If β were fixed we could maximise over α and Ω by a regression of R_{ot} on $-\beta' R_{kt}$ which gives

$$\hat{\alpha}(\beta) = - S_{ok}\beta(\beta' S_{kk}\beta)^{-1} \quad \text{(A6)}$$

and

$$\hat{\Omega}(\beta) = S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko} \quad \text{(A7)}$$

where

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it}R'_{jt} \qquad i, j = 0, k$$

and so maximising the likelihood function may be reduced to minimising

$$|S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko}| \quad \text{(A8)}$$

It may be shown that (A8) will be minimised when

$$|\beta' S_{kk}\beta - \beta' S_{ko}SS_{oo}{}^{-1}S_{ok}\beta|/|\beta' S_{kk}\beta| \quad \text{(A9)}$$

attains a minimum with respect to β.

We now define a diagonal matrix D which consists of the ordered eigenvalues $\lambda_1 > \ldots > \lambda_N$ of $(S_{ko}S_{oo}^{-1}S_{ok})$ with respect to S_{kk}. That is λ_i satisfies

$$|\lambda S_{kk} - S_{ko}S_{oo}{}^{-1}S_{ok}| = 0 \quad \text{(A10)}$$

Define E to be the corresponding matrix of eigenvectors so that

$$S_{kk}ED = S_{ko}S_{oo}{}^{-1}S_{ok}E \quad \text{(A11)}$$

where we normalise E such that $E'S_{kk}E = I$.

The maximum likelihood estimator of β is now given by the first r rows of E, that is, the first r eigenvectors of $(S_{ko}S_{oo}{}^{-1}S_{ok})$ with respect to S_{kk}. These are the canonical variates and the corresponding eigenvalues are the squared canonical correlations of R_{kt} with respect to R_{ot}. These eigenvalues may then be used in the test proposed in (A3) to test either for the existence of a cointegrating vector $r = 1$ or the number of cointegrating vectors $N > r > 1$.

Johansen (1988) calculates the critical values for the likelihood ratio test for the cases where $m \leq 5$, where $m = P - r$, and P is the number of variables in the set under consideration and r is the maximum number of cointegrating vectors being tested for.

6
Rational expectations

Over the last ten years the role of expectations formation in both theoretical and applied macroeconomics has been of central importance. New Classical models embody the assumption of rational expectations and clearing markets and may give rise to policy ineffectiveness, an issue which has influenced policy debates particularly in the US. In the UK, the treatment of expectations has been more pragmatic than in the US, but explicit modelling of expectations is now used in a wide range of large-scale macroeconometric models (see Wallis 1986 for a survey). Policy simulations of these models generally do not yield 'short-run' policy ineffectiveness but they do produce projections which differ substantially from conventional 'backward-looking' models.

At the applied level, relatively few practitioners have adopted the 'full' Muth-rational approach which requires specification of the complete macromodel. Such 'full information methods' have generally been confined to estimating 'small models' (e.g. Blake 1984, Taylor 1979). Much applied work has concentrated on estimating 'single equations' that contain expectations variables. For example, in the price expectations augmented Phillips curve (PEAPC), wage inflation depends on expected price inflation (and the excess demand for labour). The Life-Cycle hypothesis implies that consumption depends on some measure of expected future income. The risk aversion model has money and bond demand depending on expected capital gains.

The efficient markets literature is concerned with the proposition that agents use all available information to remove any known profitable opportunities in the market. For example, if uncovered interest parity holds then the interest differential in favour of the domestic currency should equal the *expected* depreciation of the domestic currency. To test such a proposition we need a framework for modelling the unobservable one-period-ahead expected spot rate of the domestic currency.

Estimation of single equations containing expectations terms has proved popular because they are more robust to potential misspecifications and simpler to estimate than full-information methods based on a 'complete' model. In this chapter we shall not be concerned with the latter case.

The literature on estimating expectations models is vast and can quickly become very complex. We have attempted to explain only the main (limited information) methods currently in use. We shall concentrate only on those problems introduced by expectations variables and shall leave 'other' problems that might also arise such as simultaneous equations problems and equations containing lagged dependent variables to be dealt with in other chapters.

The rational expectations, RE, hypothesis has featured widely in the literature and we begin in section 6.1 by discussing the basic axioms of RE which we later see are crucial in choosing an appropriate estimation procedure. We also examine equations that contain multi-period expectations. In section 6.2 we discuss the widely used errors in variables method (EVM) of estimating structural equations under the assumption that agents have rational expectations. The use of auxiliary equations (such as extrapolative predictions) to generate a suitable proxy variable for the unobservable expectations series, gives rise to two-step procedures and the pitfalls involved in such an approach are also examined in section 6.2.

In section 6.3 we highlight the problems which arise when the structural expectations equation has serially correlated errors. The generalised method of moments (GMM) estimator of Hansen (1982) and Hansen and Hodrick (1980) and the two-step two-stage least squares estimator (Cumby *et al*. 1983) provide solutions to this problem.

In sections 6.4 and 6.5 we provide illustrative empirical examples of the techniques discussed in earlier sections. We begin in section 6.4 with tests of the axioms of RE. We then discuss fixed-parameter forecasting schemes and then relax the assumption that agents have unchanged structural parameters in their model of the economy. Instead we assume that agents slowly *learn* about their economic environment and for example may utilise useful 'rules of thumb' in forming their expectations. We demonstrate how the Kalman filter and other 'variable parameter' approaches can be used to mimic simple learning processes. In section 6.5 we demonstrate how RE

models give rise to testable cross-equation restrictions. The Barro (1978) policy ineffectiveness New Classical model, the Carr–Darby (1981) unanticipated money model and tests of the efficient markets hypothesis fall into this class. We discuss such tests in the context of one-period-ahead expectations and multi-period expectations. A final section concludes.

6.1 The economics of expectations models and the RE hypothesis

In this section we analyse the various ways in which expectations variables are utilised in the applied literature and the implications of the economic assumptions for the estimation issues discussed in a later section.

Usually the applied economist is interested in estimating the structural parameters of a single behavioural equation or set of equations containing expectations terms which forms a subset of a larger model. (In a 'full' Muth–RE model (Muth 1961) we would have to specify the whole model.) The simplest *structural expectations equation* can be represented:

$$y_{1t} = bx^e_{t+j} + u_{1t} \qquad (6.1)$$

where

$$x^e_{t+j} = E(x_{t+j}|\Omega|_{t-j}) \; j \geq 0 \qquad (6.2)$$

and x^e_{t+j} is an exogenous expectations variable, E is the expectations operator conditional on the complete (relevant) information set available to the agent at time $_{t-j}$ (i.e. Ω_{t-j}). For example, in a purchasing power parity export price equation, Ex_{t+j} represents expected world prices and y_{1t} the domestic export price, both in a common currency. Also equation (6.1) could, for example, represent the wage-expected price element of the Phillips curve. In the absence of data on Ex_{t+j} (e.g. quantitative survey data) we must posit an auxiliary hypothesis for Ex_{t+j}. Whatever expectations scheme we choose, of key importance for the economics and econometrics of the model are (a) the forecast horizon, (b) the dating and content of the information set used in making the forecast, and (c) the relationship between the forecast error and the information set.

To develop these issues further it is useful to discuss the basic axioms of RE.

Basic axioms of RE

If agents have RE they act as if they know the structure of the complete model to within a set of white noise errors (i.e. the axiom of correct specification). Forecasts are unbiased on average, with constant variance and successive (one-step-ahead) forecast errors are uncorrelated with each other and with the information set used in making the forecast. Thus, the relationship between outturn x_{t+1} and the *one-step-ahead*, RE forecast ${}_tx^e_{t+1}$ using the complete information set Ω_t (or a subset Λ_t) is:

$$x_{t+1} = {}_tx^e_{t+1} + \omega_{t+1} \tag{6.3}$$

where

$$E(\omega_{t+1}|\Omega_t) = E(\omega_{t+1}|\Lambda_t) = 0 \tag{6.3a}$$

$$E(\omega^2_{t+1}|\Omega_t) = \sigma^2_\omega \tag{6.3b}$$

$$E(\omega_{t+1}\omega_{t+1-j}|\Omega_t) = 0 \qquad j = 1, 2 \ldots \infty \tag{6.3c}$$

The one-step-ahead rational expectations forecast error ω_{t+1} is 'white noise' and an 'innovation', conditional on the complete information set Ω_t and is orthogonal to a subset of the complete information set $(\Lambda_t \subset \Omega_t)$.

The *k-step-ahead* RE forecast errors $(k > 1)$ *are* serially correlated and are MA $(k-1)$. To demonstrate this in a simple case assume x_t is AR(1).

$$x_{t+1} = \phi x_t + \omega_{t+1} \quad \text{and} \quad E(\omega_{t+1}|\Omega_t) = 0 \tag{6.4}$$

Hence

$$x_{t+j} = \phi^j x_t + \omega_{t+j} + \phi\omega_{t+j-1} + \phi^2\omega_{t+j-2} + \ldots \tag{6.5}$$

From (6.5) it is easy to see that

$$(x_{t+1} - {}_tx^e_{t+1}) = \omega_{t+1} \tag{6.6}$$

while the two-period-ahead forecast error is

$$(x_{t+2} - {}_tx^e_{t+2}) = (\phi\omega_{t+1} + \omega_{t+2}) \tag{6.7}$$

The *one-step*-ahead forecast error is an independent white-noise process, ω_{t+1} but the two-period-ahead forecast error is MA(1); similarly the k-step-ahead forecast error is MA$(k-1)$. Note that *all* the multi-period forecast errors

$$(x_{t+j} - {}_tx^e_{t+j}) \qquad j \geq 1$$

are independent of (orthogonal to) the information set Ω_t (or Λ_t).

There is one further property of RE that is useful in analysing RE estimators, namely the form of revisions to expectations. The *one-period revision* to expectations

$$({}_{t+1}x^e_{t+j} - {}_tx^e_{t+j})$$

depends only on new information arriving between t and $t+1$ and hence from (6.5) is easily seen to be

$$({}_{t+1}x^e_{t+j} - {}_tx^e_{t+j}) = \phi^{j-1}\omega_{t+1} \tag{6.8}$$

The two-period revision to expectations

$$({}_{t+2}x^e_{t+j} - {}_tx^e_{t+j})$$

will of course depend on ω_{t+1} and ω_{t+2} and be MA(1); one can generalise the result for k-period revisions to expectations.

Direct tests of RE

Direct tests of the basic axioms of RE may involve multi-period expectations and this immediately raises estimation problems. For example, *if monthly* quantitative survey data is available on the *one-year-ahead* expectation, ${}_tx^e_{t+12}$, a test of the axioms often involves a regression of the form:

$$x_{t+12} = \beta_0 + \beta_1({}_tx^e_{t+12}) + \beta_2\Lambda_t + \eta_t \tag{6.9}$$

where

$$H_0\colon \beta_0 = \beta_2 = 0,\ \beta_1 = 1$$

Under the null, η_{t+12} is MA(11) and an immediate problem due to RE is the need to use some kind of generalised least squares (GLS) estimator if efficiency is to be achieved. Of course, for one-period-ahead expectations where data of the same frequency is available, the error term is white noise and independent of the regressors in (6.9); OLS therefore provides a BLUE.

An additional problem arises if the survey data on expectations is assumed to be measured with error. If the true RE expectation is ${}_tx^e_{t+12}$ and the survey data provides a measure ${}_t\tilde{x}^e_{t+12}$ where we assume a simple linear measurement model (Pesaran 1985):

$${}_t\tilde{x}^e_{t+12} = \alpha_0 + \alpha_1({}_tx^e_{t+12}) + \varepsilon_t \tag{6.10}$$

Then substituting for ${}_tx^e_{t+12}$ from (6.10) in (6.9):

$$x_{t+12} = \lambda_0 + \lambda_1({}_t\tilde{x}^e_{t+12}) + \beta_2\Lambda_t + \zeta_t \tag{6.11}$$

where

$$\lambda_0 = (\alpha_1\beta_0 - \beta_1\alpha_0)/\alpha_1$$
$$\lambda_1 = \beta_1/\alpha_1$$
$$\zeta_t = \eta_t - (\beta_1/\alpha_1)\varepsilon_t$$

The additional problem in (6.11) is that now ${}_t\tilde{x}^e_{t+12}$ is correlated with ζ_t; some form of generalised IV estimator is required for consistency and asymptotic efficiency. As we shall see the orthogonality property between the RE forecast error and the information set (Λ_t or Ω_t) is frequently used in finding a suitable instrument set. However, it is not always simply the case that Λ_t provides a valid instrument set for the problem at hand.

Multi-period expectations

Sargent's (1979) model where agents minimise a multi-period quadratic cost function provides a tractable expectations framework, much used in the applied literature. Agents are assumed to know the time-path of the 'long-run' choice variable y^*_t (as given by some static equilibrium theory) and then choose actual y_t to minimise costs of being out of equilibrium $(y_{t+i} - y^*_{t+i})^2$ and costs of adjustment $(y_{t+i} - y_{t+i-1})^2$. The cost function C is

$$C = E_t\sum_{i=0}^{\infty} D^i(a_0(y_{t+i} - y^*_{t+i})^2 + a_1(y_{t+i} - y_{t+i-1})^2) \tag{6.12}$$

where E_t is the expectation operator, D is a discount factor $0 < D < 1$ and a_0 and a_1 are weighting factors (a_0, $a_1 > 0$).

The solution to this problem is

$$y_t = \lambda_1 y_{t-1} + (1 - \lambda_1)(1 - \lambda_1 D)\sum_{i=0}^{\infty}(\lambda_1 D)^i(k_t x^e_{t+i}) \tag{6.13}$$

where we have assumed the static equilibrium relationship is

$$y^*_t = kx_t \tag{6.13a}$$

and λ_1 is the stable root of the Euler equation obtained from the first-order conditions $\partial C/\partial y_t = 0$.

Equation (6.13) has proved popular in the applied literature because it has the 'plausible' property that the weights on the future expected values of the 'forcing variables' ${}_tx^e_{t+i}$ decline, the further the expectations are into the future. In addition, it provides a rationale for the inclusion of a lagged dependent variable and many economic time series have a strong autoregressive component.

The model has been applied to the determination of employment (Hall *et al*. 1986, Hansen and Sargent 1981, 1982), export prices (Cuthbertson 1986, 1990) and the demand for money (Cuthbertson 1988a, Cuthbertson and Taylor 1987, Muscatelli 1988). A slight modification to the cost function leads to an additional lagged dependent variable (y_{t-2}) and more complex weights on the forward terms ${}_tx^e_{t+i}$ which has been used to model stockbuilding (Hall, Henry, Wren-Lewis 1986). In most of the above studies the information set is assumed to be dated at either t or $t-1$. A number of different estimation techniques have been used in applied studies utilising equation (6.13) and it is not always clear what assumptions are required to yield optimal estimators, or the relationship between the various estimation methods used. It is our aim to clarify these issues in the subsequent sections.

6.2 The EVM and extrapolative predictors

In order to motivate our discussion of the estimation problems in the next two sections it is useful at this stage to summarise some of the problems encountered when estimating a structural expectations model; problems that arise include serial correlation and correlation between regressors and the error term. For illustrative purposes assume the structural model of interest is

$$y_t = \delta_1({}_tx^e_{t+1}) + \delta_2({}_tx^e_{t+2}) + u_t \tag{6.14}$$

u_t is taken to be white noise and x_t is an exogenous expectations variable.

Under the assumption of RE we have

$$x_{t+j} = {}_tx^e_{t+j} + \omega_{t+j} \tag{6.15}$$

A method of estimation widely used (and one of the main ones discussed in this chapter) is the errors in variables method EVM, where we replace the unobservable ${}_tx^e_{t+j}$ by its realised value x_{t+j}. This method is consistent with agents being Muth–rational, but could also be taken as a condition of the relationship between outturn and forecast without invoking Muth–RE.

Substituting from (6.15) in (6.14):

$$y_t = \delta_1 x_{t+1} + \delta_2 x_{t+2} + \varepsilon_t \tag{6.16}$$
$$\varepsilon_t = u_t - \delta_1\omega_{t+1} - \delta_2\omega_{t+2} \tag{6.16a}$$

Clearly from (6.15), x_{t+j} and ω_{t+j} are correlated and hence:

$$\text{plim}\,(x'_{t+j}\varepsilon_t)/T \neq 0 \qquad (j = 1, 2)$$

and

$$E(\varepsilon\,\varepsilon') \neq \sigma_\varepsilon^2 I$$

because of the moving average error introduced by the RE forecast errors, ω_{t+j}. Hence our RE model requires some form of instrumental variables estimation procedure with a correction for serial correlation. These two general problems form a main focus for this chapter.

Fixed coefficient *extrapolative* predictors are also used widely to proxy expectations terms. Here it is explicitly recognised that the econometrician may have a subset $\Lambda_t = \{x_{t-j}\}$, say, of the complete information set used by agents $\Omega_t = (x_{t-j}, y_{t-1})$, say. Hence the econometrician posits an expectations scheme

$$x_{t+1} = \phi(L)x_t + v_t = \phi_1 x_t + \phi_2 x_{t-1} + \phi_3 x_{t-3} + \ldots + v_t \tag{6.17}$$

Given an estimate of $\phi(L)$ we generate predictions with information at time t, using the *chain rule of forecasting*:

$$\tilde{x}_{t+1} = \hat{\phi}(L)x_t \tag{6.18a}$$

$$\tilde{x}_{t+2} = \hat{\phi}_1[\hat{\phi}(L)x_t] + \sum_{j=2} \hat{\phi}_j\, x_{t+2-j} \tag{6.18b}$$

and if these replace x^e_{t+j} in (6.14) we have a structural estimation equation

$$y_t = \delta_1\,\tilde{x}_{t+1} + \delta_2\,\tilde{x}_{t+2} + q_t \tag{6.19}$$

$$q_t = \sum_1^2 \delta_i((x_{t+i} - \tilde{x}_{t+i}) - (x_{t+i} - x^e_{t+i})) + \varepsilon_t \tag{6.20}$$

The error term q_t contains the MA(1), true forecast error of agents $\omega_{t+i} = (x_{t+i} - x^e_{t+i})$ as before but there is an extra term $(x_{t+i} - \tilde{x}_{t+i})$ which may cause additional estimation problems and these issues are discussed in the next section.

There is a logical problem in using a fixed coefficient expectations equation (6.17). *All* of the data is used in estimating $\phi(L)$, yet this fixed estimate is used to predict x_{t+1} at the beginning of the sample; part of the $\tilde{x}_{t+1}$ series therefore embodies sample information that the agent could not have had at the time his forecast was made. This may not matter asymptotically if $\phi(L)$ really is constant but clearly in small samples the assumption may yield incorrect predictions (see Friedman 1979).

We may wish to relax the assumption that agents act as if they use the 'true' fixed coefficient model in forming expectations. It may be the case either that the *true* model for x_t has some time varying parameters or that agents use a limited information set (of the true fixed parameter model) and update the changing estimates of the parameters of interest in some optimal fashion. In terms of a simplified AR(1) model with time-varying parameters we have:

$$x_{t+1} = \phi_{t+1} x_t + v_t \tag{6.21}$$

The Kalman filter can be used in a wide variety of models with time-varying parameters (or unobservable components – see Chapter 7) to provide optimal estimators $\hat{\phi}_{t+1/t}$ based on information at time t. We can then generate predictors ${}_t\hat{x}_{t+j}$ to use in (6.14). Hence similar issues arise when using such 'learning models' to provide a proxy variable for ${}_tx^e_{t+j}$ in the structural equation (6.14), as in the fixed parameter model.

The precise method used in estimating single equations with expectations terms depends on whether the expectations terms are formed for the current period ${}_{t-1}x^e_t$ or for many future periods ${}_tx^e_{t+j}(j \geq 1)$ and whether the residuals are serially correlated or white noise. Serial correlation may arise because of 'omitted variables', or wrong functional form in the structural equation or the assumption of RE *per se* may induce serially correlated errors in the estimation equation. To delineate these cases and to avoid confusion we take them in turn. We can then ascertain precisely the source of the estimation problem for each case. We begin with a simple model with one-period expectations to illustrate the basic principles of the EVM, we then discuss extrapolative predictors.

The errors in variables method EVM

The EVM is a form of IV or 2SLS approach. Under RE, the unobservable expectations variable ${}_tx^e_{t+j}$ is determined by the full relevant information set Ω_t. In the EVM a subset of the true information set Λ_t $(\subset\Omega_t)$ is sufficient to generate consistent estimates. However, we begin by demonstrating that OLS yields an inconsistent estimator (see section 1.6 (Chapter 1) for a more general exposition).

One-period-ahead expectations: white noise structural error

It is important to note that here we are dealing with a very specific expectations model. The simplest structural model embodying one-period-ahead expectations is

$$y_t = \beta x^e_{t+1} + u_t \tag{6.22}$$

where u_t is white noise and x^e_{t+1} is assumed to be uncorrelated in the limit with u_t:

$$\text{plim}\,(x^e_{t+1}u'_t)/T = 0 \tag{6.23}$$

If we assume rational expectations, then

$$x_{t+1} = x^e_{t+1} + \omega_{t+1} \tag{6.24}$$

and the RE forecast error ω_{t+1} is independent of the information set Ω_t (or Λ_t)

$$E\,(\Omega'_t\omega_{t+1}) = 0 \tag{6.25}$$

Substituting (6.24) in (6.22) we obtain

$$y_t = \beta x_{t+1} + q_t \tag{6.26}$$

$$q_t = (u_t - \beta\omega_{t+1}) \tag{6.26a}$$

Consider applying OLS to (6.26), we have:

$$\hat{\beta} = \beta + (x'_{t+1}x_{t+1})^{-1}(x'_{t+1}q_t) \tag{6.27}$$

From (6.24):

$$\text{plim}\,(x'_{t+1}\,x_{t+1})/T = \text{plim}\,(x^e_{t+1}{}'x^e_{t+1})/T + \text{plim}\,(\omega'_{t+1}\,\omega_{t+1})/T \tag{6.28}$$

or rewriting this more succinctly:

$$\sigma^2_x = \sigma^2_{xe} + \sigma^2_\omega \tag{6.28b}$$

From (6.24) and (6.29) and noting that x^e_{t+1} is uncorrelated in the limit with ω_{t+1}:

$$\text{plim}\,(x'_{t+1}q_t)/T = -\beta\,\text{plim}\,(\omega'_{t+1}\omega_{t+1})/T = -\beta\sigma^2_\omega \tag{6.29}$$

Substituting these expressions in (6.27):

$$\text{plim}\,\hat{\beta} = \beta\left[1 - \frac{\sigma^2_\omega}{\sigma^2_x e + \sigma^2_\omega}\right] \tag{6.30}$$

Thus the OLS estimator for β is inconsistent and is biased downwards. The bias is smaller the smaller the variance of the 'noise' element σ^2_ω in forming expectations. The above analysis is the basis of Friedman's (1957) view that if the permanent income hypothesis of consumption is correct but the latter is proxied by measured income, then OLS yields an underestimate of the true long-run marginal propensity to

consume (out of permanent income). Similarly, many early studies of the price-expectations augmented Phillips curve used actual inflation as a proxy for expected inflation; OLS estimates are inconsistent and it was argued that the finding of the presence of money illusion and a non-vertical long-run Phillips curve is due to an inappropriate estimation technique in the presence of expectations variables.

Instrumental variables: 2SLS

OLS is inconsistent because of the correlation between the variable x_{t+1} and the error term q_t which 'contains' the RE forecast error, ω_{t+1}. The solution to this problem is to use instrumental variables, IV, on (6.26), (see Chapter 1). However, to illustrate some additional nuances when applying IV, consider the model:

$$y_t = \alpha x^e_{1t+1} + \beta x_{2t} + u_t = Q\delta + u_t \tag{6.31}$$

$$Q = \{x^e_{1t+1}, x_{2t}\} \qquad \delta = (\alpha, \beta)' \tag{6.31a}$$

where x^e_{1t+1}, x_{2t} are asymptotically uncorrelated with u_t.

Direct application of IV to (6.31) would require an instrument for x^e_{1t+1} and an obvious candidate are the OLS predictions from the regression of x_{1t+1} on a subset of the information set, Λ_t *but including* x_{2t}:

$$\hat{x}_{1t+1} = \Lambda_t\hat{\Pi} \tag{6.32a}$$

$$\hat{\Pi} = (\Lambda'_t\Lambda_t)^{-1}(\Lambda'_t x_{1t+1}) \tag{6.32b}$$

The researcher is now faced with two options. Direct application of IV would utilise the instrument matrix

$$W_1 = \{\hat{x}_{1t+1}, x_{2t}\} \tag{6.33a}$$

where x_{2t} acts as its own instrument, giving

$$\delta_1 = (W'_1Q)(W'_1y) \tag{6.33b}$$

$$\text{Var}\,(\hat{\delta}_1) = \sigma^2(W'_1Q)^{-1} \tag{6.33c}$$

This is also the 2SLS estimator since in the first stage x_{1t+1} is regressed on *all* the predetermined (or exogenous variables) in (6.31) *and* the additional instruments in Λ_t.

An alternative is to *replace* x^e_{1t+1} in (6.31) by $\hat{x}_{1t+1}$ and apply OLS to:

$$y_t = \alpha\hat{x}_{1t+1} + \beta x_{2t} + q^*_t \tag{6.34}$$

$$q^*_t = u_t - \alpha(x_{t+1} - x^e_{t+1}) - \alpha(\hat{x}_{t+1} - x_{t+1}) \tag{6.34a}$$

This yields a 'two-step estimator' but as long as x_{1t+1} is regressed on *all* the predetermined variables, the OLS on (6.34) is *numerically* equivalent to the 2SLS estimator $\hat{\delta}_1$ and is therefore consistent.

However, there is a problem with the approach. The OLS residuals from (6.34) are

$$e = y_t - \hat{\alpha}\hat{x}_{1t+1} - \hat{\beta}x_{2t} \tag{6.35}$$

but the correct (IV/2SLS) residuals use x_{1t+1} and not $\hat{x}_{1t+1}$ and are:

$$e_1 = y - \hat{\alpha}x_{1t+1} - \hat{\beta}x_{2t} \tag{6.36}$$

Hence the variance–covariance matrix of parameters from OLS on (6.34) is incorrect since $s^2 = e'e/T$ is an incorrect (inconsistent) measure of σ^2 (Pagan 1984). The remedy is straightforward however; one merely amends the OLS program to produce the correct residuals e_1 in the second stage.

Extrapolative predictors

Extrapolative predictors are those where the information set utilised by the econometrician is restricted to be lagged values of the variable itself, that is an AR(ρ) model:

$$x_{1t+1} = \phi_1 x_{1t} + \phi_2 x_{1t-1} + \phi_2 x_{1t-2} + \ldots \phi_\rho x_{t-\rho} + \varepsilon_t \tag{6.37}$$

$$x_{1t+1} = \Phi(L)x_{1t} + \varepsilon_t \tag{6.37a}$$

The maximum value of ρ is usually chosen so that ε_t is white noise. OLS applied to (6.37a) yields one-step-ahead predictions

$$\hat{x}^*_{1t+1} = \hat{\Phi}(L)x_{1t} \tag{6.37b}$$

The use of extrapolative predictors has proved popular in models with multi-period expectations and in testing RE cross-equation restrictions. (In the latter procedure a VAR rather than an AR model is normally used.)

For the moment, consider using the extrapolative predictor either *as an instrument* for x^e_{1t+1} or to *replace* x^e_{1t+1} in (6.34). Using $\hat{x}^*_{1t+1}$ as an instrument for x^e_{1t+1} *and* x_{2t} as its own instrument yields a consistent estimate of δ since x_{1t-j} ($j \geq 0$) are uncorrelated with q^*_t and therefore so is $\hat{x}_{1t+1}$. This is all we need for IV/2SLS to be consistent, but note that in this case x_{2t} also appears in the instrument matrix W_1. The latter becomes important when we consider the two-step approach. Having obtained $\hat{x}^*_{1t+1}$ in the 'first stage', the second stage regression consists of OLS on:

$$y_t = \beta\hat{x}^*_{t+1} + \gamma x_{2t} + q^*_t \tag{6.38a}$$

$$q_t = [u_t + \beta(x^e_{t+1} - x_{t+1}) - \beta(\hat{x}^*_{t+1} - x_{t+1})] \tag{6.38b}$$

Compared with the EVM/IV approach (see equations (6.26), (6.26a)), we have an additional term $(\hat{x}^*_{t+1} - x_{t+1})$ in the error term of our second-stage regression (6.38a). The term $(x_{t+1} - \hat{x}^*_{t+1})$ is the residual from the first stage regression (6.37b).

The variable x_{2t} is part of the agent's information set, at time t, and may therefore be used by the agent in predicting x_{1t+1}. If so, then $(x_{t+1} - \hat{x}^*_{t+1})$ and the 'omitted variable' from the first stage regression, namely x_{2t} are correlated. Thus in (6.38a) the correlation between the variable x_{2t} and a component of the error term q^*_t imply that OLS on (6.38a) yields inconsistent estimates of δ (Nelson 1975). This is usually expressed in the literature as follows: If x_{2t} Granger-causes x_{1t+1} then the two-step estimator is inconsistent.

This illustrates the danger in using extrapolative predictors and *replacing* x^e_{t+1} in the second stage OLS regression, rather than using $\hat{x}^*_{t+1}$ as an instrument and applying the IV formula. Viewed from the perspective of 2SLS, the inconsistency at the second stage (6.38a) arises because in the first stage regression the researcher does not use *all* the predetermined variables in the model; he erroneously excludes x_{2t}. Somewhat paradoxically then, even if x_{2t} is not used by agents in forecasting x_{1t+1} it must be included in the first stage regression if the two-step procedure is used, otherwise $(x_{1t+1} - \hat{x}_{1t+1})$ may be correlated with x_{2t}. Of course, if the two-step procedure is used and consistent estimates $(\hat{\alpha}, \hat{\beta})$ are obtained, the correct residuals calculated using x_{1t+1} and not $\hat{x}_{1t+1}$ (as in equation (6.36)) must be used in the calculation of standard errors.

6.3 Serially correlated errors and expectations variables

Up to this point in our discussion of appropriate estimators we have assumed white noise errors in the regression equation. We now relax this assumption. Serially correlated errors may arise because of multi-period expectations or because of serially correlated structural errors. In either case, we see below that two broad solutions to the problem are possible. The first method uses the generalised method of moments (GMM) approach of Hanson (1982) and 'corrects' the covariance matrix to take account of serially correlated errors. The second method is a form of generalised least squares estimator under IVs and is known as the *two-step, two-stage, least squares* estimator (2S-2SLS), (Cumby *et al*. 1983). These two solutions to the problem are by no means exhaustive but have been widely used in the literature. The estimator due to Hayashi–Sims (1983) is also briefly discussed.

The GMM approach

We demonstrate this approach by first considering serial correlation that arises in equations with multi-period expectations and then move on to consider serial correlation in the structural error.

Multi-period expectations

Suppose that the structural error u_t is white noise but we have multi-period expectations (we restrict ourselves to two-period-ahead expectations for ease of exposition):

$$y_t = \beta_1 x^e_{t+1} + \beta_2 x^e_{t+2} + u_t \tag{6.39}$$

$$x^e_{t+j} = E(x_{t+j}|\Omega_t) \quad j = 1, 2 \tag{6.39a}$$

RE implies:

$$x_{t+j} = x^e_{t+j} + \eta_{t+j} \quad (j = 1, 2) \tag{6.40}$$

and substituting (6.40) in (6.39) we have our estimating equation:

$$y_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + q_t \tag{6.41a}$$

$$q_t = u_t - \beta_1 \eta_{t+1} - \beta_2 \eta_{t+2} \tag{6.41b}$$

2SLS on (6.41a) with instrument set Λ_t will yield consistent estimates of β_1, β_2. However, the usual formula for the variance of the IV estimator is incorrect in the presence of serial correlation (see equation (1.73)) and q_t is MA(1). Hansen and Hodrick (1980) suggest a 'correction' to the formula for the variance of the usual 2SLS estimator. Putting (6.41a) in matrix notation:

$$y = X\beta + q \tag{6.42}$$

The 2SLS estimator for β is equivalent to OLS on

$$y = \hat{X} b^* + q \tag{6.43}$$

$$\hat{X} = (\hat{x}_{t+1}, \hat{x}_{t+2}) \tag{6.44}$$

and $\hat{x}_{t+j}$ are the predictions from the regression of x_{t+j} $(j = 1, 2)$ on Λ_t. The 2SLS estimator is:

$$b^* = (\hat{X}'\hat{X})^{-1} (\hat{X}'y) \tag{6.45}$$

with residuals:

$$e^* = y - Xb^* \tag{6.46}$$

Note that in the calculation of e^* we use X and not $\hat{X}$. To calculate the correct variance of β in the presence of an MA(1) error, note that the variance covariance matrix is

$$E(q \quad q') = \sigma_0^2 \begin{bmatrix} 1 & \rho_1 & 0 & \cdots & & 0 \\ \rho_1 & 1 & \rho_1 & 0 & \cdots & \vdots \\ 0 & \rho_1 & 1 & \rho_1 & & 0 \\ \vdots & & \ddots & & 1 & \rho_1 \\ 0 & & \cdots & 0 & \rho_1 & 1 \end{bmatrix} = \sigma_0^2 \Sigma \tag{6.47}$$

where ρ_1 is the correlation coefficient between the error terms.

Since e^*_t are based on the consistent estimator b^*, then consistent estimators of σ_0^2, σ_1^2 and ρ are given by the following 'sample moments':

$$\hat{\sigma}_0^2 = (n^{-1}) \sum_1^n e_t^{*2} \tag{6.48a}$$

$$\hat{\sigma}_1^2 = (n^{-1}) \sum_2^n e_t^* e_{t-1}^* \tag{6.48b}$$

$$\hat{\rho}_1 = (\hat{\sigma}_1/\hat{\sigma}_0)^2 \tag{6.48c}$$

Knowing Σ we can calculate the correct formula for Var (b^*) as follows. Substitute from (6.42) in (6.45):

$$b^* = \beta + (\hat{X}'\hat{X})^{-1}\hat{X}'q \tag{6.49}$$

Since plim$(T^{-1})(\hat{X}'q) = 0$, then b^* is consistent and the asymptotic variance of b^* is given by:

$$\text{Var}(b^*) = T^{-1}\,\text{plim}[(\hat{X}'\hat{X})^{-1}\hat{X}'(q\ q')\hat{X}(\hat{X}'\hat{X})^{-1}]$$

$$\text{Var}(b^*) = \sigma_0^2(\hat{X}'\hat{X})^{-1}(\hat{X}'\Sigma\hat{X})(\hat{X}'\hat{X})^{-1} \tag{6.50}$$

Above we assume that the population moments are consistently estimated by their sample equivalents, e.g. $(\hat{X}'\hat{X})$. Note that Var(b^*), the Hansen–Hodrick correction to the covariance matrix for b^*, reduces to the usual 2SLS formula for the variance when there is no serial correlation (i.e. $\Sigma = \sigma^2 I$). The Hansen–Hodrick correction is easily generalised to the case where we have an MA(k) error; we merely have to calculate $\hat{\rho}_s (s = 1, 2, \ldots k)$ and substitute these estimates in Σ.

Serial correlation AR(1) *in the structural error*

Our model, in this case is

$$y_t = \beta x_t^e + u_t \tag{6.51}$$

or

$$y_t = \beta x_t + (u_t - \rho\eta_t) \tag{6.51a}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{6.51b}$$

IV applied to (6.51) using Λ_{t-1} as instruments yields a consistent estimate of β but the estimator is not asymptotically efficient because it ignores the serial correlation. In conventional models (i.e. those excluding expectations terms) the solution to this problem is to apply IV to the ρ-transformed equation (see Chapter 1):

$$(y_t - \rho y_{t-1}) = \beta(x_t - \rho x_{t-1}) + q_t \tag{6.52}$$

or

$$y_t^* = \beta x_t^* + q_t \tag{6.52a}$$

$$q_t = \varepsilon_t + \beta(\eta_t - \rho\eta_{t-1}) \tag{6.52b}$$

Although Λ_{t-1} is independent of ε_t (by assumption) and of η_t, it is *not* independent of the lagged RE forecast error η_{t-1}; information arising during $t-1$ 'causes' the forecast error between $t-2$ and $t-1$, that is, η_{t-1}. The GLS transformation has destroyed the orthogonality conditions between the error term in (6.51) and the information set Λ_{t-1}. This is because the GLS transformation introduces a moving average error, MA(1), in the RE forecast errors (the term, $\eta_t - \rho\eta_{t-1}$). We have 'removed' the serially correlated structural error u_t but have introduced another serially correlation error which is MA(1) and hence q_t is MA(1).

We wish to outline two methods that can be used to circumvent the above problems. Both methods utilise the Hansen–Hodrick procedure.

In the first method we apply IV to (6.51) using Λ_{t-1} as instruments to obtain a consistent estimator of b^*. The 'consistent' residuals u^* (as in the previous section) are used to obtain an estimate for ρ:

$$\hat{\rho} = (\Sigma u_t^* u_{t-1}^*)/\Sigma u_{t-1}^2) \tag{6.53}$$

which is used to form the transformed variables y_t^*, x_t^*. Λ_{t-1} and q_t in (6.52) are correlated, as noted above, but if we move the instrument set back one period, that is use Λ_{t-2} this is independent of η_t and η_{t-1}, asymptotically.

Using Λ_{t-2} as instruments for x_t^* yields a consistent estimator b^* for β and the residuals

$$e_t^* = y_t - x_t b^* \tag{6.54}$$

can be used to form the Σ matrix. The Hansen–Hodrick variance for b is then given by (6.50) with $\hat{X}^* = (x_t - \hat{\rho}x_{t-1})$ in place of $\hat{X}$.

The second method uses the insight of Hayashi–Sims. For example, suppose we have an MA(1) error $u_t = (1 + \phi L)\,\varepsilon_t$ in our original structural model. The backward filter $(1 - \phi L)^{-1}$ removes the serial correlation in u_t, but destroys the orthogonality condition between the information set and the error term. Hayashi–Sims suggest the '*forward filter*' on the variable x_t giving:

$$\tilde{x}_t = -\phi(1 - \phi L^{-1})^{-1} x_t \tag{6.55}$$

In this case any error terms η_t introduced by the EVM are transformed into terms in η_{t+j} $(j > 0)$ which are independent of the 'original' information set at time t, Λ_t.

A two-step, two-stage least squares (2S-2SLS) estimator

So far we have been able to obtain a consistent estimator of the structural parameter β_1 in (6.39) under RE by utilising IV/2SLS or EVM method. We have then 'corrected' the usual formula for the variance of the estimator using the Hansen–Hodrick formula. Although the Hansen–Hodrick correction yields a consistent estimator of the variance it is possible to obtain an asymptotically more efficient estimator which is also consistent. Cumby *et al.* (1983) provide such an estimator which is a *specific form* of the class of generalised instrumental variables estimators. The formulae for this estimator look rather formidable. If our structural expectations equation after replacing any expectations variables by their outturn values is:

$$y = X\beta + q \tag{6.56}$$

with $E(qq') = \sigma^2\,\Sigma$ and $\text{plim}\,[T^{-1}(X'q)] \neq 0$ **(6.56a)**

then the 2S-2SLS estimator is:

$$\hat{\beta}_{g2} = [X'\Lambda(\Lambda'\Sigma\Lambda)^{-1}\Lambda'X]^{-1}[X'\Lambda(\Lambda'\Sigma\Lambda)^{-1}\Lambda'y] \tag{6.57}$$

$$\text{var}\,(\hat{\beta}_{g2}) = \sigma^2[X'\Lambda(\Lambda'\Sigma\Lambda)^{-1}\Lambda'Q]^{-1} \tag{6.58}$$

where Λ is the information set available. Clearly to make this estimator operational we need a suitable instrument set Λ and an estimate of the variance-covariance matrix of error terms Σ. We have already discussed above how to choose an appropriate instrument set and how a 'consistent' set of residuals can be used to form Σ. The 'first stage' estimate of Σ can then be substituted in the above formulae, to complete the 'second stage' of the estimation procedure (see Cuthbertson 1990).

In small or moderate size samples we cannot say whether the Hansen–Hodrick correction is 'better than' the 2S-2SLS procedure since both rely on asymptotic results. Hence at present, in practical terms either method may be used. The one clear fact which emerges however, is that the normal 2SLS estimator for $\mathrm{Var}(\hat{\beta})$ is incorrect and care must be taken in utilising Cochrane–Orcutt type transformations to eliminate AR errors since this may result in an inconsistent estimator for β.

Summary

There are two basic problems involved in estimating structural (single) equations involving expectations terms, such as equation (6.39), by the EVM. First correlation between the ex-post variables x_{t+j} and the error term which involves the use of IV (or 2SLS) estimation to obtain consistent estimates of the parameters. In addition, the error term is likely to be serially correlated, for example MA(1) in equation (6.41a), which means that the usual IV/2SLS formulae for the variances of the parameters are incorrect. Two avenues are then open. Either one can use the IV residuals to form the (non-scalar) covariance matrix $(\sigma^2\Sigma)$ and apply the 'correct' IV formula for var (b^*), see equation (6.50). Alternatively, one can take the estimate of $\sigma^2\Sigma$ and apply a variant of generalised least squares under IV, for example, the 2S-2SLS estimator $\mathrm{var}(\hat{\beta}_{g2})$ of equation (6.58).

6.4 Empirical work on expectations models

In this section we provide examples of empirical work which illustrate some of the estimation issues outlined above. We begin with tests of the axioms of RE. Alternative expectations schemes are then discussed and their use in a forward-looking money demand function is examined.

Testing the axioms of rational expectations

There has been a large number of tests of the basic axioms of RE, using survey data. Here we illustrate the methodology using the results from Taylor (1988).

Survey data of people's expectations of key economic variables are often not in the form of numerical estimates but are collected as categorical responses (such as percentage of respondents expecting inflation to go 'up', 'down' or stay the 'same'). However these categorical responses can be converted using a variety of methods, into numerical data on expectations. Taylor (1988) converts *monthly* categorical data from UK investment managers into quantitative expectations series for expected annual price inflation ${}_tp^e_{t+12}$, annual wage inflation ${}_tw^e_{t+12}$, the annual percentage change in the FTA all share index ${}_tf_{t+12}$ and the US, Standard and Poors composite share index ${}_ts_{t+12}$.

The axioms of RE imply that the forecast errors are independent of the information set used in making the forecast. Consider the regression

$$(x_{t+12} - {}_tx^e_{t+12}) = \beta'\Lambda_t + q_t \quad \textbf{(6.59)}$$

for $x = p, w, f, s$ and where Λ_t is a subset of the complete information set. If the orthogonality property of RE holds, we expect $\beta = 0$.

If *we assume no measurement error* in x^e_{t+12} then q_t is a moving average error of order 11 at most. OLS yields consistent estimates of β because Λ_t and q_t are uncorrelated asymtotically but the usual formula for the *covariance matrix* of β is incorrect. However the OLS residuals from (6.59) can be used to construct a consistent estimate of the variance–covariance matrix (White 1980) along the lines outlined in the previous section where we discussed the more general Hansen–Hodrick adjustment. In this case the OLS residuals e_t from (6.59) yield consistent estimates of the variance–covariance matrix (6.47) using (6.48a) to (6.48c), (with the OLS residuals not the 2SLS residuals). The correct OLS variance is then

$$\mathrm{Var}(b) = \sigma_0^2(\Lambda'\Lambda)^{-1}(\Lambda'\hat{\Sigma}\Lambda)(\Lambda'\Lambda)^{-1} \quad \textbf{(6.60)}$$

where $\sigma_0^2 = T^{-1}\sum_1^n e_t^2$ **(6.60a)**

Equation (6.60) has the same form as the Hansen–Hodrick correction, equation (6.50) except that the OLS rather than IV residuals are used and we do not need to instrument the information set Λ_t.

The results of this procedure are given in Table 6.1 for the information set $\Lambda_t = (x_{t-1}, x_{t-2})$. For the price inflation, wage inflation and the FT share index, the standard errors on the own lagged variables indicate that all of these variables taken individually are not significantly different from zero. This is confirmed by the Wald test W(2), which indicates that the two RHS variables in each of the first

Table 6.1 Orthogonality regressions with small information sets 1981 (7)–1985 (7), ordinary least squares with adjusted covariance matrix. (See note)

Estimated equation					R^2	SEE	W(2)
$p_{t+12} - {}_tp^e_{t+12}$ =	−0.155	$-\ 0.310p_{t-1}$	$+\ 0.214p_{t-2}$		0.06	1.131	2.96
	(0.695)	(0.255)	(0.245)				(0.23)
$w_{t+12} - {}_tw^e_{t+12}$ =	2.596	$-\ 0.492w_{t-1}$	$+\ 0.109w_{t-2}$		0.20	1.891	4.25
	(1.918)	(0.282)	(0.204)				(0.12)
$f_{t+12} - {}_tf^e_{t+12}$ =	9.842	$-\ 0.262f_{t-1}$	$-\ 0.2227f_{t-2}$		0.07	11.519	6.33
	(3.075)	(0.179)	(0.226)				(0.04)
$s_{t+12} - {}_ts^e_{t+12}$ =	15.747	$-\ 0.933s_{t-1}$	$+\ 0.463s_{t-2}$		0.21	24.17	19.99
	(9.597)	(0.233)	(0.336)				(0.00)

Note: R^2 is the coefficient of determination, SEE the standard error of the equation; W(2) is a Wald test statistic for the coefficients of the two lagged regressors to be zero and is asymptotically central chi-square under the null of orthogonality, with two degrees of freedom; figures in parentheses denote estimated standard errors or marginal significance levels for W(2).

Source: Taylor (1988)

three equations are jointly not significantly different from zero. For the S&P index the lagged values are significantly different from zero, thus rejecting the RE orthogonality axiom.

If there are measurement errors in the expectations series, see equations (6.10), (6.11), we do not expect the coefficient on ${}_t\tilde{x}^e_{t+12}$ to be unity and there is a non-zero correlation between the variable ${}_t\tilde{x}^e_{t+12}$ and the error term. The latter requires the use of IV. Taylor uses p_t, f_t, w_t, s_t as instruments for the expectations variables, ${}_t\tilde{x}^e_{t+1}$, to yield consistent estimates of the parameters, residuals and covariance matrix, see equations (6.47) and (6.48). The variance of these parameters is then given by Hansen's GMM estimator, see equation (6.50).

$$\mathrm{Var}(b) = \sigma_0^2(\Lambda'\Lambda)^{-1}(\Lambda'\ \Sigma\ \Lambda)(\Lambda'\Lambda)^{-1} \qquad \textbf{(6.61)}$$

Taylor's results using this estimator are given in Table 6.2. The results are similar to those in Table 6.1, except for the FT share price index f_{t+12}. Here the GMM estimator indicates that the forecast error for the FT share price index is not independent of the information set (W(2) = 46.9). This demonstrates that when testing the axioms of RE, correct inference may require careful choice of appropriate estimation technique.

Taylor repeats the above exercise using a larger information set:

$$\Lambda^* = (p_{t-j}, w_{t-j}, f_{t-j}, s_{t-j});\ j = 1, 2.$$

With this extended information set the GMM estimator indicates that

Table 6.2 Orthogonality regressions with small information sets 1981 (7)–1985 (7), generalised method of moments. (See note)

Estimated equation					R^2	SEE	H(3)	W(2)
p_{t+12} =	$0.550{}_tp^e_{t+12}$	$+\ 1.315$	$-\ 0.399p_{t-1}$	$+\ 0.488p_{t-2}$	0.97	1.000	0.04	6.17
	(0.202)	(1.122)	(0.286)	(0.270)			(0.99)	(0.05)
w_{t+12} =	$0.021{}_tw^e_{t+12}$	$+\ 6.151$	$+\ 0.006w_{t-1}$	$+\ 0.185w_{t-2}$	0.97	1.436	0.05	3.85
	(0.144)	(1.712)	(0.075)	(0.122)			(0.99)	(0.15)
f_{t+12} =	$0.473{}_tf^e_{t+12}$	$+\ 20.066$	$-\ 0.199f_{t-1}$	$-\ 0.124f_{t-2}$	0.89	8.004	0.06	46.49
	(0.340)	(6.925)	(0.125)	(0.175)			(0.99)	(0.00)
s_{t+12} =	$-0.725{}_ts^e_{t+12}$	$+\ 62.658$	$-\ 0.614s_{t-1}$	$-\ 0.154s_{t-2}$	0.66	19.761	0.04	18.86
	(0.468)	(16.716)	(0.179)	(0.260)			(0.99)	(0.00)

Note: Instruments used for the expectations variable were p_t, w_t, f_t and s_t; H(3) is Hansen's (1982) test statistic for the instruments, and is asymptotically central chi-square with three degrees of freedom for three *valid* over-identifying instruments. See note to Table 6.1 for other definitions.

the orthogonality condition is decisively rejected for *all* four variables.

Fixed parameter AR and VAR schemes

Cuthbertson (1988) estimates a forward-looking model in the UK demand for narrow money (M1) using a two-step procedure. The structural demand for money function (simplified somewhat) is

$$m_t = \lambda m_{t-1} + (1 - \lambda D)(1 - \lambda)(c_p\ SP^e + c_y\ SY^e + c_R SR^e) \qquad \textbf{(6.62)}$$

where

$$SX = \sum_{i=0}^{8} (\lambda D)^i (SX^e_{t+i}) \text{ and } X^e_t = (P^e, Y^e, R^e)_t \qquad \textbf{(6.62a)}$$

The agent is assumed to have information dated $t - 1$ and earlier. In order to estimate the model a data series for the expectations terms is required. Cuthbertson uses two alternative schemes; namely, fixed-parameter AR and VAR models. The AR and VAR models are given in Tables 6.3 and 6.4. The chain rule of forecasting is then applied to generate multi-period forecasts $\hat{X}^e_{t+j}$ $(j = 0, 1, 2, \ldots 8)$ for each variable. These then replace the expectations terms, X^e_{t+j} and OLS is applied to (6.62) to yield two-step estimates of the structural money demand function. Using the AR system for X^e_{t+j} yields:

$$M_t = -\underset{(1.8)}{0.86} + \underset{(20.6)}{0.89}\, M_{t-1} + \underset{(2.4)}{0.052}\, \hat{S}Y^e + \underset{(2.0)}{0.024}\, \hat{S}P^e - \underset{(3.3)}{0.176}\, \hat{S}R^e \qquad \textbf{(6.63)}$$

OLS, 64(3) − 79(4), SEE = 1.47(%), DW = 2.6, HF(12) = 13.6,

SALK(12) = 9.1, WK = 1.7.

Table 6.3 Autoregressive forecasting equations for P, Y, R

1 $\Delta P_t = \underset{(4.0)}{0.0075} + \underset{(12.3)}{0.83}\Delta P_{t-1} - \underset{(2.3)}{0.22}\Delta^2 p_{t-3} + \underset{(6.0)}{0.037}(D793)$

OLS, 64(3)–79(4), SE = 0.82(%), DW = 1.9, LM4F = 0.67, LM4 = 2.9, F(5, 52) = 0.75

2 $\Delta Y_t = \underset{(4.4)}{0.0137} - \underset{(3.3)}{0.12}(\Delta^2 Y_{t-1} + \Delta^2 Y_{t-3}) - \underset{(1.6)}{0.024}(D793)$

OLS, 64(3)–79(4), SE = 21.1(%), DW = 2.0, LM4F = 0.23, LM4 = 1.1, F(6, 52) = 0.17

3 $R_t = \underset{(42.8)}{1.00}R_{t-1} + u_t \qquad u_t = \underset{(1.65)}{0.21}u_{t-1} + \varepsilon_{t-1}$

AR, 64(3)–74(4), SE = 0.013, DW = 2.0, LM4F = 0.59, LM4 = 2.5, F(5, 55) = 1.4

Notes:

(i) SE = standard error of the regression, DW = Durbin–Watson statistic, AR = estimation subject to autoregressive errors.

(ii) LM4 is the Langrange multiplier statistic for autocorrelation up to order 4, asymptotically distributed under the null of no serial correlation, as central chi-squared with four degrees of freedom. Critical value at 5% significance level is 9.5.

(iii) LM4F is the Langrange multiplier test, expressed as an F-distribution.

(iv) $F(n_1, n_2)$ is the F-test of the restrictions in moving from the general AR(6) equations, with n_1, n_2 degrees of freedom. The critical value at 5% significance level for the above equations is (approximately) 2.4.

Table 6.4 VAR forecasting equations for Y, P, R (See notes for Table 6.3)

1 $\Delta Y_t = \underset{(2.0)}{-0.6} - \underset{(3.1)}{0.32}\Delta Y_{t-1} + \underset{(2.0)}{0.06}Y_{t-2} - \underset{(2.7)}{0.64}\Delta P_{t-2} - \underset{(4.1)}{0.82}\Delta R_{t-2}$

OLS, 64(3)–79(4), SE = 2.0(%), LM4F = 0.7, LM4 = 3.3

2 $\Delta P_t = \underset{(2.7)}{0.008} + \underset{(7.4)}{0.58}\Delta P_{t-1} - \underset{(3.7)}{0.29}\Delta P_{t-3} + \underset{(2.4)}{0.08}\Delta Y_{t-1} + \underset{(5.4)}{0.25}\Delta R_{t-1} + \underset{(6.5)}{0.032}(D793)$

OLS, 64(3)–79(4), SE = 0.6(%), LM4F = 0.32, LM4 = 1.6

3 $R_t = \underset{(4.4)}{-0.86} + \underset{(7.6)}{0.93}R_{t-1} - \underset{(1.9)}{0.23}R_{t-2} - \underset{(2.0)}{0.25}\Delta P_{t-3} + \underset{(4.4)}{0.08}Y_{t-1}$

OLS, 64(3)–79(4), SE = 0.011, LM4F = 1.2, LM4 = 5.2

A unit (expected) **nominal** income elasticity is accepted by the data (on a Wald test W(2) = 1.2, $\chi^2(2) = 6.0$), and the long-run semi-elasticity with respect to the interest rate is − 4.2. As we noted in section 6.2, the use of extrapolative AR or VAR forecasting equations and a two-step estimation procedure may result in inconsistent parameter estimates. This will occur if M_{t-1} in (6.63) Granger, causes either P_t, Y_t or R_t. Although widely used, the two-step procedure may be somewhat hazardous. Thus although the equation passes the Hendry parameter constancy test HF(12), it must be interpreted with caution.

If we apply the EVM technique to the above model, the terms SX^e must be *instrumented* and M_{t-1} acts as its 'own' instrument. Cuthbertson and Taylor (1991) employ the EVM in this forward-looking model which yields consistent parameter estimates. The instruments used for SX^e are four lagged values of P, Y, R and M. However they find serial correlation in the (IV) residuals of (6.63), of order 2 and 3. They therefore apply the Hayashi–Sims (1983) forward filter to all the variables of (6.62).

The residuals e_t^* from the IV regression (6.62), without the Hayashi–Sims correction are used to calculate consistent estimates of the 'unknown' autocorrelation coefficients ρ_j:

$$\rho_j = \sum_t e_t^* e_{t-j}^* \Big/ \sum_t e_{t-j}^* \qquad (j = 2, 3) \qquad \textbf{(6.64)}$$

Because they employ the *forward* filter, the instrument set dated $t-1$ and earlier is asymptotically uncorrelated with the error-term (see section 6.3). Hence we obtain consistent and asymptotically efficient estimators. When Cuthbertson and Taylor (1989) impose a unit long-run price level elasticity ($c_p = 1$) then representative estimates of the long-run (expected) income and interest rate elasticities (E_y, E_R, respectively) are:

$$E_y = \underset{(2.8)}{1.8} \qquad E_R = -\underset{(3.0)}{4.9} \qquad \textbf{(6.65)}$$

over the period 1968(4)–1982(4)–asymptotic t-statistics in parenthesis. Although the point estimates of the expected income elasticity exceeds unity we can easily accept a unit coefficient on a t-test ($t = 1.2$). Thus, for this particular model the results from the two-step procedure and the EVM do not differ greatly, and hence any inconsistency in the former may not be too severe.

The Lucas critique: changing expectations schemes

One of the drawbacks in using fixed coefficient AR or VAR models is that forecasts made for the early part of the data set (using the chain rule) utilise information that was not available at the time the forecasts were made. This is because in obtaining the *estimated* parameters we use *all* of the data set. Clearly it may be more realistic to assume that agents update their view about the parameters of the expectations generating equations. This applies with stronger reason after major ('regime') changes in the economy; (for example, in the 1970s, the move from low to high inflation rates in the UK and the switch towards monetary targets in the USA).

Utilising a structural forward-looking demand for money function of the form (6.62), Cuthbertson and Taylor (1990) examine the 'case of the missing money' in the USA in the context of the Lucas (1976) critique. Around 1974, conventional (e.g. partial adjustment) money demand functions in the USA overpredicted the demand for money and this was interpreted as an inexplicable shift in the money demand function. Cuthbertson and Taylor (1990) put forward the hypothesis that the underlying forward-looking demand for money function (6.62) is stable over the whole of the 1970s, but a shift in the (VAR or AR) expectations formation scheme for Y, P or R caused estimated partial adjustment models to exhibit parameter instability (and serial correlation in the residuals). This is an example of the Lucas (1976) critique. To illustrate the Lucas critique in the context of our forward-looking demand for money function (6.62), simplify somewhat and assume:

$$M_t = \lambda M_{t-1} + (1-\lambda)(1-\lambda D)c_y\left[\sum_{i=0} (\lambda D)^i Y^e_{t+i+1}\right]$$

is a stable money demand function. Now assume agents forecast 'income' according to the AR(1) model:

$$Y_{t+1} = \phi Y_t + v_t \qquad \textbf{(6.67)}$$

where v_t is white noise. Predictions from (6.67), with information dated t and earlier, are:

$${}_tY^e_{t+j} = \phi^{j+1} Y_t \qquad \textbf{(6.68)}$$

Substituting (6.68) in (6.66),

$$M_t = \lambda_1 M_{t-1} + [(1-\lambda)c_y(1-\lambda D)\phi/(1-\lambda D\phi)]Y_t \qquad \textbf{(6.69)}$$

Equation (6.69) may also be viewed as a conventional partial adjustment form of money demand function:

$$M_t = \pi_0 M_{t-1} + \pi_1 Y_t \qquad \textbf{(6.70)}$$

However if we estimate (6.70) but the way agents form their expectations alters (for example, undergoes a structural shift), then the 'conventional' partial adjustment demand function (6.70) will exhibit parameter 'shifts' even though the underlying (or 'deep') parameters λ, D and c_y of the 'true' forward-looking equation remain constant. This is the Lucas critique.

The above argument applies if the variables Y_t, P_t and R_t are assumed to be generated by a first-order vector autoregressive scheme, as assumed by Cuthbertson and Taylor (1990), when investigating the US demand function for narrow money (i.e. M1B). They find that the VAR scheme for (Y, P, R) does undergo a structural break around the 'missing money' period. They therefore estimate the ϕ parameter(s) for the pre- and post-1974 period. When these *two* separate VAR schemes are used to determine the variables SY^e, SP^e, SR^e in the forward demand for money function pre- and post-1974, they find that the demand function has relatively stable parameters and does not have serial correlation in the errors. However, if one ignores the shift in the VAR scheme then the 'solved out' form of the demand for money function, i.e. the analogue to (6.70), has 'poor' statistical and economic properties. Thus, Cuthbertson and Taylor 1990 provide some evidence that fixed parameter AR or VAR expectations schemes may be inadequate and that the Lucas critique may be of some practical relevance. (For an alternative account of the missing money episode see Baba *et al*. 1988.)

Hendry (1988) provides an interesting test to discriminate between the forward model (6.66) and the backward-looking model (6.70). Using our simple model, Hendry's argument is that if ϕ in (6.67) is found to be unstable (time varying) and the forward model (6.62) is correct, then π_1 in the 'backward-looking' model (6.70) should also be unstable. Hence a finding of a constant π_1 in (6.70) and time-varying ϕ in (6.67), leads to a refutation of the forward model (6.66) and (6.67), (and incidentally of the empirical relevance of the Lucas critique).

Another way of gaining an insight into Hendry's argument is to use the formula for the OLS estimator of the expectations model under EVM. Equation (6.30) indicates that plim $\hat{\beta}$ depends on σ^2_{xe} The variance of x^e is given by the variances of ϕY_t in (6.67) in our money demand model. If ϕ is non-constant, then σ^2_{xe} is also time-varying and hence we expect the OLS estimator, $\hat{\beta}$ to be non-constant. Hence Hendry's counterfactual argument is that a constant $\hat{\beta}$ and non-constant $\hat{\phi}$, are incompatible with the structural expectations model (6.66) and (6.67).

Cuthbertson (1991) argues that in finite samples Hendry's test does not rule out the structural forward model (6.66) and some other expectations generation equation like (6.67) that has constant parameters but which is, as yet, undiscovered by the econometrician. Also it is not clear how the Hendry's analysis deals with the issue of explicit time-varying parameters in (6.67) as discussed below. However, for any fixed parameter form for (6.67), for which it is hypothesised agents actually use in forecasting, Hendry's test is valid (even in small samples). In practice, proponents of the structural expectations model (6.66) will have to 'Hendrify' the expectations generating equation (6.67) in an attempt to obtain constant parameters in (6.67).

Variable parameter forecasting schemes

Instead of a series of discrete breaks in expectations equations, as in the above case, we may wish to assume agents *continually* update the parameters of their expectations generating equations as 'new' information becomes available. A simple yet tractable form of 'updating' is to assume agents update their AR or VAR forecasting schemes *as if* they applied recursive OLS to the model. Cuthbertson and Taylor (1991) apply a recursive VAR scheme to (Y, P, R) in the context of the forward demand for money function (6.62). At each point in time the VAR parameters are estimated (say using data from $t = 1$ to n). The chain-rule of forecasting is then applied to obtain k-period ahead forecasts for $n+1$, $n+2$, $n+k$ (with information and parameter estimates available only to period n). The VAR scheme is then re-estimated for period 1 to n_1 ($n_1 = n+1$) and the next k period ahead forecasts obtained. These forecasts provide instruments for SX^e using information actually available to the agent at the time of the forecast. The forward demand function may then be estimated using the EVM (with appropriate adjustments for any serial correlation). Cuthbertson and Taylor (1991) using a recursive VAR obtain the following long-run elasticities for UK, M1 in the forward model (6.70); for the period 1968(4)–1979(4):

$$E_y = \underset{(2.3)}{0.80} \quad E_p = \underset{(11.3)}{1.11} \quad E_R = \underset{(2.2)}{-\ 1.9} \tag{6.71}$$

(asymptotic t-statistics in parentheses). Thus under an expectations scheme that embodies a simple form of updating, the forward demand for money function continues to yield sensible long-run elasticities (note that we can accept $E_p = 1$) which are also stable over the 1980(1)–1982(4) period.

Optimal updating schemes

A Muth–rational agent is assumed to act as if his forecasts are calculated using *the* true model of economy. In reality, for any real world economy there are a number of competing models and the agent may be uncertain as to what constitutes the 'true' model. In addition it is possible that the parameters of the true model may alter through time as the economy undergoes 'regime changes' (such as from low to high inflation periods). Also, agents acting on their predictions from a false model, generate data which later may have to be explained by the econometrician. Theoretical models that embody learning by agents are relatively new and do not provide a tractable alternative for the applied econometrician (see for example, Bray and Savin 1986). Hence the applied worker either has to utilise survey data (with its own limitations, see Pesaran 1985) or has to utilise 'plausible' expectations schemes. A reasonable compromise is to assume that although costs of information (and inherent uncertainty about the true model) force agents to use sensible 'rules of thumb', nevertheless they utilise whatever information they have, in an optimal fashion as they learn about their economic environment. This leaves considerable scope for the applied worker.

In Chapter 7 we demonstrate two models which embody learning by agents and we utilise the Kalman filter to estimate these models, which are known as 'systematically varying parameters' and the 'stochastic trend' model. Both types of model can be useful in generating expectations series.

For the varying parameter model we assume agents forecast the variable x_{t+1} using:

$$x_{t+1} = (\phi_{t+1/t})x_t \tag{6.72}$$

where $\phi_{t+1/t}$ is their best guess of ϕ given information up to time, t. An explicit form of time variation in ϕ_t is assumed, the simplest being a random walk:

$$\phi_{t+1} = \phi_t + \varepsilon_{t+1} \tag{6.73}$$

We defer further discussion of the estimation of this model until Chapter 7 but merely wish to note here that such models can be used in generating expectations series where the agent continually learns about his environment and as he does so, he updates his estimate of

ϕ. Clearly such a model is not a panacea for modelling expectations, since it can only mimic the way agents form expectations. However, it is a useful alternative to assuming agents continuously know the (constant parameter) true model.

In the 'unobservable components model' the econometrician has observations on y_t (e.g. measured income) but we wish to obtain an estimate of the unobservable permanent income π_t. Measured income is assumed to consist of permanent income and (zero mean) transitory income ε_t, hence:

$$y_t = \pi_t + \varepsilon_t \tag{6.74}$$

The agent faces a 'signal extraction problem'. He has to determine how much of a change in actual income y_t can be attributed to permanent income (the 'signal') and how much is merely 'transitory' (the noise). Lucas' (1972) New Classical supply curve is derived under similar assumptions, where the firm has to decide the increase in the *aggregate* price index based on information about prices in the industry ('local prices').

To 'solve' the above signal extraction problem we have to make some assumption about the behaviour of π_t. In the stochastic trend model (Harvey and Todd 1983) the growth in π_t is itself stochastic and the model reduces to one which may be interpreted in terms of a stochastic trend for y_t and π_t:

$$y_t = \alpha_0 + \alpha_t t + u_{1t} \tag{6.75}$$

$$\Pi_t = \alpha_0^* + \alpha_t^* t + u_{2t} \tag{6.76}$$

where t = time trend but the coefficient on this variable is time varying (i.e. α_t, α_t^*).

The reader need note at this point only that the Kalman filter can be used to estimate this model and it yields optimal predictions for y_{t+j} as more information on y_t becomes available. It therefore mimics 'learning' by agents.

Cuthbertson and Taylor (1990) use the stochastic trend model to generate multi-period forecasts for (Y_t, P_t, R_t) for the UK, assuming agents 'learn' from their past forecast errors. Using these predictions $X_t^e = (y^e, P^e, R)_t$ and the 'surprise' terms $(X_t - X_t^e)$ yield the following forward-looking demand function for UK, M1:

$$\begin{aligned} M_t = &- \underset{(5.6)}{0.87} + \underset{(34.1)}{0.94}\, M_{t-1} + \underset{(1.9)}{0.0066}\,(\hat{S}P) \\ &+ \underset{(2.8)}{0.014}(\hat{S}Y) - \underset{(3.4)}{0.048}(\hat{S}R) + \underset{(0.7)}{0.11}(P - P^e) \\ &+ \underset{(3.5)}{0.20}\,(Y - Y^e) - \underset{(5.6)}{0.87}(R - R^e)_t \end{aligned} \tag{6.77}$$

1964(1)–1979(4), SEE = 1.41(%), Q(8) = 9.7, W(2) = 3.0, HF(12) = 17.1.

The Wald test W(2) = 3.0 is distributed as central chi-squared and indicates that a unit long-run price and real income elasticity is accepted by the data: the long-run interest rate semi-elasticity is – 7.1. The Hendry forecast test indicates parameter constancy over the period 1980 (1)–1982 (4). 'Surprises' in real income $(Y - Y^e)_t$ are added to money balances and unexpectedly high interest rates on alternative assets leads to a switch out of M1. The results on the demand for M1, utilising this particular optimal forecasting scheme are therefore encouraging. (Q(8) is the Ljung–Box statistic, and indicates the absence of serial correlation of up to order 8.)

6.5 Rational expectations: cross-equation restrictions

We have already noted that in order to estimate 'a structural model' containing expectations variables (such as a forward-looking demand for money function) we often require an ancillary 'weakly rational' *expectations generation equation*. However to obtain consistent estimates of the structural model we do not require knowledge of the full information set used by agents. Thus our expectations model often consists of *two* equations (even when we do *not* assume full Muth–rational expectations). So far we have used our expectations generation equation (often an AR or VAR model) to generate instruments for the unobservable expectations (for example, expected income in the demand for money). Broadly speaking, predictions from the expectations generation equation are used as 'proxy' variables for the unobservable expectations variables. In this section we show that our two-equation system, plus the assumption of rational expectations *often* implies testable *cross-equation* parameter restrictions. Cross-equation restrictions provide a test of the *joint* hypothesis of the structural model assumed and the assumption of RE. We inserted 'often' in the above sentence because in some cases cross-equation restrictions may not ensue – as in the case of 'observational equivalance'. Here, an RE model may be indistinguishable from a non-RE model. We do not discuss this aspect here (see Pesaran 1987 and Cuthbertson and Taylor 1988). In general, more efficient estimates of the parameters are obtained if the two equations that comprise our model are estimately jointly. (This applies *a fortiori* if the cross-equation restrictions *do* hold and are then imposed in estimation.)

Tests of cross-equation retrictions abound in the RE literature. For example, they have been used widely in testing (a) policy ineffectiveness and neutrality propositions (see below), (b) the efficient markets hypothesis in, for example, the foreign exchange, stock and bond markets, (c) in the Life Cycle/RE model of consumption, and (d) in forward-looking investment and employment equations (see, *inter-alia*, Cuthbertson 1985, Cuthbertson and Taylor 1988, MacDonald 1988, Pesaran 1984, 1987, and Mishkin 1983, Lucas and Sargent 1981 for surveys/readings in this area). Here we only seek to illustrate the basic issues involved. Tests of the 'policy ineffectiveness model' have been widely reported (Barro 1978, Leiderman 1980) and summaries are readily available (including Mishkin 1983, Pesaran 1987). The underlying principles behind tests of cross-equation (rationality) restrictions are very similar even though the models considered may be rather disparate. Hence we demonstrate the basic principles using empirical results on the Carr–Darby (1981) shock-absorber hypothesis of the demand for money. We contrast results obtained from two-step estimation procedures and joint estimation subject to cross-equation restrictions. The model only has one-period-ahead expectations variables. Our second main empirical example of testing cross-equation RE restrictions utilises multi-period expectations. It is based on the work of Sargent (1979) and involves our forward-looking demand for money function.

The shock-absorber model of the demand for money

An important debate in monetary economics concerns the role of money as a buffer stock (for general discussions and surveys see Laidler 1984, Goodhart 1984, Cuthbertson and Taylor 1986a). An important theme in this literature is the notion that money balances act as a short-run 'shock absorber' or buffer to unanticipated shocks to the money supply. This ideas was advance by Carr and Darby (1981) and has been examined empirically by them and by MacKinnon and Milbourne (1984). Carr and Carby (CD) argue that a proportion of unanticipated changes in the nominal money supply are willingly held in the short run, whereas anticipated changes will be fully reflected in changes in the price level and so will be neutral with respect to the level of real money holdings. The CD shock-absorber hypothesis may be represented by the two equations:

$$(m - p)_t = \beta x_t + \alpha(m - m^a)_t + \delta m_t^a + u_t \qquad \textbf{(6.78)}$$

$$m_t^a = \gamma z_{t-1} + v_t \qquad \textbf{(6.79)}$$

where m_t is the logarithm of the nominal money stock at the time t, p_t is the logarithm of the price level, x_t is the vector of determining variables observed at time t (such as income, interest rates), which also includes lagged real money balances; β is a suitably dimensioned coefficient vector and u_t is a random disturbance. m_t^a is the anticipated component of money supply and is determined as the predictions from equation (6.79). z_{t-1} is a vector the components of which are considered by agents to have a systematic influence on money supply, γ is a stable coefficient vector and v_t is the non-systematic component of the money supply process. In the CD paper, equation (6.78) is a conventional demand for money function augmented by the monetary surprise term $(m - m^a)_t$, and anticipated money. CD argue that anticipated money is fully reflected in the current price level, that is $\delta = 0$, and a proportion $(0 < \alpha < 1)$ of unanticipated money accumulates as desired money holdings.

However, Cuthbertson and Taylor (1986b, 1988) argue that the shock-absorber hypothesis makes sense only if the aggregate 'money demand' equation is interpreted as an 'inverted' price equation. Rearranging (6.78):

$$p_t = -\beta x_t + (1 - \alpha)(m - m^a)_t + (1 - \delta)m_t^a + u_t \qquad \textbf{(6.80)}$$

Expressing (6.78) in the form of a price equation (6.80) makes clear the logic of the shock-absorber hypothesis. If $\delta = 0$, anticipated money has a proportional effect on the price level, so that unanticipated shocks lead to a rise in short-run real money holdings.

Cuthbertson and Taylor (1986b, 1988) initially estimate (6.78) (or equivalently 6.80) for UK and US narrow money using a two-step procedure. Alternative expectations generation equations (6.79) (e.g. AR(4), ARIMA, and the stochastic trend model for the money supply are used to generate predictions $\hat{m}_t$ and surprises $\hat{v}_t = m_t - \hat{m}_t$. These are then used in (6.78) in place of m^a, $(m - m^a)$ and OLS is applied to (6.78). (The variables z_{t-1} are used as instruments for $\hat{m}_t$ to obtain the correct standard errors for δ. Cuthbertson and Taylor find that broadly speaking the CD hypothesis is accepted using this two-step procedure; that is, $\hat{\alpha} > 0$ and $\hat{\delta} = 0$. (See also Cuthbertson and Taylor 1987b.)

Joint estimation, imposing rationality (but not neutrality; $\delta \neq 0$), implies running the equations:

$$(m - p)_t = \beta x_t + \alpha(m_t - \gamma z_{t-1}) + \delta\gamma z_{t-1} + u_t^R \qquad \textbf{(6.80a)}$$

$$m_t = \gamma z_{t-1} + v_t^R \qquad \textbf{(6.80b)}$$

Notice that the vector of parameters 'γ' appears in both equations; this is the cross-equation restriction implied by RE. These two equations without the RE restriction imposed are:

$$(m - p)_t = \beta x_t + \alpha(m - \gamma^* z_{t-1}) + \delta\gamma^* z_{t-1} + u_t \qquad \textbf{(6.81a)}$$

$$m_t = \gamma z_{t-1} + v_t \qquad \textbf{(6.81b)}$$

where $\gamma \neq \gamma^*$. Under the assumption that u_t, v_t are normally and independently distributed with zero mean and variances σ_u^2, σ_v^2, respectively, then the log-likelihood is:

$$L = \frac{-T}{2}\ln\sigma_u^2 - \frac{T}{2}\ln\sigma_v^2 - \frac{u'u}{\sigma_u^2} - \frac{v'v}{\sigma_v^2} \qquad \textbf{(6.82)}$$

A test of the RE cross-equation restrictions is provided by a likelihood ratio test between equations (6.80a)/(6.80b) and (6.81a)/(6.81b). (Note that here this test is conditional on neutrality *not* holding, $\delta \neq 0$.) Since we assume $\sigma_{uv} = 0$ then the determinant of the covariance matrix in the unrestricted model (6.80a) + (6.80b) is

$$\det(\Sigma) = \sigma_v^2\sigma_u^2$$

Similarly, the determinant in the restricted model is obtained from the residuals u_t^R and v_t^R to give $\det(\Sigma_R)$. The likelihood ratio statistic is then:

$$\text{LR} = T\ln(\det\Sigma_R/\det\Sigma) \qquad \textbf{(6.83)}$$

which is distributed asymptotically as central chi-squared under the null that the cross-equation restrictions $\gamma = \gamma^*$ hold. (The number of degrees of freedom equals the number of independent restrictions in $\gamma = \gamma^*$.)

The above procedure is applicable to most tests of RE cross-equation restrictions and with appropriate variants (such as using instrumental variables) has been widely applied. One can also use a Wald test for $\gamma = \gamma^*$ which requires only an estimate of the unrestricted model, but we do not pursue that here (see for example Baillie *et al.* 1983).

By setting $\delta = 0$ in the above equations and repeating the LR test one can test rationality subject to neutrality. Similarly one can undertake a joint test of rationality *plus* neutrality (i.e. $\gamma = \gamma^*$ *and* $\delta = 0$) by comparing the likelihood from the completely unrestricted equations (6.81a) + (6.81b) with that from equations which impose both these restrictions. Cuthbertson and Taylor find both for UK and US (not reported) narrow money that the hypothesis of 'rationality without neutrality' and 'rationality plus neutrality' are decisively rejected by the data (see Table 6.5).

In the two-step procedure one tests the shock-absorber hypothesis while *implicitly imposing* the RE cross-equation restrictions (since $\hat{m} = \gamma Z_{t-1}$, *replaces* m^a in (6.78)); here Cuthbertson and Taylor find in favour of the shock-absorber hypothesis. However, joint estimation rejects the cross-equation restrictions. Hence, either the shock-absorber hypothesis or the assumption of RE does not hold – although we cannot determine from these tests which element of the joint hypothesis is incorrect.

Table 6.5 Results for UK, Narrow Money (See note)

Fully unconstrained model ($\gamma \neq \gamma^*$, $\delta \neq 0$)

1. LR(8) = 24.34
   (0.0020)
   Rationality imposed ($\gamma = \gamma^*$)
   Neutrality not imposed ($\delta = 0$)

1b. LR(1) = 5.53
   (0.0187)
   Neutrality test

2a. LR(9) = 29.87
   (0.0005)
   Rationality imposed
   Neutrality imposed

3. Fully restricted model
   Rationality imposed ($\gamma = \gamma^*$)
   Neutrality imposed ($\delta = 0$)

Note: Likelihood ratio test statistics for the jointly estimated model: LR(n) is the likelihood ratio statistic, asymptotically distributed as central chi-square with n degrees of freedom. Degrees of freedom are calculated as the number of identified parameters estimated in the unrestricted system, less those estimated in the restricted system, see Mishkin (1983). Figures in parenthesis below statistics values are marginal significance levels.

Forward-looking money demand function

We now address the question of how we can test cross-equation rationality restrictions when we have *multi-period* forward looking variables. The illustrative model is based on Hansen and Sargent (1982) and has been used widely elsewhere (e.g. Hall *et al.* 1986b, Kennan 1979, Cuthbertson 1988). Our forward-looking money demand function may be represented as:

$$M_t = \lambda M_{t-1} + (1 - \lambda)(1 - \lambda D)\sum_0^\infty (\lambda D)^i \gamma' Z_{t+i}^e \qquad \textbf{(6.84)}$$

where $$Z_t^e = (P_t^e, Y_t^e, R_t^e) \tag{6.84a}$$

$$\gamma' = (c_p, c_y, c_R) \tag{6.84b}$$

Assuming agents have information up to and including $t-1$, we can rearrange (6.84), (see Cuthbertson and Taylor 1987a), to yield

$$M_t = \gamma M_{t-1} + (1-\lambda)\left[\gamma' Z_{t-1} + \sum_0^\infty (\gamma D)^i (\gamma' \Delta Z_{t+i}^e)\right] + u_t \tag{6.85}$$

Suppose ΔZ_{t+1} can be represented by an rth order vector Markov process

$$\Delta Z_{t+1} \Phi(L) = v_{t+1} \tag{6.86}$$

where $\Phi(L)$ is a (3×3) rth order matrix polynominal in the lag operator L.

$$\Phi(L) = I - \sum_1^r \Phi_i L^i \tag{6.87}$$

and each Φ is a deterministic 3×3 matrix and the roots of $\det[\Phi(x)] = 0$ lie outside the unit circle. Clearly, (6.86) could be used by agents to forecast future values of Z_{t+i} which then determine the demand for money, via (6.85). Using the chain rule of forecasting on (6.86) yields a very complex expression for say Z_{t+4}^e even when we have only a VAR(1) process for $Z = (Y, P, R)$ – try it by hand! However, such an expression is required if we are to substitute for ΔZ_{t+i} in (6.85) and hence test the implicit cross-equation restrictions between (6.85) and (6.86). Sargent 1979, using the Weiner–Kolmogorov prediction formula, is able to provide a solution to this problem which results in the following 'restricted' two-equation model

$$M_t = \lambda M_{t-1} + (1-\lambda)(\gamma' Z_{t-1} + \gamma' \Pi(L) \Delta Z_t) + \zeta_t \tag{6.88a}$$

$$\Delta Z_{t+1} = \Phi'(L) \Delta Z_t + v_{t+1} \tag{6.88b}$$

where

$$\Pi(L) = \Phi(\lambda D)^1 \left[I + \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} (\lambda D)^{k-j} \Phi_k L^j\right] \tag{6.88c}$$

Thus the Φ_k elements from the VAR process (6.88b) also appear in the (reformulated) money demand function (6.88a) via the term $\Pi(L)$ given in (6.88c). These non-linear restrictions must be coded into the appropriate software and then (6.88a) and (6.88b) can be estimated jointly. Releasing the cross-equation restrictions on (6.88), gives an autoregressive distributed lag (ADL) formulation of the money demand equation which can then be estimated with (6.88b) to yield the 'unrestricted' system. An appropriate test statistic (for example, likelihood ratio, or quasi-likelihood ratio if instruments are used), can then be used to test the cross-equation rationality restrictions.

The appropriate estimation technique in this case is also not straightforward. The error term ζ_t in (6.88a) may be shown to be

$$\zeta_t = (1-\lambda) \sum_0^\infty (\lambda D)^i \{E(\gamma' \Delta Z_{t+1}/\Omega_t) - E(\gamma' \Delta Z_{t+1}/\Lambda_t)\} \tag{6.88d}$$

where Ω_t = complete information set used by the agent and Λ_t = information set available to the econometrician. Because ζ_t is a *future* convolution it is independent of a subset of the information available at time t, namely Λ_t. Also by RE, v_{t+1} is independent of Λ_t. If ζ_t is *not* serially correlated then Λ_t provide valid instruments with which to estimate the joint system (6.88a) + (6.88b). However if, for whatever reason, ζ_t is serially correlated we cannot use 'conventional adjustments' (GLS) for serial correlation (section 6.2). One of the methods outlined in section 6.3 must be used. IV on the unrestricted ADL money demand equation using Λ_t as instruments yields consistent (but not efficient) estimates of the parameters and hence the residuals. The latter can be used to estimate the (low order) AR coefficients ρ_1, ρ_2, etc. on the residuals. If we use the ρ_i to 'forward filter' the *variables* in the restricted equation (6.88a), then we can continue to use the IV set Λ_t dated at time t (Hayashi–Sims 1983).

Cuthbertson and Taylor (1987a) test the RE cross-equation restriction in the forward demand for money equation using data on UK, M1. The (quasi)-likelihood ratio statistic QLR(3) = 4.36 and does not reject these restrictions. In the restricted system of equations the elasticity of the demand for money with respect to income and price level can be constrained to unity, W(2) = 2.0, and the semi-elasticity with respect to R is -4.3. Therefore it appears as if the forward demand for money function together with the assumption of (weakly) rational expectations characterises the data reasonably well. Of course, this does not imply that other models of the demand for money might not perform better on purely statistical criteria (see the debate in Hendry 1988, Cuthbertson 1991, Muscatelli 1989, Cuthbertson and Taylor 1990). However, we hope we have demonstrated how cross-equation restrictions provide an additional test of the assumption of rational expectations (conditional in the assumed structural model) and in general provide a much more stringent test of the RE hypothesis than two-step procedures.

6.6 Summary

The implications of introducing expectations variables into both analytic and large-scale (econometric) models is now well established (see, for example, Lucas and Sargent 1981, Sargent 1979, Cuthbertson and Taylor 1988, Wallis *et al*. 1986, Fair 1979). However, there is much debate about how to model expectations variables and how important expectations actually are in influencing economic behaviour. We have presented a wide variety of econometric techniques for dealing with equations containing expectations terms. Although the rational expectations assumption has tended to dominate the applied (as well as the theoretical) literature we have also presented elementary 'learning' models of expectations formation which we believe will be of increasing importance. Also one must recognise that survey data on expectations can often be used directly in structural equations containing unobservable expectations (e.g. Pesaran 1985). Expectations variables are used widely in structural behavioural equations and we have analysed the main estimation methods used in the applied literature.

7 State-space models and the Kalman filter

State-space models were developed originally by control engineers (Wiener 1949, Kalman 1960) but are receiving increasing attention in the economics literature. There is a number of advantages in representing models in state-space form. We noted in Chapter 2 that the likelihood function can be written in terms of the one-step-ahead prediction errors $\tilde{v}_t$ and their variance f_t. The Kalman filter when applied to a model in state-space form provides an algorithm for producing $\tilde{v}_t$ and its variance. Since many models (for example all ARMA models) can be represented in state-space form, the Kalman filter provides a convenient general method of representing the likelihood function for what may be very complex models. Two types of model that are especially amenable to representation via the Kalman filter are *unobservable components* models and *time-varying parameter* models. In unobservable components models we *observe* y_t (say actual income) which we assume consists of an *unobserved* permanent component π_t plus a white noise error ε_t:

$$y_t = \pi_t + \varepsilon_t$$

The Kalman filter provides an optimal updating scheme for the unobservable π_t based on information about measured income, as it sequentially becomes available. With this interpretation the unobservable components model provides a method of generating an expectations series for permanent income π_t.

In time-varying parameter models we have

$$y_t = x_t\beta_t + \varepsilon_t$$

where (y_t, x_t) are observables. The problem is then to estimate β_t as