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 it varies through time. It is clear that the unobservable components model and time-varying parameter models are 'non-standard', that is, one cannot apply 'least squares' procedures (e.g. OLS, IV, GLS) *directly* to the above equations. However, each of the above models can be 'rewritten' in terms of two distinct types of equation (called the *measurement* and *transition* equations) which together are called *the state-space form*. The Kalman filter can be applied to the state-space form equations to yield a set of *recursive equations*; the latter are then used to generate a series for $\tilde{v}_t$ and its variance which will contain unknown parameters to be estimated. At this point, the Kalman filter recursive equations have completed the required task. Now standard maximum likelihood procedures are used to estimate the unknown parameters.

The Kalman filter is a rather versatile construct, but its derivation may involve unfamiliar concepts to some readers. Therefore we approach the issues involved from different standpoints. In section 7.1 we motivate our discussion of the state-space form and the Kalman filter recursive algorithms in terms of the modelling of expectations in a learning environment. The technical aspects are introduced via the Theil–Goldberger (1961) pure and mixed estimator in section 7.2 and the 'full' Kalman filter equations are then examined from the standpoint of Bayes theorem. Bayes theorem allows one to combine prior information with the data to yield an optimal posterior estimator. In fact, the Theil–Goldberger estimator is a special case of the 'general' updating formulae of Bayes theorem. In section 7.3 we examine how the unknown parameters in the Kalman-filter formulae can be estimated using maximum likelihood and in particular we consider time-varying parameter models. In section 7.4 we give some practical examples of the use of the Kalman filter in applied economics (see Note 1).

7.1 Expectations and learning and the state-space form

In this section we wish to motivate our discussion of the Kalman filter by interpreting it in terms of agents forming expectations. The predominant paradigm for modelling expectations is the rational expectations hypothesis REH, see Chapter 6. It assumes agents act 'as if' they know the true model of the economy (up to a set of white noise errors); a rather strong assumption that some may find a little implausible. As Friedman (1979) clearly points out, the *information exploitation* assumption of RE, namely that agents use efficiently

whatever information is available, is largely uncontentious. It is the *information availability* assumption that many economists find objectionable. For agents that are (Muth, 1960) rational their predictions are equal to the conditional mathematical expectation of the model and hence their forecast errors are independent of any information available at the time the forecast is made; the latter is the error-orthogonality property of RE. In early New Classical models (e.g. Sargent and Wallace 1975) if the parameters of the model alter (usually taken to be the parameters of the monetary policy reaction function), agents are assumed to know the 'new' parameter values immediately. The latter also applies to changes in the functional form of behavioural equations. In later work (e.g. Cyert and De Groot 1974, Bray 1982), agents are assumed to know the true structure of the model but are initially ignorant of the true values of the one (or more) of the parameters.

The results of these studies broadly suggest that agents expectations do eventually converge on the Muth rational solution. However, where agents operate with a set of possible models, then even if one of them is the true model, there is no guarantee that the learning process converges to the true model (Blume and Easley 1982). One would expect this conclusion to apply *a fortiori* when agents are not allowed the luxury of having the true model in their set of models, or where the parameters of the model vary over time. In this type of situation we have the added complication that agents, during their learning process, generate outcomes which are contaminated with 'noise' from the learning process as well as 'noise' from the underlying true model.

Consideration of the information availability assumption has led critics of the REH to label it unrealistic. However, such critics have not been able to provide an alternative 'optimising' framework to RE, particularly one that is empirically tractable. Friedman (1979) goes some way along this route when he advocates that given the true model $y_t = x_t\beta + u_t$ (u_t, white noise), agents may sequentially update their estimate of the *fixed* true parameter vector β as more information on (y_t, x_t) becomes available (e.g. time-varying parameter models). Using the Kalman filter we extend 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 alludes to the latter outcome (Friedman 1979, pp. 33–4) when he discusses the possibility that agents may perceive that a good approximation to the complex 'true' model may be a simple (linear) model but with time-varying parameters. Such a model may be analysed using the Kalman filter (Kalman 1960), and the familiar recursive least squares learning procedure is a special case of this more general procedure.

The Kalman filter, although widely used in certain branches of the engineering literature and by applied statisticians, is only just emerging as a possible useful tool of the applied economist (see, for example, Lawson 1980, Harvey *et al*. 1986, Cuthbertson and Taylor 1986). For certain models the Kalman filter may be viewed as mimicking a learning process by agents. For example, the Kalman filter may be interpreted as a form of adaptive expectations where the adjustment parameter is updated each period, based on new information. This formalises Flemming's (1976) idea of a 'change of gear' when forming expectations. It is well known that adaptive expectations is optimal (in the sense of producing unbiased forecasts) only when the data generation process is IMA(1, 1) or ARIMA(1, 1, 1). The Kalman filter, however, is optimal under more general conditions, and in fact produces minimum mean square estimators (MMSE) under normality. Therefore agents, *given the assumed information set*, do not make systematic forecast errors. The Kalman filter therefore confronts directly the question of how agents learn about the time series behaviour of economic variables; agents are not assumed to know instantaneously the 'true' model but they do use information optimally (or efficiently). The Kalman filter can also be applied to unobservable components models and it therefore formalises the 'signal extraction problem' presented in Lucas's (1972) derivation of the 'surprise supply function'. Note that the Kalman filter does not provide a panacea, it provides merely a tractable alternative to the 'extreme' information assumption of the REH, based on optimising behaviour in the face of uncertainty about the evolution of the parameters of the assumed model.

The rest of this chapter is organised as follows. We begin with a discussion of fixed coefficient adaptive expectations models that proved so popular in the empirical literature prior to the advent of rational expectations in the 1970s. We then present a simple adaptive expectations model in which the adaptive coefficient varies through time. Our final example utilises the signal extraction problem that an individual faces when trying to estimate his permanent income say, faced only with information on his measured income. This *stochastic trend model* embodies sequential learning in a time series context and allows us to demonstrate how this model is represented in *state-space* form: a prerequisite for understanding the application of the Kalman filter in more general situations.

Fixed coefficient-adaptive models

It is now well understood that if (the logarithm of) measured income y_t is accurately represented by an IMA(1, 1) process

$$y_t = y_{t-1} + \varepsilon_t - (1 - \Theta)\varepsilon_{t-1} \qquad (7.1)$$

then the optimal updating equation for expected income,

$$y^e_{t/t-1} = E(y_t|\Omega_{t-1}), \text{ where } \Omega_{t-1} \equiv \{y_{t-j}, \varepsilon_{t-j}\}_{j=1\ldots\infty}$$

is (see Note 2):

$$y^e_{t/t-1} - y^e_{t-1/t-2} = \Theta(y_{t-1} - y^e_{t-1/t-2}) \qquad (7.2)$$

This is nothing more than first-order adaptive expectations with the *fixed* updating coefficient related to the moving average parameter in the data generation process. The above approach is easily generalised to include a 'change-of-gear' (Flemming 1976). If the *growth* rather than the level of income is IMA(1, 1) then first-order adaptive expectations applied to the growth in income is optimal. 'Optimal' in this context is taken to mean that expectations are correct on average (and have minimum mean square prediction errors). Although Granger (1966) finds that a number of economic time series are adequately represented as IMA(1, 1) processes and therefore fixed coefficient adaptive expectations are optimal, nevertheless the model does not allow agents to learn slowly about their new environment as new information becomes available. For these adaptive models to be optimal, when the data generation process undergoes a 'change of gear', agents must instantaneously acquire knowledge of the 'new' moving average coefficient. Thus ironically, the above adaptive expectations model also requires a rather extreme information availability assumption when the stochastic behaviour of a variable alters.

Variable parameter adaptive expectations

Consider an agent who has sequential observations on his measured income (in logarithms) y_t which he views as consisting of an *unobserved* permanent component π_t and a zero mean (unobserved) 'surprise' element s_t. The agent has an initial or prior estimate of permanent income π_0 and wishes to update this estimate as information on measured income becomes available. Clearly to 'solve' this problem the agent must have some view (or model) of how permanent income varies over time. For expositional reasons we assume the *transition equation* describing the evolution of π_t is a random walk. Our final assumption is that the agent *perceives* that a fraction k_t of the surprise s_t in measured income, constitutes permanent income and $(1 - k_t)s_t$ is considered to be an addition to transitory income. Note that the coefficient k_t varies through time and for the moment we assume the value of k_t in each successive period is known by the agent. (The Kalman filter provides a method of estimating and optimally updating k_t as we see in section 7.2). The model assumed by the agent is therefore (see Note 3):

$$y_t = \pi_t + (1 - k_t)s_t \quad \text{'measurement equation'} \tag{7.3}$$

$$\pi_t = \pi_{t-1} + k_t s_t \quad \text{'transition equation'} \tag{7.4}$$

with

$$E_{t-1}s_t = s^e_{t/t-1} = 0; \quad \text{and} \quad E(\pi_t s_{t-j}) = 0 \quad (j = 0, \infty)$$

The measurement equation has measured income y_t as the sum of permanent π_t and transitory income $(1 - k_t)s_t$, while the transition equation represents the assumed evolution of π_t through time.

Substituting (7.4) in (7.3):

$$y_t = \pi_{t-1} + s_t \tag{7.5}$$

Multiplying (7.5) by k_t and substituting from (7.4) for $k_t s_t$ we obtain the updating equation for π_t in the form of a variable parameter adaptive model:

$$\pi_t = \pi_{t-1} + k_t(y_t - \pi_{t-1}) \tag{7.6}$$

Thus given an initial estimate of permanent income π_0, knowing k_t and y_t, the updating equation (7.6) can be used to give all future values of permanent income. The analogy with the fixed parameter adaptive model is completed by noting that the equations (7.3) and (7.4) may be written as an IMA(1, 1) model with a *time-varying moving average coefficient*. Equation (7.3) minus itself lagged one period yields:

$$\Delta y_t = \Delta\pi_t + (1 - k_t)s_t - (1 - k_{t-1})s_{t-1} \tag{7.7a}$$

Substituting for $\Delta\pi_t$ from (7.4) we obtain our IMA(1, 1) representation:

$$\Delta y_t = s_t - (1 - k_{t-1})s_{t-1} \tag{7.7b}$$

In using the updating equation (7.6) for π_t the key missing element is how the agent forms and updates the coefficient k_t which turns out to be analogous to the 'Kalman gain'. To demonstrate some preliminary

intuitive insights into how agents estimate the Kalman gain we consider the example of the generalised stochastic trend (GST) model.

Generalised stochastic trend model

Instead of assuming that the agent knows k_t, the proportion of any surprise s_t that accrues as permanent income, we adopt the weaker assumption that the shocks to permanent and measured income are statistically independent. In addition we assume that the growth in permanent income $\Delta\pi_t$ is time varying with parameter γ_{t-1} which itself evolves as a random walk (Harvey and Todd 1983). Hence the agents best approximation to his stochastic environment is assumed to be characterised as:

$$y_t = \pi_t + \varepsilon_t \tag{7.8a}$$

$$\pi_t = \pi_{t-1} + \gamma_{t-1} + \zeta_t \tag{7.8b}$$

$$\gamma_t = \gamma_{t-1} + \omega_t \tag{7.8c}$$

which may be represented in matrix form (known as the *state-space form*) as:

$$y_t = x'\beta_t + \varepsilon_t \quad \text{Measurement equation } (t = 1, 2, \dots n) \tag{7.9a}$$

$$\beta_t = T\beta_{t-1} + \eta_t \quad \text{Transition equation} \tag{7.9b}$$

where

$$x' = [1, 0]$$
$$\beta_t = (\pi_t, \gamma_t)'$$
$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$
$$\eta_t = (\zeta_t, \omega_t)'$$

ε_t, ζ_t, ω_t are zero mean, error terms independent of each other and

$$\mathrm{Var}(\varepsilon_t) = \sigma^2_\varepsilon; \ \mathrm{Var}(\zeta_t) = \sigma^2_\zeta; \ \mathrm{Var}(\omega_t) = \sigma^2_\omega$$

In the measurement equation, observed data on income y_t again consists of a permanent π_t and transitory component ε_t. The growth in permanent income $\Delta\pi_t$ is assumed to equal a stochastic growth coefficient γ_{t-1} (plus a random error term, ζ_t) and γ_{t-1} itself evolves as a random walk. The system (7.9a) and (7.9b) may appear a little strange to applied economists used to dealing with the usual fixed regression parameter model. In this unobservable components model *only* y_t is observed and the agent faces a 'signal extraction problem' to determine how much of any change in y_t can be attributed to a change in permanent income (the 'signal') and how much is due to transitory income ε_t (the 'noise'), (Lucas, 1972). However this particular unobservable components model may be given an intuitive interpretation in which y_t and π_t are perceived by the agent as being generated by a stochastic trend. This is easily seen by noting that given an initial estimate of π_0, successive substitution in (7.8b) yields

$$\pi_t = \pi_0 + \sum_{i=1}^{t} \gamma_{i-1} + \sum_{i=1}^{t} \zeta_i \tag{7.10}$$

and hence:

$$y_t = \left(\pi_0 + \sum_{i=1}^{t} \gamma_{i-1}\right) + u_t \tag{7.11}$$

where

$$u_t = \sum_{i=1}^{t} \zeta_i + \varepsilon_t = \zeta_t^* + \varepsilon_t$$

To see why (7.10) and (7.11) embody a stochastic trend, consider the special case where $\omega_t = 0$ (for all t). From (7.8c), $\gamma_t = \gamma_{t-1} = \gamma$ say, hence (7.10) and (7.11) reduce to:

$$y_t = \pi_0 + \gamma_t t + u_t \tag{7.12}$$

$$\pi_t = \pi_0 + \gamma_t t + \zeta_t^* \tag{7.13}$$

Equations (7.12) and (7.13) are global linear trend models with moving average error terms (see Note 4).

Returning to the signal extraction problem, assume for simplicity that $\omega_t = \sigma_\omega^2 = 0$ and that the agent knows the values of σ_ε^2 and σ_ζ^2. Assume also that with information on y up to period $t-1$ (which could be 'time zero') he has formed a prior estimate of the unobservable, permanent income for time t, namely $\pi_{t/t-1}$. The key question is how the agent optimally uses information to update his estimate of π when new information on y_t arrives. To gain some intuitive insights consider the two polar cases $\sigma_\varepsilon^2 = 0$ and $\sigma_\zeta^2 = 0$. In the first case there is no measurement noise ($y_t = \pi_t$) and we would expect all of his forecast error

$$\tilde{v}_t = (y_t - \tilde{y}_{t/t-1}) = (y_t - \pi_{t/t-1})$$

to be included in his estimate of permanent income, that is

$$\pi_t = \pi_{t/t-1} + (y_t - \tilde{y}_{t/t-1})$$

The converse applies for $\sigma_\zeta^2 = 0$, and here $\pi_t = \pi_{t/t-1}$. In the intermediate case ($\sigma_\varepsilon^2, \sigma_\zeta^2 \neq 0$) the proportion of the forecast error added to $\pi_{t/t-1}$ will depend upon the agents' perception of the *relative* variance of σ_ε^2 and $\mathrm{Var}(\pi_{t/t-1})$. The latter is equal to the sum of his prior estimate of the variance of π (say, σ_0^2) and his sampling error for π, (i.e. σ_ζ^2). Hence, if the updating equation is

$$\pi_t = \pi_{t/t-1} + k_t(y_t - \tilde{y}_{t/t-1}) \tag{7.14}$$

then we might expect

$$k_t = (\sigma_0^2 + \sigma_\zeta^2)/(\sigma_\varepsilon^2 + (\sigma_0^2 + \sigma_\zeta^2)) \tag{7.15}$$

It is easily seen that $k = 1$ for $\sigma_\varepsilon^2 = 0$ and $k = 0$ for $\sigma_\zeta^2 = \sigma_0^2 = 0$. Thus our intuitive arguments have led us to interpret our model both in terms of a stochastic trend and as a variable parameter partial adjustment model. The adjustment parameter k_t is known as the Kalman gain and equation (7.14) will be seen to be the updating equation for the 'unobservable' permanent income variable. Given an initial estimate $\pi_{1/0}$ and knowing k_t, equation (7.14) provides a recursion formula for updating π_t as new information on y_t arrives.

Having provided an intuitive interpretation of our unobserved components model we now turn to our main task which is to derive the general equations for the Kalman filter. These equations provide a general formula for the Kalman gain and updating equations for a wide variety of possible models.

7.2 The econometrics of the Kalman filter

The econometrics of the Kalman filter can appear rather formidable to the applied economist when reading the engineering or statistical literature. One of our aims in this section is therefore to present the econometrics of the Kalman filter using conventional procedures. We begin by deriving the formulae for one-step-ahead prediction errors in the general linear model. These results are then used to reinterpret the Theil–Goldberger 'pure and mixed' estimator in terms of a 'one-shot' Kalman filter. The prior 'guesses' for the parameters and error variances are combined with the sample data to yield an 'optimal' 'posterior' estimator based on both sets of information. We then use the stochastic trend model as a concrete example with which to develop the general formulae used in the Kalman filter. The Kalman filter is then seen to be a *useful algorithm* to generate the variables needed in the (prediction error decomposition of the) likelihood function: the key variables are the one-step-ahead prediction errors

$\tilde{v}_t$ and their variance-covariance matrix (F_t or f_t). We then present an alternative derivation of the Kalman filter equations in terms of Bayes theorem and maximum likelihood, which will reinforce the (somewhat difficult) concepts involved, when dealing with the general formulation of the state-space model.

Prediction in the general linear model

Given the true fixed parameter model

$$Y = X\beta + \varepsilon \tag{7.16}$$

where we assume a scalar covariance matrix:

$$\varepsilon \sim N(0, V) = N(0, \sigma^2 I) \tag{7.17}$$

and $E(X'\varepsilon) = 0$, X is $(n \times k)$; Y and ε are $(n \times 1)$; β is $(k \times 1)$.
The OLS estimator b_0 is BLUE:

$$b_0 = (X'X)^{-1}X'Y \tag{7.18}$$

with variance–covariance matrix:

$$\mathrm{Cov}(b_0) = P_0 = \sigma^2(X'X)^{-1} \tag{7.19}$$

and using (7.16) and (7.18) we obtain the familiar result

$$b_0 - \beta = (X'X)^{-1}X'\varepsilon \tag{7.20}$$

Of particular interest given what follows is the problem of *predicting q* 'new' observations Y_1 based on *new* information on X_1, where X_1 is $(q \times k)$, and the estimator b_0. We assume an *unchanged structural model* over the forecast horizon:

$$Y_1 = X_1\beta + \varepsilon_1 \tag{7.21}$$

$$\varepsilon_1 \sim N(0, V_1) = N(0\ \sigma^2 I_1) \tag{7.22}$$

where ε_1 is $(q \times 1)$ uncorrelated with ε. The prediction $\widetilde{Y}_1 = X_1 b_0$ is an unbiased predictor of the values of Y in the forecast period. The covariance matrix of the *one-step-ahead forecast errors* $\tilde{v}_1 = Y_1 - \widetilde{Y}_1$ is:

$$F = \mathrm{Cov}(\tilde{v}_1) = E(X_1(\beta - b_0) + \varepsilon_1)(X_1(\beta - b_0) + \varepsilon_1)' \tag{7.23}$$

where F is $(q \times q)$. Substitute from (7.20) for $(\beta - b_0)$:

$$F = E(X_1(\mathrm{Cov}\, b_0)X_1' + \varepsilon_1\varepsilon_1') = \sigma^2(X_1(X'X)^{-1}X_1' + I) \tag{7.24a}$$

$$\text{or } F = (X_1 P_0 X_1' + V_1) \tag{7.24b}$$

The variance of Y_1 around $\widetilde{Y}_1$ depends on the uncertainty in estimating the parameters in β ($\mathrm{Cov}(b_0) = P_0$) and also on the intrinsic uncertainty in equation (7.16), $V_1 = \sigma^2 I_1$.

If we have one additional observation on the x-variables, the X_1 is replaced by $x_1'(1 \times k)$ and Y_1, $\tilde{v}_1$ and F are scalars. Hence (7.24b) becomes:

$$f = [x_1' P_0 x_1 + \sigma^2] \tag{7.25}$$

which we will use in our discussion of the stochastic trend model later in this section.

Theil-Goldberger (T-G) estimation and the Kalman filter

The T-G pure and mixed estimator considers the problem of how best to combine prior information on the parameter vector β and information on β generated by our sample of observations. It is assumed that the agent (econometrician) makes an initial informed guess concerning the mean value of the true parameters β, denote this guess b_0^*. The *uncertainty* surrounding this prior 'guess' is summarised in the 'guess' about the prior covariance matrix, P_0^* Hence:

$$\beta = b_0^* + \omega_0^* \tag{7.26}$$

$$\omega_0^* \sim N(0, P_0^*) \tag{7.27}$$

where ω_0^* is a vector of 'prior' error terms and P_0^* is the (possibly non-scalar) non-diagonal prior covariance matrix. β is $(k \times 1)$, b_0^* is *known*, P_0^* is the *known* $(k \times k)$ covariance matrix (often assumed diagonal in practice (MacDonald 1988), or simplified in some way.

$$b_0^* \sim N(\beta, P_0^*) \tag{7.28}$$

Hence, the agent has both prior and sample information, the latter consists of Y which is $(n \times 1)$ and X which is $(n \times k)$, which may be represented:

$$\begin{pmatrix} Y \\ b_0^* \end{pmatrix} = \begin{pmatrix} X \\ I \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ \omega_0^* \end{pmatrix} \qquad \begin{matrix} \varepsilon \sim N(0, V) \\ \omega_0^* \sim N(0, P_0^*) \end{matrix} \tag{7.29}$$

$$\text{or } Y_* + X_*\beta + \varepsilon_* \tag{7.30}$$

where $\quad Y_* = \begin{pmatrix} Y \\ b_0^* \end{pmatrix}$

$$X_* = \begin{pmatrix} X \\ I \end{pmatrix}$$

$$V_* = E(\varepsilon_*\varepsilon_*') = \begin{pmatrix} E(\varepsilon\varepsilon') & 0 \\ 0 & E(\omega_0^*\omega_0^{*\prime}) \end{pmatrix} = \begin{pmatrix} V & 0 \\ 0 & P_0^* \end{pmatrix} \tag{7.31}$$

We have assumed zero covariance between the prior estimation error ω_0^* and the error term ε in the regression equation. (In addition we assume zero covariance between ω_0^* and X and ε and X.)

The Theil–Goldberger pure and mixed estimator may be viewed as 'one-shot' application of the Kalman filter which provides an updating equation for β and its covariance matrix based on the prior and sample information. The posterior estimates of β say b_1 is BLUE and its covariance matrix we denote $\mathrm{Cov}(b_1) = P_1$.

GLS applied to (7.30) yields:

$$b_1 = (X_*'V_*^{-1}X_*)^{-1}(X_*'V_*^{-1}Y_*) \tag{7.32}$$

with covariance matrix

$$\mathrm{Cov}(b_1) = P_1 = (X_*'V_*^{-1}X_*)^{-1} \tag{7.33}$$

Equations (7.32) and (7.33) may be interpreted as *updating equations* for β and its covariance matrix although this is not apparent from the normal textbook GLS formulae above. However, it is shown in the appendix that (7.32) and (7.33) can be rewritten in the intuitively appealing form:

$$P_1^{-1} = (P_0^*)^{-1} + (X'V^{-1}X) \tag{7.34a}$$

$$P_1 = (I - KX)P_0^* \tag{7.34b}$$

$$b_1 = b_0^* + K(Y - Xb_0^*) = b_0^* + K\tilde{v} \tag{7.35}$$

where $K = P_0^*X'F^{-1}$ **(7.36)**

$$F = \mathrm{Cov}(\tilde{v}) = (V + XP_0^*X') \tag{7.37}$$

$$\tilde{v} = Y - \tilde{Y} = (Y - Xb_0^*) \tag{7.38}$$

Equation (7.34a) is the updating equation for the inverse of the covariance of b_1. The $(k \times k)$ inverse of the posterior covariance, P_1^{-1}, is simply the sum of the inverse of the prior covariance $(P_0^*)^{-1}$ and the sample covariance (for the 'unrestricted' GLS estimator) that is $(X'V^{-1}X)$. Equation (7.3a) may be rewritten in terms of the Kalman gain and is given in (7.34b).

Equation (7.35), the updating equation for β, expresses the updated estimate b_1 as the sum of the prior estimate b_0^* and the product of the Kalman gain K and the error $\tilde{v}$ in forecasting Y using the prior estimate b_0^*. The Kalman gain depends upon 'relative variances' namely the variance of b_0^* $(= P_0^*)$ relative to the variance of the 'one-step-ahead' prediction error $\mathrm{Cov}(\tilde{v}) = F = (V + XP_0^*X')$.

We are now ready to present the complete state-space formulation which underlie most applications of the Kalman filter that are likely to be used by applied economists. The increased generality is provided by relaxing the assumption of fixed underlying true parameter vector β; the parameter vector is now assumed to vary through time but in a systematic way. It is this additional complexity that allows the Kalman filter to be used in estimating unobservable components and variable parameter models (which may be used to generate 'plausible' expectations variables without invoking the extreme RE, 'axiom of correct specification').

In the Theil–Goldberger model, β is non-stochastic and the estimator b_1 is 'optimal' where optimal is synonymous with BLUE. However, when β is stochastic in the sense that it is randomly drawn from a prior distribution before the observations on Y are generated, then b_1 retains its 'optimal' properties in that it is unbiased and is the minimum mean square estimator of β (given that Y is multi-variate normal). Thus with β stochastic, the Kalman filter will provide 'rational' predictions. The agent utilises information at time $t-1$ to provide unbiased estimators of β, which have minimum variance. As new information on Y at time t arrives, the agent combines this with his current priors to optimally update his estimate of both the parameter vector β and its covariance. It is in this sense that the Kalman filter may be viewed as mimicking a sequential optimal learning process. The predictions are 'rational' in the sense that the agent optimally exploits current and past information when learning about his stochastic environment.

State-space formulation and Kalman filtering

In developing the Kalman filter recursion formulae, conceptually, we move from our 'one-shot' Theil-Goldberger formulation to one where the estimates are updated each time period. Thus, in place of b_0, P_0 we have prior estimates $b_{1/0}$ and $P_{1/0}$ based on information at $t = 0$ and we use these to provide updated estimates b_1, P_1 as information on the scalar y_t for $t = 1$ becomes available. The recursion formulae then provide estimates $b_{t/t-1}$, $P_{t/t-1}$ $(t = 1, \ldots n)$. In order to update our estimate of the $(k \times 1)$ vector β in each period, we need to know the stochastic process by which β alters through time. This is given by the so called 'transition equation'. Our complete *state-space* model (see Note 5) is:

$$y_t = x'\beta_t + \varepsilon_t \quad (t = 1, 2, \ldots n) \quad \text{Measurement equation} \tag{7.39}$$

$$\beta_t = T\beta_{t-1} + R\eta_t \quad \text{Transition equation} \tag{7.40}$$

$$b_0 = \beta_0 + \psi_0 \quad \text{Prior estimate} \tag{7.41}$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

$$\eta_t \sim N(0, Q)$$

$$\psi_0 \sim N(0, \Psi_0)$$

where x' is $(1 \times k)$, β_t is $(k \times 1)$: T, Q, Ψ_0, R are $(k \times k)$ and we take $V = \sigma^2 I$.

We have already demonstrated how the stochastic trend model (7.8a–7.8c) may be represented in state-space form (with $R = I$). Equation (7.41) represents our initial guesses (or starting values) for the parameter vector β and its covariance matrix Ψ_0.

It is important to keep in mind what information the agent is assumed to possess. At $t = 0$, he has an initial *fixed* estimate b_0 of the true parameter vector β_0 and its covariance matrix, that is $b_0 \sim N(\beta_0, \Psi_0)$. He knows the structure of the model in the form of the *fixed* vector x, fixed matrices Q, R, T and the fixed variance of the measurement equation σ^2. The problem the agent faces is to utilise the information contained in the sequential data y_t to update optimally his estimates of β_t and its covariance matrix.

If we can reduce the three equation system (7.39–7.41) to the Theil–Goldberger formulation then we can apply the appropriate GLS formulae to produce optimal posterior (or one-step-ahead) estimates of β and its covariance matrix; these constitute the Kalman filter updating equations.

Given b_0 the unbiased predictor of β_1 is

$$b_{1/0} = Tb_0 \tag{7.42}$$

The covariance of b_1 around the true value β_1 is defined as:

$$\text{Cov}(b_{1/0}) = P_{1/0} = E(b_{1/0} - \beta_1)(b_{1/0} - \beta_1)' \tag{7.43}$$

Substituting for β_0 from (7.41) in (7.40) and using (7.42):

$$(\beta_1 - b_{1/0}) = -T\psi_0 + R\eta_1 = \omega_1, \text{ say} \tag{7.44}$$

The prediction error in forecasting β_1, namely $(b_{1/0} - \beta_1)$ is a weighted average of the 'prior uncertainty', ψ_0, and the uncertainty in the transition equation for β, namely, η_1. From (7.43) and (7.44) the covariance of this prediction error is the $(k \times k)$ matrix $P_{1/0}$:

$$P_{1/0} = E(\omega_1\omega_1') = (TP_0T' + RQR') \tag{7.45}$$

Equations (7.42) and (7.45) are the ***prediction equations*** for $t = 1$, for the state vector β_1 and its covariance matrix, which may be calculated ***without any reference to the observations*** y_t. Suppose the agent now receives a single observation y_1. The sample and prior information may now be arranged as in the Theil–Goldberger model:

$$\begin{pmatrix} y_1 \\ b_{1/0} \end{pmatrix} = \begin{pmatrix} x' \\ I \end{pmatrix} \beta_1 + \begin{pmatrix} \varepsilon_1 \\ \omega_1 \end{pmatrix} \tag{7.46}$$

where

$$\varepsilon_1 \sim N(0, \sigma^2)$$
$$\omega_1 \sim N(0, P_{1/0})$$

Comparing (7.46) with our Theil–Goldberger formulation (7.29) we have:

$$b_0^* = b_{1/0} \tag{7.47a}$$

$$P_0^* = P_{1/0} \tag{7.47b}$$

With the above substitutions, we can use the updating formulae (7.34) to (7.38) to calculate the $(k \times 1)$ vector for the Kalman gain for $t = 1$:

$$K_1 = P_{1/0}xF_1^{-1} = P_{1/0}xf_1^{-1} \tag{7.48}$$

where in this model F_1 is a scalar, denoted f_1:

$$F_1 = f_1 = (x'P_{1/0}x + \sigma^2) \tag{7.49}$$

The optimal updating equation for b_1 is:

$$b_1 = b_{1/0} + K_1(y_1 - x'b_{1/0}) \tag{7.50}$$

with $(k \times k)$ covariance matrix:

$$P_1 = (I - K_1x')P_{1/0} \tag{7.51}$$

The updated values b_1, P_1 are then used in equations (7.42) and (7.45) respectively, to generate new predictions $b_{2/1}$ and $P_{2/1}$. Estimates b_t and P_t are then updated sequentially using (7.48) and (7.51) as information on y_t becomes available. The Kalman filter also generates one-step-ahead prediction errors for y_t, that is, $\tilde{v}_t = y_t - y_{t/t-1}$ and their variance f_t (a matrix if we have a vector of observations on a *set* of variables at time t) which can be used directly in the prediction error decomposition form of the likelihood function and estimated by maximum likelihood (see section 7.3).

We have now demonstrated that the Kalman filter may be interpreted in terms of conventional least squares procedures. Furthermore, the updating equation for b may be interpreted as adaptive expectations with a time varying parameter K_t:

$$b_t = b_{t/t-1} + K_t(y_t - x'b_{t/t-1}) \tag{7.52}$$

where

$$K_t = P_{t/t-1}xf_t^{-1} \tag{7.53}$$

$$f_t = (x'P_{t/t-1}x + \sigma^2) \tag{7.54}$$

K_t may be viewed as representing the degree of uncertainty surrounding the new information y_t. For any given forecast error $\tilde{v}_t = (y_t - x'b_{t/t-1})$ the adjustment to $b_{t/t-1}$ is smaller the larger the variance of past forecast errors, since

$$f_t^{-1} = (\mathrm{Var}(\tilde{v}_t))^{-1}.$$

Throughout we have assumed that the variance–covariance matrices are known to the agent, and to the econometrician. In the practical implementation of the Kalman filter one can either assume 'plausible' values for these and conduct a sensitivity analysis (e.g. Lawson 1980) or the covariance matrices may be estimated (see below).

Two further points need to be mentioned. First, at any point in time the prediction equations (7.42) and (7.45) can be used to generate multi-period predictions based on information at t. For example

$$b_{t+n/t} = T^n b_t \quad \text{and} \quad \tilde{y}_{t+n/t} = x'b_{t+n/t}$$

and the latter can be used directly in multi-period, forward-looking models (e.g. Sargent 1979, Cuthbertson and Taylor 1986) of the form:

$$Z_t = \lambda_0 Z_{t-1} + \lambda_1 \sum_{i=1}^{m} \delta^i y^e_{t+i/t} \tag{7.55}$$

where $\tilde{y}_{t+i/t}$ replaces $y^e_{t+i/t}$.

Second, an agent at time $t = T$ may wish to use *all* past sample information to provide a 'smoothed' estimate of the unobservable (permanent income) π_t (the first element of β_t) rather than utilising his current one-step-ahead prediction. The updating equations (7.50) and (7.51) can be used in reverse to obtain $b_{t/T}$ and $P_{t/T}$. These smoothed estimates of π_t could provide a proxy for permanent income (see below).

Some special cases

We now consider how recursive least squares and our intuitive results on the stochastic trend model may be viewed as special cases of the Kalman filter equations derived in the previous section.

Our simple *unobservable components model* (with $\gamma_{t-1} = 0$) is:

$$y_t = \pi_t + \varepsilon_t \tag{7.56}$$

$$\pi_t = \pi_{t-1} + \zeta_t \tag{7.57}$$

In state-space form, the model has

$$X = T = 1,\ \beta_t = \pi_t,\ V = \sigma^2_\varepsilon I,\ Q = \sigma^2_\zeta I,\ \Psi_0 = \sigma^2_0 I \tag{7.58}$$

Substituting (7.58) in the prediction equations (7.42) and (7.45):

$$\pi_t = \pi_{t/t-1} \tag{7.59}$$

$$P_{1/0} = \sigma^2_{t/t-1} = \sigma^2_0 + \sigma^2_\zeta \tag{7.60a}$$

and the updating equations using (7.48)–(7.51) are:

$$\pi_t = \pi_{t/t-1} + k_t(y_t - \pi_{t/t-1}) \tag{7.60b}$$

$$\sigma_t = (1 - k_t)\sigma^2_{t/t-1} \tag{7.61}$$

where

$$k_t = \sigma^2_{t/t-1}(\sigma^2_{t/t-1} + \sigma^2_\varepsilon)^{-1} = (\sigma^2_0 + \sigma^2_\zeta)((\sigma^2_0 + \sigma^2_\zeta) + \sigma^2_\varepsilon)^{-1} \tag{7.62}$$

which confirm our earlier intuitive ideas on the updating equation for π_t given in equations (7.14) and (7.15).

In *recursive least squares* an initial $t - 1$ ($>k$) observations can be used to provide an initial estimate b_{t-1} with covariance matrix P_{t-1}:

$$b_{t-1} = (X'X)^{-1}_{t-1}(X'Y)_{t-1}$$

$$P_{t-1} = \sigma^2(X'X)^{-1}_{t-1}$$

The OLS model may then be represented in state-space form as

$$y_t = x_t'\beta_t + \varepsilon_t \qquad (t = 1, 2, \ldots n)$$

$$\beta_t = T\beta_{t-1} + R\eta_t$$

with

$$\varepsilon_t \sim N(0, \sigma^2)$$

$$\eta_t = 0; \quad Q, R = 0$$

$$T = I$$

and $b_0 = b_{t-1} \sim N(\beta_{t-1}, P_{t-1})$

The prediction equations are then extremely straightforward:

$$b_{t/t-1} = b_{t-1}$$

$$P_{t/t-1} = P_{t-1} = \sigma^2(X'X)^{-1}_{t-1}$$

while the updating equations, given the scalar y_t and the vector x_t are:

$$b_t = b_{t-1} + K_t(y_t - x_t'b_{t/t-1})$$

$$P_t = (I - K_t x_t')P_{t/t-1}$$

where

$$K_t = (X'X)^{-1}_{t-1}x_t' f_t^{-1}$$

and

$$f_t = \text{Var}(\tilde{v}_t) = \sigma^2(1 + x_t'(X'X)^{-1}_{t-1}x_t)$$

The series $\tilde{v}_t/f_t^{1/2}$ is also referred to as the 'recursive residuals' and forms the basis for the CUSUM and CUSUMSQ tests for parameter stability (see Chapter 4). Note that recursive least squares is *not* a variable parameter model since we do not assume a specific model of how β varies through time since we believe the *true β is constant*. The Kalman filter is merely an algorithm for 'repeating' OLS as we extend the sample. We expect to see β settle down to a constant value as more data is added, since the underlying 'true' model has β as a constant in the population.

General form of the Kalman filter using Bayes theorem

We now wish to generalise the equations for the Kalman filter and present the derivation in terms of Bayes theorem. Again it is important to focus on what is known (to the econometrician) and what is to be estimated. We have a set of m state variables = $(\beta_1, \beta_2 \ldots \beta_{mt})$ which are *not observed directly* and instead of a single series we have *n measurement variables* $y_t = (y_{1t} \ldots y_{nt})$ for time periods $t = 1, 2, 3 \ldots T$, which are observed directly. The model then has two distinct blocks.

The *measurement equation* for time t is:

$$y_t = X_t\beta_t + \varepsilon_t \qquad t = 1, 2 \ldots T \tag{7.63}$$

$$\varepsilon_t \sim N(0, V_t)$$

where X_t is an $n \times m$ *known* matrix and ε_t is an $n \times 1$ vector of error terms with mean zero and covariance matrix V_t.

As mentioned above, while the values of β_t are assumed to be unobservable we do need to make some assumption about the mechanism which governs the generation of β_t. This takes the form of the *transition equation*:

$$\beta_t = T_t\beta_{t-1} + R_t\eta_t \tag{7.64}$$

$$\eta_t \sim N(0, Q_t)$$

where T_t and R_t are again *known* $m \times m$ matrices and η_t is an $m \times 1$ vector of disturbances with mean zero and covariance matrix Q_t.

We assume η_t and ε_t are uncorrelated (for all t), that β_{t-1} is independent of the error term η_t in the transition equation and finally that β_t is uncorrelated with the measurement error ε_t:

$$E(\eta_i\varepsilon_j) = 0 \text{ for all } i, j$$

$$E(\beta_{t-1}, \eta_t) = 0 \tag{7.65}$$

$$E(\beta_t\varepsilon_t) = 0$$

Equations (7.63) and (7.64) together make up the state-space model. At first sight these two equations look fairly standard but the time subscripts must be intepreted very precisely. Equation (7.63) contains only *current* dated values of β_t while (7.64) contains only a single lagged value, β_{t-1}. These restrictions do not rule out more complex dynamic models but they do mean that such models must be reparameterised into the state-space form of (7.63) and (7.64).

Some simple examples may make this clearer. Suppose we have an AR(1) model in the scalar y_t.

$$y_t = \alpha y_{t-1} + \eta_t$$

Then the state-space form is:

$$y_t = \beta_t \qquad \text{(measurement equation)} \tag{7.66}$$

$$\beta_t = \alpha\beta_{t-1} + \eta_t \qquad \text{(transition equation)} \tag{7.67}$$

where $X_t = 1$, $T_t = \alpha$, an unknown scalar constant, $R_t = 1$, $\varepsilon_t = 0$.

Consider next the AR(2) model:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \eta_t$$

Here the reparameterisation requires the creation of an additional state variable and the state-space form is:

$$y_t = \beta_{1t} \qquad \text{(measurement equation)} \tag{7.68}$$

$$\beta_{1t} = \alpha_1\beta_{1t-1} + \alpha_2\beta_{2t-1} + \eta_{1t} \quad \text{(transition equation)} \quad \textbf{(7.69a)}$$

$$\beta_{2t} = \beta_{1t-1} \quad \text{(transition equation)} \quad \textbf{(7.69b)}$$

In matrix form we have:

$$X_t = (1, 0) \qquad \beta_t = (\beta_{1t}, \beta_{2t}) \qquad \eta_t = (\eta_{1t}, \eta_{2t})$$

$$T = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix} \qquad R = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

The above may appear a little strange but clearly such a reparameterisation allows the model to be expressed in state-space form. In the second case note that extra lags (y_{t-2}) in the original model are dealt with simply by defining extra state variables, $\beta_t = (\beta_{1t}, \beta_{2t})$. Note that although T_t, R_t have time subscripts, in many econometric applications these are constant matrices/vectors. The Kalman filter assumes T is known; estimation of T which contains the unknown parameters (α_1, α_2) is discussed in section (7.3) on maximum likelihood estimation.

The state-space form which is a first-order dynamic system (equation (7.64)) has quite wide relevance in mathematics generally. For example, the analysis of the stability of a dynamic system by the calculation of eigenvalues is carried out usually on the state-space form of a model. One final point which is worth bearing in mind when considering the relationship between structural economic models and the state-space form is that the latter is essentially a type of reduced form of the structural model. However it does not allow *simultaneous* interactions between the observed variables and so if the structural form is over-identified it is not possible to use the state-space form to represent uniquely the structural parameters of the model.

The Kalman filter

The intuition behind the Kalman filter is fairly simple. We have an initial guess of β_0 and its covariance P_0. In (7.63) and (7.64) we know the values of T_t, X_t, R_t, V_t and Q_t and further, we have observations on y_t. With this information the Kalman filter provides an 'optimal' forecast of the *unobserved* β_t ($t = 1, 2, T$). The notion of 'optimal' needs to be made a little more specific. Kalman and Bucy (1961) use the minimum mean square error (MMSE) criterion as their definition of optimality. However, if we make the additional assumption that the error terms (ε_t, η_t) are normally distributed, then the Kalman filter also provides the maximum likelihood estimator of β_t.

The latter approach is intuitively appealing and we present the derivation of the Kalman filter from the perspective of maximum likelihood using Bayes theorem.

We have an initial estimate of β_{t-1} namely b_{t-1} (at $t - 1 = 0$ say) and an initial estimate of its covariance matrix P_{t-1}. The unbiased *predictor* of β_t based on information at $t - 1$, that is $\beta_{t/t-1}$, is given by the transition equation

$$\beta_{t/t-1} = T_t\beta_{t-1} \qquad \textbf{(7.70a)}$$

We discussed earlier, see equation (7.45), the estimate of the covariance matrix based on information at $t - 1$:

$$\text{Cov}(\beta_{t/t-1}) = P_{t/t-1} = (T_tP_{t-1}T_t' + R_tQ_tR_t') \qquad \textbf{(7.70b)}$$

Equations (7.70a) and (7.70b) are the *prediction equations* for the state vector β_t and its covariance which may be calculated without any reference to the observations y_t. We can use this information at $t - 1$ to *predict* y_t at time t, and the covariance matrix of the one-step-ahead prediction errors F_t:

$$y_{t/t-1} = X_t\beta_{t/t-1}$$

The one-step-ahead prediction error $\tilde{v}_t$ is

$$\tilde{v}_t = y_t - y_{t/t-1}$$

with covariance matrix:

$$F_t = \text{Cov}(y_t) = (X_tP_{t/t-1}X_t' + V_t) \qquad \textbf{(7.70c)}$$

We can now state the probability density functions for β_t, ε_t and y_t:

$$p_1(\beta_t) = C_1 \exp[-0.5(\beta_t - \beta_{t/t-1})'(P_{t/t-1})^{-1}(\beta_t - \beta_{t/t-1})] \qquad \textbf{(7.71)}$$

$$p_2(\varepsilon_t) = C_2 \exp[-0.5(y_t - X_t\beta_t)V_t^{-1}(y_t - X_t\beta_t)] \qquad \textbf{(7.72)}$$

$$p_3(y_t) = C_3 \exp[-0.5(y_t - X_t\beta_{t/t-1})'F_t^{-1}(y_t - X_t\beta_{t/t-1})] \qquad \textbf{(7.73)}$$

where C_1, C_2 and C_3 can be evaluated but are complex and add nothing to the present exposition.

Now the 'optimal' estimate of β_t is taken to be that value which maximises the conditional probability of β_t, given the observed values of y_t. As β_t and ε_t are uncorrelated their joint probability density function is simply

$$p_4(\beta_t, \varepsilon_t) = p_1(\beta_t).p_2(\varepsilon_t) \qquad \textbf{(7.74)}$$

It is also possible to show that the joint probability density function of β_t and y_t is

$$p_5(\beta_t, y_t) = p_4(\beta_t, \varepsilon_t) = p_1(\beta_t).p_2(y_t - X_t\beta_t) \qquad \textbf{(7.75)}$$

Finally using Bayes' decision rule we can state the probability density function of β_t conditional on y_t as:

$$\begin{aligned} p_6(\beta_t/y_t) &= p(\beta_t, y_t)/Pr(y_t) \\ &= p_1(\beta_t).p_2(y_t - X_t\beta_t)/p_3(y_t) \end{aligned} \qquad \textbf{(7.76)}$$

and it is this quantity which we wish to maximise by a suitable 'estimate' of β_t, which we denote b_t. The first- and second-order conditions for a maximum are that

$$\frac{\partial p_6(\beta_t|y_t)}{\partial \beta_t} = 0 \quad \text{and} \quad \frac{\partial}{\partial \beta_t}\left[\frac{\partial p_6(\beta_t|y_t)}{\partial \beta_t}\right] \text{ is negative definite.}$$

Now differentiating (7.76):

$$\begin{aligned} \frac{\partial p_6}{\partial \beta_t} &= \frac{p_1(\beta_t)\frac{\partial}{\partial \beta_t}[p_2(y_t - X_t\beta_t)] + \left[\frac{\partial}{\partial \beta_t} p_1(\beta_t)\right] p_2(y_t - X_t\beta_t)}{p_3(y_t)} \\ &= 0 \end{aligned} \qquad \textbf{(7.77)}$$

which implies that

$$[p_1(\beta_t)] \frac{\partial}{\partial \beta_t} [p_2(y_t - X_t\beta_t)] = -\frac{\partial}{\partial \beta_t} [p_1(\beta_t)][p_2(y_t - X_t\beta_t)] \qquad \textbf{(7.78)}$$

Using (7.71) and (7.72) respectively we have:

$$\partial[p_1(\beta_t)]/\partial\beta_t = -P_{t|t-1}^{-1}(\beta_t - \beta_{t|t-1})p_1(\beta_t) \qquad \textbf{(7.79a)}$$

$$\partial[p_2(y_t - X_t\beta_t)]/\partial\beta_t = X_t'V_t'(y_t - x_t\beta_t)[p_2(y_t - X_t\beta_t)] \qquad \textbf{(7.79b)}$$

Substituting (7.79a) and (7.79b) in (7.78) and rearranging:

$$X_t'V_t^{-1}(y_t - X_t b_{t|t-1}) = P_{t|t-1}(b_t - b_{t|t-1}) \qquad \textbf{(7.80)}$$

Rearranging terms we get

$$b_t = b_{t|t-1} + [P_{t|t-1}^{-1} + X_t'V_t^{-1}X_t]^{-1}X_t'V_t^{-1}(y_t - X_t b_{t|t-1}) \qquad \textbf{(7.81)}$$

or by defining an updating matrix P_t as

$$P_t = (P_{t|t-1}^{-1} + X_t'V_t^{-1}X_t)^{-1} \qquad \textbf{(7.82)}$$

then (7.81) is rewritten as

$$b_t = b_{t|t-1} + P_t X_t' V_t (y_t - X_t b_{t|t-1}) \qquad \textbf{(7.83)}$$

Equation (7.82) may be put into a slightly more convenient form by using the matrix inversion lemma (see Harvey (1983) page 118) to yield an equivalent formula:

$$P_t = P_{t|t-1} - P_{t|t-1}X_t'F_t^{-1}X_t P_{t|t-1} \qquad \textbf{(7.84)}$$

and this may be substituted into (7.83) which upon rearranging gives

$$b_t = b_{t|t-1} + P_{t|t-1}X_t'F_t^{-1}(y_t - X_t b_{t|t-1}) \qquad \textbf{(7.85)}$$

Equations (7.84) and (7.85) are the standard *updating equations* of the Kalman filter. The filter works recursively through time: given an initial estimate of the state β_{t-1} and the covariance matrix P_{t-1} we can form predictions of β_t and P_t as new information on y_t becomes available. The one-step-ahead prediction errors $\tilde{v}_t = (y_t - X_t b_{t|t-1})$ and their covariance F_t, can then be used as an input into the prediction error decomposition of the likelihood function.

7.3 Maximum likelihood and the Kalman filter

Let us turn now to a specific practical example. Consider using the Kalman filter and the state-space form to estimate the AR(2) model above. When we derive the prediction and updating equations for the Kalman filter we assume the covariance matrices (Q_t, V_t) are known as are the matrices X_t, T_t and R_t and we have observations on y_t. In our AR(2) example we will assume Q and V are fixed scalar covariance matrices $Q = \sigma_q^2 I$, $V = \sigma_\varepsilon^2 I$, and we noted that $T = (\alpha_1, \alpha_2)$. Clearly, σ_q^2, σ_ε^2 and T are *not* known, these are precisely what we wish to estimate. (Note that X_t and R_t *are* fixed and known.) However, in using the Kalman filter we can assume any initial values for α_1, α_2, σ_q^2 and σ_ε^2 and derive recursive values for b_t, P_t and $\tilde{v}_t$, *conditional on these initial guesses*. Hence, these $\tilde{v}_t$, F_t, Q_t, V_t and of course y_t can be fed into the prediction error decomposition of the likelihood function and a suitable maximisation routine used to choose that combination of α_1, α_2, σ_q^2, σ_ε^2 that maximises the likelihood. The Kalman filter here merely acts as a useful *algorithm* to yield the likelihood function. This procedure is summarised in Figure 7.1 for our AR(2) model.

What is known and fixed (e.g. X_t, R_t for our AR(2) model) throughout the maximisation procedure and what is to be maximised (i.e. $Q_t = \sigma_q^2 I$, $\Omega_t = \sigma_\varepsilon^2 I$, $T_t = f(\alpha_1, \alpha_2)$) varies depending on the problem under consideration. But the procedure remains the same. Once the model is cast in state-space form and starting values provided, the Kalman filter generates the inputs to the likelihood function which can then be maximised by the numerical optimisation procedures described in Chapter 2. This makes the Kalman filter a particularly powerful tool, as it is possible to estimate both the unobserved part of a model (such as a time-varying parameter) or an expectations variable (e.g. stochastic trend) plus the parameters of the system, (e.g. the variances) in one operation.

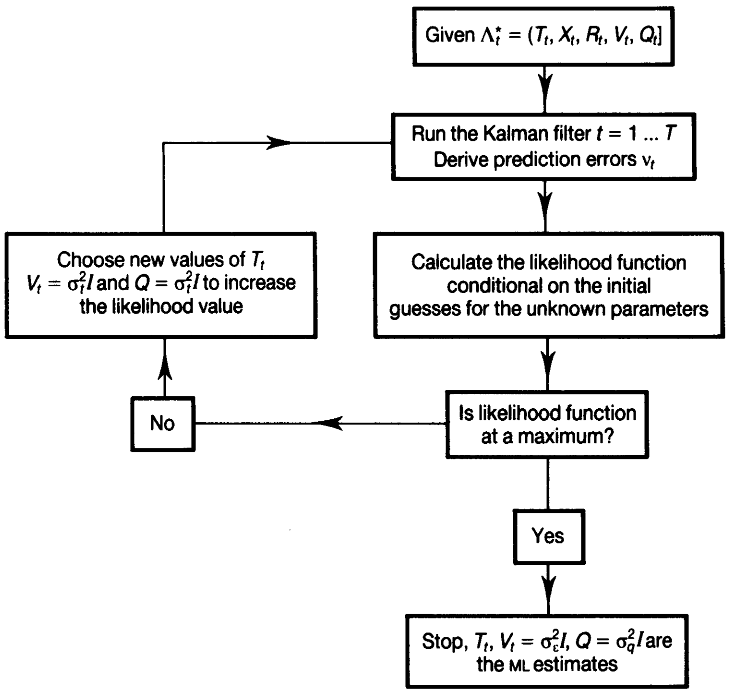

Figure 7.1 The logical structure of maximum likelihood estimation with the Kalman filter.

Smoothing

We have discussed how the Kalman filter yields 'optimal' predictors of the state vector $b_{t+n|t-1}$ based on information at $t-1$. The Kalman filter may also be used to produce *smoothed estimates*. These are the best estimates of β_t, given *all information*, $t = 1, 2, \ldots T$ in the sample. 'Smoothing' is a process whereby we 'look back' from $t = T$, to obtain best estimates for $T-1$, $T-2$, etc. On the last 'round' of the Kalman filter we obtain b_T and its covariance matrix P_T (at $t = T$). The smoothing equations are recursive equations that work backwards from b_T, P_T. If $b_{t|T}$ and $P_{t|T}$ denote the smoothed estimator and its covariance then the smoothing equations are:

$$b_{t|T} = b_t + P_t^*(b_{t+1|T} - T_{t+1}b_t) \qquad \textbf{(7.86)}$$

and

$$P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^* \qquad \textbf{(7.87)}$$

where

$$P_t^* = P_t T'_{t+1|t} P_{t+1|t}^{-1} \qquad t = T-1, T-2, \ldots 1$$

and $b_{T|T} = b_T$ and $P_{T|T} = P_T$. There is little or no 'intuitive feel' one can give to these smoothing recursions. However, in terms of the stochastic trend model, say for income y_t, then smoothed estimates of π_t may be viewed as a measure of permanent income. In the case of time-varying parameters, the smoothed estimates may be interpreted as the best estimates obtainable with all the data available, even though the parameters are still assumed to vary over time.

Time-varying parameter models and the state-space form

It is possible to cast many models in state-space form. We have already demonstrated this for some ARMA models, the unobservable components/stochastic trend model and for recursive least squares. We now analyse how two forms of time-varying parameter model may be cast in state-space form. The Kalman filter and the prediction error decomposition of the likelihood function may then be used to estimate these models.

The state-space formulation (7.3) and (7.4) provides a very general form of time-varying parameter model and in fact this may be further generalised by adding a matrix of constant terms to the transition equation. In practical applications such a general model will usually prove unmanageable and some simplification is required. Consider first, the *random coefficient model*. Here the parameters have a constant mean value but are allowed to stochastically deviate around the mean. The state-space form is:

Random coefficient model

(a) *Measurement equation*

$$y_t = X_t\beta_t + \varepsilon_t \tag{7.88}$$

where β_t is an $(m \times 1)$ unknown state vector: $\beta_t = (\beta_{1t}, \beta_{2t}, \ldots \beta_{mt})$, X_t is $(T \times m)$ data matrix.

(b) *Transition equations*

$$\begin{aligned} \beta_{1t} &= \phi_1 + \eta_{1t} \\ &\vdots \\ \beta_{mt} &= \phi_2 + \eta_{mt} \end{aligned} \tag{7.89}$$

Note that $\phi_i = (i = 1, 2, \ldots m)$ are constants and $\varepsilon_t, \eta_{1t} \ldots \eta_{mt}$ are normally distributed error terms with zero mean, constant variance (and are independent of each other). This formulation allows the parameters to depart from their expected values of $\beta_i (i = 1, 2, \ldots m)$, but this departure is temporary, as at any point in time the expected value of β_{it} is ϕ_i, a constant. So trend-like behaviour in the parameters is ruled out. The transition equation is straightforward and for $m = 2$ is:

$$\beta_t = T\beta_{t-1} + \phi + \eta_t$$

where

$$T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \beta_t = (\beta_{1t}, \beta_{2t}), \quad \phi = (\phi_1, \phi_2)$$

Our second-time varying parameter model might be referred to as a '*systematically varying parameter model*'. In this case the parameters follow a random walk, which is much less restrictive than the random coefficient model. We have:

Systematically varying parameters

(a) *Measurement equation*

$$y_t = X_t\beta_t + \varepsilon_t \tag{7.90}$$

(b) *Transition equation*

$$\begin{aligned} \beta_{1t} &= \beta_{1t-1} + \eta_{1t} \\ &\vdots \\ \beta_{mt} &= \beta_{mt-1} + \eta_{mt} \end{aligned} \tag{7.91}$$

This model allows considerable scope for systematic variation in the parameters but note that β_{it-1} does not affect β_{jt} $(i \neq j)$, so the T matrix in (7.4) is assumed to be diagonal. It is also usual to assume that Q_t is diagonal. The assumption of diagonality may be easily relaxed, although it is often not wise to do so. A further restriction in (7.91) is that the variation in β_t is random rather than being caused by some observed variable. If for example we have a prior belief that the random parameter β_{1t} is related to some observed variable y_{2t} then we should build this into the estimation process. This can be done by making one of the measurement equations non-stochastic. A simple two-variable example will demonstrate this.

Measurement equation

$$\begin{aligned} y_{1t} &= X_{1t}\beta_{1t} + \varepsilon_t \\ y_{2t} &= \beta_{2t} \end{aligned} \tag{7.92}$$

Transition equation

$$\begin{aligned} \beta_{1t} &= T_1\beta_{2t-1} + \eta_{1t} \\ \beta_{2t} &= \beta_{2t-1} + \eta_{2t} \end{aligned} \tag{7.93}$$

Note that the second measurement equation in (7.92) has no error term (or its variance is zero) so $\beta_{2t} = y_{2t}$. In (7.93) β_{1t} is a function of β_{2t-1}, so by including the extra measurement equation we are able to build the prior information about the dependence of β_{1t} on y_{2t} (i.e. $\beta_{1t} = T_1 y_{2t-1} + \eta_{1t}$) into the transition equation.

7.4 Applied work using the Kalman filter

A model of the exchange rate

Our first example of estimation using time-varying parameters is a model of the exchange rate. Hall (1987) presents a structural equation for the log of the real exchange rate:

$$E_t = A_1 E^e_{t+1} + A_2 E_{t-1} + A_3 r_t + A_4 r_{t-1} + A_5 TB_t + A_6 TB_{t-1} \tag{7.94}$$

where E_t is the log of the real (Sterling) effective exchange rate, r_t is the real interest rate differential between UK short-term rates and world rates (proxied by the real three-month Eurodollar rate and the real three-month Treasury Bill rate) and TB the log of the ratio of exports to imports which is a measure of the real trade balance. The theoretical derivation of this equation will not be repeated here, it may be derived in a number of ways. For example, Hall (1987b) uses a stock equilibrium model with government intervention whereas Curry and Hall (1989) use a model which characterises capital markets as exhibiting both stock and flow elements in equilibrium. At a pragmatic level it may even be thought of as a general encompassing model of a wide range of models, for example if $A_6 = A_5 = A_2 = 0$, then the model reduces to the open arbitrage model (see Cuthbertson and Taylor 1987, Chapter 5, or Cuthbertson and Gripaios 1992 Chapters 5 and 6).

Earlier work estimated (7.94) under the Rational Expectations Hypothesis (REH) using the errors in variables (IV) methods described in Chapter 6 for the expectations variable E^e_{t+1}. Here we concentrate on constructing a learning model for the expected exchange rate which has time-varying parameters. We can then use the forecasts of E_{t+1} as inputs into the structural exchange rate equation (7.94).

We may rearrange (7.94) to give

$$E^e_{t+1} = B_1 E_t + B_2 E_{t-1} + B_3 r_t + B_4 r_{t-1} + B_5 TB_t + B_6 TB_{t-1} \tag{7.95}$$

Hall (1987) assumes that the reduced form equations for r_t and T_t are:

$$r_t = C_1(L) r_{t-1} + C_2(L)\text{GDP}_{t-1} + C_3(L) P_{t-1} \tag{7.96}$$

$$TB_t = D_1(L) TB_{t-1} + D_2(L)\text{GDP}_{t-1} + D_3(L) OP_{t-1} + D_4(L) E_{t-1} \tag{7.97}$$

where OP is the log of real oil prices, P is the rate of domestic inflation (i.e. the change in the log of the RPI) and GDP is the log of the real output measurement of GDP. C_i, D_i are polynomial lag operators.

Hall (1987) then demonstrates that equations (7.95)–(7.97) yield the following equation for the evolution of the exchange rate:

$$(E_t - E_{t-1}) = B_{ot} + B_{it}(OP_{t-2} - OP_{t-3}) + B_{2t}(r_{t-2}) + B_{3t}(\dot{P}_{t-2} - \dot{P}_{t-3}) + B_{4t}(\text{GDP}_{t-2} - \text{GDP}_{t-3}) + B_{5t}(T_{t-2} - T_{t-3}) + B_{7t} E_{t-2} \tag{7.98}$$

Because equations (7.96) and (7.97) are 'rules of thumb' used by agents when forecasting τ_t and TB_t, Hall assumes the parameters are likely to be time-varying and hence the B_i coefficients in (7.98) have 't' subscripts.

Note that all lagged information is dated $t-2$ or greater so that when we forecast E_{t+1} in (7.95) the information set will be dated at $t-1$. The time-varying parameters are assumed to be generated by a random walk:

$$B_{it} = B_{it-1} + \eta_{it} \tag{7.99}$$

The measurement equation is (7.98) with $y_t = (E_t - E_{t-1})$ as the dependent variable and the known X_t matrix consisting of the RHS variables in (7.98). The state vector β_t is the vector of time varying parameters B_i ($i = 0, 1, 2, \ldots 7$) and the transition equation(s) are given in (7.93).

As demonstrated in equations (7.84) and (7.85) above we can apply the Kalman filter to (7.98) and (7.99), conditional on the variance of the error term in (7.98) and the covariance matrix for (7.99) which is assumed to be diagonal. In fact the likelihood function may be concentrated so that only the *ratio* of the variance of (each of) the state equation(s) to the measurement equation is estimated. Hall finds that the residuals from the measurement equation (7.98) are reasonably well behaved, the Ljung–Box tests for serial correlation are $\text{LB}(1) = 0.1$, $\text{LB}(2) = 2.4$, $\text{LB}(4) = 2.5$, $\text{LB}(8) = 5.6$, $\text{LB}(16) = 17.3$ which indicates a lack of serial correlation in the error process. The latter suggests that there are no important variables omitted from (7.98).

Hall (1987) shows the graphs of some of the time-varying parameters which will not be included here to save space. The overall conclusions are that all the parameters exhibit marked variation over time with no strong tendency to converge on a stable parameter value; they also show a tendency to jump markedly in 1978. Interpreting the movement in the parameter values is not straightforward as we must remember that they reflect market expectations not underlying structural parameters. For example, in the early part of the period a positive interest rate differential seems to be associated with an expected rise in the exchange rate; this effect seems to disappear during the 1980s. Part of the explanation for this may be given by a corresponding movement in the coefficient on the lagged exchange rate from zero to nearly minus one. When this coefficient is zero the exchange rate is a first difference formulation so that it is essentially a random walk. When it is minus one the equation determines the level of the exchange rate rather than its change. An intepretation of these coefficients movements is as follows. As the commitment of the government towards controlling inflation strengthened in the 1980s then the foreign exchange (FOREX) market interpreted this as a change in the exchange rate regime such that a particular level of the exchange rate was seen as a target, in order to aid the fight against inflation stemming from exchange rate changes.

The fact that there is no serial correlation in the errors is clearly one requirement for the forecast from the learning model to be weakly rational, but we clearly need to check that the expectations series generated by the model is not consistently biased. We may do this by first generating the one-step-ahead forecast of the model and then testing this for biasedness relative to the outturn. The one-step-ahead forecast of the model is generated as:

$$E^e_{t+1} = E_{t-1} + \sum_{i=1}^{7} B_{it} X_{it-i} + B_{ot} \qquad \textbf{(7.100)}$$

where the X_i are all the variables given in (7.98). This series for E^e_{t+1} was then subject to the following tests:

$$\underset{}{E_{t+1}} = \underset{(0.0017)}{1.000898}\, E^e_{t+1} \qquad \textbf{(7.101)}$$

$$E_{t+1} - E^e_{t+1} = \underset{(0.0080)}{0.00451} \qquad \textbf{(7.102)}$$

$$E_{t+1} = \underset{(0.45)}{1.49} + \underset{(0.098)}{0.678}\, E^e_{t+1} \qquad \textbf{(7.103)}$$

where () = standard error of the coefficient. Equations (7.101) and (7.102) are simple tests of unbiasedness. In (7.101) the coefficient is not significantly different from one and hence we do not reject unbiasedness. The latter conclusion is reinforced by (7.102) where the constant is not significantly different from zero. Equation (7.103) is a little more complex under the null hypothesis that E^e_{t+1} is an unbiased *and* efficient forecast of E_{t+1}; the constant in (7.103) should equal zero and the coefficient on E^e_{t+1} should equal unity (Wallis 1989, Mincer and Zarnowitz 1969). Both of these conditions are statistically rejected so we may conclude that while the learning model is unbiased it is not fully efficient. This is a satisfactory result since weak REH requires unbiasedness but only the strong form of REH implies efficiency. It is therefore not surprising that a '*partial information*' learning model as used here would fail to meet the efficiency requirement.

Having derived the expectations series from our 'learning model', the structural exchange rate equation (7.94) is estimated. This is done by estimating a three-equation system using three-stage least squares where the three equations are the exchange rate equation itself, the interest rate (r) and trade balance equation (T). In addition, E^e_{t+1} is specified as endogenous in the estimation. The trade balance and interest rate equations are not a central concern of this paper so they will not be discussed here, they should rather be thought of as instrumenting equations which help to give consistent and efficient estimates of the exchange rate parameters. Applying this system estimation technique then gives the parameter estimates shown in Table 7.1 for the exchange rate equation: a restricted and an unrestricted model are presented.

The two restrictions on the model, $A_2 = 1 - A_1$ and $A_3 = 0$ are accepted easily with a quasi likelihood rate test statistic of 1.32 (distributed as $\chi^2(2)$). Both the interest rate effect (A_4) and the trade effect ($A_5 + A_6$) are correctly signed and significant. The summary statistics indicate an absence of serial correlation and heteroscedasticity in the error process. Structural stability is clearly an important

Table 7.1 Estimation of a structural model of the exchange rate

	Unrestricted model	Restricted model
A_1	0.55 (4.8)	0.53 (4.8)
A_2	0.45 (3.9)	$(1 - A_1)$
A_3	−0.14 (0.4)	— —
A_4	0.73 (2.8)	0.66 (3.8)
A_5	0.35 (3.3)	$(T_{-2} - T_{-3})$ 0.35 (3.6)
A_6	−0.20 (1.9)	0.16 (2.9)
σ	0.022	0.017
DW	1.92	2.06
$BP(1)^1$	0.03	0.07
$BP(2)^1$	2.8	2.0
$BP(4)^1$	4.5	4.0
$BP(8)^1$	12.9	11.5
$BP(1)^2$	0.8	0.8
$BP(2)^2$	1.6	1.2
$BP(4)^2$	2.9	2.7
$BP(8)^2$	4.2	5.0

Data period: 1978 Q2–1988 Q1

Note: $BP(.)^1$ is the Box–Pierce test carried out on the residuals of the equation. $BP(.)^2$ is the Box–Pierce test carried out on the squared residuals of the equation.

requirement of any equation although it is not often found in exchange rate models. Assessing structural stability is not straightforward when the estimation process is 3SLS and the number of observations is fairly limited. In order to gain some insight into the stability of the model recursive 3SLS estimation is performed over the period 1985 Q1–1988 Q1. The overall impression is that the model is reasonably stable, with the parameter estimates never moving outside their standard error bounds.

Thus the use of a learning model based on a time-varying parameter model for the exchange rate yields reasonable results when incorporated in a structural exchange-rate equation.

7.5 Summary

The Kalman filter involves some specialist terminology and concepts which have been discussed widely in this section and are sumarised below. The Kalman filter recursive algorithms may be interpreted in a number of ways because they constitute an optimal updating procedure for a wide class of models. The Kalman filter itself consists of a set of convenient recursive formulae which allow one to calculate the one-step-ahead prediction errors $\tilde{v}_t$ and their variance-covariance matrix F_t (or scalar, $\sigma^2 f_t$). However, to apply these recursive algorithms (i.e. updating and prediction equations) one must be able to express the model in state-space form (the measurement and transition equations). The Kalman filter itself does not estimate the unknown parameters of the model; it merely provides $\tilde{v}_t$ and F_t *conditional on* these unknown parameters. However, the prediction error decomposition of the likelihood function utilises $\tilde{v}_t$ and F_t and hence conventional maximisation routines can then be used to determine the unknown parameters. For certain models (for example generalised stochastic trend model) the Kalman filter recursive algorithms also provide an intuitive insight into the working of the statistical model.

The procedure used when estimating a model with the *aid of* the Kalman filter is (a) express the model in state-space form, (b) generate $\tilde{v}_t$ and F_t using the Kalman filter recursions, (c) use $\tilde{v}_t$ and F_t to set up the prediction error decomposition of the likelihood functions, and (d) maximise the latter with respect to the unknown parameters. We have seen that the Kalman filter is useful in estimating variable parameter models, unobservable components, standard ARMA and least squares problems.

Notes

1. Our aim is to bring together different strands of a diverse literature, so that applied economists can understand and utilise the Kalman filter. The main results are all available in the technical literature, indeed the Kalman filter first appeared as early as 1960 (Kalman 1960). The basic 'source material' for this chapter is to be found in Lawson (1980, 1984), Athans (1974), Duncan and Horn (1972), Diderrich (1985), Harrison and Stevens (1976), Harvey and Todd (1983), Harvey (1984a, 1984b), and, most notably, Harvey (1984c).
2. The proof is as follows:

$$\Delta y_t = \varepsilon_t - (1-\theta)\varepsilon_{t-1} \qquad \text{(i)}$$

Taking expectations of (i):

$$y^e_{t/t-1} = y_{t-1} - (1-\theta)\varepsilon_{t-1} \qquad \text{(ii)}$$

Rearranging (i) using the lag operator L:

$$\varepsilon_t = \Delta y_t / [1-(1-\theta)L] \qquad \text{(iii)}$$

Substituting for ε_{t-1} from (iii) in (ii) and rearranging:

$$y^e_{t/t-1} = y_{t-1} - (1-\theta)\Delta y_{t-1}/[1-(1-\theta)L]$$

or

$$y^e_{t/t-1} - y^e_{t-1/t-2} = \theta(y_{t-1} - y^e_{t-1/t-2})$$

3. This example is taken directly from Lawson (1984).
4. An alternative method of illustrating the stochastic trend nature of the model is to take first differences of (7.8a) and substitute for $\Delta\pi_t$ from (7.8b), yielding:

$$\Delta y_t = \gamma_{t-1} + (\zeta_t + \varepsilon_t)$$

where γ_{t-1} is the stochastic trend growth in y.
5. In the most general form of the Kalman filter the matrices X, T, R, V and Q may be time-varying. This makes little difference to the analytics of the derivation of the Kalman filter as will be seen in section (7.2).

Appendix

Lemma 1

To show:

$$P_1^{-1} = (P_0^*)^{-1} + (X'V^{-1}X) \qquad \textbf{(A1)}$$

$$b_1 = b_0^* + K(Y - Xb_0^*) \qquad \textbf{(A2)}$$

Given:

$$K = P_1 X' V^{-1} \tag{A3}$$

$$P_1 = (X_*' V_*^{-1} X_*)^{-1} \tag{A4}$$

$$b_1 = P_1 (X_*' V_*^{-1} Y_*) \tag{A5}$$

$$\begin{pmatrix} Y \\ b_0^* \end{pmatrix} = \begin{pmatrix} X \\ I \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ \omega_0 \end{pmatrix} \tag{A6}$$

or

$$Y_* = X_* \beta + \varepsilon_* \tag{A7}$$

where

$$V_* = \begin{pmatrix} V & O \\ O & P_0^* \end{pmatrix} \tag{A8}$$

Given (A4) and the definitions (A6)–(A8), the (A1) is easily derived:

$$P_1 = \left((X', I') \begin{pmatrix} V^{-1} & O \\ O & P_0^{*-1} \end{pmatrix} \begin{pmatrix} X \\ I \end{pmatrix} \right)^{-1}$$

$$= (X' V^{-1} X + P_0^{*-1})^{-1} \tag{A9}$$

To derive (A2), note that using (A5) and the definitions (A3), (A4) and (A6)–(A8) we have:

$$b_1 = P_1 (X', I') \begin{pmatrix} V^{-1} & O \\ O & P_0^{*-1} \end{pmatrix} \begin{pmatrix} Y \\ b_0^* \end{pmatrix}$$

$$= (P_1 X' V^{-1}) Y + P_1 P_0^{*-1} b_0^*$$

$$= KY - P_1 P_0^{*-1} b_0^* \tag{A10}$$

Concentrating on the term $P_1 P_0^{*-1}$, using (A1) and (A3):

$$P_1 P_0^{*-1} = P_1 (P_1^{-1} - X' V^{-1} X) = (I - KX) \tag{A11}$$

Substituting (A11) in (A10) completes the proof:

$$b_1 = KY - (I - KX) b_0^* = b_0^* + K(Y - X b_0^*) \tag{A12}$$

Lemma 2

To show:

$$K = P_1 X' V^{-1} = P_0^* X' (V + X P_0^* X') \tag{A13}$$

From (A11):

$$P_1 = (I - KX) P_0^* \tag{A14}$$

Substitute (A14) in (A3):

$$K = (I - KX) P_0^* X' V^{-1}$$

$$K (I + X P_0^* X' V^{-1}) = P_0^* X' V^{-1} \tag{A15}$$

Rearranging (A15) completes the proof:

$$K = P_0^* X' (V + X P_0^* X')^{-1} = P_0^* X' F^{-1}$$

where

$$F = (V + X P_0^* X')$$