4

Dynamic modelling – the general to specific methodology

In this chapter we give an account of one particular approach to econometric modelling. This approach has been developed largely by individuals associated with the London School of Economics and we will term it the LSE tradition. This approach does not of course enjoy universal support; alternative modelling frameworks have for example been proposed by Leamer (1978), Sims (1980) or Zellner (1971), to name but a few. Nonetheless, we regard the LSE tradition to be of sufficient importance that it warrants a detailed exposition. The founder of the LSE tradition is without a doubt Sargan who both through his own research Sargan (1964) and through the work of his students, Davidson *et al.* (1978), Davidson and Hendry (1981), Hendry and Von Ungern-Sternberg (1981), Hendry and Mizon (1978), Mizon and Richard (1986), etc. have had an enormous influence on applied econometrics.

In the first section we will outline the conceptual approach of the LSE tradition. Section 4.2 will examine a range of test procedures which have come to play a central role in applying the approach. The final section will give an example of the practical application of the approach using the demand for money.

4.1 The conceptual approach

At the centre of the LSE approach lies the concept of the data generation process (DGP), see Hendry Pagan and Sargan (1984), which represents a totally general statement of the joint probability distribution of all variables. As such it is too general to have any

direct practical use but its importance lies in providing a benchmark against which more simple models may be measured. It also allows us to formalise the assumptions and steps we need to make when we construct actual models for estimation.

Suppose x_t is a vector of observations on all variables in period t, and $X_{t-1} = (x_{t-1} \ldots x_1)'$, then the joint probability of the sample x_t, the DGP, may be stated as

$$\prod_{t=1}^{t} D(x_t | X_{t-1}; \Theta) \qquad \textbf{(4.1)}$$

where Θ is a vector of unknown parameters. The process of econometric modelling then consists of simplifying this very general formulation by imposing a set of restrictions. We therefore 'simplify' the DGP to yield a set of explicit equations complete with numerical parameter estimates. These simplifying assumptions may be categorised into four types.

1. Marginalise the DGP. The full DGP contains far more variables than we are normally interested in, or can possibly deal with. We therefore select a subset of 'variables of interest' and relegate the rest to a set of variables which are of no interest given the problem at hand.
2. Conditioning assumptions. Given the choice of 'variables of interest' we must now select a subset of these variables to be the endogenous variables (Y_t). These are then 'conditioned' or determined by the remaining variables (Z_t) of interest. The Z_t should be, at least, weakly exogenous for this 'conditioning' to be valid.
3. Selection of functional form. The full DGP is a general functional specification and before any estimation can be done a specific functional form for the model must be assumed.
4. Estimation. Finally, the unknown parameters in the assumed functional form must be replaced by a set of estimated numerical values.

It is wrong to think of these stages as being sequential. As Spanos (1986) has emphasised, the early stages of marginalising and conditioning are often done with a sharp eye on how the data perform at stage 4. It is therefore best to view the process of applied econometrics as an interaction among these stages until an adequate model is achieved.

Given the general DGP in (4.1) we may represent the first two assumptions by the following factorisation, where the function 'B' represents what one might usually refer to as the structural equations of interest.

$$D(x_t|X_{t-1};\Theta) = A(W_t|X_t:\alpha)\ B(Y_t|Y_{t-1}, Z_t: \beta) \times C(Z_t|Y_{t-1}, Z_{t-1}: \gamma) \quad (4.2)$$

The first component, A, specifies the determination of W, the variables of *no* interest, as a function of all the variables X_t. The second term B gives the endogenous variables of interest Y_t as a function of lagged Y and the exogenous variables Z_t. The final term C gives the determination of the exogenous variables Z_t as a function of the lagged endogenous and exogenous variables.

For the conditioning assumptions of the model to be valid we require that the Z_t variables are at least weakly exogenous. This means that Z_t is independent of Y_t, as is assumed in term 'C' in (4.2). It also requires that the parameters of interest of the model to be finally estimated (Θ) are a function of β only and the β and γ are variation free.

Other, more general, forms of exogeneity are strong exogeneity and super exogeneity: Strong exogeneity is given by the assumption that the third term in (4.2) takes the form $(Z_t|Z_{t-1}: \gamma)$ that is, the exogenous variables are determined without any reference to any lagged values of the endogenous variables Y_t. The strong exogeneity assumption therefore amounts to the assumption of weak exogeneity plus the assumption that Y does not 'Granger cause' Z. Super exogeneity is related to the Lucas (1976) critique. Lucas points out that when we model expectations by functions of lagged variables then the parameters of these functions may vary as the regime for determining the expectation variable changes. Super exogeneity rules out this possibility by assuming that the parameter vectors β and γ are independent. Under this assumption a change in the β vector will not influence γ. Super exogeneity is strong exogeneity plus this assumption of independence of β and γ. In general, weak exogeneity is all that is needed for estimation and testing, strong exogeneity is necessary for forecasting and super exogeneity for policy analysis.

Having made our assumption about the conditioning and marginalisation we may then state a partial log likelihood function for our model as

$$\log[L(\Theta)] = \sum_{t=1}^{t} L(\Theta; y_t|z_t, y_{t-1}) \quad (4.3)$$

and this may form the basis for estimation. It is important to realise, at this stage, that the assumptions needed to produce (4.3) are virtually never satisfied, in particular the chance of producing a correct and complete marginalisation of the data set is vanishingly small. As a result we can characterise the situation reached by equation (4.3) by

the statement that 'all models are false'. We do not therefore want to determine whether our model is the true model but rather we want to test the model to see if it is an 'adequate' model. A model derived as above cannot be regarded as correct or valid in an absolute sense but rather as a useful tentative hypothesis. A good model will be congruent with all the evidence, that is to say it will be a statistically acceptable representation of the data which cannot be unambiguously outperformed by any other known model.

The LSE tradition in practice

Given an economic variable to be explained, say Y, the first step is to use economic theory to determine a set of m explanatory variables $X_t = (X_1 \ldots X_m)_t$. This is the marginalisation of the complete set of all variables into the set of variables to be considered $(Y, X)_t$. The conditioning of the data is determined by which, if any, elements of X_t are deemed not to be weakly exogenous. Given that agents will normally be operating within an inherently dynamic environment, it is likely that the X_{it} will influence Y_t with a certain lag structure. Let n be the maximum lag with which an element of X influences the current value of Y, and suppose also that Y_t be a function of its own lagged values. (In practice, n will be determined on such considerations as the available degrees of freedom and the nature of the data, for example four lags perhaps for seasonally unadjusted quarterly data.) The real-world process generating Y_t is then assumed to be contained or nested within the linear model:

$$Y_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i Y_{t-1} + \sum_{k=1}^{m} \sum_{i=0}^{n} \beta_{ki} X_{kt-i} + u_t \quad (4.4)$$

where u_t is a white noise disturbance. Since economic theory generally has little to say about short-run dynamics, the LSE tradition starts with the general unrestricted form (4.4) as the maintained hypothesis. Having estimated (4.4) the next step is to sequentially impose economically meaningful restrictions on the maintained hypothesis, each restriction being tested for significance against the slightly less restricted specification which precedes it in the sequence.

Hendry *et al*. (1984) provide a typology of the various dynamic specifications which are nested within (4.4), an exposition of which can be given in terms of the simplest form of (4.4), when $m = n = 1$:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t \quad (4.5)$$

Imposing $\alpha_1 = \beta_1 = 0$ in (4.5) yields a static regression model, while setting $\beta_1 = 0$ yields the standard partial adjustment form. Setting $\alpha_1 = \beta_0 = 0$ indicates that X acts as a 'leading indicator' for Y_t. Imposing $\alpha_1 = 0$ makes Y depend on a finite distributed lag of X. If both of the restrictions $\alpha_1 = 1$ and $\beta_0 + \beta_1 = 0$ are accepted, then (4.4) can be reduced to a first difference formulation (if Y and X are logarithms, this yields an equation in the growth rates of Y and X). Setting $\beta_0 = 0$ in (4.5) yields Y_t as a function of lagged values of itself and X_t, and is termed by Hendry *et al.* a reduced form or 'dead start' equation, for obvious reasons.

An interesting reparameterisation of (4.5) may be used to show how the above dynamic equation can be represented as a 'static equation' with an AR(1) error. This is the basis of the so called 'common factor' test. If one believes that a static equation has an AR(1) error, one cannot simply perform Cochrane–Orcutt (1949) estimation and accept the ensuing parameter estimates; one must also 'pass' the common factor test. To illustrate the latter, rewrite (4.5) as:

$$[1 - \alpha_1 L]Y_t = \alpha_0 + \beta_0[1 + (\beta_1/\beta_0)L]X_t + u_t \tag{4.6}$$

If the restriction $\alpha_1 = -\beta_1/\beta_0$ is not rejected by the data, then the polynomial in the lag operator (in square brackets) contains a common element (factor) namely the coefficient in front of 'L'. Multiplying (4.6) by $(1 - \alpha_1 L)^{-1}$ and assuming the common factor restriction holds, we have

$$Y_t = \alpha_0^* + \beta_0 X_t + \varepsilon_t \tag{4.7}$$

where

$$(1 - \alpha_1 L)^{-1}u_t = \varepsilon_t \quad \text{or } \varepsilon_t = \alpha_1\varepsilon_{t-1} + u_t \tag{4.8a}$$

$$(1 - \alpha_1 L)^{-1}\alpha_0^* = \alpha_0 \quad \text{or } \alpha_0^* = \alpha_0/(1 - \alpha_1) \tag{4.8b}$$

Hence, imposing the non-linear restriction $\alpha_1\beta_0 + \beta_1 = 0$ in (4.5) is equivalent to assuming a first-order serially correlated error AR(1) in the static model (4.7). Equations (4.7)–(4.8) contain one less parameter than (4.5) and it is in this sense that serial correlation can be a convenient simplification rather than a nuisance (Hendry and Mizon 1978). To present the argument in a different vein, if the naïve researcher runs a highly restricted equation (4.7) and finds evidence of first-order serial correlation and performs Cochrane–Orcutt, the results from the latter cannot be accepted unless the researcher also runs equation (4.5) and tests the common factor restriction. The Cochrane–Orcutt regression is

$$(Y_t - \alpha_1 Y_{t-1}) = \alpha_0(1 - \alpha_1) + \beta_0(X_t - \alpha_1 X_{t-1}) + v_t \tag{4.8c}$$

with α_1 being the AR(1) parameter – obtained from the residuals in (4.7). The common factor likelihood ratio test is then

$$\text{LR}(k) = T \ln [\text{RSS}(4.8c)/\text{RSS}(4.5)]$$

where RSS is the residual sum of squares, k is the number of restrictions in (4.8c) – in this case 1 – and T is the number of observations.

$\text{LR}(k) < \chi_k^2$ leads to non-rejection of the common factor restriction and supports (4.7) and (4.8a). If the common factor restriction is rejected, then even if α_1, α_0^*, β_0 are statistically significant, they are inconsistent estimates. In the latter case we must assume some other form of serial correlation in (4.7), say MA(q), AR(p); $p > 1$, or accept (4.5), the dynamic equation as our new maintained model rather than (4.7) + (4.8).

Another reparameterisation of (4.5) introduces an error correction mechanism (ECM).

$$\Delta_1 Y_t = \alpha_0 + \beta_0\Delta_1 X_t - (1 - \alpha_1)(Y_{t-1} - X_{t-1}) + \gamma X_{t-1} + u_t \tag{4.9}$$

where

$$\gamma = \alpha_1 + \beta_0 + \beta_1 - 1 \tag{4.9a}$$

Equations (4.9) and (4.5) are just different ways of expressing the same equation. However, proponents of the LSE tradition would probably argue that (4.9), a form of error correction model (ECM), is more intuitively appealing than (4.5). To illustrate this point, note that the static equilibrium solution from either equation is given when $\Delta Y_{t-j} = \Delta X_{t-j} = 0$, and $Y_{t-j} = Y$; $X_{t-1} = X$ are constant (and $u_t = 0$):

$$Y = [\alpha_0/(1 - \alpha_1)] + [(\beta_0 + \beta_1)/(1 - \alpha_1)]X$$

However if $\gamma = 0$ in (4.9), then $\beta_0 + \beta_1 = (1 - \alpha_1)$ and the long-run static equilibrium solution is

$$Y = [\alpha_0/(1 - \alpha_1)] + X$$

Hence a t-test on γ in (4.9) provides a very simple way of testing for a long-run unit elasticity. Suppose $\gamma = 0$ is not rejected in (4.9), the dynamic equation then becomes

$$\Delta Y_t = \beta_0\Delta X_t - (1 - \alpha_1)(Y_{t-1} - Y_{t-1}^*) \tag{4.10}$$

where

$$Y_{t-1}^* = \alpha_0/(1 - \alpha_1) + X_{t-1} \tag{4.10a}$$

If $(1 - \alpha_1) > 0$ then if actual Y_{t-1} is above its long-run equilibrium value Y^*_{t-1}, we expect ΔY_t to fall in the next period, which brings actual Y_t closer to Y^*. Also, in (4.10) the growth in Y_t depends on the growth in X_t with a coefficient β_0. Hence $\beta_0 \Delta X_t$ and the ECM terms are 'sensible' dynamic decision variables given the long-run equilibrium condition of a unit elasticity.

In estimation the constant term in the equation for Y^*_{t-1} (4.10) would be estimated 'separately':

$$\Delta Y_t = \alpha_0 + \beta_0 \Delta X_t + \beta_1 (Y - X)_{t-1} + u_t \tag{4.11}$$

where $\beta_1 = -(1 - \alpha_1)$. Equation (4.11) *imposes* the long-run unit elasticity for any (non-zero) value of β_1 (and for dynamic stability $-2 < \beta_1 < 0$). One can use the ECM formation in a number of useful ways; for example, consider the following two equations:

$$\Delta Y_t = \beta_0 \Delta X_t - \beta_1 (Y - 0.9X)_{t-1} \tag{4.12a}$$

$$\Delta Y_t = \beta_0 \Delta X_{1t} + \beta_1 \Delta X_{2t} - \beta_3 (Y - X_1)_{t-1} - \beta_4 (Y - X_2)_{t-1} \tag{4.13a}$$

with long-run static equilibrium solutions

$$Y = 0.9X \tag{4.12b}$$

$$Y = [\beta_3/(\beta_3 + \beta_4)]\, X_{1t} + [\beta_4/(\beta_3 + \beta_4)] X_{2t} \tag{4.13b}$$

Thus equation (4.12a) *imposes* long-run elasticity of 0.9 while equation (4.13) imposes homogeneity between Y and X_1 *plus* X_2 (since the coefficients in square brackets *sum* to unity for *any* non-zero values of β_3 and β_4). The latter might represent a production function with constant returns to scale or a price mark-up equation on wage costs (X_{1t}) and raw materials costs (X_{2t}). Note that the restrictions apply to the long run and the dynamic response in the estimated equations (4.12) and (4.13) is reasonably general. To *test* for the above restrictions one merely adds γX_{t-1} to (4.12) and either $\gamma_1 X_{1t-1}$ or $\gamma_2 X_{2t-1}$ to (4.13) and performs a simple t-test on the appropriate γ.

Let us now consider 'growth effects' in the auto-regressive distributed lag (ADL) framework, our unrestricted ADL model is

$$Y_t = \alpha_0 + \sum_{i=1}^{4} \alpha_i Y_{t-1} + \sum_{i=0}^{4} \beta_1 X_{t-1} + u_t \tag{4.14}$$

and this unrestricted ADL form embodies 'growth effects': that is the level of Y depends not only on the *level of the* X *variables* but also on the *rate of growth* of X. To illustrate this point suppose we sequentially test and impose the restrictions on (4.14), namely

$$\alpha_i = 0, \qquad i = 2, 3, 4$$

$$\beta_i = 0, \qquad i = 3, 4$$

$$\alpha_1 + \beta_0 + \beta_1 = 1$$

then the restricted form of (4.14) is

$$\Delta_1 Y_t = \alpha_0 + \beta_0 \Delta_1 X_t - (1 - \alpha_1)(Y_{t-1} - X_{t-1}) + u_t \tag{4.15}$$

The steady-state 'growth solution' for (4.15) is obtained by using $\Delta X_t = g_x$, $\Delta Y_t = g_y$, $X_{t-1} = X_t - g_x$, etc. in (4.15)

$$Y_t = \frac{\alpha_0}{1 - \alpha_1} + X + \left[\frac{\beta_0 - (1 - \alpha_1)}{(1 - \alpha_1)}\right] g_x - \left(\frac{\alpha_1}{1 - \alpha_1}\right) g_y \tag{4.16}$$

Taking first differences of (4.16) and noting that by assumption $\Delta g_x = \Delta g_y = 0$, we obtain $g_y = g_x$ and hence (4.16) becomes

$$Y_t = \frac{\alpha_0}{(1 - \alpha_1)} + X + \left[\frac{(\beta_0 - 1)}{(1 - \alpha_1)}\right] g_x \tag{4.17}$$

Unless $(\beta_0 - 1)/(1 - \alpha_1) = 0$ then Y depends on the growth in X, in steady state. The impact of g_x on Y can often be large empirically. Although growth effects are usually not implied by economic theory, however, if we impose the restriction of a zero growth effect we may severely distort the lag structure. In our extremely simple illustrative example a zero growth effect implies $\beta_0 = 1$, that is the short-run and long-run response is the same: a strong restriction compared with the lagged effects in the unrestricted ECM, (4.15). In practice, growth effects are often ignored and in any case one would not expect a 'constant growth solution' from an equation estimated over data that is quite short and volatile to yield very precise estimates of 'growth effects'.

The ECM specification can be justified theoretically within (finite or infinite horizon) quadratic costs of adjustment framework (Hendry and von Ungern-Sternberg 1981; Nickell 1985). Less formally, the specification captures the idea that agents alter their behaviour according to 'signals' that they are out of equilibrium. For example, if Y is the logarithm of (real or nominal) money stock and X is the logarithm of (real or nominal) income, then the error correction term $(Y_{t-1} - X_{t-1})$ is the logarithm of the money income ratio lagged once. Deviation in the money–income ratio from its long-run value will lead to future changes in money holdings by agents, in order to move closer to their desired long-run position.

The ECM specification has worked well in a number of empirical studies of the demand for money in the UK, both for broad definitions (Hendry and Mizon 1978) and a narrow definition (Hendry 1980) and also for the US demand for money (Baba *et al.* 1987). An ECM consumption function (Davidson *et al.* 1978, Hendry 1983) for the UK also performs well statistically.

Two further points should be noted about the econometric methodology outlined in this section. Firstly, general-to-specific methodology will inevitably involve a certain amount of 'data mining' (or 'lag mining' perhaps). For this reason it has become customary to subject the final 'preferred' equation to a number of diagnostic checks. Whilst these checks will usually have greatest power against a specific alternative hypothesis (higher-order serial correlation, heteroskedasticity, etc.), they will usually also give some idea of the general adequacy of the specification.

Secondly, when agents have forward-looking expectations, the parameters of dynamic models of this kind will generally be functions of the parameters of agents' objective functions and of the historically given stochastic environment (see Hansen and Sargent 1980; Sargent 1981; Cuthbertson and Taylor 1987). They may therefore be subject to the Lucas (1976) critique.

4.2 Testing the dynamic model

Clearly when we move from the general dynamic model (4.4) to a restricted parameterisation of it such as (4.15) it is important to test the model in a number of ways. It is important to test the general unrestricted model for a homoscedastic serial uncorrelated error process at an early stage since all further testing is (usually) dependent on white noise errors in the maintained hypothesis. We must test the restrictions directly and we must also check that the assumptions made about the residuals are not violated in the restricted model.

A range of test statistics is used to assess the validity of a model and this section will outline some of the most common. The underlying theoretical derivation of these statistics will not be given here as Chapter 2 discusses the construction of the three main classes of test procedures and the relationship between them. Often an ECM model is contructed or 'designed' so that one ensures it passes a set of diagnostic 'tests'. Tests for parameter constancy and encompassing tests then become of increasing importance in testing competing models.

Testing the restrictions (F-test)

At each stage in moving from our general equation to our 'best' equation we need to test the acceptability of the restrictions we are imposing on the model. In the case of a simple exclusion restriction this can be done using the standard t test. In the case of a combination of restrictions or a set of linear restriction involving more than one parameter we must use a more general procedure. The most commonly used test is the F-test which is a special version of the likelihood ratio test. Suppose we have a general 'unrestricted' model $Y_t = \beta_0' X_t + u_t$ and a restricted set of parameters β which gives $Y_t = \beta_1' X_t + v_t$, where β_0 contains fewer non-zero coefficients than β_1. Then we may construct a test of these restrictions by estimating both the unrestricted and the restricted models (as both are needed for the construction of the test it is a likelihood ratio test).

We define RSS_1 to be the residual sum of squares from the unrestricted regression and RSS_2 to be the residual sum of squares from the restricted regression. The F-test may then be most conveniently calculated as

$$F(m, T-k) = \left[\frac{\text{RSS}_2 - \text{RSS}_1}{\text{RSS}_1}\right]\left(\frac{T-k}{m}\right) \tag{4.18}$$

where T is the total sample size, k is the number of parameters in the unrestricted model and m is the number of restrictions. This is then distributed as $F(m, T-k)$. This test allows a wide range of restrictions and combinations of restrictions to be tested although it must be remembered that when a number of restrictions are tested jointly, rejection may be due to only one of the restrictions being invalid.

The intuition behind this test is simple; if the restriction is valid then we would expect RSS_2 to be only slightly larger than RSS_1. We are therefore testing for an increase in RSS_2 which is 'too' large to be due to chance.

The Durbin–Watson statistic (DW)

One of the earliest tests for serial correlation in the error process is due to Durbin and Watson (1950); this test is still used widely so we present it here, although it does have a number of disadvantages. In particular it is known to be inappropriate when the model contains a lagged dependent variable and also the rejection criteria consists of a region rather than an actual point. The formula for the DW statistic is

$$\text{DW} = \frac{\sum_{t=2}^{T} (u_t - u_{t-1})^2}{\sum_{t=1}^{T} u_t^2} \tag{4.19}$$

where u_t is the residual from the estimated equation, $Y_t = \beta' X_t + u_t$. It may easily be shown that

$$\text{DW} \simeq 2 - 2\rho$$

where ρ is the first-order serial correlation coefficient in the residual process $u_t = \rho u_{t-1} + v_t$. When there is no serial correlation, $\rho = 0$ and the DW statistic takes a value of 2. Positive serial correlation ($\rho_{max} = 1$) produces a DW < 2 while negative serial correlation produces a DW > 2. We set up the null hypothesis H_0: $\rho = 0$, and a DW value sufficiently far away from 2 rejects this hypothesis (in favour of the assumption that serial correlation is present). The DW statistic can be generalised to tests of higher-order serial correlation but other tests are more frequently used in such cases.

The Lagrange multiplier (LM) test for serial correlation

A more satisfactory test for serial correlation may be constructed using the Lagrange multiplier approach discussed in Chapter 3. This test has an asymptotically exact distribution and is valid in the presence of lagged dependent variables. It can also be constructed to test for any order of serial correlation. We begin by setting up two general models of the error process, an AR(m) model.

$$u_t = \rho_1 u_{t-1} \ldots + \rho_m u_{t-m} + \varepsilon_t \tag{4.20}$$

and a MA(m) one

$$u_t = v_t + \rho_1 v_{t-1} \ldots \rho_m v_{t-m}$$

where ε_t and v_t are white noise errors and u_t is the error term from the structural equation $y_t = \Sigma\alpha_i y_{t-i} + \beta' X_t + u_t$. The null hypothesis H_0: $\rho_1 \ldots \rho_m = 0$, is that there is no serial correlation. The LM statistic is based on the R^2 from the auxiliary regression.

$$\hat{u}_t = \gamma_1 \hat{u}_{t-1} \ldots + \gamma_m \hat{u}_{t-m} + \sum_{i=1}^{K} \alpha_i y_{t-i} + \beta' X_t \tag{4.21}$$

where $\hat{u}_t$ is the residual from the structural equation; $\hat{u}_t$ is the residual utilising consistent parameter estimates $(\hat{\alpha}_i, \hat{\beta})$.

The LM test statistic with m degrees of freedom is then given by LM(m) = TR^2, where T is the sample size, and this is asymptotically distributed as $\chi^2(m)$, under the null. Intuitively if H_0 is true we expect γ_i in (4.21) to be zero, for the R^2 from (4.21) to be low and hence LM(m) to be 'small' and less than $\chi^2(m)$.

One difficulty with tests of this form, based on the LM procedure using an auxiliary regression, is that they are valid only when the estimation procedure is OLS. If any form of IV estimation is used then this LM test (and the Breusch–Pagan, Arch or Reset tests given below) is invalid. Breusch and Godfrey (1981) have however suggested a generalisation of the auxiliary regression LM procedure in the case of instrumental variable estimation. The null hypothesis is H_0: $\rho_1 = \ldots \rho_m = 0$, and the modified LM(MLM) test is then given as

$$\text{MLM} = T(R_1^2 - R_2^2)$$

where R_1^2 is the R^2 statistic of the OLS regression of $\hat{u}_t$ on the full set of instruments used in the estimation process (and $\hat{u}_t$ are the structural errors generated by (4.4) with $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_{ki}$ the IV estimates). R_2^2 is the R^2 statistic of the OLS regression of q_t on the same set of instruments (where q_t is the residuals generated by (4.4) when $\hat{u}_{t-1} \ldots \hat{u}_{t-m}$ is also added to the set of explanatory variables in (4.4)). This test is again disributed as $\chi^2(m)$.

Instrument validity test

Much of this chapter has proceeded on the assumption that the estimation technique being used is OLS; section 1.6 (Chapter 1) demonstrated that when the right-hand side variables are not all weakly exogenous then a suitable estimation strategy is instrumental variables (IV). The choice of a correct set of instruments in the absence of a complete knowledge of the system is difficult and even if we know the full system the full set of instruments may be too large given the available data set, so a subset may have to be used. We would then naturally wish to test our chosen set of instruments to see if they are independent of the structural error term ε_1. Under the null hypothesis that the instruments are independent of the error term, the IV/2SLS residuals are consistent (see Chapter 1). Define the instruments as

$$W = (w_1, x_1)$$

where x_1 is the weakly exogenous variables in the equation and w_1 are the instruments for the endogenous variables. If W is independent of ε_1 we would expect a regression of ε_1 on W to yield a low R^2. This intuitive argument is consistent with the Sargan instrument validity test. In place of the unobservable ε_{1t}, we use the IV residuals $\hat{\varepsilon}_{1t}$.

The required OLS regression is:

$$\hat{\varepsilon}_{1t} = W\hat{\alpha}$$

The R^2 from this regression is then used to form the Sargan test:

$$\text{SARG} = (T - k)R^2 \sim \chi^2(r)$$

where

T = number of observations

k = number of parameters in the structural equation

r = number of over-identifying restrictions (the number of instruments in w_1 minus the number of endogenous variables on the right-hand side of the equation)

Under the null(H_0) of independence of the instruments and errors SARG is asymptotically distributed as $\chi^2(r)$ and hence for SARG $< \chi^2_c$ we 'accept' H_0. If SARG is greater than the chosen critical value then we conclude that at least one of the instruments is correlated with the error term and the IV estimates are invalid. The Sargan test may be written in an alternative form which often appears in the literature:

$$\text{SARG} = (\hat{\varepsilon}_1{}' P_w \hat{\varepsilon}_1)/s^2$$

where P_w = projection matrix of instruments = $W(W'W)^{-1}W'$

$$s^2 = (\hat{\varepsilon}_1{}'\hat{\varepsilon}_1)/(T - k)$$

These two forms of the test may easily be shown to be equivalent.

The Box–Pierce and Ljung–Box test

Clearly an important source of information in detecting the presence and form of serial correlation, for example AR(1) versus MA(1), is the correlogram, discussed in Chapter 3. Qualitative examination of the correlogram is an important diagnostic tool but it does not constitute a formal statistical test. The Box–Pierce and its related test the Ljung–Box test are both portmanteau tests which allow us to test the hypothesis that the first m points on the correlogram are random with a true value of zero. If we define r_i as the ith autocorrelation coefficient (or point on the correlogram) then it may be shown that asymptotically, r_i is approximately $N(0, T^{-1/2})$ under the null of no serial correlation of order i. Hence for $T = 64$ observations any individual $|r_i| > 0.23$ is indicative of serial correlation of order i (this test is very approximate for $i = 1 \ldots 4$ and more precise for $i > 4$).

Box–Pierce test (generally denoted Q) is defined as

$$Q = T \sum_{i=1}^{m} r_i^2$$

and asymptotically this will be distributed as $\chi^2(m)$. In fact it has been noted that the Q statistic has rather poor small sample properties and a better small sample statistic is given by Ljung–Box (often denoted Q^*) statistic which is defined as

$$Q^* = T(T + 2) \sum_{i=1}^{m} (T - i)^{-i} r_i^2 \qquad \textbf{(4.23)}$$

This is again distributed as $\chi^2(m)$ under the null hypothesis of no serial correlation. Intuitively, if a subset of r_i^2 are 'large' then Q (or Q^*) will be 'large' indicating the presence of serial correlation. For both LM and $Q(Q^*)$ acceptance of H_0 for say $m = 8$, requires one to check the individual r_i to see if a large number of r_i close to zero do not mask the presence of a highly significant individual or subset of r_i.

Heteroscedasticity

The general Breusch–Pagan procedure

The most general forms of heteroscedasticity considered in the econometric time series literature usually take the form

$$\sigma_t^2 = \sigma^2 \alpha' X_t = \sigma^2(\alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \ldots) \qquad \textbf{(4.24)}$$

$$\sigma_t^2 = \sigma^2(\alpha' X_t)^2 = \sigma^2 [\alpha_0^2 + \alpha_1^2 X_{1t}^2 + \ldots + \sum_{i \neq j} \sum \alpha_i \alpha_j X_{it} X_{jt}] \qquad \textbf{(4.25)}$$

$$\sigma_t^2 = \sigma^2 \exp(\alpha' X_t)^2 = \sigma^2 \exp[\alpha_0^2 + \alpha_1^2 X_{1t}^2 + \ldots + \sum_{i \neq j} \sum \alpha_i \alpha_j X_{it} X_{jt}] \qquad \textbf{(4.26)}$$

where X_t is a vector of variables which is assumed to be associated with the changing variance of the errors u_t. (The first element of X_t is a constant, and α is a suitably dimensioned vector of parameters.) Often X_t consists of a *subset* of the variables of the 'structural equation' $Y_t = \beta X_t + u_t$ (where X_t may contain lagged dependent variables but this is not necessary for the procedure to be valid).

Breusch and Pagan (1979) point out that the assumption of homoscedastic errors is equivalent to the null hypothesis

$$H_0: \alpha_1 = \alpha_2 = \ldots \alpha_m = 0$$

Under H_0, $\alpha_t = k\sigma^2$ (where k is a constant) and is therefore constant and homoscedastic. They propose a standard LM test of this hypothesis based on the auxiliary regression for (4.24) for example:

$$(\hat{u}_t^2/\hat{\sigma}^2) = \alpha_1 + \alpha_2 X_{2t} + \ldots \alpha_m X_{mt} \quad \textbf{(4.27)}$$

where $\hat{\sigma}^2$ is the standard error of the structural equation $Y_t = \beta' X_t + u_t$. Once again the LM test in this case is $HT(m) = TR^2$, where the R^2 is from equation (4.27). Under H_0, $HT(m)$ is asymptotically distributed as $\chi^2(m)$. The intuition behind this test is as follows. Under the null $\alpha_2 = \ldots \alpha_m = 0$ and so the R^2 of this regression should be zero. If the R^2 is high then it says there is a systematic movement in u_t^2 which is highly correlated with one or more of the X variables and so $E(u_t^2) \neq \sigma^2$ (a constant).

Testing for an ARCH process

An alternative form of heteroscedasticity is termed auto-regressive conditional heteroscedasticity (ARCH). Instead of relating σ_t^2 to a vector of variables (X) as above, u_t^2 is assumed to depend on past squared errors $u_{t-1}^2, u_{t-2}^2 \ldots$ The ARCH process is autoregressive in the second moment. Engle (1982) proposed a LM test for the presence of an ARCH process. The appropriate auxiliary regression in this case is:

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 \hat{u}_{t-1}^2 + \ldots + \alpha_m \hat{u}_{t-m}^2 \quad \textbf{(4.28)}$$

and again the test statistic $\text{ARCH} = TR^2$ from (4.28). Under $H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$ ARCH is asymptotically distributed as $\chi^2(m-1)$. The most common form of this test considers only the first order autoregressive model ($m = 1$).

Parameter stability tests

Two types of Chow test (denoted c_1 and c_2 below) are used to test for statistical parameter stability, that is whether parameters remain stable given that they are always estimated with error. The general idea of parameter stability tests is that we have some known data T_1, after which we believe a structural break may have occurred in the model. So there is the possibility that the general model has the form

$$Y_t = B'_1 X_t + u_t, \qquad u_t \sim N(0, \sigma_1^2): t < T \quad \textbf{(4.29)}$$

and

$$Y_t = B_2' X_t + u_t, \qquad u_t \sim N(0, \sigma_2^2): t \geqslant T_1 \quad \textbf{(4.30)}$$

The total number of observations is $T = T_1 + T_2$. $T_1 = 1 \ldots T_1$ $T_2 = T_1 + 1 \ldots T$

The null hypothesis that the model is structurally stable is H_0: $B_1 = B_2$ *and* $\sigma_1^2 = \sigma_2^2$. This of course involves two separate hypotheses H_0^1: $B_1 = B_2$ and H_0^2: $\sigma_1^2 = \sigma_2^2$ where we are generally more interested in testing H_0^1 than H_0^2.

A complication which arises in constructing tests of this hypothesis lies in the choice of T_1. In order to estimate both of the models (4.29) and (4.30) we require $T_1 > k$ and $T - T_1 > k$, where k is the number of regressors in the model. This is simply a requirement that there are sufficient degrees of freedom in both sub-samples to estimate the models. We need to consider a test statistic for the case where both $T_1 > k$ and $T - T_1 > k$ holds, and when it does not.

Case A: $T_1 > k$ and $(T - T_1) > k$

This is an analysis of variance (ANOVA) test. In this case we can estimate the model over the whole period and each of the sub-samples. We define RSS_T as the residual sum of squares for the model estimated over the whole period, RSS_1 as the residual sum of squares over the period with T_1 observations, and RSS_2 as the residual sum of squares for the second period with T_2 observations. Then under the null H_0 (the joint hypothesis), the statistic c_1:

$$c_1 = \left(\frac{\text{RSS}_T - (\text{RSS}_1 - \text{RSS}_2)}{\text{RSS}_1 + \text{RSS}_2}\right)\left(\frac{T - 2k}{k}\right) \quad \textbf{(4.31)}$$

is distributed as $F(k, T - 2k)$. c_1 is commonly called the Chow test (Chow 1960). We can also separately test H_0^2, namely $\sigma_1^2 = \sigma_2^2$ using the statistic

$$V_1 = \frac{s_2^2}{s_1^2} \equiv \left(\frac{\text{RSS}_2}{\text{RSS}_1}\right)\frac{(T_1 - k)}{(T_2 - k)} \quad \textbf{(4.32)}$$

where s_i is the standard error of the appropriate regression in periods T_1 and T_2 $s_i = \text{RSS}_i/(T_i - k)$. V_1 is distributed as $F(T_2 - k, T_1 - k)$ under the null that $\sigma_1^2 = \sigma_2^2$. Intuitively the test (V_1) for equality of variances in the two sub-samples is straightforward. If we have equal variances across sub-samples then $V_1 = 1$ and it will be less than the critical value of the F distribution.

Since c_1 tests for the *joint* hypothesis H_0 it is useful to first test V_1. If V_1 is not rejected (i.e. $\sigma_1^2 = \sigma_2^2$) then we test c_1. Rejection of c_1 then implies $B_1 \neq B_2$. If V_1 is rejected we would also expect c_1 to be rejected but we cannot say whether the latter implies that $B_1 \neq B_2$. Inference on $B_1 = B_2$ in such circumstances must remain inconclusive.

Case B: $T_2 < k$

It is usual to consider only the case where $T_2 < k$ since the case of $T_1 < k$ may be dealt with in an exactly analogous fashion. When there are not enough degrees of freedom to estimate B_2 or RSS_2 directly, a second version of the Chow test is possible:

$$C_2 = \left(\frac{\text{RSS}_T - \text{RSS}_1}{\text{RSS}_1}\right)\left(\frac{T_1 - k}{T_2}\right) \quad \textbf{(4.33)}$$

This is distributed as $F(T_2, T_1 - k)$ under the null that $B_1 = B_2$ against the alternative that $B_1 \neq B_2$ *and* $\sigma_0^2 = \sigma_2^2$. C_2 is a joint hypothesis and to test separately for constant error variance $\sigma_1^2 = \sigma_2^2$ we proceed as follows. Estimate over the first T_1 observations to obtain $\hat{B}'$. If we denote the values of (Y_t, X_t) over the second period as Y_t^2, X_t^2, then the one-step-ahead forecast errors (using $\hat{B}'$) are $\tilde{u}_{2t} = Y_t^2 - \hat{B}'X_t^2$ (there are T_2 of these) under the null $\sigma_1^2 = \sigma_2^2$, the variance of these one-step-ahead forecast errors in the second period, should equal those in the first period – as measured by $s_1^2 = \text{RSS}_1/(T_1 - k)$. Under the null that $\sigma_1^2 = \sigma_2^2$,

$$\text{HF}(T_2) = (s_1^2)^{-1} \sum \tilde{u}_{2t}^2 \quad \textbf{(4.34)}$$

is distributed as $\chi^2(T_2)$. This test is sometimes referred to as the Hendry forecast test. Again, the sequence of testing should be first to use HF to check that $\sigma_1^2 = \sigma_2^2$ cannot be rejected, and then C_2 to check that $B_1 = B_2$ cannot be rejected.

Although HF is a test of constant error variances, it may also be viewed either as an indicator of *numerical* parameter constancy or equivalently as a test of the *relative* accuracy of out-of-sample point forecasts. s_1^2 is a measure of the *within* sample variance of the errors or *within sample* forecast accuracy since $\hat{u}_t^1 = y_t^1 - \hat{y}_{2t}'$. The $\tilde{u}_{2t}$ series measure out-of-sample forecast errors (using the estimate of $\hat{B}$ based on the first T_1 observation.) If B is numerically the same in T_1 and T_2 periods we would expect $\tilde{u}_{2t}$ to be of the same order of magnitude as $\hat{u}_{1t}$. Hence HF would be unity for each of the T_2 periods, and $\text{HF} < \chi_c^2$, that is we do not reject numerical parameter constancy. A word of caution: if the equation fits badly within sample (s_1^2 large) then one may have $\text{HF} < \chi_c^2$ but the *absolute* value of the out-of-sample forecast errors $\tilde{u}_{2t}$ may be large. Here we have a 'bad' fit within the sample and equally poor predictions out-of-sample. It is therefore worth looking at individual $\tilde{u}_{2t}$ values.

These structural stability tests may be used more powerfully in a recursive setting by computing a sequence of tests where the 'break' period moves through time (see below). Another useful test procedure is the Salkever (1976) test which is similar in approach to the Chow tests. In this test a set of dummies ($\text{DV}_i = 00 \ldots 0100$, each with unity in the ith period) are added to the equation for j sub-periods. Then a joint F-test of the significance of the set of dummies is constructed to test for a structural break over the sub-period. The size of each dummy coefficient is equal to the out-of-sample forecast error and the 't' statistics on individual coefficients indicate those periods for which the equation undergoes a statistically significant shift in its parameters.

Recursive estimation and testing structural stability

Recursive estimation may be viewed as a special case of the Kalman filter and as such it is a powerful and interesting technique of its own right. In this section, however, we will be considering the more limited use of recursive estimation in testing the stability of structural models. One of the difficulties of the formal stability tests presented above is that we make the assumption that a possible break point is known '*a priori*' and we simply wish to test this known point. In general, however, we have no strong prior knowledge of specific structural breaks and so it is useful to have a general framework to investigate the stability of a model. Recursive estimation provides such a framework.

Recursive estimation may be thought of as a series of conventional OLS estimation of the same model where the data period is increased successively by one period in each estimate. It therefore produces a time series of estimates of β, $\hat{\beta}_t$ from the estimated equation:

$$Y_i = \hat{\beta}_t X_i + \hat{u}_i \qquad i = 1 \ldots t; \quad t = k \ldots T \quad \textbf{(4.35)}$$

It must be stressed that while $\hat{\beta}_t$ varies, the underlying β is assumed to be constant, so this is not a time-varying parameter model. We simply derive varying estimates of the constant β from different data sets. It is intuitively clear that if our model is structurally stable the variation in $\hat{\beta}_t$, as we move through time, should be small and random. So sudden large changes in $\hat{\beta}_t$ may indicate periods of structural break, while non-random or trend movements in $\hat{\beta}_t$ may indicate some underlying misspecification. Once a specific period of instability is detected we could then turn to one of the structural tests above. However, the usefulness of recursive estimation does not end with the estimation of $\hat{\beta}_t$ as we see below. The recursive residuals are defined as:

$$v_t = Y_t - \hat{\beta}_{t-1}' X_t \qquad t = k + 1 \ldots T$$

This amounts to the one-step-ahead forecasting error made by the

OLS estimation procedure. Under the null hypothesis that β is constant and $u_t \sim N(0, \sigma^2)$ then $v_t \sim N(0, \sigma^2 d_t^2)$ where $d_t = (1 + x_r'(X_{r-1}'X_{r-1})^{-1}x_r)^{-1}$ defining $X_{r-1}' = (x_1, \ldots, x_{r-1})$, and so we may define the *standardised recursive residuals* as

$$w_t = v_t/d_t \sim N(0, \sigma^2)$$

While the standardised recursive residuals follow the same distribution as the OLS residuals they have a number of advantages. The first is that the OLS residuals are constrained (when a constant is included in the regression) to sum to zero. So, by definition there can be no overall departure of the residuals from zero. This is not true of the recursive residuals and so they will often show systematic departures from zero if there is any misspecification of time variation in the parameters. The second important property of the recursive residuals is that it may be shown that

$$\text{RSS}_t = \text{RSS}_{t-1} + w_t^2$$

That is, the residual sum of squares for an OLS estimation over period 1 to t is given by the residual sum of squares for an OLS estimation over the period 1 to $(t-1)$ plus the squared standardised recursive residual for time t. So given w_t it is possible to construct a wide variety of alternative Chow tests. For example, we could construct a series of one-period Chow tests, each testing the hypothesis that a structural break occurs in a successively later period.

Two test procedures which take special advantage of the properties of the recursive residuals are the CUSUM and CUSUMSQ tests of Brown, Durbin and Evans (1975). Both tests consist of a series of statistics, defined as:

$$\text{CUSUM}_t = (1/s) \sum_{i=k+1}^{t} w_i$$

where s is the full sample estimate of the standard error of the regression

$$\text{CUSUMSQ}_t = \left(\sum_{i=k+1}^{t} w_i^2\right) \Big/ \left(\sum_{j=k+1}^{T} w_j^2\right) = \frac{\text{RSS}_t}{\text{RSS}_T}$$

The CUSUM test is therefore simply the sum of the recursive residuals normalised by the standard error of the residuals. If the residuals are random we would expect the CUSUM statistic to remain close to zero; any systematic departure from zero would suggest misspecification. The CUSUMSQ statistic is simply the sum of the *squared* recursive residuals normalised by the residual sum of squared errors for the full period, so at T, CUSUMSQ = 1. Both of these tests are used generally

in the form of a plot of either the CUSUM or CUSUMSQ statistics against time and critical values may be found in Harvey (1981). It is generally recognised however that the formal power of the tests is rather low and in practice they are often used as an informal diagnostic tool.

It is perhaps finally worth noting that the CUSUMSQ test may be put into the form of a recursive Chow test since the Chow (c_2) test given in (4.33) may be written as

$$c_{2t} = \left(\frac{1}{\text{CUSUMSQ}_t} - 1\right)\left(\frac{T_1 - K}{T - T_1}\right)$$

Hence the CUSUMSQ test may be interpreted as a particular form of sequential Chow test.

Testing functional form

An important simplification in the move from the *general* DGP to an actual maintained hypothesis that is estimable is the assumption of a particular functional form. The Box–Cox (1964) procedure provides one method of assessing functional form, but a simple yet fairly general test is that due to Ramsey (1974). In Ramsey's test the alternative model involves a high-order polynomial to represent a different functional form. The RESET test (Ramsey 1974) in its most common form consists of the following regression

$$Y_t = \beta' X_t + \alpha_1 \hat{Y}_t^2 + \alpha_2 \hat{Y}_t^3 + \ldots + \alpha_m \hat{Y}_t^m \qquad \textbf{(4.36)}$$

where $\hat{Y}_t = \hat{\beta} X_t$ are the predictions from the preferred structural model. The higher order powers in $\hat{Y}_t$ implicitly involve higher order terms in X_t as well as cross terms (such as $X_{1t}X_{2t}$) and hence embody a functional form different from $Y = B'X$.

Subtracting $\hat{\beta}' X_t$ from both sides of (4.36) we obtain

$$\hat{u}_t = \gamma' X + \sum_{i=1}^{m} \alpha_i \hat{Y}_t^{i+1} \qquad \textbf{(4.37)}$$

where $\gamma' = (B' - \hat{B})'$. Under the null H_0: $\alpha_1 = \alpha_2 = \ldots \alpha_m = 0$, the RESET test is RESET$(m) = TR^2$ and is distributed as $\chi^2(m)$.

Testing for normality

An important assumption underlying the use of OLS, and most test statistics, is that the residuals of the model are normally distributed. When this assumption and the others regarding marginalisation and conditioning are valid then OLS is the maximum likelihood estimator. There are several non-parametric tests for normality (the Kolmogorov–Smirnov test and the Shapiro–Wilk test are examples) which we will not discuss here. The most widely used parametric test is based on testing the third and fourth moments, skewness and kurtosis, for departures from normality. Skewness is given by the formula

$$\text{SK} = \left(\frac{1}{T}\sum_{t=1}^{T} u_t^3\right) \Big/ \left(\frac{1}{T}\sum_{t=1}^{T} u_t^2\right)^{3/2} \qquad \textbf{(4.38)}$$

SK is centred on zero and, when standardised by $T^{0.5}$ has a variance of 6. Kurtosis is given by

$$\text{EK} = \left(\frac{1}{T}\sum_{t=1}^{T} u_t^4\right) \Big/ \left(\frac{1}{T}\sum_{t=1}^{T} u_t^2\right)^{2} \qquad \textbf{(4.39)}$$

When this is standardised by $T^{1/2}$ it has a mean value of 3 and a variance of 24. Given those properties it is possible to construct the following test for normality, due to Bera and Jarque (1982):

$$\text{BJ} = \left[\frac{T}{6}\text{SK}^2 + \frac{T}{24}(\text{EK} - 3)^2\right] \qquad \textbf{(4.40)}$$

and under the null that the error term is normally distributed this will be distributed as $\chi^2(2)$.

While testing for normality is obviously important in practical applications the BJ test is perhaps even more useful as a test of outliers. It is very sensitive to the presence of outlier observations and so failing the BJ test is often simply a signal to look for one or two large errors and see if there are data problems or specific effects (such as strikes, incomes policy periods) which can be 'eliminated' with dummy variables.

Encompassing test

The idea of a model being adequate in the sense of being congruent with the data is an important one. It involves passing all the tests outlined above, but it also involves the model being one which cannot be dominated in all senses by some other model. To implement the latter point we need a framework for testing models against each other; this is the encompassing principle, see Mizon and Richard (1986). In general terms the notion of encompassing is a simple one. A model M_1 may be said to encompass another model M_2 if it can explain the results of that model. As an example, suppose M_1 contains an important weakly exogenous variable which behaves erratically at some point in time. If M_1 represents the DGP fairly well and M_2 excludes this variable we might expect to see structural instability of M_2 at the point in the data set where the omitted variable changes. In this case M_1 would predict the structural failure of M_2 and we would say that M_1 encompasses M_2. More formally we may follow the definition of Mizon and Richard. Let $\tilde{\Theta}$ denote some statistic which we are using to assess M_2 and let $\Theta_1 = E_1(\tilde{\Theta})$ denote the expectation of Θ when it is applied to M_1. Then under a suitable set of exogeneity assumptions we may consider the statistic.

$$\phi = \tilde{\Theta} - \Theta_1$$

which compares the *observed* value of $\tilde{\Theta}$ with its expectation under M_1. It may be shown that M_1 encompasses M_2 with respect to $\tilde{\Theta}$ if ϕ does not differ significantly from zero. Clearly in order to implement this we must derive forms of ϕ with a known distribution. One of the advantages of the encompassing principle is that it provides a framework for linking many existing test procedures. In particular, when we are dealing with nested pairs of models (i.e. when either $M_1 \subset M_2$ or $M_2 \subset M_1$) then Mizon and Richard show that the standard test procedures may be given an encompassing interpretation. So we may use F-tests or likelihood ratio tests in the usual way, giving the results an encompassing interpretation. Similarly in a non-nested framework (when neither M_1 is contained in M_2 nor M_2 is contained in M_1) then many of the non-nested tests may be applied as encompassing tests, for example the J test of Davidson and Mackinnon (1981) or the Hausman (1978) specification test.

To illustrate the case of *variance encompassing* consider the following two competing explanations of Y:

$$M_1{:}\ Y = X\alpha + u \qquad u \sim (0, \sigma_u^2) \qquad \textbf{(4.41)}$$

$$M_2{:}\ Y = ZB + w \qquad w \sim (0, \sigma_w^2) \qquad \textbf{(4.42)}$$

For any given sample of data we have the following relationship between the variables X and Z

$$X = Z\gamma + v \qquad \textbf{(4.43)}$$

On the assumption that M_1 is true we would expect M_2 to be estimated as

$$Y = Z(\gamma\alpha) + (v\alpha + u) \qquad \textbf{(4.44)}$$

Comparing (4.44) and (4.42) under M_1 we expect

$$\sigma_w^2 = \sigma_u^2 + \alpha^2\sigma_v^2 \qquad \textbf{(4.45)}$$

and $\sigma_u^2 < \sigma_w^2$, asymptotically. Hence if M_1 is true we expect it to have a lower standard error than a competing model M_2; this is a variance-encompassing test.

In the case where one model M_2 is nested within another larger model M_1 (i.e. $Z \subset X$) the M_1 will automatically encompass M_2. This means that a model can always be made to encompass rival models simply by adding variables so as to nest the rival models. This approach is of little value and to rule out this trivial form of encompassing the concept of *parsimonious* encompassing is used. A model is said to be parsimonious when it uses the minimum number of estimated parameters to adequately represent the DGP. So we may say that M_2 parsimoniously encompasses M_1 if and only if M_2 encompasses M_1 and M_2 is nested within M_1. (Note that it is possible for both M_1 to encompass M_2 and M_2 to encompass M_1.)

Where we are dealing with two non-nested hypotheses the encompassing principle offers a new approach to the standard non-nested tests which is intuitively appealing. Hendry and Richard (1987) demonstrate that if we define a model M_c as an artificial model which nests both M_1 and M_2 within it, then M_1 encompasses M_2 if and only if M_1 encompasses M_c. So a conventional F-test against the artificial nesting model M_c may be given an encompassing interpretation. The Hendry and Richard result is however based on a moderately strong set of assumptions including fixed regressors and strong exogeneity.

4.3 An application to the demand for M2 in three European countries

In this section we will illustrate the general to specific and ECM modelling strategy discussed above using an example taken from Taylor (1986) of the estimation of broad money (M2) demand functions for three European Countries – West Germany, France and the Netherlands.

The data

A common problem encountered in investigating money demand in European countries is that data definitions, particularly for the broader measure of the money stock, are not consistent across the various countries concerned (OECD 1977). Partly in order to attenuate this problem, and partly because some of the required data series are not available in published sources, Den Butter and Fase (1981) (BF) asked the Central Banks of eight European countries to provide data on the relevant variables. The data series published in BF therefore constitute a fairly consistent data bank which is highly desirable in comparative studies of this kind. Even within this data bank, however, unbroken series on all variables for the whole of the sample period is available for only three countries: West Germany, the Netherlands and France. Also, all series on these three countries start at 1960(1) and terminate at 1978(4).

The series used were nominal M2, nominal GNP (GDP for France), the implicit GNP (GDP) deflator (1970 = 100), the long-term interest rate, the short-term interest rate (for West Germany and the Netherlands only, the three-month interbank rate for the former, the local authority three-month rate for the latter), and a business cycle indicator (derived from industrial output indices for France and West Germany and from the labour utilisation rate for the Netherlands). All data except those for M2 are seasonally adjusted.

The implications of using seasonally adjusted/unadjusted data should be pointed out at this point. For ease of exposition, consider a two-variable relationship using polynomials in the lag operator L(i.e. $L^i x_t = x_{t-i}$), and suppress the constant term:

$$\alpha(L)y_t = \beta(L)x_t + u_t \tag{4.46}$$

where

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_n L^n$$
$$\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_n L^n$$

Suppose that y_t is seasonally adjusted to y_t^a by means of the filter $\lambda(L)$ (a scalar polynomial in the lag operator):

$$y_t^a = \lambda(L)y_t \tag{4.47}$$

and similarly, x_t is seasonally adjusted by applying the filter $\mu(L)$ (a scalar polynomial in the lag operator):

$$x_t^a = \mu(L)x_t \tag{4.48}$$

Substituting (4.47) and (4.48) in (4.46):

$$\alpha(L)y_t^a = \beta(L)x_t^a + v_t \tag{4.49}$$

where

$$v_t = [\lambda(L) - \mu(L)]\beta(L)x_t + \lambda(L)u_t \tag{4.50}$$

From this we can note the following (see also Hendry and Mizon 1978). Firstly, if u_t in (4.46) is 'seasonally serially correlated and is 'whitened' by applying the filter $\lambda(L)$ and if, further, the same filter is applied to both y_t and x_t i.e. $\lambda(L) = \mu(L)$, then the disturbance in (4.49) will be white noise. Secondly, seasonal adjustment does not alter the appropriate lag structure for the equation. Thirdly, although it may seem odd to adjust seasonal variables such as interest rates, the above algebra makes clear that this is reasonable in the context of estimation since the whole equation is seasonally adjusted. Fourthly, problems may arise when (as in the present context) the same filter has *not* been applied to both the left- and right-hand side variables ($\lambda(L) \neq \mu(L)$). As expression (4.50) makes clear, this may introduce serial correlation into the disturbance term and distort the testing and estimation procedures. Since all the data used in this section are obtained from the BF data base and are not readily available elsewhere, this appears to be an insuperable problem.

However, the following method was applied in mitigation. Suppose the seasonal filter for the x variables, $\mu(L)$, can be approximated closely by the standard method of regressing the unadjusted variable on to seasonal dummies and using the residual as the adjusted series. Since, as is well known, including seasonal dummies in a regression is identically equivalent to adjusting all of the (left- and right-hand side) variables prior to estimation, this will have the effect of seasonally adjusting the dependent variable in the same fashion as the right-hand side variables (see Frisch and Waugh 1933; Malinvaud 1970, pp 486–9). Accordingly, seasonal dummies were included in all regressions. It should be noted, however, that previous empirical applications in the LSE tradition to money demand often use seasonally adjusted data.

Estimation results

Since the data on M2 was seasonally unadjusted and all series were quarterly, we decided to set the length of the lag structure for the maintained hypothesis at four periods. The maintained hypothesis for each of the countries was therefore:

$$\begin{aligned} m_t = \alpha_0 &+ \sum_{i=1}^{4} \alpha_i m_{t-i} + \sum_{i=0}^{4} \beta_i P_{t-i} + \sum_{i=0}^{4} \gamma_i y_{t-i} \\ &+ \sum_{i=0}^{4} \sigma_i r^l_{t-i} \sum_{i=0}^{4} k_i r^s_{t-i} + \sum_{i=0}^{4} \lambda_i c_{t-i} + u_t \end{aligned} \qquad \textbf{(4.51)}$$

where m denotes M2, p the price level, y real income, r^l the long interest rate, r^s the short interest rate, and c the business cycle indicator. All variables are in natural logarithms except r^s and r^l.

The business cycle indicator was included, following BF, on the argument that precautionary balances should rise as economic activity slows down. The short-term interest rate was included for the Netherlands and West Germany, again following BF, in order to pick up switching between components of M2 and less liquid assets that occurred in these countries during the 1970s. As in BF, r^s is entered as zero up to the fourth quarter of 1969 and a dummy variable taking the value one for 1960 (1)–1969 (4) and zero otherwise was included to adjust for the discontinuity. The short rate was entered in levels in order to allow its elasticity to vary, since switching becomes more likely as short rates rise. Three seasonal dummies were also included in all regressions (but are not reported) as well as a dummy in the French equations to account for the student riots of May 1968 (see BF).

The first four observations of each series were lost because of lags in (4.51). In common with BF, we reserved the last eight observations for post-estimation stability tests. The specification search therefore took place using data for 1961 (1)–1976 (4), a total of sixty-four observations. We used ordinary least squares for estimation purposes, and tested for the validity of this procedure rather than use an estimator such as instrumental variables (see below). In what follows we use a nominal test size of five per cent (unless stated to the contrary).

Our final, parsimonious short-run money demand functions are listed in Table 4.1, together with a set of diagnostic statistics for each equation. The equation for demand for M2 in the Netherlands is particularly encouraging. It relates short-run growth in real M2 demand to an error correction term of the kind discussed above (with one lag), implying a highly significant 'inverse velocity' effect on short-run money balances. The current rate of inflation ($\Delta_1 p_t$) and lagged values of the long and short rates are also found to be significant explanatory variables and have coefficients of the expected signs. This indicates significant switching between components of M2 and less liquid securities and real assets over the period. The current value of the cyclical indicator also was high explanatory power and indicates a significant level of precautionary demand. These and the other terms in the Netherlands equation allow an extremely rich pattern for the short-run dynamics of money demand.

Turning to the diagnostic for the Netherlands equation, 'RESET' is the F-statistic for the restrictions imposed on the general unrestricted form (4.51) in order to arrive at the final specification, and is highly insignificant – as one should expect given the data-based nature of the specification search (see Note 1). We can see that the equation explains nearly ninety per cent of the variation in real money growth

Table 4.1 Final parsimonious equations for money demand

Netherlands:

$$
\begin{aligned}
\Delta_1(m-p)_t = {} & \underset{(6.52)}{4.99} - \underset{(7.34)}{0.26(m-p-y)_{t-1}} - \underset{(8.43)}{0.95\Delta_1 p_t} - \underset{(4.34)}{0.059 r_{t-3}} \\
& \underset{(2.40)}{-0.082\Delta_1^2 r^1_{r-2}} + \underset{(5.75)}{0.0043 r^s_{t-1}} - \underset{(6.74)}{1.13 c_t} + \underset{(6.01)}{5.17\Delta_1 c_t} \\
& - \underset{(3.21)}{5.38\Delta_1^2 c_{t-1}}
\end{aligned}
$$

$R^2 = 0.88$, DW = 2.09, BP(12) = 8.23, SER = 0.010,
RESET(21, 30) = 0.38, SK = −0.07,
EK = −0.42, BJ(2) = 0.42, LM4(4, 43) = 1.02, Q(16) = 13.47,
ARCH(1) = 0.13, RESET(4, 47) = 2.17, EX(1, 50) = 1.44,
CHOW(7, 44) = 0.48, HF(5, 51) = 2.00

Germany:

$$
\begin{aligned}
\Delta_1(m-p)_t) = {} & \underset{(2.11)}{-0.17} + \underset{(2.67)}{0.23\Delta_2 y_t} - \underset{(3.6)}{0.15(m-p-y)_{t-1})} - \underset{(3.20)}{0.40\Delta_4 P_t} \\
& - \underset{(2.92)}{0.056 r_t} + \underset{(3.18)}{0.052 r_{t-4}} + \underset{(3.55)}{0.0031 r^s_t} \underset{(3.43)}{-0.38\Delta_1(m-p)_{t-3}}
\end{aligned}
$$

$R^2 = 0.81$, DW = 2.12, SER = 0.012, RESET(21, 20) = 1.15,
SK = −0.36, EK = 0.39, BJ(2) = 1.55, BP(10) = 12.40,
LM(44, 45) = 1.48, Q(16) = 15.77, ARCH(1) = 0.032,
RESET(4, 45) = 2.22, EX(4, 49) = 1.17,
CHOW(7, 46) = 1.47, PF(6, 53) = 1.64.

France:

$$
\begin{aligned}
\Delta_1(m-p)_t = {} & \underset{(1.96)}{0.64} + \underset{(2.78)}{0.13 y_t} + \underset{(2.59)}{0.17\Delta_4 y_t} + \underset{(3.98)}{0.52\Delta_2 y_{t-1}} - \underset{(3.56)}{0.20(m-p-y)_{t-4}} \\
& \underset{(1.91)}{-0.34\Delta_1 p_t} - \underset{(2.21)}{0.11\Delta_1^2 r^1_t} + \underset{(1.61)}{0.052 rl_{t-1}} + \underset{(1.91)}{0.11\Delta_1 r^l_{t-3}} + \underset{(2.20)}{0.29\Delta_1(m-p)_{t-2}} \\
& \underset{(2.46)}{0.21c} - \underset{(4.49)}{0.67\Delta_2 c_{t-1}}
\end{aligned}
$$

$R^2 = 0.73$, DW = 2.08, SER = 0.009, RESET(13, 33) = 0.43,
SK = −0.61, EK = 0.65, BJ(2) = 3.84, LM4(4, 40) = 2.40,
Q(16) = 22.89, ARCH(1) = 0.047, RESET(4, 44) = 0.53,
EX(4, 44) = 1.30, CHOW(12, 36) = 1.73, HF(5.48) = 1.34.

Note: Figures in parentheses below coefficient estimates denote *t*-ratios.

($R^2 = 0.88$) with an equation standard error (SER) of one per cent. DW is the Durbin–Watson statistic which, together with the Lagrange multiplier statistic for up to fourth-order (moving average or autoregressive) serial correlation (LM4), indicates that the non-systematic dynamics of the equation are white noise (see Note 2). This impression is echoed by the value of the Box–Pierce portmanteau statistic for sixteen lagges ($Q(16)$). SK and EK are the moment coefficients of

skewness and excess kurtosis and should be approximately zero for normally distributed errors. Their size should also give an indication of any significant outliers in the residuals. BJ tests for the joint significance of SK and EK and is thus a test for normality of the residuals.

In order to examine possible heteroskedasticity in the residuals, we computed a test for a non-scalar covariance matrix due to Breusch and Pagan (1979) (BP), which was found to be insignificant. We also calculated a Lagrange multiplier test for possible first-order autoregressive conditional heteroscedasticity (ARCH) effects in the residuals and this statistic (ARCH) was also insignificant. Since the equation was estimated by ordinary least squares, we implicitly assumed the econometric exogeneity of the current-dated, right-hand side variables. This hypothesis was tested using a test due to Hausman (1978) (EX) and we were unable to reject the hypothesis of exogeneity of the right-hand side variables (see Note 3). The general test for misspecification of the model, (RESET) is insignificant at the five per cent level.

Finally, we performed two tests for parameter stability on the model. CHOW is the analysis of covariance test for parameter stability due to Chow (1960), and tests for a structural break from the first quarter of 1970 onwards; it is insignificant. HF tests for the predictive accuracy of the equation of the model over the period 1977 (1)–1978 (4), which was not included in the estimation period. We used the indicator variable method due to Salkever (1976) and Pagan and Nicholls (1984) to perform this test. This essentially involves defining a dummy variable for each of the post-estimation data points and testing the joint significance of these dummies when the equation is run over the whole sample including the prediction period. Salkever (1976) shows that the coefficients of these dummies are the prediction errors with confidence intervals which can be calculated from the estimated standard errors.

A major advantage of this method is that it controls for sampling variations in the parameter estimates. It was found, however, that some of these dummies were individually significant when added into the general unrestricted form (5) when the whole sample was used. This indicates that the maintained hypothesis itself is incapable of explaining these observations and may be indicative of extraordinary circumstances in these periods or that the maintained hypothesis is itself incorrect (Baba *et al.* 1985, for example, include variables to control for items such as risk to long-term bond holding). Since there was some degree of overlap in the significance of the dummies in the maintained hypothesis for each of the three countries (1977 (3), 1978 (3) and 1978 (4) for the Netherlands, 1977 (1) and 1978 (4) for

West Germany, and 1977 (1), 1978 (2) and 1978 (3) for France) and because the maintained hypothesis could not readily be expanded because of data limitations, the predictive failure tests were computed without testing for the significance of dummies which were found to be individually significant in the unrestricted form. The resulting value of HF for the Netherlands is insignificant.

Similar comments apply to short-run equations obtained for West Germany and France. Good fits were obtained and all of the diagnostic statistics are insignificant at nominal test sizes greater than five per cent. In particular, both equations pass the Chow test for in-sample parameter stability. In the German equation, the short-term interest rate again showed the significant explanatory power and yielded a coefficient of the expected sign. However, the business cycle indicator dropped out of the German regressions during the sequential specification search, indicating the absence of any significant precautionary elements in German money demand over the period.

Following BF, the short interest rate was not included in the French regressions, but the business cycle indicator does appear with a significant coefficient of the expected sign in the final equation. Another interesting feature of the French equation is that the error correction term appears with a lag of four periods, indicating a slower response to the 'inverse velocity effect' than in the other two countries. Also, the French equation includes a significant value of the current level of real income. As discussed above, this destroys the property of long-run unit elasticity of money demand with respect to real income.

The long-run or steady-state solutions to the short-run demand functions are given in Table 4.2. Long-run unit real income elasticities are found for the Netherlands and West Germany. This contrasts with the results of BF (and also of Boughton 1979) who find real income elasticities in excess of unity for these countries – 1.19 and 1.21 respectively for the Netherlands and Germany (see Note 4). On the

Table 4.2 Steady-state solutions for money demand equations

Netherlands:	$m_t = \kappa_1 + p_t + y_t - 0.23r^l_t + 0.017r^s_t - 4.34c_t$
	$(\kappa_1 = 19.20 + 0.92g_p - 1.71g_y)$
Germany:	$m_t = \kappa_2 + p_t + y_t - 0.026r^l_t + 0.02r^s_t$
	$(\kappa_2 = -1.11 - 2.67g_p - 0.27g_y)$
France:	$m_t = \kappa_3 + p_t + 1.64y_t - 0.26r^l_t - 1.02c_t$
	$(\kappa_3 = 3.19 - 0.43g_p + 1.33g_y)$

Note: g_p and g_y denote the annualised steady-state growth rates of prices and real income respectively.

other hand, we estimated the long-run income elasticity for France to be 1.64, which is very close to BF's estimate of 1.61. The steady-state, long-term interest rate elasticity for France is indentical to BF's long-run elasticity at −0.26, and the long-run coefficients of the business cycle indicator are also very close (−1.02 on our estimate against −1.05 on BF's).

In the Netherlands long-run equation the long-term interest elasticity of −0.23 is very slightly lower than BF's estimate of −0.30, and the size of our long-run coefficient on the business cycle indicator (−4.34) compares with that of BF (−3.61). In the German steady-state equation, on the other hand, our long-term interest state elasticity of −0.026 is much smaller than the value reported by BF (−0.20), and we found the business cycle indicator to be insignificant altogether in explaining German money demand.

We find a long-run elasticity of 0.02 for the short interest rate in the German equation. This translates into an elasticity of 0.12 at the mean interest rate of about six per cent, and compares with the corresponding figure reported by BF of 0.15. In the Netherlands steady-state equation, at a mean short interest rate of approximately six per cent over the period, the semi-elasticity of 0.017 becomes 0.102, comparing with 0.13 reported by BF.

Overall, therefore, our steady-state money demand equations compare well with the long-run solutions to the transfer functions estimated by BF. A major difference between the two sets of results being that we find long-run real income elasticities of unity for West Germany and the Netherlands, in contrast to BF who find long-run elasticities or real income substantially in excess on unity. We believe that this may be due to the arbitrary (and untested) restrictions which BF impose on the lag structure of their equations.

4.4 Conclusion

In this chapter we have developed the methodology of dynamic modelling which has grown out the LSE tradition of econometrics and we have illustrated its power and usefulness by presenting a study of the demand for money. Dynamic modelling is a flexible tool which allows a complex interaction of economic theory and time series data so that both theory coherence and data coherence can be achieved. We would end on a note of warning, as dynamic modelling is sometimes presented as an almost mechanical rule for model building; this is almost never the case. Dynamic modelling is a framework for bringing together data and economic theory which requires skill and understanding on the part of the user; if this is absent then dynamic modelling can be little more use than step-wise regression and it is unlikely to yield insights into the real world.

Notes

1. Test statistics which appear in Table 4.1 with two figures in brackets (e.g. REST (21, 30)) should be referred to the F-distribution with the indicated degrees of freedom, while those appearing with one figure (e.g. BJ(2)) should be referred to the chi-square distribution with the indicated degree of freedom.
2. We calculated the Langrange multiplier statistic for serial correlation as an F rather than a chi-square statistic in the light of the Monte Carlo evidence of Kiviet (1983).
3. The Hausman exogeneity test requires an estimator which is consistent even under the alternative hypothesis of exogeneity of the current-dated right-hand-side variables. For this purpose we used an instrumental variables estimator with the once-lagged 'foreign' values of the putative endogenous variables as instuments (e.g. the French and German lagged inflation rates were used as instruments for the Netherlands inflation). In each case the instruments set was tested and accepted on the basis of Sargan's (1964) test for the validity of the instruments.
4. We refer to BF's estimates of real money demand (1981, Table 3).