables and assume that they are unrelated with each other, like for example

$$x_{1,t} = x_{1,t-1} + \varepsilon_{1t}$$
$$x_{2,t} = x_{2,t-1} + \varepsilon_{2t}$$

where ε_{1t} and ε_{2t} are independent. If we run a static regression between these two variables, it is quite likely to obtain a "significant" regression coefficient – which is just the sample covariance of the variables divided be the variance of the regressor- and large R^2 and t-ratios, even when the variables are in fact unrelated and their only common feature is that in both variables values above/below the mean tend to be followed by values above/below the mean. This behaviour entails a high sample covariance between the two variables, $x_{1,t}$ and $x_{2,t}$, which in turn generates large values of the R^2 and t-ratios in the regression of $x_{1,t}$ on $x_{2,t}$. This phenomenon is usually called spurious regression and is known in the literature since the article on nonsense correlations by Udny Yule, published in 1926.

In these situations, the apparently satisfactory or even rather high values of the R^2 are usually accompanied by extremely low values for the Durbin-Watson d statistic, which indicates the presence of strongly autocorrelated residuals and hence a possibly misspecified model. In these cases, the message of the d statistic turns out to be essential to avoid drawing wrong conclusions about the relationship between the variables considered and also to point out through the possible cause of the misleading results, that is, an inappropriate error structure which is not white noise (in fact, it is not even stationary). As we have emphasized in previous sections, one of the problems raised by serially correlated errors is that the usual tests of significance are invalid, and its use might lead to spurious relations if also ignoring the message of other statistics such as the Durbin-Watson d or the sequence of values displayed in the correlogram.

A key contribution to this question was the paper of Granger and Newbold (1974) and the subsequent of Phillips (1986), who gave insights into the potential severity of the problem. In particular, they show that when fitting a simple regression model on two independent nonstationary variables, as in the example above, it is very likely to obtain rather high R^2 values that may suggest a very good fit, and very high *t*-ratios, i.e. very high probability of rejecting (wrongly) the hypothesis of no relationship between the two variables. This means that the exogenous variable may wrongly appear to contribute significantly toward explanation of the dependent variable. However this is nothing but a consequence of an incorrect use of the conventional linear regression methods which turn out to be highly inappropriate when the error series is not stationary. Indeed, the main conclusion from those papers is that when random walks or integrated processes in general are involved, the chances of "discovering" apparently significant relationships between them using conventional test procedures are very high. In fact, in models relating the levels of variables it will be the rule rather than the exception to find spurious regressions. It is also clear that a high value of R^2 or R_c^2 associated with a rather low value of the *d* statistic is not an indication of a good fit but that the model is some way misspecified.

Example 2: Consumer Price Index (Bread) and Industrial Production Index (Equipment parts)

In this example we want to illustrate the problem of spurious regression when relating unrelated integrated variables. The data used are monthly Consumer Price Index (CPI) of bread and Industrial Production Index (IPI) of Equipment parts, both considered in logs, from 1998:1 to 2008:7. Both series are plotted in Figure 4.8 and their autocorrelation functions are reported in Table 4.3. Clearly, both series are upward trendy and display positive sample autocorrelations. Also, the sample coefficient of correlation between them is rather high, namely 0.945.

FIGURE 4.8 + TABLE 4.3

The regression between these two variables (see Table 4.4) generates the typical results described above, that is a large value of the R^2 (0.89) and large *t*-ratios with associated zero *p*-values that wrongly suggest a good fit, together with a very low Durbin-Watson statistic (0.057) that indicates the presence of strongly autocorrelated residuals and hence a possibly misspecified model. The residual plot (Figure 4.9) and the residual correlogram (Table 4.5) confirm that the residuals are not white noise. In fact, the application of a unit root test (to be explained in section 4.7.3) indicates that these residuals actually posses a unit root.

TABLE 4.4 + FIGURE 4.9 + TABLE 4.5

If we now run a regression between the first differences of both variables, which are actually stationary, it turns out that the exogenous variable (Industrial Production Index) no longer appears to contribute significantly toward explanation of CPI of Bread; see Table 4.6. Therefore, this is a case of spurious regression.

TABLE 4.12

■

The problem of spurious regressions is really serious because by increasing the length of the series, the problem does not disappear but it worsens. To face this issue Granger and Newbold (1974) recommend first differencing the series to achieve stationarity before applying regression methods. If the relationship is genuine, it will appear with a significant regression on the differenced series while if it is spurious, the regression on the differences will not be significant, as in the previous example.

However, some econometricians argued that although the differencing strategy may actually alleviate the problem of non-stationary errors, it is not a general solution. In particular, as it is mentioned in Granger (2003), David Hendry first argued that the difference between two integrated variables could be stationary. Thus, sometimes a regression model between two nonstationary variables could lead to stationary residuals (see Example 3 in next section). In this case, a model for the first differences of these variables will yield residuals with a non-invertible moving average structure, leaving the problem unresolved. Moreover, taking first differences previous to modelling would eliminate the information about the long run behaviour of the variables, an issue that may become very important in many economic relationships.

The paper of Granger (1981) gave a new insight on this debate discovering the key concept of *cointegration* which provides a general framework for dealing with some of the problems stated above. The main aspects of cointegration and the related *error correction models* are discussed in next sections.

4.7.2. Cointegration

For ease of exposition and to keep notation simple, in this section we will concentrate on models with only two variables y_t and x_t both of them being $I(1)$. Extensions to more than two variables will probably come up along this section while the generalization to variables integrated of higher order, i.e. variables $I(d)$ with $d>1$, will rarely be mentioned. A further discussion on this topic can be found in more advanced econometrics books; see Enders (2004).

As it was explained in Chapter 3, when the values of an integrated series are plotted against time its evolution appears to be rather smooth, moving usually with local trends or with long swings (see, for instance, Figures 4.2 and 4.8), in contrast to the behaviour of a stationary series, which moves sharply around a constant mean (see Figure 4.3). Falling back on the simile used by Granger (2003) in his Nobel Prize lecture, we can think of the pattern of an integrated time series as the result of throwing down, gently, onto a hard table a loosely strung string of pearls. As the pearls are linked, the position of one pearl will affect the position of the next one and the pearls will be smoothly stretched out on the surface of the table. Unlike, following Espasa (2007), who tries to extend the Granger's simile, the pattern of a stationary series could be imagined as the result of a trembling hand throwing unlinked pearls one by one along a hard non-slipping table. In this case, the pearls will be scattered around an imaginary central line and the dispersion around it will be limited by the dimension of the table.

Following with the simile in Granger (2003), suppose now that instead of throwing a single string of linked pearls, we throw two similar strings separately on the table in such a way that they do not cross one another. Both would display smooth but different patterns with no relationship between them. In fact, if we measure the distances between

the two strings and plot them, it is very likely that they also make up a smooth integrated series. However, if we ensemble the pearls of the two strings with small strong magnets and throw them on the table, there would probably be an attraction between the two strings so that they display two smooth similar paths, although not identical. The distances between the pearls would now be a stationary series with values being up and below around their global mean. This would be an example of what it is called *cointegrated* series. These series are such that they are both individually integrated with important trend components but these long-run components cancel out when taking its difference and what it remains is stationary.

As an illustration Figure 4.10 plots, in panel (a), the series of ????? and ??????? and in panel (b), the difference between both variables.

PUT HERE TWO GRAPHS: ONE WITH THE TWO COINTEGRATED SERIES + ANOTHER WITH THEIR DIFFERENCES,

More formally, we say that two series x_t and y_t, both of which are $I(1)$, are *cointegrated* if there is a linear combination of them,

$$ax_t + by_t = n_t,$$

which is stationary, i.e. with n_t being $I(0)$. Since this equality still holds if we multiply it by a nonzero scalar, it is clear that the linear combination is not unique and so, to choose the more appropriate one, we will call on its economic interpretation. In fact, the economic theory will usually indicate us which of the two variables can be normalized with coefficient one in the linear combination and hence, after adjusting for a possible non-zero mean, the linear combination will take the form

$$y_t - \mu - \beta x_t = m_t, \tag{4.27}$$

where m_t is an $I(0)$ process with zero mean. The vector with the coefficients $(1, -\beta)$ is called the *cointegrating vector* and β is the coefficient which represents the long-run relationship between x_t and y_t so that we expect these variables to obey an equation of the form

$$y = \mu + \beta x$$

in equilibrium. But equilibrium rarely occurs and in most time periods this equality is not expected to hold and thus the quantity m_t, defined in (4.27), will measure the deviation from equilibrium at time t. The m_t variable is by definition stationary and may be called the *equilibrium error*.

In practice, cointegration corresponds to situations in which pairs of nonstationary variables move similarly in such a way that they are tied together in the long run, i.e., there exists a stationary equilibrium relationship that keep them together in the long run. Hence although the variables may drift away from equilibrium – m_t are the deviations from equilibrium –, they are expected to tend to come back to it eventually. In other words, we could say that two cointegrated variables share one hidden common factor that generates their smooth evolution and ties them in the long run equilibrium. This link in the long run, given by the linear combination (4.27), implies a restriction on the

joint evolution of y_t and x_t. In fact, the smooth paths of y_t and x_t along time are not independent, but constrained to keep a difference which is stationary, as in the case of ¿¿interest rates?? **(¿¿OR IN THE EXAMPLE TO BE FOUND??).** Unlike, if two variables $I(1)$ are not cointegrated, there will be two different trend factors, each one determining the smooth trend behaviour of each variable, with no link between them in the long run, as in the dividends-earnings example. Therefore, when the cointegration property between two integrated series exists, it implies a reduction of the dimension of the long-term dynamics of the system, which passes from two (the number of variables) to one (the only common stochastic trend that causes the nonstationarity of both variables).

In business and economics, it is not quite unlikely to face situations where cointegrated variables arise. Actually, there are many examples in the economic and econometric literature of pairs of cointegrated variables. Just to recall some of them we can mention disposable income and consumption; prices of the same commodity in different markets; nominal exchange rate and relative prices; and short and long term interest rates, among others. Sometimes the cointegration relationship between these variables will not appear in the variables themselves, but in some transformation of them, like its logarithm. For instance, in the case of disposable income and consumption, it is the difference between its logarithms – the ratio between consumption and income or mean propensity to consume – which turns out to be stationary. This means that, in the long run, there exists an equilibrium relationship such that consumption is proportional to income.

From equation (4.27), it turns out that if y_t and x_t are cointegrated and y_t is our variable of interest, it can be decomposed as

$$y_t = \mu + \beta x_t + m_t \tag{4.28}$$

where y_t and x_t are both $I(1)$ while m_t is a stationary disturbance. Therefore, the existence of cointegration between the explanatory and the endogenous variable ensures that the nonstationarity of the former completely explains the nonstationary behaviour of the latter, and the residual of the model, m_t, is stationary. Thus, in this system there is no need to take first differences to make the residuals white noise. Actually, taking first differences would be a wrong way to proceed because it would have undesirable consequences, as the long run properties of the relationship would be spoiled and a noninvertible moving average structure would be induced in the residuals. Therefore, it seems clear that testing for cointegration will be essential in order to properly model the relationship between integrated variables. This issue will be addressed in next section.

4.7.3. Testing for cointegration

Testing for cointegration can be carried out in different ways. The most popular tests are those proposed by Engle and Granger (1987) and Johansen (1988, 1991). The Johansen's methodology is established in the framework of the Vector AutoRegressive (VAR) models to be discussed in Chapter 5, so we postpone its explanation to that chapter, and we only introduce here the test proposed by Engle and Granger. Hereafter, this test will be denoted as EG test.

The basic idea behind the EG test is quite simple and is closely related to unit root tests. As it was stated above, if the variables y_t and x_t are cointegrated, the equilibrium error term m_t in (4.28) must be stationary. But if the variables are not cointegrated, regression (4.28) is a spurious one and m_t will be integrated $I(1)$, i.e. it will have a unit root. Therefore a test for the null hypothesis of no-cointegration between y_t and x_t can be carried out by testing for the existence of a unit root in the error m_t in regression (4.28). The rejection of the null is then interpreted as the existence of cointegration between the variables y_t and x_t. Otherwise, the residuals will have a unit root implying that the variables are not cointegrated and then the model should be better estimated in first differences.

Since the variable m_t is not observable, the unit root test is performed on the estimated residuals, $\hat{m}_t = y_t - \hat{\mu} - \hat{\beta} x_t$, where $\hat{\beta}$ is the OLS estimator from the cointegration regression (4.28). In particular, the EG test consists of applying an ordinary or an augmented Dickey-Fuller test for unit roots (Dickey and Fuller, 1979) to the residuals $\hat{m}_t$. For this reason, this test is usually called *residual-based cointegration test*.

The simplest version of the ordinary Dickey-Fuller test on $\hat{m}_t$ would be performed as follows. Let us consider an AR(1) representation for $\hat{m}_t$ given by:

$$\hat{m}_t = \phi\ \hat{m}_{t-1} + e_t.$$

Then, testing for a unit root in $\hat{m}_t$ amounts to testing the hypothesis that ϕ=1. A suitable transformation to perform such test consists of subtracting $\hat{m}_{t-1}$ from both sides of the equation above, so that we get the following auxiliary regression:

$$\Delta \hat{m}_t = (\phi\text{-}1)\ \hat{m}_{t-1} + e_t. \qquad (4.29)$$

If we define $\phi^+=\phi$-1, the null hypothesis of a unit root (ϕ=1) becomes ϕ^+=0, and this will be rejected against the one-side alternative that ϕ^+<0 (ϕ<1), if the appropriate test statistic is well below the corresponding critical value. The test statistic, denoted by τ, is the usual *t*-statistic for the slope coefficient in regression (4.29), but its asymptotic distribution under the null (ϕ=1) is non-standard. Moreover, since the variable $\hat{m}_t$ involved in the regression equation (4.29) is not actually observed, but it is a residual series that depends on some estimated parameters, then, in this case, the asymptotic distribution of the test statistic τ is not the same as that corresponding to the ordinary Dickey-Fuller test and its usual critical values can not be directly applied here. Tables with appropriate critical values for this problem can be found in Mackkinon (1991) or in advanced econometric textbooks (see, for example, Davidson and Mackkinon, 1993). Usually, these tables include asymptotic critical values at the usual rejection significance levels, 1%, 5% and 10%, for different values of the number of variables being cointegrated (these are two in the simple case we are discussing but, as we said before, there could be more than two integrated variables). The tables also include other test statistics which arises when we include in (4.39) a constant intercept or a constant and a linear trend. These statistics are usually denoted as τ_c and τ_{ct}, respectively.

The Dickey-Fuller test for unit roots explained above can be generalize to cope with possible positive serial correlation in the residuals of the auxiliary regression. In this

case, it is usually called augmented Dickey-Fuller (ADF) test. Engle and Granger (1987) propose to apply ADF as a unit root test for $\hat{m}_t$, by testing that ϕ-1=0 in a regression similar to (4.29) but including enough lags of $\Delta\hat{m}_t$ as regressors to eliminate any evidence of serial correlation. As noted before, since the variables involved are not observed, the asymptotic distribution of the corresponding test statistic is not the same as the ordinary ADF. However, proper critical values are already available in the references mentioned above.

Example 1: Dividends and Earnings (continued)

In the example on Earnings and Dividends introduced in section 4.3.1, we said that both variables D_t and E_t were nonstationary and their relationship was formulated on their first differences, ΔD_t and ΔE_t. Now we will see that this was done because the variables in levels, D_t and E_t, are actually $I(1)$ and the residuals of a regression between them are nonstationary; in fact, they have a unit autoregressive root.

Table 4.7 and Table 4.8 display the correlograms and the Box-Ljung statistics for the series of Dividends and Earnings, respectively. As expected, the correlations at any lag are significant and hardly decrease as the lag increases, suggesting that these two series are not stationary. Obviously, the null hypothesis of uncorrelation is always rejected. The Augmented Dickey Fuller (ADF) test for unit root in each series, reported in Table 4.9 and Table 4.10, confirms the results and clearly indicates that Dividends and Earnings have both a unit root. The same result is obtained when performing the ADF test that includes only a constant (rather than a constant and trend) and other unit root tests, such as Phillips-Perron, have also been applied providing the same conclusions. Moreover, after removing such unit root by taking first differences, the resulting series ΔD_t and ΔE_t seem stationary (see Figure 4.3) and it has been checked that the ADF test rejects the presence of a unit root in both differenced series. Therefore, D_t and E_t can be taken as nonstationary $I(1)$.

INCLUDE TABLES 4.7 – 4.10

Let us see what happens when running a linear regression between these two integrated variables. Table 4.11 displays the results of such regression model and Figure 4.11 plots the actual and fitted values of Dividends together with the residuals. The residual correlogram together with the Box-Ljung statistic values for several lags are reported in Table 4.12. Clearly, the residuals are not white noise. Actually, they are not even stationary. In applying the EG cointegration test explained above, by performing an ADF test on the residuals, we get the value of the test-statistic τ_c=-2.4903. If we compare such value with the 5% and 10% asymptotic critical values given in Mackinnon (1991), namely -3.34 and -3.04, respectively, we can not reject the null and conclude that the residuals have a unit root and the series D_t and E_t are not cointegrated. Consequently, a model relating the first differences, rather than levels, of Earnings and Dividends, as that estimated in previous section 4.6, is the appropriate one.

TABLES 4.11 + 4.12 + Figure 4.11

■

Example 3:
INCLUDE HERE AN EXAMPLE OF REGRESSION BETWEEN INTEGRATED VARIABLES WITH STATIONARY RESIDUALS, I.E. COINTEGRATION. (The series should be those that have been plotted previously to motivate cointegration and Granger's simile)

The example should include:

- Table 4.13: ADF unit root test on the endogenous
- Table 4.14: ADF unit root test on th exogenous
- Table 4.15: Regression estimation output (with stationary residuals)
- Figure 4.12 residuals plus fitted and actual values of endogenous variable
- Table 4.16: Residual correlogram + Box-Ljung on residuals
- See model in first differences: large negative moving average structure

■

4.7.4. Equilibrium correction mechanism: EqCM representation

One of the main features of the systems which involve cointegrated variables is that they can be represented in what is called *Equilibrium Correction Model* (EqCM). This type of model, initially known as Error Correction Model, was first proposed by Sargan (1964) and then developed in several articles by David Hendry and co-authors; see e.g. Davidson *et al*. (1978). This formulation has a very appealing interpretation in later. The idea behind this type of formulation, in its simplest form, is that if at time (t-1) the economy is out of equilibrium so that $m_{t-1} = y_{t-1} - \mu - \beta x_{t-1}$ is not zero, then the increments of y_t in the next period, Δy_t, will respond to it in order to push y_t to the equilibrium value. Therefore, the equilibrium error m_{t-1} should enter in the dynamic regression model for Δy_t. Thus this model will capture the short run dynamics (lag increments in y_t and x_t) and the long run relationship in levels (through m_t) between both variables y_t and x_t.

The EqCM representation for a general dynamic regression model is given in the Granger Representation Theorem (see Engle and Granger, 1987). This will be further explained in next chapter. Herein, we confine our discussion to the bivariate case, i.e. to models with only two cointegrated variables y_t and x_t both of them being $I(1)$ and we derive the EqCM representation for ADL models involving such variables.

To keep matters simple, we focus on the ADL(1,1) model with a contemporaneous relationship between the variables.[6] The formulation of this model has already been given in section 4.3 but we reproduce it here for ease of exposition. In particular, its equation is as follows:

$$y_t = c + a_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + \varepsilon_t \qquad (4.30)$$

where $|a_1|<1$ and ε_t is white noise. The transformed representation for this model is obtained through the following steps:

[6] The EqCM representation for a more general ADL model is given in Appendix B.

1) Reformulate the model in order to have Δy_t as dependent variable. This implies to subtract y_{t-1} in both sides of the equation, so that we get:

$$\Delta y_t = c + (a_1\text{-}1)\, y_{t-1} + b_0\, x_t + b_1\, x_{t-1} + \varepsilon_t$$

2) Reformulate the model to have Δx_t as a regressor. This is achieved by adding and subtracting $b_0\, x_{t-1}$ in the right hand side of the equation:

$$\Delta y_t = c + (a_1\text{-}1)\, y_{t-1} + b_0\, \Delta x_t + (b_0 + b_1)\, x_{t-1} + \varepsilon_t$$

3) Rearrange the first lagged terms of y_t and x_t to make m_{t-1} appears as another regressor. This is done by multiplying and dividing x_{t-1} by the coefficient of y_{t-1}, namely $(a_1\text{-}1)$, to get:

$$\Delta y_t = c + \alpha\, (y_{t-1} - \beta\, x_{t-1}) + b_0\, \Delta x_t + \varepsilon_t$$

where $\alpha = (a_1 - 1)$ and $\beta = (b_0 + b_1)/(1 - a_1)$.

4) Redefine the constant term, as the expected value of Δy_t. For that purpose recall that the constant term in a regression is the expected value of the regresand minus the expected value of all the regressors multiplied by their corresponding coefficients. Consequently,

$$c = \mathrm{E}(\Delta y_t) - b_0\, \mathrm{E}(\Delta x_t) - \alpha\, \mathrm{E}(y_{t-1} - \beta\, x_{t-1}).$$

Denoting by γ, ρ and μ the above three mentioned expected values, respectively, and subtracting ρ and μ from their corresponding regressors we end up with the following representation for Δy_t:

$$\Delta y_t = \gamma + b_0\, (\Delta x_t - \rho) + \alpha\, (y_{t-1} - \mu - \beta\, x_{t-1}) + \varepsilon_t. \qquad (4.31)$$

Note that the second term in brackets on the right hand side of this equation is just the equilibrium error lagged once, i.e. $y_{t-1} - \mu - \beta\, x_{t-1} = m_{t-1}$, therefore, (4.31) can also be written as

$$\Delta y_t = \gamma + b_0 (\Delta x_t - \rho) + \alpha\, m_{t-1} + \varepsilon_t. \qquad (4.32)$$

This equation states that the value of Δy_t at a given time t is explained by: a transitory dynamic component, given by the increment in the explanatory variable; an adjustment to the past equilibrium error, in a proportion α; and a white noise disturbance. This formulation of the model is known as the equilibrium correction model (EqCM), because the disequilibrium term m_{t-1} appears as a regressor. [7] Note that if b_0 is zero in (4.30), the step 2 described above is not required and there will not be a transitory dynamic component in (4.32).

[7] When the initial model has more lags of y_t and x_t than in equation (4.30), the steps to obtain the EqCM are those stated above but with a more complex step 2 which also affects to the endogenous lags. The resulting EqCM representation, derived in appendix B, will have the same structure as (4.32) but with more lags on Δy_t and Δx_t.

It should be remarked that if the variables y_t and x_t are cointegrated all components of equation (4.32) are stationary. Moreover, although this equation could apparently be viewed as a relationship between the first differences of the variables y_t and x_t, this is not the case. The equilibrium correction term involves levels, rather than changes, of the variables and therefore this model actually represents the relationship between the variables y_t and x_t in levels, as it can be seen by formulating (4.32) as (4.30). If the variables are not cointegrated, then m_{t-1} is non-stationary and since all the others terms in (4.32) are stationary, the coefficient of m_{t-1} must be necessarily zero for the equation to be balanced and so the term involving m_{t-1} will disappear. In these cases, a model in first differences would be an appropriate specification, where differences of the endogenous variable are explained by differences of the explanatory variable and only short-run effects would be modelled; recall Example 1 on Earnings and Dividends. Therefore, as we noted before, if two variables are cointegrated, a model in first differences will be misspecified because it omits the long-run equilibrium relationship suggested by economic theory. Thus the presence of cointegration requires a stationary formulation that includes the long run disequilibrium as a further explanatory variable which contributes to explain the changes of the endogenous variable.

Equation (4.32) has two appealing features. First it enables an important interpretation in terms of equilibrium relationships and second it has a suitable form to undertake estimation and hypothesis testing, as we will see in next section. With regards to its economic interpretation, note that the third term on the right hand side of (4.32) represents the equilibrium error, i.e. the deviation from equilibrium which may influence the future evolution of the variables in the system. When m_{t-1} is positive, y_{t-1} is above its value at equilibrium, $\mu+\beta x_{t-1}$, and a proportion α of this disequilibrium is corrected in the next period. In fact, the change in y_t over y_{t-1} is corrected downwards to take the system back to its long run path and this correction is made through the term α m_{t-1}, whose coefficient $\alpha=(a_1-1)$ is negative.[8]

Note also that equation (4.31) involves all the parameters of interest, namely $\{\gamma, b_0, \alpha, \mu, \beta\}$. In fact, in this equation we can distinguish the following elements:

(a) The long-run relationship between y_t and x_t given by $(y_{t-1} - \mu - \beta x_{t-1})$, where μ is the intercept and the parameter β captures the long-run gain in y_t with respect to x.

(b) The adjustment to equilibrium given by $\alpha(y_{t-1} - \mu - \beta x_{t-1})$, where α is the proportion of adjustment at time t, i.e. the proportion of disequilibrium that will affect the evolution of y_t;

(c) The transitory dynamics given by the effect on Δy_t due to Δx_t, being b_0 the only transitory dynamic parameter. In models with more lags of the endogenous and exogenous variables, this element will also include the effect on Δy_t from past increments of both Δy_t and Δx_t.

(d) Parameter γ, which is the expected value of Δy_t .Whether this parameter is or not zero determines whether there is or not growth in the system, and this requires a more detail discussion.

[8] Recall that the parameter a_1 must be less than one, in absolute value; otherwise the process will show an explosive behaviour.

Suppose that y_t and x_t are $I(1,0)$ and consequently their increments have zero mean, i.e. $E(\Delta y_t)=0$ and $E(\Delta x_t)=0$, and so there is not systematic growth in the variables but just local oscillations in time. In this case the intercept γ in (4.31) is zero and so is the parameter ρ. However, an interesting case arises when dealing with economic series that are systematically increasing so that Δy_t and Δx_t have no zero means. We will refer to these series as being $I(1,1)$. This type of series is characterised by having two trend components, the stochastic one, represented by the unit root that produces local oscillations in level, and another deterministic trend that is responsible of the steady increase in the series; see Chapter…, section…. In this context, y_t and x_t can be represented as follows

$$\begin{aligned} y_t &= \tau\, t + y_t^* \\ x_t &= \rho\, t + x_t^* \end{aligned} \qquad (4.33)$$

where y_t^* and x_t^* are both $I(1,0)$. Note that $E(\Delta y_t)= \tau$ and $E(\Delta x_t)= \rho$. If y_t and x_t are cointegrated, then their unit roots disappear in the residuals of a regression like (4.28) in the sense that $m_t = y_t - \mu - \beta x_t$ has not unit roots. But let us see what happens with the deterministic components. In order to do that, we substitute in the expression of m_t the values of y_t and x_t given in (4.33) so that m_t can be written as:

$$m_t = [(\tau - \beta \rho)t - \mu] + (y_t^* - \beta x_t^*) \qquad (4.34)$$

If y_t and x_t are cointegrated, the second term in (4.34) is stationary. The first term is a linear trend unless the parameter β, which cancels the stochastic non-stationarity of y_t and x_t in the cointegration relationship, also cancels their linear trends ($\tau - \beta \rho=0$), in which case model (4.32) will have a non-zero intercept ($\gamma \neq 0$) but will not contain a linear trend in the cointegration relationship. In this case, we have $\gamma=\tau$, which is the equilibrium rate of growth of y_t. Otherwise, there will be also a non-zero intercept in the equation and a trend in the cointegration relationship in model (4.32), implying that y_t and x_t will diverge in the long-run by this deterministic trend. The presence of this trend component could be indicating that there could be some important variables that have been erroneously not included in the model. Therefore, when dealing with $I(1,1)$ variables, the cointegration relationship to be sought for will be

$$y_t = \mu + \delta t + \beta x_t + m_t.$$

If y_t and x_t are cointegrated, m_t will have not unit roots, but depending on whether δ is zero or not, we will end up with two different situations. If $\delta \neq 0$, there is a linear trend in the long-run relationship and hence the differences $y_t - \mu - \beta x_t$ will become larger along time. Unlike, if $\delta=0$, $y_t - \mu - \beta x_t$ will be stationary as this cointegration relationship cancel both the stochastic trends (unit roots) and the deterministic trends of y_t and x_t.

<u>Example 4</u>: Include an example of cointegration with linear trend: prices of the same good in different markets or geographic areas.
Figure 4.10: time series plot of both series: prices show and evolution with systematic growth.
Table regression (with LINEAR TREND stationary residuals)
Table 4.10 Residual correlogram + Box-Ljung

4.7.5. Estimation of EqCM representation

As we noted before, inference on the parameters on a model involving integrated variables can not be carried out based on the formulation in levels, because the asymptotic theory on the parameter estimates requires the variables to be stationary and this is not the case if the variables are integrated. Actually, another advantage of the EqCM representation, besides its interpretation in economic terms, is that it is a model on stationary variables and so it enables to carry out usual estimation and tests on the parameters of the model that could not be performed otherwise; see Sims, Stock and Watson (1990). Therefore, a convenient way to carry out estimation of cointegrated systems is by using its equilibrium correction representation.

Engle and Granger (1987) proposed a two step estimator for this representation which is obtained as follows. In the first step the parameter β of the cointegrating vector is estimated by ordinary least squares (OLS) applied to the simple (static) regression of y_t on x_t in (4.28). This regression will be called the "cointegrated regression" as it attempts to fit the long run equilibrium relationship. In the second step, the residual term from this regression, $\hat{m}_{t-1} = y_{t-1} - \hat{\mu} - \hat{\beta} x_{t-1}$, which is the equilibrium error, is plugged in the equilibrium correction form (4.32). Then all the elements in this equation are stationary and its parameters are estimated by OLS.

The OLS estimator of the long run parameter in the first step, say $\hat{\beta}$, is super-consistent, meaning that as the sample size increases it converges even faster to the true value than standard estimates in regression over stationary data. In fact, its rate of convergence is $1/n$ rather than the usual $1/\sqrt{n}$. However, its asymptotic distribution is not normal and the standard inference procedures are not valid. Moreover, in a dynamic system, the static regression (4.28) could provide biased estimates of β in small samples and inferences may be misleading in some cases. Banerjee *et al* (1993) suggest that the biases can be diminished by specifying dynamic rather than static regressions to estimate the cointegrating parameter β, i.e. by including lagged values of the first differences of both the dependent and exogenous variable in regression (4.28) and estimating β from this dynamic regression model. With respect to the OLS estimators of the short-run parameters at the second step, it turns out that they have the usual asymptotic normal distribution. Thus they are as efficient as the maximum likelihood estimators based on the known value of β. However, the bias in $\hat{\beta}$ may cause these estimators to be also biased in small samples.

Alternative, all the parameters can be estimated from the full model in (4.31) by nonlinear least squares. In this case, the limiting distribution of $\hat{\beta}$ is again non-standard but different from that obtained with the two stage procedure and it also displays small sample bias, although less severe than one might expect. On the other hand, the estimators of the short run dynamics have still the same asymptotic normal distribution as in the two-step procedure.

As a further remark, it must be noted that when the explanatory variables are exogenous, as it is the case which we are considering in this chapter, a single-equation model for the variables in levels, as that in (4.29), is still valid even if the variables

involved are integrated, and this representation is indeed very useful for forecasting purposes, although it is inappropriate for estimation and testing.

Example 3: (continued)
HERE THE ESTIMATION OF THE EqCM representation of the EXAMPLE WITH COINTEGRATED VARIABLES SHOULD BE INCLUDED.

4.8. REGRESSION MODELLING: SOME FURTHER EXAMPLES

4.9. FORECASTING WITH REGRESSION MODELS: CONDITIONAL AND UNCONDITIONAL FORECASTS

Once a dynamic regression model has been estimated it can be used to forecast the future evolution of the endogenous variable. The main feature of forecasting with econometric models is that they take into account not only the own past of the variable of interest, as it happens in the univariate time series models in Chapter 3, but also the values of other variables related to it. Therefore, it is to be expected that if the econometric model is well specified and correctly estimated and the future values of the exogenous variables are known or can be accurately forecast, then the econometric model will generate better forecasts than the univariate models. For example, forecasts of future unemployment based on past and present unemployment and past and present production figures and labour costs seem to be, *a priori*, more reliable than those based only on unemployment figures. Since more information related to the phenomenon being predicted has been included, the prediction is expected to be done more efficiently. However, it should be recall that building an econometric model is not an easy task, as it has emerged from the discussion in previous sections of this chapter.

Forecasting from a dynamic regression model is carried out recursively along the same lines described in Chapter 3 for univariate time series models. That is, forecasts of future events will be made by extrapolating the regression model beyond the time period of the last observation available; if this period is n, forecasts will be made for the next periods $n+1$, $n+2$, and so on until period $n+h$. As it was discussed in Chapter 1, making an error in the forecast will have a cost. Generally, the larger the magnitude of the error in absolute terms, the larger will be the cost. Therefore, a good criterion to compare several forecasting methods will be to select that method which results in the minimum average cost. Costs are often asymmetric functions of the errors, so that the cost of underprediction will be different to the cost of overprediction. For example, the cost of arriving one hour before the time of departure of a flight is quite different from arriving one hour late. As another example, we could think that the cost of an inefficient use of extra electricity production due to an excessively high forecast will be less grave than the cost of a blackout due to a forecast lower than the consumption actually demanded. In practice, it is unlikely to know the cost function and what is usually done is assuming that the cost is a quadratic function of the error, although this is obviously symmetric (see the discussion in Chapter 1). This will be the approach adopted here, although allowance for asymmetry will be roughly discussed when introducing the *fan chart*.

In general, there are two different ways in which econometric models can be used for forecasting purposes. The first is to use them to provide *conditional* or *ex-post* forecasts,

that is, to forecast the value of the endogenous variable given some specific values of the explanatory variable(s). Secondly, models for the explanatory variables can be used to provide *unconditional* or *ex-ante* predictions of the endogenous variable based on forecasts of the explanatory variables themselves. Both options are explained below. To keep notation simple, we will assume a dynamic regression model with only one explanatory variable. Extensions to cases with more explanatory variables are straightforward.

The forecasts defined above, and forecasts of any kind, are always subject to error since around future events there is always a certain level of uncertainty which can not be eliminated or forecast. In econometric models, the forecast error comes from several sources[9]. First, we find what is usually called mis-specification uncertainty, that is, our model is a simplified representation of reality and if this is not an accurate one we will be exposed to errors. Of course, this source of uncertainty could be minimized by constructing a "good" model and how to do it has been dealt with in previous sections. A second source of error is the innovation in the regression model. When the forecast is made at time t, the future innovation ε_{t+1} is unknown and unforecastable given the information set used in the construction of the econometric model. In fact, the randomness of ε_t makes the forecast error to be inevitable, since the forecast of y_t could differ from its actual value eventually observed, even if a correctly specified model with known parameters were used. Third, there is uncertainty due to estimating the parameters of the model. In general, the values of these parameters are unknown and the coefficients used to compute the forecasts are actually their estimates. Estimates are point values of random variables and therefore are subject to sampling errors. However, as far as good estimators are used and sample size is large enough the effect of this error will diminish. Finally, another important source of error, the most pernicious one, as it is shown in Clements and Hendry (1998, 1999), is the presence of structural breaks in the deterministic elements of the model, such as shifts in the equilibrium mean and parameter changes over time. In this case, it could be difficult to correctly model the data generating process and hence to provide accurate forecasts. However, some strategies like intercept corrections and differencing seem to help to alleviate the problem and therefore to improve efficacy in forecasting. We will make some additional comments on the last alternative in the following chapter. In what follows the only source of forecasting error that we consider is the one coming from the innovations.

4.9.1. Conditional forecasting

Conditional forecasts of the variable of interest, say y, are obtained directly by inserting plausible or actual values of the explanatory variables $x_1,..., x_k$ in the corresponding econometric model that relates y with $x_1,..., x_k$. This kind of forecast turns out to be interesting in different situations.

First, let us imagine that we have information up to time $n+h$, so the values of $y_{n+1},\ldots, y_{n+h}$ are already known, but we are still interested in finding out which would have been the estimation of such values if we had forecast them at time n using the actual values of the explanatory variables. This kind of forecasts is called conditional because they are obtained by fixing the values – which in this case are known – of the explanatory variables corresponding to the forecasting period. In this case, and assuming that the

[9] See Clements and Hendry (1998, 1999) for a further discussion on this topic.

parameters of the model are known, the forecast errors of $y_{n+1},\ldots, y_{n+h}$ are only due to the unknown values of the innovations of the model, $\varepsilon_{n+1},\ldots, \varepsilon_{n+h}$, through their contemporaneous effect on y_{n+j} and through the lag effects due to the presence of lagged values of y_t as explanatory variables. These conditional or ex-post forecasts should dominate the ex-ante forecasts, to be explained later, and thus estimating this dominance by the reduction in the mean squared forecasting errors we could evaluate the importance of knowing the realizations of the exogenous variables when forecasting the variable of interest. For instance, in forecasting the evening peak in electricity consumption of the day t, the maximum temperature of this day could be relevant. Therefore, estimating the reduction in the mean squared error due to knowing such temperature may be important to decide whether to invest in the implementation of a procedure which automatically incorporates the temperature data into the forecasting system. In this case, we are making use of the advantage that the explanatory variable is usually observed some hours earlier than the value of the endogenous variable to be forecast.

Conditional forecast can also be done with information up to time n, by fixing the future (unknown) realizations of the explanatory variables to certain plausible values. This exercise is sometimes called *simulation* or *scenario analysis*, because it enables the possibility of emulating the response of the dependent variable y_t to different scenarios defined by possible values of the explanatory variables that are fixed by the analyst. For a policy maker or a firm manager this could be very useful because it enables to predict what the future value of a certain variable, such as employment or profits, would be under various policy or investment options. For instance, using these conditional predictions, we can answer questions as the following: if a company invested a certain sum x_t in advertising, what would it be its expected increase in sales?; if a government expends a given sum x_t on a certain employment policy, what could it be the expected increment on the employment rate? Turning back to our previous example on electricity consumption, a useful simulation analysis could be to forecast, at day n, the evening peak of electricity consumption the day after, say y_{n+1}, given a possible sharp change in the temperature. Obviously, at day n the maximum temperature of the following day (x_{n+1}) is unknown, but we could be interested in knowing the effect that a sudden drop of say 16ºF in such temperature with respect to that of the previous day (x_n=50ºF) will have in electricity consumption. Thus, we can emulate the value of y_{n+1} in such scenario, i.e. we can make a conditional forecast of y_{n+1} by fixing the value of x_{n+1} to be 34ºF. This value does not need to be the best forecast of x_{n+1}, but simply a value defining a range around it, say 33.5ºF-34.4ºF, within which the maximum temperature has a not negligible probability of falling. This simulation exercise is important if the system operator faces an asymmetric cost function with higher costs for unpredicting electricity demand than for overpredicting it.

Finally, conditional forecasting may also be relevant for control purposes, as will be explained later in section 4.12.

In order to illustrate some of the issues described above, let us first consider the ADL(1,1) model given in (4.29). To forecast y_{n+1} in this model, we first replace t by n+1 so that we can write

$$y_{n+1} = c + a_1 y_n + b_0 x_{n+1} + b_1 x_n + \varepsilon_{n+1}. \qquad (4.36)$$

If information up to time n is available for all the variables in the model and we also assume that future values of x, like x_{n+1}, are known with certainty, then formulation (4.36) can be regarded as an AR(1) model for y with a deterministic known component, given by $\{c + b_0 x_{n+1} + b_1 x_n\}$. Therefore, the forecast errors for y will have the properties of those from an univariate AR(1) model (see Chapter 3). In particular, the forecast of y_{n+1} at time n, denoted by $f_{n,1}^{y}$, is obtained from (4.36) as

$$f_{n,1}^{y} = c + a_1 y_n + b_0 x_{n+1} + b_1 x_n , \tag{4.37}$$

and the one-step forecast error for y, denoted by $e_{n,1}^{y}$, is computed by subtracting (4.37) from (4.36) so that

$$e_{n,1}^{y} = y_{n+1} - f_{n,1}^{y} = \varepsilon_{n+1}. \tag{4.38}$$

Thus, in this case, the one-step forecast error is just the innovation of the model at time $n+1$. Therefore, since ε_{n+1} is white noise, also is $e_{n,1}^{y}$ and the errors for different n are uncorrelated. Moreover, the error variance is simply

$$\mathrm{Var}(e_{n,1}^{y}) = \mathrm{Var}(\varepsilon_{n+1}) = \sigma_{\varepsilon}^{2} . \tag{4.39}$$

Then, under the assumption of normality for the innovations, it can be shown that the conditional density of y_{n+1} is Normal with mean $f_{n,1}^{y}$ and variance (4.39). From this distribution, the 80% prediction interval for y_{n+1} can be directly constructed as

$$f_{n,1}^{y} \pm 1.28 \sqrt{\mathrm{Var}(e_{n,1}^{y})} , \tag{4.40}$$

which, in this case, yields the following simple form

$$f_{n,1}^{y} \pm 1.28\ \sigma_{\varepsilon}. \tag{4.41}$$

Obviously, the larger the variance σ_{ε}^{2}, the wider the forecast interval and the larger the uncertainty in forecasting the dependent variable y.

Forecasting beyond one-step ahead is carried out in a similar way, but it requires more information to be available. For instance, the two-step forecast, that is, the forecast of y_{n+2} at time n, is

$$f_{n,2}^{y} = c + a_1\ f_{n,1}^{y} + b_0 x_{n+2} + b_1 x_{n+1},$$

provided that x_{n+2} and x_{n+1} are fixed at time n. The corresponding forecast error is

$$e_{n,2}^{y} = y_{n+2} - f_{n,2}^{y} = a_1\ e_{n,1}^{y} + \varepsilon_{n+2} = \varepsilon_{n+2} + a_1\ \varepsilon_{n+1}.$$

Hence the two-step error has a stationary MA(1) structure and its variance is given by

$$\text{Var}(e^{y}_{n,2}) = (1 + a_1^2)\,\sigma_\varepsilon^2 . \tag{4.42}$$

The corresponding 80% forecast interval for y_{n+2} will in turn be

$$f^{y}_{n,2} \pm 1.28\ \sigma_\varepsilon \sqrt{1 + a_1^2}\ ,$$

which is clearly wider than the one-step ahead forecast interval given in (4.41).

Following the same argument as above, it is easily shown that the *h*-step forecast error will take the form

$$e^{y}_{n,h} = \varepsilon_{n+h} + a_1\,\varepsilon_{n+h-1} + a_1^2\,\varepsilon_{n+h-2} + \ldots + a_1^{h-1}\,\varepsilon_{n+1},$$

which is an MA(*h*-1) process and hence, its variance is given by

$$\text{Var}(e^{y}_{n,h}) = (1 + a_1^2 + \ldots + a_1^{2(h-1)})\,\sigma_\varepsilon^2 . \tag{4.43}$$

Comparing this formula with those in (4.39) and (4.42) makes it clear that the further the forecast horizon, the larger the error variance and therefore, the wider the forecast interval. That is to say that uncertainty increases as the forecast horizon increases.

However, the uncertainty about the future of *y* is limited by the variance of its dynamic stochastic component which is, in this case, the residual term gathering the dynamic of *y* not explained by the exogenous variable. In other words, if the ADL(1,1) model (4.29) is alternatively written as

$$y_t = \frac{c}{1 - a_1} + \frac{b_0 + b_1 L}{1 - a_1 L}\,x_t + \frac{1}{1 - a_1 L}\,\varepsilon_t, \tag{4.44}$$

the residual term, that will be denoted by η_{yt}, is the following AR(1) process

$$\eta_{yt} = \frac{1}{1 - a_1 L}\,\varepsilon_t, \tag{4.45}$$

Then, it is clear from (4.44) and (4.45) that in ex-post forecasting the uncertainty about the forecast of y_t only depends on the uncertainty about η_{yt}, since the exogenous variable x_t is taken as known. Hence, the variance of the forecast error will be limited by the variance of η_{yt}, which in this case is equals to $\sigma_\varepsilon^2/(1 - a_1^2)$.

The results above can be easily extended to the general ADL(*r*,*s*) model in (4.6). Since in conditional forecasting, the future values of the explanatory variable *x* are known, they can be considered as if they were deterministic, and the ADL(*r*,*s*) model can be formulated for y_{n+h} as

$$y_{n+h} = D_{n+h} + a_1 y_{n+h-1} + \ldots + a_r y_{n+h-r} + \varepsilon_t, \tag{4.46}$$

where $D_{n+h} = \{c + b_0 x_{n+h} + b_1 x_{n+h-1} \ldots + b_s x_{n+h-s}\}$ denotes this "deterministic" part of the model given that the values of the explanatory variables after n are fixed at time n. Hence, formulation (4.46) can be regarded as an AR(r) model with a deterministic component D_t and, therefore, the forecast errors will have the properties of those from a univariate AR(r) model (see Chapter 3). In particular, the formula for the h-step forecast error, which can be easily derived from (4.46), is given by

$$e^{y}_{n,h} = \varepsilon_{n+h} + a_1 e^{y}_{n,h-1} + a_2 e^{y}_{n,h-2} + \ldots + a_{h^*} e^{y}_{n,h-h^*},$$

where $h^*=\min(h-1,r)$. Proceeding recursively in this equation, the forecast error can be written as a function of the future innovations, which are stationary, and therefore its variance will depend on the variance of the innovation process and the coefficients of the stochastic part of the model (the lagged values of the endogenous variable). On the other hand, and following a similar argument as in the ADL(1,1) case, it can be proved that the value of $\text{Var}(e^{y}_{n,h})$ is limited by the variance of an AR(r) process. This confirms that in ex-post forecasting, the uncertainty about the future of y actually increases with the horizon but is bounded by the variance of its dynamic stochastic component η_{yt}.

In the discussion so far, we have implicitly assumed that the variables x_t and y_t are stationary. However, since the explanatory variable is exogenous, y_t can be non-stationary and the errors in conditional forecasting can be still stationary. This will happen if the variables y_t and x_t are cointegrated. Hence, in this case, the dynamics of y_t not explained by the explanatory variables, gathered in the residual η_{yt}, are still stationary and the non-stationarity of the explanatory variables does not add uncertainty about the future of y_t because these variables are taken as known. Therefore, the conditional forecasting procedure sketched above still applies and so does the formulae of the variance of the forecast error. This means that, if both variables x_t and y_t are $I(1)$ but are cointegrated, the variance of the forecast error in conditional forecasting increases with the horizon but with a limit given by the variance of the stationary residual dynamics, η_{yt}.

However, if both the variables y_t and x_t are integrated $I(1)$ but are not cointegrated, then the residual term η_{yt} of the ADL model is also non-stationary. In particular, for the ADL(1,1) model given in (4.44), non-stationarity of y_t implies $a_1=1$ and then the residual term in (4.45) becomes $\eta_{yt} = \dfrac{1}{1-L}\varepsilon_t$, which is now non-stationary. Nevertheless, forecasts can still be performed on the model in levels and the formulae above for the forecasts themselves and the forecast errors still hold in this case. However, as the forecast errors depend on the residual term of the model, $\eta_{yt,}$ and this is now $I(1)$, the variance of the forecast error tends to infinite with h, as it happens with the ARI(1,p) models that we saw in Chapter 3. A summary of the properties of the forecast errors in the three situations described above is displayed in Table 4.1 under the column heading "ex-post forecasting".

All the forecast errors formulae discussed previously correspond to a situation where the values of the parameters are known. In practice, these are unknown and they will be replaced by their estimates. As expected, this has implications on the forecast errors.

For instance, if we focus on the ADL(1,1) model, a feasible version of the one-step forecast equation (4.37) is obtained by replacing the parameters by their estimates, to get the following expression

$$\hat{f}_{n,1}^{y} = \hat{c} + \hat{a}_1 y_n + \hat{b}_0 x_{n+1} + \hat{b}_1 x_n .$$

Subtracting this equation from equation (4.36) gives the corresponding forecast error

$$\hat{e}_{n,1}^{y} = (c - \hat{c}) + (a_1 - \hat{a}_1) y_n + (b_0 - \hat{b}_0) x_{n+1} + (b_1 - \hat{b}_1) x_n + \varepsilon_{n+1}. \qquad (4.47)$$

The comparison of this expression with (4.38) highlights the effect of estimating the parameters on the forecast error. In particular, the larger the deviance of estimated values from the true parameter values, the larger the forecast errors and therefore, the less reliable the forecast. However, if "good" estimators are used, i.e. consistent and asymptotically efficient ones, it is unlikely that the estimated values are far apart from the true ones. Moreover, as the sample size increases, the higher is the probability that estimates and true values are very close, and so the more reliable are the forecasts. Forecast error variances for the whole sequences of forecasts, in the stationary and non-stationary cases, can be found in Clements and Hendry (1998, 1999), who also discuss in detail the different sources of forecast error. As we said before, these authors claim that, although mis-estimation is actually one source of error, it is not the most important one and shifts in the deterministic elements of the model turn out to be the most dangerous cause of forecasting failure. For example, in model (4.36) in section 4.5, this will imply a shift in the parameter γ, whose consequences in forecasting y_t will not be negligible.

4.9.2. Unconditional forecasting

Econometric models are especially useful to provide forecasts of the variable being explained by linking it with forecasts of the explanatory variables. In this case, forecasts are said to be *unconditional* or *ex-ante* predictions, because they are not assuming any given value of the explanatory variables but they require predictions of them in order to predict variable y_t. Therefore, to achieve unconditional forecasts, separate models for each of the exogenous explanatory variables are needed to provide forecasts for these variables and these forecasts are then used to form a forecast for the variable of interest. The exogeneity property allows to forecast the regressors independently of the endogenous variable and then to forecast the latter given the forecasts of the former.

In an ADL(1,1) model with a contemporaneous effect of the exogenous variable, the one-step ahead ex-ante forecast of y at time n must also be obtained from equation (4.36), as in conditional forecasting, but taking into account that the value of x_{n+1} on the right hand side of that equation is no longer known and must be substituted by an estimation (forecast). This forecast will be a value drawn from the probability density of x_{n+1} conditional on its own past. In fact, the forecast of x_{n+1}, that will be denoted by $f_{n,1}^{x}$, is the conditional mean of all possible realizations of x_{n+1} according to a certain probability distribution. Thus, in ex-ante forecast we do not have a value for x_{n+1} which is taken as unquestionable and really fixed. Instead, we are forced to use a value $f_{n,1}^{x}$ which will presumably differ from the fixed value x_{n+1} that will be eventually observed.

In general, $f^{x}_{n,1}$ will differ from x_{n+1} by a forecasting error term, $e^{x}_{n,1} = x_{n+1} - f^{x}_{n,1}$, and hence there will be uncertainty around it due to this error.

Once the forecast of x_{n+1} is available, then the ex-ante forecast of y_{n+1} at time n is obtained by replacing x_{n+1} in (4.37) by its forecast $f^{x}_{n,1}$, so that we have

$$f^{y}_{n,1} = c + a_1 y_n + b_0 f^{x}_{n,1} + b_1 x_n. \tag{4.48}$$

For this formula to be applied, an aside model for x is required to provide the forecast $f^{x}_{n,1}$. The usual way to tackle this problem is to fit a univariate time series model to x, such as an ARIMA model, use this model to predict x and then plug this prediction, $f^{x}_{n,1}$, into the corresponding forecasting equation for y given in (4.48).

As expected, the inclusion of the forecast of the explanatory variable in the forecasting equation for y_t will entail additional uncertainty in the forecast of y_t and will in turn influence the magnitude of the forecast error and the corresponding confidence interval. In particular, the one-step unconditional forecast error for y_t in the ADL(1,1) model that we are considering, assuming known values of the parameters, is computed from (4.36) and (4.48) as

$$e^{y}_{n,1} = y_{n+1} - f^{y}_{n,1} = b_0 (x_{n+1} - f^{x}_{n,1}) + \varepsilon_{n+1} = b_0 e^{x}_{n,1} + \varepsilon_{n+1}, \tag{4.49}$$

where $e^{x}_{n,1}$ is the one-step forecast error for x_t. A comparison of this formula with (4.38) makes it clear that, in ex-ante forecasting, the error due to forecasting the explanatory variable is an additional source of uncertainty besides the randomness of the innovation.

To forecast two or more steps ahead, forecasts of all the terms in the right hand side of (4.36) will be needed. In general, the h-step forecast, that is, the forecast of y_{n+h} at time n, will be

$$f^{y}_{n,h} = c + a_1 f^{y}_{n,h-1} + b_0 f^{x}_{n,h} + b_1 f^{x}_{n,h-1}$$

where $f^{x}_{n,h}$, $f^{y}_{n,h-1}$ and $f^{x}_{n,h-1}$ are the forecasts of x_{n+h}, y_{n+h-1} and x_{n+h-1} at time n, respectively. Then, the h-step forecast error is computed as

$$e^{y}_{n,h} = y_{n+h} - f^{y}_{n,h} = a_1 e^{y}_{n,h-1} + b_0 e^{x}_{n,h} + b_1 e^{x}_{n,h-1} + \varepsilon_{n+h} \tag{4.50}$$

where $e^{y}_{n,h-1}$ is the (h-1)-step forecast error for y_t and $e^{x}_{n,h}$ and $e^{x}_{n,h-1}$ are the h-step and (h-1)-step forecast errors for x_t, respectively.

The computation of the variance of the forecast errors will be further illustrated through some examples, distinguishing between stationary and non-stationary models. If the variables y_t and x_t are both stationary it will be seen (Example 1) that, although the variance of the error increases with the horizon, it is limited by the marginal variance of

y_t. However, if both y_t and x_t are $I(1)$, the uncertainty about the future values of y increases steadily without limit as the horizon increases (Example 2).

Example 1: Suppose that x_t follows an AR(1) model

$$x_t = \phi\, x_{t-1} + \upsilon_t \tag{4.51}$$

where υ_t is white noise. Then the forecast of x_{n+1} at time n is $f^x_{n,1} = \phi\, x_n$ and the corresponding forecast error for x_{n+1} is

$$e^x_{n,1} = x_{n+1} - f^x_{n,1} = \upsilon_{n+1}. \tag{4.52}$$

Substituting this expression in (4.49) gives the one-step error for y as

$$e^y_{n,1} = b_0\, \upsilon_{n+1} + \varepsilon_{n+1}. \tag{4.53}$$

Since the disturbances υ_t and ε_t are uncorrelated, the variance of the one-step forecast error can be immediately worked out from this expression as

$$\mathrm{Var}(e^y_{n,1}) = b_0^2\, \sigma_\upsilon^2 + \sigma_\varepsilon^2 \tag{4.54}$$

where σ_υ^2 and σ_ε^2 are the variances of υ_t and ε_t, respectively. Then, under the assumption of normality, the 80% forecast interval for y_{n+1} is computed by substituting (4.54) in (4.40) to obtain the following

$$f^y_{n,1} \pm 1.28\sqrt{b_0^2\, \sigma_\upsilon^2 + \sigma_\varepsilon^2}\ .$$

As expected, this interval is wider than that in (4.41) for ex-ante forecasting. Hence, the uncertainty about future forecasts is to be smaller when forecast is made conditional on given values of the explanatory variable than when this variable has to be forecast. Obviously, this is so because the inclusion of forecasts of the explanatory variables is an additional source of error in the forecasts for y and entails a reduction in accuracy.

Confidence intervals for more than one-step ahead forecasts are performed in a similar way. In particular, the two-step forecast error is computed by applying formula (4.50), with h=2, so that we come up with the following expression

$$e^y_{n,2} = y_{n+2} - f^y_{n,2} = a_1\, e^y_{n,1} + b_0\, e^x_{n,2} + b_1\, e^x_{n,1} + \varepsilon_{n+2} \tag{4.55}$$

where $e^y_{n,1}$ is given in (4.53) and $e^x_{n,1}$ and $e^x_{n,2}$ are the one-step and two-step forecast errors for x_t, respectively. Given that these two errors are correlated, it is convenient to write them down in terms of the disturbances υ_j's, which are independent white noise. In particular, the value of $e^x_{n,1}$ is given in (4.52) and the expression of $e^x_{n,2}$ is obtained as follows. Putting t=n+2 in (4.51) gives the forecast of x_{n+2} as $f^x_{n,2} = \phi\, f^x_{n,1} = \phi^2 x_n$, and then the corresponding two-step forecast error for x is

$$e^x_{n,2} = x_{n+2} - f^x_{n,2} = \phi\,(x_{n+1} - \phi\, x_n) + \upsilon_{n+2} = \phi\,\upsilon_{n+1} + \upsilon_{n+2}\,.$$

Putting back in (4.55) this expression and expressions (4.52) and (4.53) for $e^x_{n,1}$ and $e^y_{n,1}$ respectively, yields the following formulation for the two-step forecast error for *y*

$$e^y_{n,2} = b_0\,\upsilon_{n+2} + b_0\,\phi\,\upsilon_{n+1} + b_1\upsilon_{n+1} + a_1\,b_0\,\upsilon_{n+1} + a_1\,\varepsilon_{n+1} + \varepsilon_{n+2}\,.$$

Since all the innovations in this equation are uncorrelated, the variance of the two-step forecast error is computed as

$$\mathrm{Var}(e^y_{n,2}) = b_0^2\ \mathrm{Var}(\upsilon_{n+2}) + (a_1\,b_0 + b_0\,\phi + b_1)^2\,\mathrm{Var}(\upsilon_{n+1}) + a_1^2\ \mathrm{Var}(\varepsilon_{n+1}) + \mathrm{Var}(\varepsilon_{n+2}),$$

which becomes:

$$\mathrm{Var}(e^y_{n,2}) = [\,b_0^2 + (a_1\,b_0 + b_0\,\phi + b_1)^2]\,\sigma_\upsilon^2 + (1 + a_1^2)\,\sigma_\varepsilon^2\,. \qquad (4.56)$$

As expected, this variance is larger than that of the one-step forecast error in (4.54).

In general, it can be shown that for model (4.36), with x_t being an AR(1) process, the *h*-step forecast error in (4.50) becomes

$$e^y_{n,h} = a_1\ e^y_{n,h-1} + (b_0\,\phi + b_1)\ e^x_{n,h-1} + b_0\,\upsilon_{n+h} + \varepsilon_{n+h},\ \text{for } h \geq 2.$$

After some easy but tedious algebra, this expression can be written as a function of the innovations of both the regression and the AR(1) models. In particular, it turns out that

$$e^y_{n,h} = \{\,b_0^*\,\upsilon_{n+h} + b_1^*\,\upsilon_{n+h-1} + \ldots + b_{h-1}^*\,\upsilon_{n+1}\} + \{\varepsilon_{n+h} + a_1\,\varepsilon_{n+h-1} + \ldots + a_1^{h-1}\,\varepsilon_{n+1}\}, \quad (4.57)$$

where

$$b_0^* = b_0$$
$$b_i^* = a_1\,b_{i-1}^* + \phi^{\,i-1}\,(b_0\,\phi + b_1),\ \text{for } i \geq 1.$$

Since the value of $e^y_{n,h}$ in (4.57) has been written down by grouping terms of independent innovations, the variance of the forecast error can be computed as:

$$\mathrm{Var}(e^y_{n,h}) = (b_0^{*2} + b_1^{*2} + \ldots + b_{h-1}^{*\,2})\,\sigma_\upsilon^2 + (1 + a_1^2 + \ldots + a_1^{2(h-1)})\,\sigma_\varepsilon^2\,. \qquad (4.58)$$

The comparison of this expression with the equivalent error variance in conditional forecasting, given in (4.43), highlights how the uncertainty in unconditional forecasting is enlarged by the forecasting of the explanatory variables. Furthermore, expression (4.58) also enhances that the further the forecast horizon the larger the error variance of the forecast and, therefore, the more uncertainty about the point forecast and the wider the confidence interval. ■

In the discussion above it is important to understand how the structure of the forecast error from forecasting y_t is increasing along time. A summary of it is given in Table 4.xxa, where it can be seen that the forecast error has three components: (i) one due to the dynamic structure $b(L)$ that x_t incorporates in the model; (ii) one due to the lag structure $a(L)$ of y_t in the model; (iii) and the component due to the contemporaneous innovation ε_t. The first component (block I) is as complex as the contemporaneous and lagged values of x_t entering in the model. In the example above an ADL(1,1) model is considered and so only x_t and x_{t-1} are involved and we only have two columns in this block. However, if x_t entered in the model contemporaneously and with s lags, the number of columns in block I would be (s+1). In each column of block I, the additional complexity of the forecast error as the horizon increases, diagonals [a] and [b], depends on the dynamic univariate structure of x_t -in this case, it is relatively simple, as x_t is AR(1) - and it operates through columns from h equal to the order of the lag of x_t in the corresponding column. The second component of the forecast error, block II in the table, is due to the lag structure of y_t in the model and has two effects, one operating on x_t and another on the innovations ε_t, as in equation (4.44), so in the table we have the sub-blocks II.A and II.B. In the first case, the increase in complexity, diagonal [c], is due to the interaction of $b(L)$, $\phi(L)$ and $a(L)$ while in the sub-block II.B, diagonal [d] is only due to the dynamics on $a(L)$.

Table 4.xxa shows that if there were not dynamics in the model, in which case $b(L)$ and $a(L)$ were zero, y_t would be a white noise and the only source of error in forecasting will come from block III. If we had a dynamic relationship between y_t and x_t but without endogenous lags in the equation ($b(L)\neq 0, a(L)=0$), then the sources of errors would come from blocks I and III. Finally, if there was no dynamic relationship between y_t and x_t ($b(L)=0$), but y_t followed an autoregressive structure, the error term will be made up of the elements of sub-block II.B and block III. A summary of all these situations that can be encountered in different dynamic regression models is reported in Table 4.xxb.

At this point, a word of caution is required on the results above. Table 4.xxb is not saying that the more structure the model, the greater the absolute mean forecasting error, but it is pointing out that the complexity of the errors increases with the complexity of the model. In general, models with more structure that include really relevant explanatory variables, have been constructed with good data for all variables, have been well specified and estimated and their explanatory variables can be accurately forecast, will produce better forecasts than simpler models, because the variance of their innovation errors, ε_t and υ_t, will be considerably lower than the variance of the error term of, say, an ARIMA model for the dependent variable.

In the previous example, the variables x_t and y_t involved in the ADL(1,1) model are stationary. If this is not the case, the formulae for the forecasts of y_t and their forecast errors, given in (4.48)-(4.50), still apply but the forecast errors of x_t, x_{t-1} and y_{t-1} are quite different. In fact, as it was seen in Chapter 3, when forecasting an integrated variable, the error variance increases with the horizon and tends to infinite, rather than to a finite limit, as in the stationary case. For example, if the explanatory variable x_t in the ADL(1,1) model is nonstationary (and so is y_t) but it becomes stationary after differencing, then its first differences, i.e. Δx_t, will follow a stationary ARMA model and forecasts of x_t can be obtained from the forecasts of Δx_t. In particular, taking into

account that $x_t = \Delta x_t + x_{t-1}$, a forecast of x_{n+1} at time n will be obtained from the forecast of the first differenced series as follows

$$f^{x}_{n,1} = x_n + \text{forecast of } \Delta x_{n+1}$$

and the one-step forecast error of x_t will be

$$e^{x}_{n,1} = x_{n+1} - f^{x}_{n,1} = e^{\Delta x}_{n,1}$$

where $e^{\Delta x}_{n,1}$ is the one-step forecast error of Δx_t. Therefore, the one-step forecast errors of the integrated variable x_t and its stationary transformation Δx_t are the same, because they are just the contemporaneous innovation of the model.

To forecast h-steps ahead, we write x_{n+h} as the sum of its value at time n plus the values of its stationary transformation from (n+1) till (n+h) with unitary coefficients, i.e.

$$x_{n+h} = x_n + \Delta x_{n+1} + \Delta x_{n+2} + \ldots + \Delta x_{n+h}.$$

Then, the forecast of x_{n+h} at time n, say $f^{x}_{n,h}$, is given by the following expression:

$$f^{x}_{n,h} = x_n + f^{\Delta x}_{n,1} + \ldots + f^{\Delta x}_{n,h},$$

where $f^{\Delta x}_{n,j}$ is the forecast of Δx_{n+j} at time n. Then the h-step forecast error of x becomes

$$e^{x}_{n,h} = x_{n+h} - f^{x}_{n,h} = e^{\Delta x}_{n,1} + \ldots + e^{\Delta x}_{n,h}, \qquad (4.59)$$

where $e^{\Delta x}_{n,j}$ is the j-step forecast error of the stationary series Δx_t. Thus, the forecast error of the integrated variable x_t at time (n+h) depends on the forecast errors of its stationary transformation from (n+1) till (n+h). As the series Δx_t is stationary, the variance of its forecast errors tends to the marginal variance of the series itself, that is, as $h \to \infty$, Var($e^{\Delta x}_{n,h}$) tends to a constant given by the variance of Δx_t. Therefore, as $h \to \infty$, the variance of the h-step forecast error of x, given in (4.59), will tend to infinite because it will become an infinite sum of finite terms. As a consequence, the variance of the h-step forecast error of y_t, which depends on the forecast error variance of x_t, will also tend to infinite in this case.

We will further illustrate the results with non-stationary variables through an example, distinguishing whether the variables y_t and x_t are cointegrated or not.

Example 2: Let us again focus on the ADL(1,1) model, as in Example 1, but assuming now that the explanatory variable x_t in such model is nonstationary. In particular, let us suppose that x_t follows a random walk model

$$x_t = x_{t-1} + \upsilon_t$$

where υ_t is white noise. The forecast errors formulae in (4.49)-(4.50) still apply but the variance of the *h*-step forecast error will be different from that of the stationary case.

Let us start with the one-step forecasts. In Chapter 3 it was seen that in a random walk model the forecast of x_{n+1} at time n is $f^{x}_{n,1} = x_n$, and the corresponding forecast error is

$$e^{x}_{n,1} = \upsilon_{n+1}.$$

Then, substituting this expression in (4.49) provides the one-step error for y_{n+1} which is

$$e^{y}_{n,1} = b_0\, \upsilon_{n+1} + \varepsilon_{n+1,}$$

i.e., the same as that in (4.53) from the previous example where x_t was stationary. This is so because, as it was pointed out in Chapter 3, in one-step forecast, the only unknown and unpredictable terms are the contemporaneous innovations, and in the ADL(1,1) model these are the innovations of the endogenous and explanatory variables, that is, ε_{n+1}, multiplied by the b_0 coefficient of x_t in the model, and υ_{n+1}, respectively.

However, forecast errors beyond one-step ahead will be different from those in the stationary case, as it was highlighted in equation (4.59). In our particular example, it can be shown that the forecast error for x_{n+2} comes to be $e^{x}_{n,2} = e^{x}_{n,1} + \upsilon_{n+2} = \upsilon_{n+1} + \upsilon_{n+2}$, and putting this expression into equation (4.49), together with those of $e^{x}_{n,1}$ and $e^{y}_{n,1}$ given above, the corresponding two-step forecast error for *y* becomes

$$e^{y}_{n,2} = (b_0\, \upsilon_{n+2} + b_0\, \upsilon_{n+1}) + b_1 \upsilon_{n+1} + a_1\, b_0\, \upsilon_{n+1} + a_1\, \varepsilon_{n+1} + \varepsilon_{n+2}\,.$$

As before, the structure of the forecast error for y_t has three components (see Table xx). However, in this example, x_t follows a random walk, which is an AR(1) with coefficient ϕ=1, and hence the terms in the diagonals [a] and [b] of Table xx, which were decreasing to zero with the powers of ϕ in Example 1, no longer do. Actually, in this case, the variance of the two-step forecast error for *y* becomes

$$\mathrm{Var}(e^{y}_{n,2}) = [\,b_0^2 + (a_1 b_0 + b_0 + b_1)^2]\,\sigma_{\upsilon}^2 + (1 + a_1^2)\,\sigma_{\varepsilon}^2\,. \qquad (4.60)$$

Therefore, in ex-post forecasting, the assumption of stationarity for x_t makes a big difference with respect to the nonstationary assumption. In fact, a thorough comparison of (4.60) with its equivalent expression for the stationary case, given in (4.56), reveals that they only differ in one term, which is $b_0\phi$ in Example 1 and becomes b_0 in the present example (ϕ=1). As the forecast horizon goes further ahead, this difference between the stationary and the non-stationary cases becomes more remarkable, because a term involving powers of ϕ will eventually vanish in the stationary case ($|\phi|<1$) but it will not otherwise. As a consequence, the error variance in the non-stationary case will increase without limit as the horizon increases. In fact it can be proved that, in our particular example, the *h*-step forecast error admits the following expression

$$e_{n,h}^{y} = \{ b_0' \upsilon_{n+h} + b_1' \upsilon_{n+h-1} + \ldots + b_{h-1}' \upsilon_{n+1}\} + \{\varepsilon_{n+h} + a_1 \varepsilon_{n+h-1} + \ldots + a_1^{h-1} \varepsilon_{n+1}\},$$

where

$$b_0' = b_0$$

$$b_i' = a_1^i b_0 + \frac{1-a_1^i}{1-a_1}(b_0 + b_1), \text{ for } i \geq 1.$$

Then, since all the disturbances are uncorrelated, the *h*-step error variance is

$$\text{Var}(e_{n,h}^{y}) = (b_0'^2 + b_1'^2 + \ldots + b_{h-1}'^2)\sigma_{\upsilon}^2 + (1 + a_1^2 + \ldots + a_1^{2(h-1)})\sigma_{\varepsilon}^2 .$$

This expression highlights that as the horizon goes further ahead, the variance of the forecast error becomes larger and its limit is given, in this case, by

$$\lim_{h\to\infty} \text{Var}(e_{n,h}^{y}) = \sigma_{\upsilon}^2 \sum_{n=0}^{\infty} b_n'^2 + \sigma_{\varepsilon}^2 \sum_{n=0}^{\infty} a_1^{2n} . \qquad (4.61)$$

Recalling Table 4.xxa, it turns out that first term on the right hand side of equation (4.61) represents the variance of the dynamics of y_t due to the explanatory variable x_t, i.e. Block I and II.A in the table, while the second term in (4.61) accounts for the variance of the residual dynamics, i.e. Blocks II.B and III in the table. Taking into account that, in our example, the variables y_t and x_t are non-stationary and that the residual behaviour depends on whether y_t and x_t are or not cointegrated, we analyze next both situations separately.

First, if the variables x_t and y_t are cointegrated, the residual term η_{yt} defined in (4.45) is stationary ($|a_1|<1$) and therefore, the second term on the right hand side of (4.61) – columns II.B and III in Table xx– tends to a finite constant, namely its variance $\text{Var}(\eta_{yt}) = \sigma_{\varepsilon}^2/(1-a_1^2)$. Unlike, since the variable x_t is $I(1)$, i.e. $\phi=1$, the first summation on the right hand side of (4.61) –block I and column II.A in Table 4.xxa– diverges. In fact, its general term, $b_n'^2$, does not converge to zero, but to a non-zero limit given by

$$\lim_{n\to\infty} b_n'^2 = \frac{(b_0+b_1)^2}{(1-a_1)^2} \neq 0.$$

Thus the global limit in (4.61) becomes infinite and we can conclude that the variance of the forecast error of y_{t+h} tends to infinite with *h*.

On the other hand, if the variables x_t and y_t are $I(1)$ but are not cointegrated, then the residual term η_{yt} also becomes $I(1)$, because $a_1=1$. Therefore, in this case, both terms on the right hand side of (4.61) will tend to infinity and so will the global limit, leading to the same conclusion about the limit variance of $e_{n,h}^{y}$ as in the cointegrated situation.

Consequently, the variance of the forecast error of y_{t+h} will tend to infinite in any case, regardless the variables are cointegrated or not, that is, regardless of the behaviour of

the residual dynamics. Therefore, in ex-ante forecasting (see the last column of table 4.??), if x_t and y_t are $I(1)$ we will always have

$$\lim_{h\to\infty} \text{Var}(e^{y}_{n,h}) = \infty.\blacksquare$$

The discussion above assumes that the model parameters are known. In practice, these parameters are replaced by their estimates and this introduces another source of error in forecasting. In particular, equation (4.48) of the one-step forecast for y becomes

$$\hat{f}^{y}_{n,1} = \hat{c} + \hat{a}_1 y_n + \hat{b}_0 \; f^{x}_{n,1} + \hat{b}_1 x_n .$$

Subtracting this equation from (4.36) gives the following estimated forecast error

$$\hat{e}^{y}_{n,1} = (c-\hat{c}) + (a_1-\hat{a}_1) y_n + (b_0-\hat{b}_0) x_{n+1} + \hat{b}_0 e^{x}_{n,1} + (b_1-\hat{b}_1) x_n + \varepsilon_{n+1}. \quad (4.62)$$

This expression, as compared to equation (4.47), enhances the effect that estimating the parameters has on the forecast error. It seems clear that as far as the point estimates differ from the true parameter values, the larger is the forecast error and therefore the more uncertain is the forecast. However, using consistent estimators and sufficiently large sample sizes will diminish the effect of this source of error and expression (4.62) will collapse to (4.49).

4.10. DENSITY FORECAST AND THE FAN CHART

So far we have been able to obtain point forecasts and interval forecasts, either conditionally or unconditionally. However, to fully characterize the value of y to occur at a future time $t+h$, one needs a complete probability density function of the possible future values of y, so that the conditional probability of y_{n+h} taking values in any interval can be calculated. This probability density function will provide a complete description of the uncertainty about the forecast of y, in contrast to the point forecast, which does not contain any information about its associated uncertainty. Moreover, the density forecast will enlarge the information provided by the prediction interval, which only specifies the probability of the forecast falling within a certain range.

One way of producing a density forecast consists of assuming a functional form for the probability density and then estimating the parameters of that density based on the current and past information of the exogenous variables and possibly on some kind of judgemental adjustment from experts. This is the approach adopted by the Bank of England for predicting inflation and other macroeconomic magnitudes, whose density forecasts are then displayed in a very useful graphical device called the *fan chart.*[10]

If the density forecast was assumed to follow a Normal distribution, the most likely point forecast of y would be the conditional expectation. Hence, under the normality assumption, the mode, which is the value where the density forecast peaks its highest point, coincides with the mean and the density forecast spreads out symmetrically around it, with the values close to it being more probable than those further apart. The

[10] For a more detailed and technical explanation about the fan chart see the February 1998 *Inflation Report* from the Bank of England.

degree of dispersion of the future values around the central forecast (mean/mode) is measured by the other parameter of the Normal distribution, namely the variance.

As stated above, the mode corresponds to the most likely point forecast and it is estimated using an econometric model for the relationship between the variable to be forecast and its possible explanatory variables. From the information of the explanatory variables, the most likely forecast in the future for each of them is computed and with these values, other relevant information, and the impulse response functions from the estimated model, the most likely value of endogenous variable in the future is made up. The estimation of the variance is based on the record of forecast errors over the past periods. Anyhow, the estimated parameters values could always be subject to final adjustment from the experts' judgement.

The *fan chart* represents a picture of the density forecast describe above. As an example, Figure 4.?? displays a fan chart for the inflation in the Euro zone with information up to the *third* quarter of 2007. The solid line up to this point is the observed inflation. From this point onwards, this chart depicts, for each horizon, the point forecast and selected quantiles of the density forecast conditional to the information available at current time. These values constitute further lines after the period of the last observation. In our example, the added central line represents the most likely path that inflation is expected to take in the future. This line is made up by connecting the point forecasts at different horizons, i.e. by connecting the modes of the forecast densities for all the forecast horizons. Therefore each point on this line represents the most likely value of inflation at that period conditional on the current information available. Under the assumption of normality, this point forecast is the conditional mean.

INSERT FIGURE 4.??

Together with the solid central line, there is an open fan from the 4rth quarter of 2007 onwards. This fan represents for each horizon, several forecast intervals defined from different quantiles on the conditional distribution. These quantiles are chosen so that they define forecast intervals with an increasing coverage, starting, for example, with two quantiles that covers α% probability in the center of the distribution, following by those with coverage of 2α% and so on until a given coverage, say 80% or 90%, is reached. If the density forecast is assumed to be Normal the confidence bands will be symmetric around the central line. These successive intervals are displayed in the chart as different shaded zones, going gradually from the darkest one in the centre, associated with the narrowest interval, to the lightest ones in the extremes corresponding to the widest interval. In our example, the darkest central zone covers a 20% probability so there is a 20% chance that future inflation will lie within this band, given the current information. Moving away from the center we find pairs of identically shaded bands, each of them covering a 20% probability, so the probabilities being covered by successive bands are 20%, 40% and 60% up to the full coverage of all the coloured bands which is 80%. Therefore, if we made 100 point forecasts of future inflation, it would be expected that 80 out of them would lie somewhere within the entire fan. This means that there is only a 20% chance of observing a future inflation outside the fan.

As expected, the bands widen (as a fan) into the future indicating that the variance of the forecast density increases as the time horizon increases, that is, predicting many

periods ahead entails more uncertainty than predicting the closest future, a property that we already underlined in previous section. In fact, the width of the shaded bands is a measure of the uncertainty about the forecast in the sense that the wider the bands the more uncertain the forecast.

A remarkable property of the fan chart is that it allows estimating the probability of the variable being forecast to exceed a certain value. This can be very useful for monetary policy in an inflation targeting regime, for example. In this case, knowing that the probability of inflation being larger than the target value for a certain period is 50% has obviously no the same implications as knowing that this probability is 20%.

The assumption of a symmetric distribution for the forecast density, as the normal one explained above, rules out the possibility of some forecast errors being more likely in one direction than in the other, a situation that arises quite frequently in practice. In fact, the Bank of England does not use the normal distribution as the model for the density forecast, but the 'two-piece' normal distribution instead. This distribution spreads out around a central most likely value (the mode), which does not necessary coincides with the mean, and allows for a degree of asymmetry that is not accounted for by the normal distribution. The 'two-piece' normal is characterized by three parameters associated with central location, dispersion and skewness. The central location (mode) and the dispersion (variance) parameters are estimated as explained before. The estimation of the skewness parameter, which allows for asymmetry in the forecast density, mainly relies on the impulse response functions from the exogenous to the endogenous variable in the estimated econometric model and on the judgement of the experts about the balance of risks around the mode projection. For instance, if it is expected a higher probability that inflation would be above the mode than below, then the forecast density would be skewed to the right. Skewness to the left appears if values below the mode are expected to be more probable. From this "two-piece" normal density, the fan chart is built up in the same way as explained above but taking into account that as the density forecast is assumed to be asymmetric the same colour bands representing a certain coverage may be wider above the central path and narrower below it or *vice versa*. Wallis (1999) proposes a modification of the Bank of England's fan chart in such a way that the confidence bands are built up around the median with equal tail probabilities.

4.11. FORECASTING WITH REGRESSION MODELS: SOME EXAMPLES

4.12. FORECASTING AND CONTROL

As it was advanced at the beginning of this section, forecasting can also be used for control purposes. Actually, one of the main objectives of governments, companies, universities, and other institutions is to control their environments. A government may decide to try to bring down unemployment or inflation, to reduce a balance of trade or budget deficit, or to improve investment. A company will try to increase sales and profits by controlling prices and advertising expenditures. Provided that the number of control variables, whose values can be strongly influenced by the institution, is equal to the number of variables to be controlled, then in theory control can be achieved. However, the relationships between variables have to be well understood and various forecasts have to be formed carefully.

Suppose, for example, that y_t is the gross national product (GNP) of some economy, being a measure of the size of the total flow of the economy. It will be assumed that y_t is well modelled by an equation of the form

$$y_t = a + b\, y_{t-1} + c\, G_t + d\, x_t + \varepsilon_t \tag{4.63}$$

where G_t is government expenditure and x_t is GNP of an important neighbouring country. For example, y_t could be the GNP of Canada and x_t the GNP of U.S.A. Suppose for the moment that the government can choose the future value of G_t so that it decides at time t-1 what value G_t will take. Suppose that we are now at time t and that the government has a target value y^*_{t+1} that it would like y_{t+1} to take. The problem that it faces is how to choose the control variable G_{t+1} so that y_{t+1} is as near as possible to the target y^*_{t+1}. It is necessary to have a measure of nearness corresponding to the cost function discussed in Chapter 1. As was true there, a convenient measure is the squared difference between the variable one is trying to control and its target, i.e.,

$$C = E_t\,(y_{t+1} - y^*_{t+1})^2 \tag{4.64}$$

where E_t is the expectation taken at time t so that everything that occurs at or before time t can be treated as a constant.

Replacing t by t+1 in (4.63) and substituting into (4.64) gives

$$C = E_t\,(a + b\, y_t + c\, G_{t+1} + d\, x_{t+1} + \varepsilon_{t+1} - y^*_{t+1})^2$$

Both G_{t+1} and y^*_{t+1} are selected by the government at time t in this assumed situation. Let $f^x_{t,1}$ be the optimum forecast of x_{t+1} made at time t, then expanding the squared expression and taking expectations gives

$$C = (a + b\, y_t + c\, G_{t+1} + d\ \ f^x_{t,1} - y^*_{t+1})^2 + \mathrm{var}(\varepsilon_{t+1} + e^x_{t,1}) \tag{4.65}$$

where $e^x_{t,1} = x_{t+1} - f^x_{t,1}$ is the one-step forecast error for x and is assumed unforecastable at time t. The government needs to select G_{t+1} so that C is minimized, and this is easily seen to be achieved by making the first term in (4.65) zero, so that

$$G_{t+1} = -\,c^{-1}\,(a + b\, y_t + d\, f^x_{t,1} - y^*_{t+1}).$$

This analysis is an example of a more general theory that proves that proper forecasting of important variables in the area of consideration is necessary for successful control. If x_{t+1} is poorly forecast, the control value selected will be suboptimal and unfortunate surprises can occur. In practice, control by governments is less simple as they typically have great difficulty in achieving the required values of their control variables, such as government expenditure or money supply, because of political and institutional factors. See Granger (2002) for a further discussion on economic policy and forecasting.

FURTHER READINGS

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.F. Hendry (1993) *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Clements, M.P. and D.F. Hendry (1998) *Forecasting Economic Time Series*. Cambridge University Press.

Clements, M.P. and D.F. Hendry (1999) *Forecasting Non-stationary Economic Time Series*. The MIT Press.

Davidson, R. and J.G. MacKinnon (1993) *Estimation and Inference in Econometrics*. Oxford University Press. Chapters 19 and 20.

Dickey, D.A. and W.A. Fuller (1979) "Distribution of the estimators for autoregressivetime series with unit root", *Journal of the American Statistical association*, 74, 427-431.

Durbin, J. (1970) "Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables", *Econometrica*, 38, 410-421.

Enders, W. (2004) *Applied Econometric Time Series*, 2nd. Ed. Wiley

Engle, R.F. and C.W.J. Granger (1987) "Co-integration and error correction: representation, estimation and testing", *Econometrica*, 55, 251-276.

Granger, C.W.J. (1981) "Some properties of time series data and their use in econometric model specification", *Journal of Econometrics*, 16, 121-130.

Granger, C.W.J. (1999) *Empirical Modeling in Economics.* Cambridge University Press.

Granger, C.W.J. (2002) "Economic Policy and Uncertainty", manuscript.

Granger, C.W.J. (2003) "Time Series Analysis, cointegration and applications", Nobel Lecture.

Granger, C.W.J. and P. Newbold (1974) "Spurious regressions in econometrics", *Journal of Econometrics*, 2, 111-120.

ESTE NO SE CITA Granger, C.W.J. and P. Newbold (1986) *Forecasting Economic Time Series*. 2nd. Ed. Academic Press

Hendry, D.H. (1995) *Dynamic Econometrics*. Oxford University Press

Hendry, D.H., A.P. Pagan and J.D. Sargan (1984) "Dynamic Specification", In Z. Grilliches and m.D. Intrilligator (eds.) *Handbook of Econometrics*, vol. II, p. 1021-1099

Hendry, Doornik and Krolzig (??). PCGETS

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Dynamics and Control*, 12, 231-254.

Johansen, S. (1991) "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". *Econometrica*, 59, 1551-1580.

MacKinnon, J.G. (1991) "Critical Values for Co-integration Tests", in R.F. Engle and C.W.J. Granger (eds.), *Long-Run Economic Relationships*, Oxford University Press, 267-276.

Phillips, P.C.B. (1986) "Understanding spurious regressions in econometrics", *Journal of Econometrics*, 33, 311-340.

Sims, C.A., J.H. Stock and M.W. Watson (1990) "Inference in linear time series models with some unit roots". *Econometrica*, 58, 113-144.

Wallis, K.F. (1999) "Asymmetric Density Forecasts of Inflation anf the Bank of England's Fan Chart", *National Institute Economic Review*, January 1999, 0(167), 106-112.

APPENDIX A: GENERAL TO SPECIFIC METHODOLOGY AND PCGETS

A computer program to implement automatic model selection procedures from a general-to-specific econometric methodology is the *PcGets* program developed by **Hendry, Doornik and Krolzig (??)**. This program proceeds from what is called a General Unrestricted Model (GUM) which is formulated by the user and should include all the relevant information for the problem being tackled. This GUM is first estimated and then tested against possible mis-specifications to ensure data coherence. If no mis-specification problem is apparent, then the reduction procedure starts by eliminating statistically insignificant variables. At each simplification step, PcGets adopts a multi-path search strategy that carries out diagnostic tests on every initially-feasible path to ensure that no invalid reduction is performed (relevant information is not lost) and that the resulting model is congruentEXPLICAR EL TERMINO . Once all paths have been searched, it is expected that a set of admissible models emerges, although it could also happen that only a single model is selected. In this fortunate case, the final model is already found so that no further reduction is possible. Otherwise, the reduction process continues with encompassing testsCLARIFICAR ESTO to achieve a parsimonious model. If this step reveals several congruent and mutually ecompasing models, then some information criteria are used to select the final model. All this is automatically done in PcGets through a procedure called *Testimation* which also allows the user some flexibility to change the default settings for model selection and output reporting. After the estimation process, some graphical devices are available to perform residual analysis and parameters constancy tests on the final fitted model. Forecasts can also be done using a procedure called *Forecast*. Many of these issues will be discussed and illustrated in next sections.ESTO NECESITA DESARROLLARSE MUCHO MAS.

APPENDIX B: *ECM* REPRESENTATION OF THE GENERAL *ADL* MODEL

Let us consider the general ADL model with only one explanatory variable and without intercept given by the following equation

$$a(L)\, y_t = b(L)\, x_t + \varepsilon_t \tag{A.1}$$

where $a(L)=(1-\sum_{j=1}^{r} a_j L^j)$, $b(L)=(\sum_{j=0}^{s} b_j L^j)$ and ε_t is white noise. The long run relationship between y_t and x_t is given by

$$y = \frac{b(1)}{a(1)} x = \frac{b_0 + b_1 + ... + b_s}{1 - a_1 - ... - a_r} x.$$

where the coefficient $b(1)/a(1)$ is the long-run gain in y_t with respect to x_t.

To derive the ECM representation of (A.1) we will make use of a known result which states that any polynomial $\delta(L)= \delta_0+\delta_1 L +\ldots+\delta_p L^p$ can be decomposed as

$$\delta(L)= \delta+ \delta^*(L)(1- L) \tag{A.2}$$

where $\delta=\delta(1)$ and $\delta^*(L)$ is a polynomial of order p-1 whose coefficients are obtained by equating powers of L in both sides of the equation (A.2).

Let us write the polynomial $a(L)$ in (A.1) as

$$a(L)=1-L\,\bar{a}\,(L) \tag{A.3}$$

where $\bar{a}\,(L)= a_1+a_2L+\ldots+a_{r-1}L^{r-1}$. Applying formula (A.2) to this polynomial $\bar{a}\,(L)$ and substituting in (A.3), $a(L)$ can be alternatively written as:

$$a(L)=1-L\,\bar{a}\, - L\,\bar{a}\,^*(L)(1-L). \tag{A.4}$$

where $\bar{a} = \bar{a}\,(1)$. Applying now formula (A.2) to the polynomial $b(L)$ it can be written as

$$b(L)= b+ b^*(L)(1- L). \tag{A.5}$$

where $b = b(1)$. Putting back expressions (A.4) and (A.5) into equation (A.1) yields

$$y_t = \bar{a}\; y_{t-1} + \bar{a}\,^*(L)\, \Delta y_{t-1} + b x_t + b^*(L)\, \Delta x_t + \varepsilon_t.$$

If y_{t-1} is subtracted from both sides of (A.1) and bx_{t-1} is subtracted and added on the right hand side of it, then the above equation becomes

$$\Delta y_t = (\bar{a} - 1)\, y_{t-1} + \bar{a}\,^*(L)\, \Delta y_{t-1} + [b + b^*(L)]\, \Delta x_t + b x_{t-1} + \varepsilon_t,$$

and this can be alternatively written as

$$\Delta y_t = \bar{a}\,^*(L)\,\Delta y_{t-1} + b^{**}(L)\,\Delta x_t + \alpha\,(\,y_{t-1} - \beta x_{t-1}) + \varepsilon_t, \qquad \text{(A.6)}$$

where $\alpha=(\bar{a}-1)$, $\beta=b/(1-\bar{a})$ and $b^{**}(L)=b+b^{*}(L)$. However, from (A.3) it can be immediately seen that $(1-\bar{a})=a$, so that $\beta=b/a=b(1)/a(1)$, i.e. the long-run gain in y_t with respect to x_t.

Therefore, equation (A.6) is the EqCM formulation of model (A.1). This formulation explains changes in the current endogenous variable in terms of changes in its own past and changes in the explanatory variable, plus an adjustment to the past equilibrium errors, gathered in the term $\alpha(y_{t-1} - \beta x_{t-1})$, and a white noise disturbance ε_t. Thus, this model is a stationary formulation for a relationship between non-stationary variables, making use of the restriction that ties them in the long run as they are cointegrated.