CHAPTER 3

EXPECTATIONS, LEARNING AND THE KALMAN FILTER

Expectations are formed by intelligent agents who are not fully informed and who, as a consequence, must learn about their environment.
(Hall and Garratt 1992a:52)

3.1 INTRODUCTION

Individuals frequently form expectations about the future level of prices e.g. when making consumption expenditure decisions and during wage bargaining. Expectations are formed conditional on economic agents' perceptions of the current economic environment or regime as well as possible time-related changes in that environment.

Expectations, i.e. anticipations or views of the future, have featured prominently in macroeconomic literature from the inception of the concept. Since 1930, when Irving Fisher introduced the 'anticipated rate of inflation' as the difference between the nominal and real interest rates, expectations have played an important role in economic theory. Formal analytical treatment of expectations formation, however, only emerged over the last quarter of this century. As a result, important advances in this area have occurred.

The development of macroeconomic models over the past thirty years has forced economists to recognise that expectations are not to be treated as exogenous - or to be ignored at will - but instead are central to our understanding of the functioning of the economy (Holden *et al.* 1985:1).

Historically, there have been two distinct methods of dealing with expectations in economic analysis - one is the direct measuring of expectations by conducting surveys to determine expectations, the other is to provide a simple model of expectations formation.

Much early empirical work on expectations centred around attempts to provide direct measures of agents' expectations (e.g. Katona 1951, 1958; Tobin 1959 and Eisner 1965).

Much of this research was directed towards understanding the psychological underpinnings of individual expectations formation.

Direct measures are, for obvious reasons, not a very plausible method of obtaining expected values for future outcomes of economic variables. Although undoubtedly useful in preparing economic forecasts, gathering direct measures of expectations is costly and time consuming; in addition the data become outdated rather quickly. Furthermore, and perhaps more importantly, direct measures of agents' expectations provide little insight into how expectations would change as policy changes. The breakthrough that led to a more general approach to expectations modelling came with the realisation that expectations could be treated as an unobservable component. In this study, the latter method is adopted: a simple model of expectations formation is specified and the coefficient vector of the expectations rule is treated as an unobservable component.

This chapter surveys the theoretical developments on expectations formation by economic agents, starting with the *adaptive expectations hypothesis*. This is followed by *rational expectations* (or *model-consistent expectations*), and finally the process of *learning* (or *boundedly rational expectations*). A variable parameter estimation technique, namely the Kalman filter, which will be utilised in the estimation of consumer price expectations in South Africa, is discussed in section 3.4. Section 3.5 deals with the implementation of expectations in macro-models and section 3.6 reports on a number of international studies concerning the learning approach. The empirical result of application of the variable parameter estimation technique to the price expectations rule, is reported in Chapter 5. Price expectations are then implemented in a set of 'forward-looking' consumption functions, the result of which is also documented in Chapter 5.

3.2 EXPECTATIONS IN MACRO MODELS

Any macro model containing expectations can, in general terms, be defined as

$$y_{it} = f_i(Y_t, X, {}_tY^e_{t+1}) \qquad i = 1 \ldots n \qquad t = 1 \ldots T \tag{3.1}$$

where

Y	=	the vector of current and lagged values of the n endogenous variables ($y_{it}...y_{it-k}$, where k is the lag depth of the model)
X	=	the vector of exogenous variables over all time periods ($x_0...x_T$) and
${}_tY^e_{t+1}$	=	expectation of Y for period t+1, based upon information available at period t.

This expectation may, in general, be viewed as being derived from another set of relationships, namely

$$ {}_ty^e_{t+1} = g_k(Y_t, X) \qquad k = 1...m. \qquad (3.2)$$

The next section provides an exposition of different approaches in theory and practice to find expressions for equation (3.2).

3.3 MODELS OF EXPECTATIONS FORMATION

This section briefly discusses the theoretical development of expectations analysis and also refers to its implementation in empirical models.

3.3.1 Adaptive expectations mechanisms

During the twenty years after the Second World War, the adaptive expectations hypothesis (AEH), first proposed by Cagan (1956), enjoyed considerable popularity amongst econometricians as a simple and apparently sensible model of how economic agents form expectations.

In modelling expectations of, say, future price levels, the hypothesis of adaptive expectations states that

$$({}_{t-1}p^e_t - {}_{t-2}\,p^e_{t-1}) = \Phi(p_{t-1} - {}_{t-2}p^e_{t-1}) \qquad 0 < \Phi < 1 \qquad (3.3)$$

where ${}_{t-1}p_t^e$ represents the expected value of the price level in period t formed in period t–1.

The individual therefore harbours a series of expectations for future outcomes of the price level and, in each period, the expectation for the future is revised in a way proportional with the most recently observed error.

Rearranging (3.3) yields

$$ {}_{t-1}p_t^e = \Phi p_{t-1} + (1-\Phi)\,{}_{t-2}p_{t-1}^e \tag{3.4}$$

and by successively substituting for the lagged expectations we get

$$ {}_{t-1}p_t^e = \Phi p_{t-1} + \Phi(1-\Phi)p_{t-2} + \Phi(1-\Phi)^2 p_{t-3} + \dots \tag{3.5}$$

Unobservable expectations with respect to the price level may thus be measured purely in terms of past observations of the actual price level.

The above seemed intuitively appealing, as it states that expectations of the future are a simple extrapolation of the past. The deficiencies of this approach, however, gradually became apparent. The adaptive expectations rule would, for example, cause consistent and growing mistakes in periods when prices grow at a constant rate, of say 10 per cent. The adaptive expectations model would then consistently under-estimate the level of p, and it would do so at an increasing absolute amount over time, implying that an economic agent would make entirely predictable errors, even in the very long run. It is hard to believe that such a feature could be true of an intelligent economic agent.

One way of addressing this problem is to generalise the rule to an extrapolative one, which can cope with growing variables. Fixed parameter rules in general are, however, liable to perform poorly in one circumstance or the other. An example of this is a change in the regime generating the underlying variable. The Lucas critique (Lucas 1976) pointed out that when expectations are modelled by functions of lagged variables, the parameters of these functions may vary as the regime for determining the expectations variable changes. So if we assume that agents are intelligent and that they avoid consistent expectations errors, then any fixed parameter extrapolative expectations model will be unable to accommodate policy or other regime changes (Currie and Hall 1994:93).

3.3.2 The concept of rational expectations

The predominant paradigm for modelling expectations is the rational expectations hypothesis. The concept of rational expectations (RE) was originally formulated by Muth (1961). He suggested that agents form their expectations in the same way that they undertake other activities – that is, they use economic theory to predict the value of the variable and this is their 'rational' expectation. Rational expectations are thus simply predictions based on economic theory, using the information available at the time the predictions are made (Holden *et al.* 1985:18).

Walters (1971) preferred the term 'consistent' to 'rational' since the expectations are consistent with the relevant economic theory – it assumes that agents use information efficiently in such a way that the aggregate of all agents' expectations will not be systematically wrong. The seventies witnessed an explosion in theoretical work incorporating rational expectations, which then became closely linked with the neo-classical approach to macroeconomics, to the extent that the two approaches were widely regarded as synonymous.

In the full or strong form of the rational expectations hypothesis, it is assumed that economic agents have a complete knowledge of the economic system about which they need to form expectations. This knowledge includes the functional form, the parameters of the system, as well as any exogenous process governing the system.

Formally, the rational expectations hypothesis (REH) may be stated as follows:

$$ {}_{t}p^{e}_{t+s} = E_{t}(p_{t+s}) \qquad (3.6)$$

where

p	=	the variable being forecasted (i.e. the price level)
${}_{t}p^{e}_{t+s}$	=	the prediction of the price level for time t+s, formed at time t and
E_t	=	the statistical expectation conditional on information available at time t, when the forecast is made.

The rational expectations hypothesis requires that the prediction made by the forecaster be consistent with the prediction generated by the model, conditional on information available at that time. Setting s=1, equation (3.6) implies:

$$ p_{t+1} = ({}_{t}p^{e}_{t+1}) + \varepsilon_{t+1}. \qquad (3.7)$$

This equation asserts that the price fluctuates about its forecast level with a purely random error ε_{t+1} that has a zero mean. This relationship between the price and its prediction, according to Turnovsky (1996:59), characterises an efficient market. It means that prices fully reflect available information, thus eliminating any systematic opportunities for making supernormal profits. As an empirical description, this assumption is an appealing one for asset and financial markets in which information is generally readily available. But it is probably less appealing for forecasting such quantities as the expectations of the consumer price index, which are likely to be based on far inferior information.

Several arguments for and against the rational expectations hypothesis are found in the literature. Some authors still maintain that one of the most compelling arguments in favour of the rational expectations hypothesis, is the relative weakness of the alternatives. As already noted, traditional expectations schemes, such as the adaptive expectations hypothesis, involve systematic forecasting errors. This is not particularly desirable, since one would expect individuals to learn this eventually and to abandon such rules or to modify them in some way. By contrast, the rational expectations hypothesis generates expectations that are self-fulfilling to within a random error, which cannot be predicted on the basis of information available at the date when the expectations are formed.

The most important objections to the rational expectations hypothesis concern the fact that the application thereof requires not only knowledge of the underlying structure of the model, but also of the relevant parameters. Specialists are unable to agree on a model, but even if they should find agreement, they usually obtain varying estimates for relevant coefficients. Much less will consumers, who are presumably less sophisticated in economic theory, but whose expectations we are attempting to model, have such information. This informational argument clearly has merit.

The first estimated macroeconomic models incorporating rational expectations were those of Anderson (1979) and Fair (1979). The latter was a fairly large model with 84 equations, including expectations of future prices of the stock and bond markets. The solution technique in these applications (Fair and Taylor 1983) consists of a two part iteration scheme: first, values for the expectations variable are considered given and conventional Gauss-Seidel solution methods are used to solve the model conditional on these given values for expectations; second, the expectations variables are set equal to the solution values from the model derived in the first stage. The whole process is then repeated until the expectations variables used in the first stage are consistent with the updated values of the second stage.

The rational expectations hypothesis has failed to become general practice in a forecasting context, due to a number of practical problems associated with this approach. Hall (1995:979) points out that the presence of rational expectations tends to cause jumps in the initial period value of a range of variables, for example the exchange rate, which is considered implausible by forecasters. However, it is in the large empirical models where the weakness of the rational expectations assumption has become most apparent. As a tool for analysing long-run behaviour, an assertion of rational expectations is both powerful and useful, but it is not a good representation of the short-run dynamic behaviour of many markets. Implementation in large models has in fact emphasised this weakness (Currie and Hall 1994:109). Another practical problem related to the above, is that many policy options may be impossible to analyse under rational expectations. Hall and Garratt

(1998:257), e.g. illustrated how a permanent rise in interest rates would lead to an infinitely large jump in the exchange rate under the open arbitrage condition and the model then fails to yield a solution.

Another fact to consider is the possibility that the information set may contain errors. In such instances, rational expectations assumptions will, by exploiting the information set to the full, make larger overall errors than the adaptive expectations assumption which makes less efficient use of the information.

One conclusion to be drawn from the above, may be that agents do not have full information and full model-consistent expectations but that they learn about regime changes over time and assimilate new information. It would, therefore, be natural to progress from *strong* rational expectations to a *weaker* model, which allows for the possibility of making errors in the short run, while ruling out long-run systematic or predictable errors. This gives rise to learning models. Learning models represent a shift from the assumption of full rationality on the part of economic agents in forming expectations, towards the idea of bounded rationality which is more consistent with the psychology literature on learning processes.

3.3.3 Learning

Learning models of expectations have increasingly received attention in the theoretical literature over the past decade or more. In response to the limitations of the rational expectations hypothesis imposed by the stringent assumption of full knowledge of the true structural model, a number of economists have introduced processes describing the way in which economic agents learn about the underlying economic structure over time.

Cuthbertson (1988:224) reports on an early attempt by Friedman (1979) to provide an alternative optimising framework to rational expectations. Friedman advocates that given the true model $y_t = x_t\beta + u_t$ (where u_t is a white noise error term), agents may sequentially update their estimate of the fixed true parameter vector β as more information on (y_t, x_t) becomes available – a process of recursive least squares. Cuthbertson extends Friedman's framework to include the case where (i) agents have some prior information about β (at time $t=0$) and (ii) β is allowed to vary stochastically. Friedman (1979:33-34) alludes to the latter outcome when he discusses the possibility that agents may perceive a simple linear model with time varying parameters as a good approximation to the complex 'true' model. Such a model would then represent a boundedly rational learning model of expectations.

The concept of boundedly rational learning is the result of a slightly weaker informational assumption. It implies that agents use some 'reasonable' rule to form expectations and that the form of this rule remains constant over time while agents 'learn' the parameters of this rule. Agents are therefore not assumed to instantaneously know the 'true' model but they do use information optimally (or efficiently).

Implementing boundedly rational learning in a macro model entails the specification of a relatively simple expectations rule and then allowing the parameters of this rule to vary, thus correcting previous errors made by that rule. Over time, the expectations rule would be expected to adjust to the particular regime under which the model is operating and a rational expectations solution will be established. However, in the short term, the learning rule allows for errors and hopefully generates a more plausible path to equilibrium.

A serious issue with regard to the learning approach, and to a certain extent unresolved, is the selection of an appropriate expectations rule. In many instances, the reduced form of the structural equation containing the expectations variable, is the best vantage point for formulating the expectations rule (e.g. DeCanio 1979; Radner 1982; Bray and Slavin 1986; Hall and Garratt 1992a, 1992b).

The specification of an expectations rule, such as equation (3.2), is based on the assumption that some element of the rule – usually the parameter vector – is not known with certainty. The basic idea is that over time, the economic agent will methodically increase his knowledge about the true values of these parameters. Equation (3.2) may therefore be restated to explicitly include the parameters of the rule:

$$ {}_t y^e_{t+1} = g_k(Y_t, X, \xi_t) \qquad k = 1 \ldots m \tag{3.8} $$

where ξ_t is the vector of agents' estimates of the parameters at period t. Having specified the learning rule or the expectations rule, a mechanism that governs the evolution of these parameters through time needs to be specified.

The way in which the parameters change can be determined in a number of ways. At the simplest level, agents could simply perform a set of 'rolling' regressions using ordinary least squares, each period adding the latest expectation error to the data set. According to Hall and Garratt (1992a:52), this simple form of learning does not respond well to structural changes. The more sophisticated mechanism based on the Kalman filter, is therefore preferable, although the steps are conceptually similar.

The Kalman filter, proposed as an optimal method of implementing the learning process, will be discussed in the next section. 'Optimal' in this context would mean that expectations are correct on average and have minimum mean square prediction errors.

To standardise on notation, note that ${}_{t-1}p^e_t$ and ${}_tp^e_{t+1}$ in equations (3.5), (3.7) and (3.8) may also be written as $p^e_{t|t-1}$ and $p^e_{t+1|t}$. The latter corresponds to the notation used in the next section where a technical exposition of the Kalman filter algorithm is presented – the Kalman filter may be viewed as a form of adaptive expectations where the adjustment parameter, and thus the price expectation, is updated each period based on new information.

3.4 A TECHNICAL EXPOSITION OF STATE-SPACE MODELS AND THE KALMAN FILTER

State-space models were originally developed by control engineers (Wiener 1949, Kalman 1960, 1963). Kalman filtering, named after the contributions of R.E. Kalman, found applications in, for example, the technology of radars, aircraft stabilisation, the determination of coordinates in nautical or aerospace applications and chemical processes. It was not until the 1980s that state-space models received attention in economics literature (amongst others, Lawson 1980; Harvey *et al.* 1986; Cuthbertson 1988; Barrel *et al.* 1994; Currie and Hall 1994).

There are two major benefits to representing a dynamic system in state-space form. First, the state-space allows unobserved variables (known as the state variables) to be incorporated into, and estimated along with, the observable model. Second, state-space models can be estimated using the Kalman filter, a powerful recursive algorithm. Apart from estimating vector autoregressions with coefficients that change over time, this algorithm also provides a way to calculate exact finite-sample forecasts and the exact likelihood function for Gaussian ARMA processes. Also, factor matrix autocovariance-generating functions or spectral densities may be determined by means of the Kalman filter (Hamilton 1994:372).

State-space models have been applied in the economics and econometrics literature to model unobserved variables such as permanent income, expectations, measurement errors, missing observations, unobserved components (cycles and trends) and the natural rate of unemployment. Extensive surveys of applications of state-space models in econometrics can be found in Hamilton (1994:372-408) and Cuthbertson *et al.* (1992:191-225). Cuthbertson *et al.* (*op cit.*:191) distinguish between two types of models especially amenable to representation via the Kalman filter, namely *unobservable components* models and *time-varying parameter* models. In this study, the state-space model with stochastically time-varying parameters has been applied to a linear regression, i.e. the price expectations rule, in which the coefficient vector changes over time.

The next section describes how a dynamic system can be written in state-space form, which form is suitable for the application of the Kalman filter.

3.4.1 The state-space representation of a dynamic system

The state-space representation of the dynamics of an $(n\times1)$ vector, y_t, is given by the following system of equations:

$$y_t = A'x_t + H'\xi_t + w_t \tag{3.9}$$

$$\xi_{t+1} = F\xi_t + v_{t+1} \tag{3.10}$$

where A', H' and F are matrices of parameters of dimension (n×k), (n×r) and (r×r), respectively, and x_t is a (k×1) vector of exogenous or predetermined variables. ξ_t is a (r×1) vector of possibly unobserved state variables, known as the *state vector*. The first equation is known as the *observation* (or measurement) equation and the second is known as the *state* (or transition) equation. The (n×1) and (r×1) disturbance vectors w_t and v_t are assumed to be independent white noise with

$$E(v_t v'_\tau) = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

$$E(w_t w'_\tau) = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

where Q and R are (r×r) and (n×n) matrices, respectively. The disturbances v_t and w_t are assumed to be uncorrelated at all lags:

$$E(v_t w'_\tau) = 0 \qquad \text{for all } t \text{ and } \tau. \tag{3.13}$$

The statement that x_t is predetermined or exogenous means that x_t provides no information about ξ_{t+s} or w_{t+s} for s=0,1,2,... beyond what is contained in $y_{t-1}, y_{t-2},\ldots,y_1$. Thus, x_t could include lagged values of y or variables which are uncorrelated with ξ_τ and w_τ for all τ.

The system of equations (3.9) through (3.13) is typically used to describe a finite series of observations $\{y_1, y_2,\ldots,y_T\}$ for which assumptions about the initial value of the state vector ξ_t are needed.

The various parameter matrices (F, Q, A, H or R) could be functions of time, as will be discussed in section 3.4.3.

Substituting (3.17) into (3.18),

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1}) . \qquad (3.21)$$

The coefficient matrix in (3.21) is known as the *gain matrix* and is denoted K_t:

$$K_t \equiv FP_{t|t-1} + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1} . \qquad (3.22)$$

For a derivation of these equations, see Hamilton (1994:379-380). Equations (3.17) and (3.18) may be called the updating equations and the value of $\hat{\xi}_{t+1|t}$ denotes the best forecast of ξ_{t+1} based on a constant and a linear trend function of $(y_t, y_{t-1}, \ldots, y_1, x_t, x_{t-2}, \ldots x_1)$. The matrix $P_{t|t+1}$ gives the MSE of this forecast.

The forecast of y_{t+1} is then given by

$$\hat{y}_{t+1|t} \equiv \hat{E}(y_{t+1|t} \mid x_{t+1|t}, \vartheta_t) = A'x_{t+1} + H'\hat{\xi}_{t+1|t} \qquad (3.25)$$

with the associated MSE

$$E[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})'] = H'P_{t+1|t}H + R. \qquad (3.26)$$

What still remains outstanding in the exposition above, is the estimation of the unknown parameter matrices F, Q, A, H and R, which are determined through maximum likelihood estimation of these parameters. The Kalman filter was motivated in the above discussion in terms of linear projections. The forecasts $\hat{\xi}_{t|t-1}$ and $\hat{y}_{t|t-1}$ are thus optimal within the set of forecasts that are linear in (x_t, ϑ_{t-1}), where $\vartheta_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, x'_{t-1}, x'_{t-2}, \ldots, x'_1)'$.

If the initial state ξ_1 and the innovations $\{w_t, v_t\}_{t-1}^{T}$ are multivariate Gaussian[3], then the stronger claim can be made that the forecasts $\hat{\xi}_{t|t-1}$ and $\hat{y}_{t|t-1}$ calculated by the Kalman filter are optimal among any functions of (x_t, ϑ_{t-1}). Moreover, if ξ_1 and $\{w_t, v_t\}_{t-1}^{T}$ are

---

[3] The vector $Y = (Y_1, Y_2, \ldots, Y_n)'$ has a multivariate Gaussian or multivariate normal distribution if $f_Y(y) = (2\pi)^{-n/2}|\Omega|^{-1/2} \exp[(-\frac{1}{2})(y-\mu)'\Omega^{-1}(y-\mu)]$ and is written as $Y_t \sim N(\mu, \Omega)$ with the mean μ an $(n\times1)$ vector and the variance-covariance a matrix Ω of dimension $(n\times n)$ which is symmetric and positive semidefinite.

Gaussian, then the *distribution* of y_t conditional on (x_t, ϑ_{t-1}) is Gaussian with mean given by (3.23) and variance given by (3.24):

$$y_t \mid x_t, \vartheta_{t-1} \sim N((A'x_t + H'\hat{\xi}_{t|t-1}), (H'P_{t|t-1}H + R));$$

that is

$$\begin{aligned} & f_{y_t|x_t,\vartheta_{t-1}}(y_t \mid x_t, \vartheta_{t-1}) \\ & \quad = (2\pi)^{-n/2} \mid H'P_{t|t-1}H + R \mid^{-1/2} \\ & \quad\quad \times \exp\{-\tfrac{1}{2}(y_t - A'x_t - H'\xi_{t|t-1})'(H'P_{t|t-1}H + R)^{-1} \\ & \quad\quad \times (y_t - A'x_t - H'\xi_{t|t-1})\} \qquad \text{for } t = 1,2,\ldots,T. \end{aligned} \tag{3.25}$$

From the above equation, the sample log likelihood can be constructed, namely

$$\sum_{t=1}^{T} \log f_{y_t|x_t,\vartheta_{t-1}}(y_t \mid x_t, \vartheta_{t-1}). \tag{3.26}$$

Expression (3.26) can then be maximised numerically with respect to the unknown parameters in the matrices F, Q, A, H and R. Burmeister and Wall (1982) illustrate this application.

To summarise, the state vector ξ_t and its mean squared error P_t are recursively estimated by the set of equations (3.17) through (3.20), which also represents the key equations for the Kalman filter. $\xi_{s|v}$ is the forecast of the state vector at time period s, given information available at time v. Note that the recursion for P does not depend on the forecasted state vector, or on the observed data (y_t,x_t).

To implement the Kalman filter, the starting values must be specified and the unknown matrices must be replaced by their estimates. By default, a software package like EViews obtains starting values by treating these matrices as fixed coefficients and estimating them using OLS.

Once starting values are obtained, the parameters A, H, F, R and Q are estimated by maximising the log likelihood function under the assumption that the distribution of y_t, conditional on x_t and past values of (y_t,x_t), is multivariate normal (Gaussian).

3.4.3 State-space model with stochastically varying coefficients

In the above discussion, it was assumed that the matrices F, Q, A, H and R were all constant. The Kalman filter can also be adapted for more general state-space models in which the values of these matrices depend on the exogenous or lagged dependent variables included in the vector x_t.

Equations (3.9) and (3.10), i.e. the state-space representation may thus be altered to:

$$y_t = a(x_t) + [H(x_t)]'\xi_t + w_t \tag{3.27}$$

$$\xi_{t+1} = F(x_t)\xi_t + v_{t+1}. \tag{3.28}$$

Here $F(x_t)$ denotes an $(r\times r)$ matrix whose elements are functions of x_t; $a(x_t)$ similarly describes an $(n\times 1)$ vector-valued function and $H(x_t)$ an $(r\times n)$ matrix-valued function.

It is assumed that conditional on x_t and on the data observed through date $t-1$, denoted $\vartheta_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, x'_{t-1}, x'_{t-2}, \ldots, x'_1)'$, the vector $(v'_{t+1}, w'_t)'$ has the Gaussian distribution

$$\begin{bmatrix} v_{t+1} \\ w_t \end{bmatrix} \mid x_t, \vartheta_{t-1} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q(x_t) & 0 \\ 0 & R(x_t) \end{bmatrix}\right). \tag{3.29}$$

Equations (3.27) to (3.29) generalise the earlier framework by allowing for stochastically varying parameters, but it is more restrictive in the sense that a Gaussian distribution is assumed in (3.29). Hamilton (1994:399-400) explains the role of the Gaussian requirement, as set out below.

Suppose it is taken as given that $\xi_t \mid \vartheta_{t-1} \sim N(\hat{\xi}_{t|t-1}, P_{t|t-1})$. Assuming as before that x_t contains only strictly exogenous variables or lagged values of y, this also describes the distribution of $\xi_t \mid x_t, \vartheta_{t-1}$. It follows from the assumptions in (3.27) to (3.29) that

$$\begin{bmatrix} \xi_t \\ y_t \end{bmatrix} \mid x_t, \vartheta_{t-1} \sim N\left(\begin{bmatrix} \hat{\xi}_{t|t-1} \\ a(x_t) + [H(x_t)]'\hat{\xi}_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}H(x_t) \\ H'(x_t)P_{t|t-1} & [H(x_t)]'P_{t|t-1}H(x_t) + R(x_t) \end{bmatrix}\right). \tag{3.30}$$

Conditional on x_t, the terms $a(x_t)$, $H(x_t)$ and $R(x_t)$ can all be treated as deterministic. Thus, the formula for the conditional distribution of Gaussian vectors[4] can be used to deduce that

$$\xi_t \mid y_t, x_t, \vartheta_{t-1} \equiv \xi_t \mid \vartheta_t \sim N(\hat{\xi}_{t|t}, P_{t|t}). \tag{3.31}$$

where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + \left\{P_{t|t-1}H(x_t)\left[[H(x_t)]'P_{t|t-1}H(x_t) + R(x_t)\right]^{-1} \times \left[y_t - a(x_t) - [H(x_t)]'\xi_{t|t-1}\right]\right\} \tag{3.32}$$

and

$$P_{t|t} = P_{t|t-1} - \left\{P_{t|t-1}H(x_t) \times \left[[H(x_t)]'P_{t|t-1}H(x_t) + R(x_t)\right]^{-1}[H(x_t)]'P_{t|t-1}\right\}. \tag{3.33}$$

It then follows from (3.27) and (3.29) that $\xi_{t+1} \mid \vartheta_t \sim N(\xi_{t+1|t}, P_{t+1|t})$, where

$$\hat{\xi}_{t+1|t} = F(x_t)\hat{\xi}_{t|t} \tag{3.34}$$

and

$$P_{t+1|t} = F(x_t)P_{t|t}[F(x_t)]' + Q(x_t). \tag{3.35}$$

Equations (3.32) through (3.35) are the Kalman filter equations (3.17) to (3.20), with the parameter matrices replaced with their time-varying analogs. Thus, as long as the initial state ξ_1 is treated as $N(\hat{\xi}_{1|0}, P_{1|0})$, the Kalman filter iterations proceed as before. The obvious generalisation of (3.25) may still be used to calculate the likelihood function.

The only difference from the constant-parameter case is that the inference (3.32) is a non-linear function of x_t. This means that although (3.32) gives the optimal inference if the disturbances and initial state are Gaussian, it cannot be interpreted as the linear projection of ξ_t on ϑ_t, with non-Gaussian disturbances.

3.4.4 The Kalman filter in terms of conventional least squares procedures

The exposition above is mainly in applied statistical terms. Cuthbertson (1988:234-241) presents the econometrics of the Kalman filter using conventional least squares procedures. He begins by reinterpreting the familiar problem of estimation subject to linear restrictions

---

[4] The conditional density of ξ_t given y_t, is found by dividing the joint density by the marginal density (Hamilton 1994:101-102, 399).

in terms of a Kalman filter updating equation, utilising the Kalman gain. This is followed by a reinterpretation of the Theil-Goldberger pure and mixed estimator as a 'one-shot' Kalman filter updating equation combining *a priori* and sample information. In both of the above cases, Cuthbertson reinterprets the standard formulae for the covariance matrix of the parameters as Kalman filter variance updating equations. He also concludes by presenting the relatively complex case of the state-space form with variable parameters.

3.5 IMPLEMENTATION OF THE LEARNING PROCESS IN A MACROECONOMIC MODEL

The learning process is generally implemented in model context as follows. The model essentially contains three blocks of equations, namely:

$$y_{it} = f_i(Y_t, X, {}_tY^e_{t+1}) \qquad i = 1\ldots n, \qquad t = 1\ldots T \tag{3.36}$$

$$ {}_ty^e_{t+1} = g_k(Y_t, X, \xi_t) \qquad k = 1\ldots m \tag{3.37}$$

$$\xi_t = \xi_{t-1} + v_t \qquad v_t \sim N(0,Q). \tag{3.38}$$

Equation (3.37) represents the measurement (or observation) equation(s) of the state-space model in Kalman filter terms, while equation (3.38) represents the state (or transition) equation(s).

Assuming that the value of ξ_{t-1} is known, the last block of equations for the expected value of ξ_t can be solved, which is simply the Kalman filter prediction equations for ξ, called the state vector. Given ξ, the second block of equations for the expected value of ${}_tY^e_{t+1}$ can be solved, and given this, the first block of equations for Y_t can finally be solved.

Q, the covariance matrix of the errors in the equations governing the evolution of the parameters (or in Kalman filter terms, the state equation error terms) is given by the original estimation and an estimation for P_{t-1} can be obtained (the uncertainty of the parameters or state variables). The Kalman filter prediction equations for P can then be used to derive an estimate of ${}_{t-1}P_t$. Having solved the complete model for Y_t, we can determine the one-step-ahead prediction error, that is the error that occurs between the expectation of the vector Y_t derived from the learning model and the model's final solution for Y_t. The one-step-ahead prediction error is a combination of stochastic error terms of the measurement and state equations. Given this and the estimate of $P_{t|t-1}$, the Kalman filter updating equations can be used to derive revised estimates of P_t and ξ_t. The updating is done on the basis of the observed errors between the whole model solution and the original expectations model forecast.

The process is then repeated for the next period, starting from the new updated estimates of ξ_t to predict ξ_{t+1}, and so on. In this way, the learning model will adjust its own parameters to cope with any change in structure or regime of the whole model. In the forecasting period, the final values of the state vector are used, in addition to other exogenous input, to solve the model for Y_t.

The underlying assumptions of this process are still quite strong, as agents are still assumed to process all available information in an optimal fashion and a substantial degree of sophistication on the part of the economic agent is still assumed. The learning model may, however, fulfil the criteria for weak rational expectations, since agents are not assumed to have full information. They will most likely make mistakes in the short run, but systematic errors over an extended period of time are ruled out.

Considering the model's response to any regime change reveals that the parameters of the learning rule respond over time to that change. The behaviour of the parameters of the rule, provides an important insight into the form of equilibria which may emerge from the system. Marcet and Sargent (1988) summarise the main theoretical results. The concepts of learning is characterised as the process of changing the parameters of the rule. If these parameters settle down to some fixed level, learning may be regarded as having ceased and this is sometimes called an expectational equilibrium (or E-equilibrium). Marcet and Sargent demonstrate that this stable condition is also a full rational expectations equilibrium.

3.6 EMPIRICAL LITERATURE ON THE LEARNING APPROACH

Although the theoretical literature on learning has grown from the early work of Friedman (1975 and 1979), the empirical literature has remained relatively sparse. Two empirical implementations of the learning approach in large-scale macro models will be presented as representative examples of the application of the Kalman filter to an expectations rule.

The learning approach to the treatment of expectations was first adopted in the exchange rate sector of the London Business School (LBS) model of the UK economy. Hall and Garratt (1992b) present an empirical model for the Sterling/Deutchmark real exchange rate assigning an important role to exchange rate expectations, which are assumed to be formed through a Kalman filter-based learning process.

The structural form of the exchange rate equation, derived from a capital stock model with government intervention is:

$$E_t = \alpha_0 + \alpha_1 E^e_{t+1} + \alpha_2 E_{t-1} + \alpha_3 r_t + \alpha_4 r_{t-1} + \alpha_5 T_t + \alpha_6 T_{t-1} \tag{3.39}$$

with

E_t	=	log of the real Sterling/Deutchmark exchange rate
r_t	=	real interest rate differential between UK and German short-term rates and
T_t	=	log of the ratio of exports to imports which is a measure of the real trade balance.

The end result of a search is reported as a restricted form of the above equation, estimated by three stage least squares, namely

$$E_t = 0.0329 + 0.675E^e_{t+1} + 0.299E_{t-1} + 0.352r_t \tag{3.40}$$

which, in the long-run, exactly equals uncovered interest rate parity (UIP).

The time-varying rule for exchange rate expectations is derived from the structural equation (*op. cit.*:10) by rearranging equation (3.39) to give:

$$E^e_{t+1} = \beta_0 + \beta_1 E_t + \beta_2 E_{t-1} + \beta_3 r_t + \beta_4 r_{t-1} + \beta_5 T_t + \beta_6 T_{t-1}. \tag{3.41}$$

Lagging this equation by one period and using it to substitute out the term in E_t, after collecting terms, gives:

$$\begin{aligned} E^e_{t+1} &= \beta_0 + (\beta_1^2 + \beta_2)E_{t-1} + \beta_1\beta_2 E_{t-2} + (\beta_1\beta_3 + \beta_4)r_{t-1} \\ &+ \beta_1\beta_4 r_{t-2} + \beta_3 r_t + (\beta_1\beta_5 + \beta_6)T_{t-1} + \beta_1\beta_6 T_{t-2} + \beta_5 T_t. \end{aligned} \tag{3.42}$$

The β coefficients are combinations of the α coefficients, and the contemporaneous terms in the above equation are replaced by assuming partial reduced form equations for r_t and T_t, respectively, as follows:

$$r_t = C_1(L)r_{t-1} + C_2(L)GDP_{t-1} + C_3(L)INF_{t-1} \tag{3.43}$$

$$T_t = D_1(L)T_{t-1} + D_2(L)GDP_{t-1} + D_3(L)PO_{t-1} + D_4(L)E_{t-1} \tag{3.44}$$

where PO is the log of the oil price and INF is the rate of inflation which could include both UK and German inflation or the differential and GDP is the log of real GDP. C_i and D_i are polynomial lag operators.

The terms r_t and T_t my thus be eliminated and collection of terms yields:

$$\begin{aligned} E^e_{t+1} &= \beta_0 + (\beta_1^2 + \beta_2 + \beta_5 D_4(L))E_{t-1} + \beta_1\beta_2 E_{t-2} + (\beta_1\beta_2 + \beta_4 + \beta_3 C_1(L))r_{t-1} \\ &+ \beta_1\beta_4 r_{t-2} + (\beta_1\beta_5 + \beta_6 + \beta_5 D_1(L))T_{t-1} + \beta_1\beta_6 T_{t-2} + \beta_3 C_2(L)GDP_{t-1} \\ &+ C_3(L)INF_{t-1} + \beta_5(D_2(L)GDP_{T-1} + D_3(L)PO_{t-1}). \end{aligned} \tag{3.45}$$

Equation (3.45) is regarded as the basic partial reduced form rule which agents use to form their expectations. The above was simplified by dropping any lagged terms greater than t-3 by introducing a stochastic constant (*op. cit.*:11).

The basic structure was then used in a specification search to produce the following equation for expectations formation:

$$E^e_{t+1} = \gamma_{0t} + \gamma_{1t}E_{t-1} + \gamma_{2t}r_{t-1} + \gamma_{3t}INF^{UK}_{t-1} + \gamma_{4t}T^{UK}_{t-1} + \gamma_{5t}PO_{t-1}. \tag{3.46}$$

Note that all lagged information is dated t−1 when the equation is used to forecast E_{t+1}, i.e. the information set does not include current information. The parameters γ are restricted forms of the β coefficients. Hall and Garratt (*op. cit.*:11) note that their derivation of the expectations equation is by no means exclusive. In principle, formation of expectations could be the result of information from anywhere in the model, provided that it is relevant.

The time-varying parameters are then assumed to be generated by the following process:

$$\gamma_{it} = \gamma_{it-1} + \varepsilon_{it} \ . \tag{3.47}$$

Equation (3.46) may be regarded as the standard measurement equation and the set of equations (3.47) as the state equations. The model has thus been formulated in state-space form and the Kalman filter can be applied to estimate the time-varying parameters conditional on the variance of the error terms of (3.46) and the covariance matrix of (3.47), which is assumed to be diagonal. The estimation reported (*op. cit.*:14-18) suggests that the most important determinants of the forward exchange rate are real interest rate differentials and the previous period's exchange rate. All parameter values displayed a reasonable degree of time variance, with the main movements in the coefficients on the real interest rate differential and the ratio of exports to imports for the UK, reflecting a faster rate of learning on these variables.

Another example where the Kalman filter was employed to estimate the time-varying parameter rule of expectations is a study conducted by Barrel *et al.* (1994:173). In this study, the boundedly rational learning approach was applied to wage behaviour in three countries, namely the UK, France and Italy. A time-varying parameter model for forecasting prices was first estimated. This defined the information set of the economic agents. The expectations rule together with the structural equations in which the expectations are embedded were then incorporated into the global econometric model (GEM), developed by the National Institute of Economic and Social Research (NIESR) in the UK and jointly maintained by the Institute and the London Business School.

Barrel *et al.* (*op. cit.*:174) in deriving the time-varying rule for price expectations, noted that by definition, it did not follow from a tightly formulated theory. The selection was however also not completely *ad hoc* and variables were selected to capture important endogenous linkages operating in the model.

The dependent variable for the price expectations equation in each of the three countries is the change in the log of consumer price inflation one period ahead. The information set includes the change of the log of the home country inflation, lagged by one or two periods, a short-term interest rate lagged by one period, the change of the log of capacity utilisation, lagged by one period, and the change of the log of the relevant spot nominal exchange rate, with respect to the Deutschmark. The only exception is France, where German inflation was also included.

Barrel *et al.* (*op. cit.*:175) first report an OLS estimation of the expectations rule, proving that in all three cases price expectations are autoregressive with lagged home country inflation the most important variable. The other variables are for the most part insignificant but are retained for the reason that they might play an important role in capturing endogenous linkages in the model. The estimates of time-varying parameters reported demonstrate a reasonable degree of variation over the period. The hyperparameters associated with the autoregressive terms imply that learning will occur rapidly with respect to changes in these variables and, by contrast, learning with respect to the other variables will be slow.

Finally, Barrel *et al.* (*op. cit.*:183) report on the outcomes of a set of simulations where the learning mechanism is in operation. These are then compared with those under rational expectations and a fixed parameter adaptive expectations mechanism. The simulations are the realignment of the franc, lira and pound within the ERM, and an oil price shock. The authors conclude from the results of these simulations that filter-based learning models caused prices to rise more sharply than they do under model-consistent or strongly rational expectations. This suggests that policy analysis within models relying only on model-consistent expectations, can be seriously misleading.

3.7 CONCLUSION

In this chapter, the theoretical development on the formation of expectations by economic agents was surveyed, starting with the adaptive expectations hypothesis, rational expectations (also called model-consistent expectations), and, most recently, boundedly rational learning. Practical problems with the implementation of the adaptive expectations hypothesis and the rational expectations hypothesis in large-scale macro models have been highlighted. The process of learning has been proposed as an alternative where the possibility that economic agents can make consistent prediction errors, even in the very long run, is ruled out. Boundedly rational learning seems to be an appealing alternative; full information on the part of the economic agent is not required, but it is accepted that the intelligent agent, knowing the structure of the expectations model, will assimilate new information as time progresses and adjust the parameters of the model accordingly.

An exposition of the state-space representation of a system and the Kalman filter as time-varying parameter estimation technique were presented. A brief account of a number of international studies in this regard concludes this chapter. Chapter 5 describes the application of the Kalman filter to a price expectations rule for expectations formation by South African consumers, as well as the incorporation of the price expectations variable into a set of behavioural equations.

3.4.2 Estimation by the Kalman filter

Given observations (y_t,x_t) for $t = 1,2,\ldots,T$, one of the ultimate objectives is to estimate the unknown parameters in the system based on these observations. The parameters to estimate would include the matrices A, H, F, R, and Q, and the state vector ξ_t of which inferences need to be made. The Kalman filter is a recursive algorithm for sequentially updating the state vector given past information. More technically, it is an algorithm for calculating linear least squares forecasts of the state vector on the basis of data observed up to date t,

$$\hat{\xi}_{t+1|t} \equiv \hat{E}(\xi_{t+1} \mid \vartheta_t),$$

where

$$\vartheta_t \equiv (y'_t, y'_{t-1}, \ldots, y'_1, x'_t, x'_{t-1}, \ldots, x'_1)' \tag{3.14}$$

and $\hat{E}(\xi_{t+1} \mid \vartheta_t)$ denotes the linear projections of ξ_{t+1} on ϑ_t and a constant. The Kalman filter calculates these forecasts recursively, generating $\xi_{1|0}, \xi_{2|1}, \ldots, \xi_{T|T-1}$ in succession. Associated with each of these forecasts is a mean squared error (MSE) matrix, represented by the following $(r \times r)$ matrix:

$$P_{t+1|t} = E[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})']. \tag{3.15}$$

The recursion would begin with $\hat{\xi}_{1|0}$, which denotes a forecast of ξ_1 based on no observations of y or x. This is just the unconditional mean of ξ_1,

$$\hat{\xi}_{1|0} = E(\xi_1),$$

with associated MSE (variance of ξ_1),

$$P_{1|0} = E\{[\xi_1 - E(\xi_1)][\xi_1 - E(\xi_1)]'\}. \tag{3.16}$$

Then, iteration on the following set of equations takes place:

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1}) \tag{3.17}$$

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t} \tag{3.18}$$

and

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1} \tag{3.19}$$

$$P_{t+1|t} = FP_{t|t}F' + Q \qquad \text{for } t=1,2,\ldots,T. \tag{3.20}$$

Substituting (3.17) into (3.18),

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1}) . \tag{3.21}$$

The coefficient matrix in (3.21) is known as the *gain matrix* and is denoted K_t:

$$K_t \equiv FP_{t|t-1} + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1} . \tag{3.22}$$

For a derivation of these equations, see Hamilton (1994:379-380). Equations (3.17) and (3.18) may be called the updating equations and the value of $\hat{\xi}_{t+1|t}$ denotes the best forecast of ξ_{t+1} based on a constant and a linear trend function of $(y_t, y_{t-1}, \ldots, y_1, x_t, x_{t-2}, \ldots x_1)$. The matrix $P_{t|t+1}$ gives the MSE of this forecast.

The forecast of y_{t+1} is then given by

$$\hat{y}_{t+1|t} \equiv \hat{E}(y_{t+1|t} \mid x_{t+1|t}, \vartheta_t) = A'x_{t+1} + H'\hat{\xi}_{t+1|t} \tag{3.25}$$

with the associated MSE

$$E[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})'] = H'P_{t+1|t}H + R. \tag{3.26}$$

What still remains outstanding in the exposition above, is the estimation of the unknown parameter matrices F, Q, A, H and R, which are determined through maximum likelihood estimation of these parameters. The Kalman filter was motivated in the above discussion in terms of linear projections. The forecasts $\hat{\xi}_{t|t-1}$ and $\hat{y}_{t|t-1}$ are thus optimal within the set of forecasts that are linear in (x_t, ϑ_{t-1}), where $\vartheta_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, x'_{t-1}, x'_{t-2}, \ldots, x'_1)'$.

If the initial state ξ_1 and the innovations $\{w_t, v_t\}_{t-1}^{T}$ are multivariate Gaussian[3], then the stronger claim can be made that the forecasts $\hat{\xi}_{t|t-1}$ and $\hat{y}_{t|t-1}$ calculated by the Kalman filter are optimal among any functions of (x_t, ϑ_{t-1}). Moreover, if ξ_1 and $\{w_t, v_t\}_{t-1}^{T}$ are

[3] The vector $Y = (Y_1, Y_2, \ldots, Y_n)'$ has a multivariate Gaussian or multivariate normal distribution if $f_Y(y) = (2\pi)^{-n/2}|\Omega|^{-1/2}\exp[(-\frac{1}{2})(y-\mu)'\Omega^{-1}(y-\mu)]$ and is written as $Y_t \sim N(\mu,\Omega)$ with the mean μ an $(n\times 1)$ vector and the variance-covariance a matrix Ω of dimension $(n\times n)$ which is symmetric and positive semidefinite.

Gaussian, then the *distribution* of y_t conditional on (x_t, ϑ_{t-1}) is Gaussian with mean given by (3.23) and variance given by (3.24):

$$y_t \mid x_t, \vartheta_{t-1} \sim N((A'x_t + H'\hat{\xi}_{t|t-1}), (H'P_{t|t-1}H + R));$$

that is

$$\begin{aligned} & f_{y_t|x_t,\vartheta_{t-1}}(y_t \mid x_t, \vartheta_{t-1}) \\ & \quad = (2\pi)^{-n/2} \mid H'P_{t|t-1}H + R \mid^{-1/2} \\ & \qquad \times \exp\{-\tfrac{1}{2}(y_t - A'x_t - H'\xi_{t|t-1})'(H'P_{t|t-1}H + R)^{-1} \\ & \qquad \times (y_t - A'x_t - H'\xi_{t|t-1})\} \qquad \text{for } t = 1, 2, \ldots, T. \end{aligned} \tag{3.25}$$

From the above equation, the sample log likelihood can be constructed, namely

$$\sum_{t=1}^{T} \log f_{y_t|x_t,\vartheta_{t-1}}(y_t \mid x_t, \vartheta_{t-1}). \tag{3.26}$$

Expression (3.26) can then be maximised numerically with respect to the unknown parameters in the matrices F, Q, A, H and R. Burmeister and Wall (1982) illustrate this application.

To summarise, the state vector ξ_t and its mean squared error P_t are recursively estimated by the set of equations (3.17) through (3.20), which also represents the key equations for the Kalman filter. $\xi_{s|v}$ is the forecast of the state vector at time period s, given information available at time v. Note that the recursion for P does not depend on the forecasted state vector, or on the observed data (y_t, x_t).

To implement the Kalman filter, the starting values must be specified and the unknown matrices must be replaced by their estimates. By default, a software package like EViews obtains starting values by treating these matrices as fixed coefficients and estimating them using OLS.

Once starting values are obtained, the parameters A, H, F, R and Q are estimated by maximising the log likelihood function under the assumption that the distribution of y_t, conditional on x_t and past values of (y_t, x_t), is multivariate normal (Gaussian).

3.4.3 State-space model with stochastically varying coefficients

In the above discussion, it was assumed that the matrices F, Q, A, H and R were all constant. The Kalman filter can also be adapted for more general state-space models in which the values of these matrices depend on the exogenous or lagged dependent variables included in the vector x_t.

Equations (3.9) and (3.10), i.e. the state-space representation may thus be altered to:

$$y_t = a(x_t) + [H(x_t)]'\xi_t + w_t \tag{3.27}$$

$$\xi_{t+1} = F(x_t)\xi_t + v_{t+1}. \tag{3.28}$$

Here $F(x_t)$ denotes an $(r \times r)$ matrix whose elements are functions of x_t; $a(x_t)$ similarly describes an $(n \times 1)$ vector-valued function and $H(x_t)$ an $(r \times n)$ matrix-valued function.

It is assumed that conditional on x_t and on the data observed through date t−1, denoted $\vartheta_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, x'_{t-1}, x'_{t-2}, \ldots, x'_1)'$, the vector $(v'_{t+1}, w'_t)'$ has the Gaussian distribution

$$\begin{bmatrix} v_{t+1} \\ w_t \end{bmatrix} \mid x_t, \vartheta_{t-1} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q(x_t) & 0 \\ 0 & R(x_t) \end{bmatrix}\right). \tag{3.29}$$

Equations (3.27) to (3.29) generalise the earlier framework by allowing for stochastically varying parameters, but it is more restrictive in the sense that a Gaussian distribution is assumed in (3.29). Hamilton (1994:399-400) explains the role of the Gaussian requirement, as set out below.

Suppose it is taken as given that $\xi_t \mid \vartheta_{t-1} \sim N(\hat{\xi}_{t|t-1}, P_{t|t-1})$. Assuming as before that x_t contains only strictly exogenous variables or lagged values of y, this also describes the distribution of $\xi_t \mid x_t, \vartheta_{t-1}$. It follows from the assumptions in (3.27) to (3.29) that

$$\begin{bmatrix} \xi_t \\ y_t \end{bmatrix} \mid x_t, \vartheta_{t-1} \sim N\left(\begin{bmatrix} \hat{\xi}_{t|t-1} \\ a(x_t) + [H(x_t)]'\hat{\xi}_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}H(x_t) \\ H'(x_t)P_{t|t-1} & [H(x_t)]'P_{t|t-1}H(x_t) + R(x_t) \end{bmatrix}\right). \tag{3.30}$$

Conditional on x_t, the terms $a(x_t)$, $H(x_t)$ and $R(x_t)$ can all be treated as deterministic. Thus, the formula for the conditional distribution of Gaussian vectors[4] can be used to deduce that

$$\xi_t \mid y_t, x_t, \vartheta_{t-1} \equiv \xi_t \mid \vartheta_t \sim N(\hat{\xi}_{t|t}, P_{t|t}). \tag{3.31}$$

where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + \left\{P_{t|t-1}H(x_t)\left[[H(x_t)]'P_{t|t-1}H(x_t) + R(x_t)\right]^{-1} \times \left[y_t - a(x_t) - [H(x_t)]'\xi_{t|t-1}\right]\right\} \tag{3.32}$$

and

$$P_{t|t} = P_{t|t-1} - \left\{P_{t|t-1}H(x_t) \times \left[[H(x_t)]'P_{t|t-1}H(x_t) + R(x_t)\right]^{-1}[H(x_t)]'P_{t|t-1}\right\}. \tag{3.33}$$

It then follows from (3.27) and (3.29) that $\xi_{t+1} \mid \vartheta_t \sim N(\xi_{t+1|t}, P_{t+1|t})$, where

$$\hat{\xi}_{t+1|t} = F(x_t)\hat{\xi}_{t|t} \tag{3.34}$$

and

$$P_{t+1|t} = F(x_t)P_{t|t}[F(x_t)]' + Q(x_t). \tag{3.35}$$

Equations (3.32) through (3.35) are the Kalman filter equations (3.17) to (3.20), with the parameter matrices replaced with their time-varying analogs. Thus, as long as the initial state ξ_1 is treated as $N(\hat{\xi}_{1|0}, P_{1|0})$, the Kalman filter iterations proceed as before. The obvious generalisation of (3.25) may still be used to calculate the likelihood function.

The only difference from the constant-parameter case is that the inference (3.32) is a non-linear function of x_t. This means that although (3.32) gives the optimal inference if the disturbances and initial state are Gaussian, it cannot be interpreted as the linear projection of ξ_t on ϑ_t, with non-Gaussian disturbances.

3.4.4 The Kalman filter in terms of conventional least squares procedures

The exposition above is mainly in applied statistical terms. Cuthbertson (1988:234-241) presents the econometrics of the Kalman filter using conventional least squares procedures. He begins by reinterpreting the familiar problem of estimation subject to linear restrictions

[4] The conditional density of ξ_t given y_t, is found by dividing the joint density by the marginal density (Hamilton 1994:101-102, 399).

in terms of a Kalman filter updating equation, utilising the Kalman gain. This is followed by a reinterpretation of the Theil-Goldberger pure and mixed estimator as a 'one-shot' Kalman filter updating equation combining *a priori* and sample information. In both of the above cases, Cuthbertson reinterprets the standard formulae for the covariance matrix of the parameters as Kalman filter variance updating equations. He also concludes by presenting the relatively complex case of the state-space form with variable parameters.

3.5 IMPLEMENTATION OF THE LEARNING PROCESS IN A MACROECONOMIC MODEL

The learning process is generally implemented in model context as follows. The model essentially contains three blocks of equations, namely:

$$y_{it} = f_i(Y_t, X, {}_tY^e_{t+1}) \qquad i = 1\ldots n, \qquad t = 1\ldots T \tag{3.36}$$

$${}_ty^e_{t+1} = g_k(Y_t, X, \xi_t) \qquad k = 1\ldots m \tag{3.37}$$

$$\xi_t = \xi_{t-1} + v_t \qquad v_t \sim N(0,Q). \tag{3.38}$$

Equation (3.37) represents the measurement (or observation) equation(s) of the state-space model in Kalman filter terms, while equation (3.38) represents the state (or transition) equation(s).

Assuming that the value of ξ_{t-1} is known, the last block of equations for the expected value of ξ_t can be solved, which is simply the Kalman filter prediction equations for ξ, called the state vector. Given ξ, the second block of equations for the expected value of ${}_tY^e_{t+1}$ can be solved, and given this, the first block of equations for Y_t can finally be solved.

Q, the covariance matrix of the errors in the equations governing the evolution of the parameters (or in Kalman filter terms, the state equation error terms) is given by the original estimation and an estimation for P_{t-1} can be obtained (the uncertainty of the parameters or state variables). The Kalman filter prediction equations for P can then be used to derive an estimate of ${}_{t-1}P_t$. Having solved the complete model for Y_t, we can determine the one-step-ahead prediction error, that is the error that occurs between the expectation of the vector Y_t derived from the learning model and the model's final solution for Y_t. The one-step-ahead prediction error is a combination of stochastic error terms of the measurement and state equations. Given this and the estimate of $P_{t|t-1}$, the Kalman filter updating equations can be used to derive revised estimates of P_t and ξ_t. The updating is done on the basis of the observed errors between the whole model solution and the original expectations model forecast.

The process is then repeated for the next period, starting from the new updated estimates of ξ_t to predict ξ_{t+1}, and so on. In this way, the learning model will adjust its own parameters to cope with any change in structure or regime of the whole model. In the forecasting period, the final values of the state vector are used, in addition to other exogenous input, to solve the model for Y_t.

The underlying assumptions of this process are still quite strong, as agents are still assumed to process all available information in an optimal fashion and a substantial degree of sophistication on the part of the economic agent is still assumed. The learning model may, however, fulfil the criteria for weak rational expectations, since agents are not assumed to have full information. They will most likely make mistakes in the short run, but systematic errors over an extended period of time are ruled out.

Considering the model's response to any regime change reveals that the parameters of the learning rule respond over time to that change. The behaviour of the parameters of the rule, provides an important insight into the form of equilibria which may emerge from the system. Marcet and Sargent (1988) summarise the main theoretical results. The concepts of learning is characterised as the process of changing the parameters of the rule. If these parameters settle down to some fixed level, learning may be regarded as having ceased and this is sometimes called an expectational equilibrium (or E-equilibrium). Marcet and Sargent demonstrate that this stable condition is also a full rational expectations equilibrium.

3.6 EMPIRICAL LITERATURE ON THE LEARNING APPROACH

Although the theoretical literature on learning has grown from the early work of Friedman (1975 and 1979), the empirical literature has remained relatively sparse. Two empirical implementations of the learning approach in large-scale macro models will be presented as representative examples of the application of the Kalman filter to an expectations rule.

The learning approach to the treatment of expectations was first adopted in the exchange rate sector of the London Business School (LBS) model of the UK economy. Hall and Garratt (1992b) present an empirical model for the Sterling/Deutchmark real exchange rate assigning an important role to exchange rate expectations, which are assumed to be formed through a Kalman filter-based learning process.

The structural form of the exchange rate equation, derived from a capital stock model with government intervention is:

$$E_t = \alpha_0 + \alpha_1 E^e_{t+1} + \alpha_2 E_{t-1} + \alpha_3 r_t + \alpha_4 r_{t-1} + \alpha_5 T_t + \alpha_6 T_{t-1} \tag{3.39}$$

with

E_t = log of the real Sterling/Deutchmark exchange rate

r_t = real interest rate differential between UK and German short-term rates and

T_t = log of the ratio of exports to imports which is a measure of the real trade balance.

The end result of a search is reported as a restricted form of the above equation, estimated by three stage least squares, namely

$$E_t = 0.0329 + 0.675 E^e_{t+1} + 0.299 E_{t-1} + 0.352 r_t \tag{3.40}$$

which, in the long-run, exactly equals uncovered interest rate parity (UIP).

The time-varying rule for exchange rate expectations is derived from the structural equation (*op. cit.*:10) by rearranging equation (3.39) to give:

$$E^e_{t+1} = \beta_0 + \beta_1 E_t + \beta_2 E_{t-1} + \beta_3 r_t + \beta_4 r_{t-1} + \beta_5 T_t + \beta_6 T_{t-1}\,. \tag{3.41}$$

Lagging this equation by one period and using it to substitute out the term in E_t, after collecting terms, gives:

$$\begin{aligned} E^e_{t+1} &= \beta_0 + (\beta_1^2 + \beta_2)E_{t-1} + \beta_1\beta_2 E_{t-2} + (\beta_1\beta_3 + \beta_4)r_{t-1} \\ &+ \beta_1\beta_4 r_{t-2} + \beta_3 r_t + (\beta_1\beta_5 + \beta_6)T_{t-1} + \beta_1\beta_6 T_{t-2} + \beta_5 T_t. \end{aligned} \quad (3.42)$$

The β coefficients are combinations of the α coefficients, and the contemporaneous terms in the above equation are replaced by assuming partial reduced form equations for r_t and T_t, respectively, as follows:

$$r_t = C_1(L)r_{t-1} + C_2(L)GDP_{t-1} + C_3(L)INF_{t-1} \quad (3.43)$$

$$T_t = D_1(L)T_{t-1} + D_2(L)GDP_{t-1} + D_3(L)PO_{t-1} + D_4(L)E_{t-1} \quad (3.44)$$

where PO is the log of the oil price and INF is the rate of inflation which could include both UK and German inflation or the differential and GDP is the log of real GDP. C_i and D_i are polynomial lag operators.

The terms r_t and T_t my thus be eliminated and collection of terms yields:

$$\begin{aligned} E^e_{t+1} &= \beta_0 + (\beta_1^2 + \beta_2 + \beta_5 D_4(L))E_{t-1} + \beta_1\beta_2 E_{t-2} + (\beta_1\beta_2 + \beta_4 + \beta_3 C_1(L))r_{t-1} \\ &+ \beta_1\beta_4 r_{t-2} + (\beta_1\beta_5 + \beta_6 + \beta_5 D_1(L))T_{t-1} + \beta_1\beta_6 T_{t-2} + \beta_3 C_2(L)GDP_{t-1} \\ &+ C_3(L)INF_{t-1} + \beta_5(D_2(L)GDP_{T-1} + D_3(L)PO_{t-1}). \end{aligned}$$

(3.45)

Equation (3.45) is regarded as the basic partial reduced form rule which agents use to form their expectations. The above was simplified by dropping any lagged terms greater than t-3 by introducing a stochastic constant (*op. cit.*:11).

The basic structure was then used in a specification search to produce the following equation for expectations formation:

$$E^e_{t+1} = \gamma_{0t} + \gamma_{1t}E_{t-1} + \gamma_{2t}r_{t-1} + \gamma_{3t}INF^{UK}_{t-1} + \gamma_{4t}T^{UK}_{t-1} + \gamma_{5t}PO_{t-1}. \quad (3.46)$$

Note that all lagged information is dated t–1 when the equation is used to forecast E_{t+1}, i.e. the information set does not include current information. The parameters γ are restricted forms of the β coefficients. Hall and Garratt (*op. cit.*:11) note that their derivation of the expectations equation is by no means exclusive. In principle, formation of expectations could be the result of information from anywhere in the model, provided that it is relevant.

The time-varying parameters are then assumed to be generated by the following process:

$$\gamma_{it} = \gamma_{it-1} + \varepsilon_{it} \ . \tag{3.47}$$

Equation (3.46) may be regarded as the standard measurement equation and the set of equations (3.47) as the state equations. The model has thus been formulated in state-space form and the Kalman filter can be applied to estimate the time-varying parameters conditional on the variance of the error terms of (3.46) and the covariance matrix of (3.47), which is assumed to be diagonal. The estimation reported (*op. cit.*:14-18) suggests that the most important determinants of the forward exchange rate are real interest rate differentials and the previous period's exchange rate. All parameter values displayed a reasonable degree of time variance, with the main movements in the coefficients on the real interest rate differential and the ratio of exports to imports for the UK, reflecting a faster rate of learning on these variables.

Another example where the Kalman filter was employed to estimate the time-varying parameter rule of expectations is a study conducted by Barrel *et al.* (1994:173). In this study, the boundedly rational learning approach was applied to wage behaviour in three countries, namely the UK, France and Italy. A time-varying parameter model for forecasting prices was first estimated. This defined the information set of the economic agents. The expectations rule together with the structural equations in which the expectations are embedded were then incorporated into the global econometric model (GEM), developed by the National Institute of Economic and Social Research (NIESR) in the UK and jointly maintained by the Institute and the London Business School.

Barrel *et al.* (*op. cit.*:174) in deriving the time-varying rule for price expectations, noted that by definition, it did not follow from a tightly formulated theory. The selection was however also not completely *ad hoc* and variables were selected to capture important endogenous linkages operating in the model.

The dependent variable for the price expectations equation in each of the three countries is the change in the log of consumer price inflation one period ahead. The information set includes the change of the log of the home country inflation, lagged by one or two periods, a short-term interest rate lagged by one period, the change of the log of capacity utilisation, lagged by one period, and the change of the log of the relevant spot nominal exchange rate, with respect to the Deutschmark. The only exception is France, where German inflation was also included.

Barrel *et al.* (*op. cit.*:175) first report an OLS estimation of the expectations rule, proving that in all three cases price expectations are autoregressive with lagged home country inflation the most important variable. The other variables are for the most part insignificant but are retained for the reason that they might play an important role in capturing endogenous linkages in the model. The estimates of time-varying parameters reported demonstrate a reasonable degree of variation over the period. The hyperparameters associated with the autoregressive terms imply that learning will occur rapidly with respect to changes in these variables and, by contrast, learning with respect to the other variables will be slow.

Finally, Barrel *et al.* (*op. cit.*:183) report on the outcomes of a set of simulations where the learning mechanism is in operation. These are then compared with those under rational expectations and a fixed parameter adaptive expectations mechanism. The simulations are the realignment of the franc, lira and pound within the ERM, and an oil price shock. The authors conclude from the results of these simulations that filter-based learning models caused prices to rise more sharply than they do under model-consistent or strongly rational expectations. This suggests that policy analysis within models relying only on model-consistent expectations, can be seriously misleading.

3.7 CONCLUSION

In this chapter, the theoretical development on the formation of expectations by economic agents was surveyed, starting with the adaptive expectations hypothesis, rational expectations (also called model-consistent expectations), and, most recently, boundedly rational learning. Practical problems with the implementation of the adaptive expectations hypothesis and the rational expectations hypothesis in large-scale macro models have been highlighted. The process of learning has been proposed as an alternative where the possibility that economic agents can make consistent prediction errors, even in the very long run, is ruled out. Boundedly rational learning seems to be an appealing alternative; full information on the part of the economic agent is not required, but it is accepted that the intelligent agent, knowing the structure of the expectations model, will assimilate new information as time progresses and adjust the parameters of the model accordingly.

An exposition of the state-space representation of a system and the Kalman filter as time-varying parameter estimation technique were presented. A brief account of a number of international studies in this regard concludes this chapter. Chapter 5 describes the application of the Kalman filter to a price expectations rule for expectations formation by South African consumers, as well as the incorporation of the price expectations variable into a set of behavioural equations.