BANK OF ENGLAND

Economic modelling and forecasting

2-6 February 2015

© Bank of England

State-space models and the Kalman filter

Ole Rummel

Adviser, CCBS at the Bank of England

ole.rummel@bankofengland.co.uk

BANK OF ENGLAND

Outline

- Modelling and forecasting (problems) at central banks
- What is the state-space representation?
- State-space models and statistical inference
- The Kalman filter
- Maximum-likelihood estimation of the Kalman filter
- A simplified Kalman filter in action
- State-space models and the Kalman filter in EViews
- Summary

Modelling and forecasting at central banks

- Monetary policymakers need to forecast in order to make policy
- This is because the monetary policy transmission mechanism has ‘long and variable’ lags (Friedman (1960))
- But a lot of quantities that we routinely use to analyse monetary policy (and, indeed, financial stability) are **unobservable**
- They often correspond to concepts with economic meaning rather than being inherently measurable, such as the business cycle, the output gap, shocks to productivity or preferences, credit conditions, the natural rate of interest or the non-accelerating inflation rate of unemployment (NAIRU)

Other modelling and forecasting issues

- Moreover, we are frequently interested in models with time-varying parameters:
  - prominent example are models with slowly evolving coefficients, such as the Phillips curve or models with regime switches − the coefficients of these models are both unobserved and potentially time-varying
- We may also face additional problems, such as missing data, structural change and mixed-frequency data
- All in all, econometricians face major problems in dealing with these issues and estimating such models
- But there is a method we can use for estimation given whatever data we have to hand – this approach has become known as the Kalman filter (Kalman (1960, 1963))

The basics of the Kalman filter

- The idea underlying the Kalman filter approach is that a dynamic system can be described by changes in the state of its **components**
- The variables of concern in the system, which are **observable**, are represented by dynamic functions of these components, which are themselves **unobservable**
- The unobserved components, also called **state** variables, transit from one state to another or evolve according to certain rules which are neither easy nor straightforward to apply to the observed variables themselves
- This kind of dynamic modelling of systems is called the **state-space method**…
- …which can be estimated with the aid of the **Kalman filter**

What is a state-space model?

- **State-space models** is a rather loose term given to time-series models, usually formulated in terms of unobserved components, that make use of the state-space representation for their statistical treatment
- A wide range of dynamic time-series models in economics and finance, such as ARIMA models, time-varying regression models, dynamic linear models with unobserved components, factor models and stochastic volatility models, can all be written and estimated as special cases of a state-space specification
- In addition, it is becoming customary to use the Kalman filter to calculate the likelihood function for DSGE models

Representing state-space models (1)

- There are two main benefits to representing a dynamic system in state-space form:
  - the state-space representation allows unobserved variables (known as the state variables) to be incorporated, and estimated along with, the observable model; and
  - state-space models can be analysed using a powerful recursive algorithm known as the Kalman filter
- Since many 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
- What is the state-space form?

Representing state-space models (2)

- The linear state-space form is a simple device whereby a dynamic model involving observed and unobserved (state) variables is written in terms of just **two** equations (rather confusingly, though, this can be done in a variety of ways):
  - the **measurement**, **observation** or **signal equation** describes the relationship between observed and unobserved (state) variables; and
  - the **transition** or **state equation** describes the dynamics of the unobserved (state) variables over time (using a first-order Markov process)

Representing state-space models (3)

- **Measurement, observation** or **signal equation**:

$$y_t = H_t \beta_t + A_t Z_t + e_t \quad (1)$$

  where y_t is a $(n \times 1)$ vector of observations, β_t is a $(m \times 1)$ vector of unobservable **state** variables, H_t is a $(n \times m)$ matrix, Z_t is a $(n \times k)$ matrix of deterministic, lagged endogenous or exogenous variables, A_t is a $(n \times n)$ matrix and e_t is a $(n \times 1)$ error vector
- Written this way, the measurement equation (1) is familiar from our understanding of regression models
- It relates the vector of observations, y_t, to the state vector, β_t, the explanatory variables, Z_t, and the measurement error, e_t (noise)
- The (deterministic) matrices H_t and A_t are called **system matrices** and are usually sparse selection matrices

Representing state-space models (4)

- **Transition** or **state equation**:

$$\beta_t = \mu_t + F_t \beta_{t-1} + v_t \qquad (2)$$

  where F_t is a $(m \times m)$ transition matrix, μ_t is a $(m \times 1)$ vector and v_t is a $(m \times 1)$ error vector
- The state equation (2) is perhaps more familiar from our analysis of dynamic models
- It governs the evolution of the state variables, β_t, over time – in particular, it describes a first-order Markov chain that governs the state transitions with innovations, v_t
- In (2), F_t is again a (sparse) **system matrix**

Representing state-space models (5)

- One big difference to conventional models is the **combination** of the measurement and the transition equation, which formalises the role of the signal, v_t, and the noise, e_t
- The strength of the signal relative to the random deviation is measured by the **signal-to-noise ratio** of variances, $\lambda = \sigma_v^2/\sigma_e^2$
- The model allows for the possibility that we neither directly observe the driving variables of the system nor that we measure them correctly:

$$y_t = H_t \beta_t + A_t Z_t + e_t$$

$$\beta_t = \mu_t + F_t \beta_{t-1} + v_t$$

Representing state-space models (6)

- Consider the shocks in the state-space model:

$$y_t = H_t \beta_t + A_t Z_t + \mathbf{e_t} \quad (3)$$

$$\beta_t = \mu_t + F_t \beta_{t\text{-}1} + \mathbf{v_t} \quad (4)$$

- The assumptions underlying the errors are as follows:

$$e_t \sim \text{iid N}(0,R_t)$$

$$v_t \sim \text{iid N}(0,Q_t)$$

$$\text{E}[e_t v_t] = 0$$

- For most applications, it is assumed that the measurement equation errors, e_t, are independent of the transition equation errors, v_t, but this assumption is not strictly required
- The system matrices H_t, A_t, R_t, F_t and Q_t can themselves be functions of some parameters that can be estimated

The time-varying parameter model (1)

- Consider a model linking import price inflation, Δp_t, and the nominal exchange rate, E_t, where both the intercept and the pass-through coefficient from the exchange rate to import price inflation are allowed to vary over time:

$$\Delta p_t = \alpha_t + \delta_t E_t + e_t \tag{5}$$

- Equation (5) has two unobserved state variables, α_t and δ_t
- What does that mean for the set-up of the state-space model in terms of the **measurement** and the **state** equation?

The time-varying parameter model (2)

- We have the following state-space system for the time-varying parameter model:

$$\Delta p_t = \alpha_t + \delta_t E_t + e_t$$

$$\underbrace{\Delta p_t}_{y_t} = \underbrace{(1 \quad E_t)}_{H_t} \underbrace{\begin{pmatrix} \alpha_t \\ \delta_t \end{pmatrix}}_{\beta_t} + e_t \qquad (6)$$

- Note that matrix H_t now becomes a function of E_t

$$\underbrace{\begin{pmatrix} \alpha_t \\ \delta_t \end{pmatrix}}_{\beta_t} = \underbrace{\begin{pmatrix} \alpha_{t-1} \\ \delta_{t-1} \end{pmatrix}}_{\beta_{t-1}} + \underbrace{\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}}_{v_t} \qquad (7)$$

- In the observation equation (6), $A_t = 0$ and $Z_t = 0$
- In the transition equation (7), $\mu_t = 0$ and $F_t = I_2$

The trend-cycle model (1)

- Assume you want to decompose GDP into a permanent (trend) and transitory (cycle) component, both of which are unobserved:

$$GDP_t = T_t + C_t + e_t \tag{8}$$

- We assume that the (unobserved) trend component follows a random walk:

$$T_t = c + T_{t-1} + v_{1t} \tag{9}$$

and that the (unobserved) cyclical component is a stationary AR(1) process ($|\rho| < 1$):

$$C_t = \rho C_{t-1} + v_{2t} \tag{10}$$

The trend-cycle model (2)

- We have the following state-space system for the trend-cycle model:

$$\underbrace{GDP_t}_{y_t} = \underbrace{\begin{pmatrix} 1 & 1 \end{pmatrix}}_{H_t} \underbrace{\begin{pmatrix} T_t \\ C_t \end{pmatrix}}_{\beta_t} + e_t \qquad (11)$$

$$\underbrace{\begin{pmatrix} T_t \\ C_t \end{pmatrix}}_{\beta_t} = \underbrace{\begin{pmatrix} c \\ 0 \end{pmatrix}}_{\mu_t} + \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix}}_{F_t} \underbrace{\begin{pmatrix} T_{t-1} \\ C_{t-1} \end{pmatrix}}_{\beta_{t-1}} + \underbrace{\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}}_{v_t} \qquad (12)$$

The trend-cycle model (3)

- We can also set up the state-space model in (11) and (12) if the cyclical component follows a stationary AR(2) process ($\rho_1 + \rho_2 < 1$):

$$C_t = \rho_1 C_{t-1} + \rho_2 C_{t-2} + v_{2t} \quad (13)$$

- The following approach naturally extends to the general stationary AR(p) model ($\rho_1 + \ldots + \rho_p < 1$):

$$C_t = \rho_1 C_{t-1} + \rho_2 C_{t-2} + \ldots + \rho_p C_{t-p} + v_{2t}$$

The trend-cycle model (4)

- We have the following state-space system for the trend-cycle model with an AR(2) cycle component:

$$\underbrace{GDP_t}_{y_t} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}}_{H_t} \underbrace{\begin{pmatrix} T_t \\ C_t \\ C_{t-1} \end{pmatrix}}_{\beta_t} + e_t \qquad (14)$$

$$\underbrace{\begin{pmatrix} T_t \\ C_t \\ C_{t-1} \end{pmatrix}}_{\beta_t} = \underbrace{\begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix}}_{\mu_t} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho_1 & \rho_2 \\ 0 & 1 & 0 \end{pmatrix}}_{F_t} \underbrace{\begin{pmatrix} T_{t-1} \\ C_{t-1} \\ C_{t-2} \end{pmatrix}}_{\beta_{t-1}} + \underbrace{\begin{pmatrix} v_{1t} \\ v_{2t} \\ 0 \end{pmatrix}}_{v_t} \qquad (15)$$

- In other words, extra lags in the state-space representation are dealt with simply by creating an additional state variable, such that $\beta_t = (T_t, C_t, C_{t-1})'$

The dynamic factor model (1)

- Assume we want to examine an (unobserved) common factor, G_t, in GDP growth rates across five African countries (i = 1, 2, 3, 4):

$$\Delta GDP_{it} = \beta_i G_t + e_{it} \qquad (16)$$

- In addition, assume that the common factor in GDP growth across four African countries follows a stationary AR(2) process ($\rho_1 + \rho_2 < 1$):

$$G_t = c + \rho_1 G_{t-1} + \rho_2 G_{t-2} + v_t \qquad (17)$$

The dynamic factor model (2)

- We have the following state-space system for the dynamic factor model with an AR(2) factor:

$$\underbrace{\begin{pmatrix} \Delta GDP_{1t} \\ \Delta GDP_{2t} \\ \Delta GDP_{3t} \\ \Delta GDP_{4t} \end{pmatrix}}_{y_t} = \underbrace{\begin{pmatrix} \beta_1 & 0 \\ \beta_2 & 0 \\ \beta_3 & 0 \\ \beta_4 & 0 \end{pmatrix}}_{H_t} \underbrace{\begin{pmatrix} G_t \\ G_{t-1} \end{pmatrix}}_{\beta_t} + \underbrace{\begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{pmatrix}}_{e_t} \quad (18)$$

$$\underbrace{\begin{pmatrix} G_t \\ G_{t-1} \end{pmatrix}}_{\beta_t} = \underbrace{\begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix}}_{F_t} \underbrace{\begin{pmatrix} G_{t-1} \\ G_{t-2} \end{pmatrix}}_{\beta_{t-1}} + \underbrace{\begin{pmatrix} v_t \\ 0 \end{pmatrix}}_{v_t} \quad (19)$$

What are we trying to estimate? (1)

- Both the parameters of the state-space model and the unobservable state variables are unknown
- Our first aim is to estimate the **parameters** of the state-space model:

$$y_t = \boldsymbol{H_t}\beta_t + \boldsymbol{A_t}Z_t + e_t$$

$$\beta_t = \boldsymbol{\mu_t} + \boldsymbol{F_t}\beta_{t-1} + v_t$$

$$e_t \sim \text{iid N}(0,\boldsymbol{R_t})$$

$$v_t \sim \text{iid N}(0,\boldsymbol{Q_t})$$

given observations on $y_1, y_2, \ldots, y_T$ (and, possibly, $Z_1, Z_2, \ldots, Z_T$)

What are we trying to estimate? (2)

- Our second aim is to recover the **unobserved state** variable:

$$y_t = H_t \boldsymbol{\beta_t} + A_t Z_t + e_t$$

$$\boldsymbol{\beta_t} = \mu_t + F_t \boldsymbol{\beta_{t-1}} + v_t$$

$$e_t \sim \text{iid N}(0,R_t)$$

$$v_t \sim \text{iid N}(0,Q_t)$$

given observations on $y_1, y_2, \ldots , y_T$ (and, possibly, $Z_1, Z_2, \ldots , Z_T$)

Statistical inference (1)

- One aim of the exercise is to infer properties of the unobservable state variable, β_t, from the observed data $\{y_t | t = 1, \ldots, T\}$ and the state-space model
- Let $Y_T = \{y_0, y_1, \ldots, y_T\}$ be the information available at time T (inclusive) and assume that the model is known (including all parameters)
- The aim of the Kalman filter is to update knowledge of the (unobservable) state variables recursively when a new data point, y_t, becomes available…
- …that is, knowing the conditional distribution of the parameters of interest given the history of y up to time t-1, $Y_{t-1} = \{y_0, y_1, \ldots, y_{t-1}\}$, and the new observation, y_t, we would like to obtain the conditional distribution of the parameters of interest given Y_t

Statistical inference (2)

- Three types of inference are commonly discussed in the literature, depending on the exact nature of T:
  - **filtering**, which aims to update our knowledge of the system in general and that of the state variable, β_t, in particular as each observation, y_t, comes in, that is, given Y_t;
  - **prediction**, which aims to forecast β_{t+h} or y_{t+h} for $h > 0$ given Y_t, where t is the forecast origin; and
  - **smoothing**, which aims to estimate β_t given the entire sample, Y_T, where $T > t$

Estimation (filtering)

- We use the **Kalman filter** as an aid for estimation
- In particular, the Kalman filter is a recursive algorithm that provides the inputs for the (optimal) maximum-likelihood estimate of the moments of the normally distributed state vector, β_t , conditional on an information set (the observed data) and knowledge of the parameters of the state-space model: H_t, A_t, μ_t, F_t, R_t and Q_t
- The Kalman filter accomplishes this by linear least-squares projections

Set-up of the Kalman filter

- The Kalman filter operates within a state-space representation, where the transition and measurement equations are **linear** and where the shocks to the system, e_t and v_t, as well as the initial state, are all **normally** distributed (Gaussian)
- Because a normal distribution is characterised by its first two moments, the Kalman filter can be interpreted as updating the mean and the variance-covariance matrix of the conditional distribution of the state vector as new observations become available
- The equations of the filter are:
  - the conditional expectation (forecast) depending on all available data; and
  - an update that uses the news contained in the latest data point to obtain the best prediction

The Kalman filter: example

- Assume a time-varying parameter model:

$$y_t = x_t \beta_t + e_t \qquad (20)$$

$$\beta_t = \mu + F\beta_{t-1} + v_t \qquad (21)$$

$$e_t \sim \text{iid N}(0,R)$$

$$v_t \sim \text{iid N}(0,Q)$$

The Kalman filter: notation (1)

- To describe inference using the Kalman filter more precisely, we need to introduce some notation
- The estimate of β_t conditional on information up to time t-1 is denoted $\beta_{t|t-1}$
- The estimate of β_t conditional on information up to time t (i.e., the current sample estimate) is denoted $\beta_{t|t}$
- The variance-covariance matrix of β_t conditional on information up to time t-1 is denoted $P_{t|t-1}$
- The variance-covariance matrix of β_t conditional on information up to time t is denoted $P_{t|t}$

The Kalman filter: notation (2)

- The forecast of y_t given information up to time t-1 is denoted $y_{t|t\text{-}1}$
- The (one-step ahead) prediction error, $\eta_{t|t\text{-}1}$, is therefore given by:

$$\eta_{t|t\text{-}1} = y_t - y_{t|t\text{-}1} \qquad (22)$$

- The variance of the prediction error, $\eta_{t|t\text{-}1}$, is denoted by $f_{t|t\text{-}1}$

The Kalman filter: how does it work? (1)

- Assume that the parameters μ, F, R and Q are **known**
- The Kalman filter recursion consists of three steps:
  1. start with a guess of the initial state at time $t = 0$, i.e., $\beta_{0|0}$ and $P_{0|0}$
  2. **prediction**: at time $t = 0$, form an (optimal) prediction $y_{1|0}$ using an estimated value for $\beta_{1|0}$
  3. **updating**: using the observed value of y_t at time $t = 1$, y_1, calculate the prediction error $\eta_{1|0} = y_1 - y_{1|0}$

The Kalman filter: how does it work? (2)

- This prediction error, $\eta_{1|0}$, contains information that we can use to refine our guess about $\beta_{1|0}$
- In particular, the current sample estimate of β_t at time $t = 1$ is given by:

$$\beta_{1|1} = \beta_{1|0} + K_t \eta_{1|0} \tag{23}$$

- K_t is the weight assigned to new information and is called the **Kalman gain** (with determinants yet to be derived)

The Kalman filter prediction equations (1)

- We can derive the following **prediction** equations for β, P, η and f:

$$\beta_{t|t\text{-}1} = \mu + F\beta_{t\text{-}1|t\text{-}1} \qquad (24)$$

$$P_{t|t\text{-}1} = FP_{t\text{-}1|t\text{-}1}F\,' + Q \qquad (25)$$

$$\eta_{t|t\text{-}1} = y_t - y_{t|t\text{-}1} = y_t - x_t\,\beta_{t|t\text{-}1} \qquad (26)$$

$$f_{t|t\text{-}1} = x_t\,P_{t|t\text{-}1}x_t' + R \qquad (27)$$

The Kalman filter prediction equations (2)

- Where do the prediction equations come from?
- Start with the first equation:

$$\beta_{t|t\text{-}1} = \mu + F\beta_{t\text{-}1|t\text{-}1} \quad (24)$$

- Examine the original transition equation (21):

$$\beta_t = \mu + F\beta_{t\text{-}1} + v_t$$

take expectations (at time t-1), note that $E_{t\text{-}1}[v_t] = 0$ and we end up with prediction equation (24)

The Kalman filter prediction equations (3)

- The second equation:

$$P_{t|t\text{-}1} = FP_{t\text{-}1|t\text{-}1}F' + Q \qquad (25)$$

follows from calculating the variance of the transition equation (21) at time t-1:

$$\mathrm{Var}(\beta) = \mathrm{Var}(\mu + F\beta_{t\text{-}1} + v_t)$$
$$= \mathrm{Var}(\mu) + F \times \mathrm{Var}(\beta_{t\text{-}1}) \times F' + \mathrm{Var}(v_t)$$
$$= FP_{t\text{-}1|t\text{-}1}F' + Q$$

- An alternative would be to start from:

$$\mathrm{Var}(\beta) = \mathrm{E}_{t\text{-}1}[\beta\beta']$$
$$= \mathrm{E}_{t\text{-}1}[(\mu + F\beta_{t\text{-}1} + v_t)(\mu + F\beta_{t\text{-}1} + v_t)']$$
$$= \mathrm{E}_{t\text{-}1}[(\mu + F\beta_{t\text{-}1} + v_t)(\mu' + \beta'_{t\text{-}1}F' + v_t')]$$

which – after some algebra – yields the same result

The Kalman filter prediction equations (4)

- The third equation (for the forecast error):

$$\eta_{t|t\text{-}1} = y_t - y_{t|t\text{-}1} = y_t - x_t \beta_{t|t\text{-}1} \qquad (26)$$

  is pretty self-explanatory from (22) and (20)

The Kalman filter prediction equations (5)

- The fourth equation:

$$f_{t|t\text{-}1} = x_t P_{t|t\text{-}1} x_t' + R \qquad (27)$$

follows from the variance of the prediction error (at time t-1):

$$f_{t|t\text{-}1} = \text{Var}(y_t - y_{t|t\text{-}1})$$
$$= \text{Var}([x_t \beta_t + e_t] - x_t \beta_{t|t\text{-}1})$$
$$= \text{E}[(x_t (\beta_t - \beta_{t|t\text{-}1}) + e_t)^2]$$
$$= \text{E}[x_t \beta_t - \beta_{t|t\text{-}1}]^2 + 2\text{E}[e_t x_t(\beta_t - \beta_{t|t\text{-}1})] + \text{E}[e_t^2]$$
$$= x_t P_{t|t\text{-}1} x_t' + R$$

The Kalman filter updating equations (1)

- We can further derive **updating** equations for β, P and K:

$$\beta_{t|t} = \beta_{t|t-1} + K_t \, \eta_{t|t-1} \qquad (28)$$

$$P_{t|t} = P_{t|t-1} - K_t \, x_t \, P_{t|t-1}$$

$$= P_{t|t-1} \, (1 - K_t \, x_t) \qquad (29)$$

$$K_t = P_{t|t-1} x_t' (f_{t|t-1})^{-1} \qquad (30)$$

The Kalman filter updating equations (2)

- Where do the updating equations come from?
- We need to update the forecast of β_t based on new information contained in the prediction error, η_t
- The prediction error may contain information that is **new** relative to the past data
- Recall the general formula for updating a linear projection for some random variable Y_3 (Hamilton (1994), Section 4.5):

$$F(Y_3|Y_2,Y_1) = F(Y_3|Y_1) + \Omega_{32}/\Omega_{22}[Y_2 - F(Y_2|Y_1)] \qquad (31)$$

where:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix}\right) \qquad (32)$$

The Kalman filter updating equations (3)

- Equation (31) provides a convenient formula for updating a linear projection
- Suppose we are interested in forecasting the value of Y_3, and let Y_1 be some initial information on which this forecast might be based
- A forecast of Y_3 on the basis of Y_1 alone takes the form $F(Y_3|Y_1)$

The Kalman filter updating equations (4)

- Let Y_2 represent some new information with which we could update this forecast
- If we were asked to guess the magnitude of this second variable on the basis of Y_1 alone, the answer would be $F(Y_2|Y_1)$
- But we can use equation (31) to optimally update the initial forecast $F(Y_3|Y_1)$ by adding to it a multiple, given by $\Omega_{32}\Omega_{22}^{-1}$, of the unanticipated component of the new information, $[Y_2 - F(Y_2|Y_1)]$

The Kalman filter updating equations (5)

- In our case:

$$Y_1 = x_t,\ Y_2 = y_t \text{ and } Y_3 = \beta_t$$

  whence it follows that:

$$\beta_{t|t} = \beta_{t|t\text{-}1} + [\mathrm{Cov}(\beta_t, v_t)/\mathrm{Var}(y_t)][y_t - y_{t|t\text{-}1}]$$
$$= \beta_{t|t\text{-}1} + K_t\,\eta_{t|t\text{-}1} \qquad (28)$$

- Note that (28) provides a possible definition of the **Kalman gain**:

$$K_t = \mathrm{Cov}(\beta_t, v_t)/\mathrm{Var}(y_t) \qquad (33)$$

The Kalman filter updating equations (6)

- Updating the variance of the forecast error, $P_{t|t}$, involves the formula (Hamilton (1994), Section 4.5):

$$\mathrm{Var}(Y_3 - F(Y_3|Y_2,Y_1)) = \Omega_{33} - \Omega_{32}\Omega_{22}^{-1}\Omega_{23} \qquad (34)$$

or:

$$P_{t|t} = P_{t|t\text{-}1} - K_t\, x_t\, P_{t|t\text{-}1} \qquad (29)$$

The Kalman gain

- The **Kalman gain** – or weight given to new information (about the state) contained in the prediction error – is given by:

$$K_t = P_{t|t\text{-}1}\, x_t'(f_{t|t\text{-}1})^{-1}$$

$$= P_{t|t\text{-}1} x_t' x_t\, (P_{t|t\text{-}1} x_t' + R)^{-1} \qquad (30)$$

- This expression shows the determinants of the Kalman gain, K_t, which are a positive function of the uncertainty associated with $\beta_{t|t\text{-}1}$ ($P_{t|t\text{-}1}$) and the variance of the new information, R

The Kalman filter: overview (1)

Starting values (time 0)

$$\beta_{0/0}, P_{0/0}$$

$$\downarrow$$

Predict state vector (time 1….)

$$\beta_{t/t-1} = \mu + F\beta_{t-1/t-1}$$

$$P_{t/t-1} = FP_{t-1/t-1}F' + Q$$

$$\downarrow$$

Calculate prediction error

$$\eta_{t/t-1} = y_t - y_{t/t-1} = y_t - x_t\beta_{t/t-1}$$

$$f_{t/t-1} = x_t P_{t/t-1} x_t' + R$$

$$\downarrow$$

Update states

$$K_t = P_{t/t-1} x_t' f_{t/t-1}^{-1}$$

$$\beta_{t/t} = \beta_{t/t-1} + K_t \eta_{t/t-1}$$

$$P_{t/t} = P_{t/t-1} - K_t x_t P_{t/t-1}$$

Maximum-likelihood estimation of the Kalman filter (1)

- Recall that up to now we have had to assume that the parameters of the state space, namely H_t, A_t, μ, F_t, R_t and Q_t, are known
- Generally, this is not the case and these parameters have to be estimated as well
- But the Kalman filter provides us with input for a likelihood function which can be maximised with respect to the unknown parameters

Maximum-likelihood estimation of the Kalman filter (2)

- Recall our time-varying parameter model:

$$y_t = x_t \beta_t + e_t$$

$$\beta_t = \mu + F\beta_{t-1} + v_t$$

$$e_t \sim \text{iid N}(0,R)$$

$$v_t \sim \text{iid N}(0,Q)$$

- Given normal error terms and the state vector, the data is distributed as:

$$y_t \mid x_t \sim \text{N}(x_t \beta_{t|t-1},\, x_t P_{t|t-1} x_t' + R) \qquad (35)$$

Maximum-likelihood estimation of the Kalman filter (3)

- The log-likelihood function is given by:

$$\ln L = -\frac{1}{2}\sum_{t=1}^{T} \ln(2\pi^{n}[\det(x_t P_{t|t-1} x_t' + R]) \quad (36)$$
$$= -\frac{1}{2}\sum_{t=1}^{T} (y_t - x_t \beta_{t|t-1})'(x_t P_{t|t-1} x_t' + R)^{-1}(y_t - x_t \beta_{t|t-1})$$

or:

$$\ln L = -\frac{1}{2}\sum_{t=1}^{T} \ln(2\pi^{n} \det(f_{t|t-1})) - \frac{1}{2}\sum_{t=1}^{T} \eta_{t|t-1}' f_{t|t-1}^{-1} \eta_{t|t-1} \quad (37)$$

- Representation (37) of the log-likelihood in terms of the prediction errors was first given by Schweppe (1965) and is generally referred to as the **prediction error decomposition**

The Kalman filter: overview (2)

Starting values (time 0)
Initial conditions for filter and guess for SS parameters

$$\beta_{0/0}, P_{0/0}, Q_0, \mu_0, F_0, R_0$$

$$\downarrow$$

$$\beta_{t/t-1} = \mu + F\beta_{t-1/t-1}$$

$$P_{t/t-1} = FP_{t-1/t-1}F' + Q$$

$$\downarrow$$

$$\eta_{t/t-1} = y_t - y_{t/t-1} = y_t - x_t\beta_{t/t-1}$$

$$f_{t/t-1} = x_t P_{t/t-1} x_t' + R$$

$$\downarrow$$

$$K_t = P_{t/t-1} x_t' f_{t/t-1}^{-1}$$

$$\beta_{t/t} = \beta_{t/t-1} + K_t \eta_{t/t-1}$$

$$P_{t/t} = P_{t/t-1} - K_t x_t P_{t/t-1}$$

$$\downarrow$$

Calculate likelihood function for each t

$$\ln L = \ln L + \left[-\tfrac{1}{2}\ln(2\pi^n \det(f_{t/t-1})) - \tfrac{1}{2}\eta_{t/t-1}' f_{t/t-1}^{-1} \eta_{t/t-1}\right]$$

$$\downarrow$$

$$\ln L = -\tfrac{1}{2}\sum_{t=1}^{T} \ln(2\pi^n \det(f_{t/t-1})) - \tfrac{1}{2}\sum_{t=1}^{T} \eta_{t/t-1}' f_{t/t-1}^{-1} \eta_{t/t-1}$$

A simplified Kalman filter in action (1)

- Assume a basic Kalman filtering problem of the form:

$$y_t = \beta x_t + \eta_t \tag{38}$$

$$x_t = \rho x_{t-1} + \varepsilon_t \tag{39}$$

- Although y_t is observed, the values of x_t, η_t and ε_t cannot be directly observed
- But it is known that the influence of x_t on y_t is β and that η_t and ε_t are orthogonal to each other (E[$\eta_t \varepsilon_t$] = 0)
- The issue is to form optimal predictors of x_t and y_t

A simplified Kalman filter in action (2)

- If we assume that $\beta = 1$, $\rho = 0.5$ and that $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$, the model becomes:

$$y_t = x_t + \eta_t$$

$$x_t = 0.5x_{t-1} + \varepsilon_t$$

and the prediction and updating equations are given by (from (24), (25) and (28), (29), (30)):

- Prediction: $$x_{t|t-1} = 0.5x_{t-1|t-1} \quad (40)$$

$$P_{t|t-1} = 0.25P_{t-1|t-1} + 1 \quad (41)$$

- Updating: $$K_t = P_{t|t-1}/(P_{t|t-1} + 1) \quad (42)$$

$$x_{t|t} = x_{t|t-1} + K_t(y_t - x_{t|t-1}) \quad (43)$$

$$P_{t|t} = (1 - K_t)P_{t|t-1} \quad (44)$$

A simplified Kalman filter in action (3)

- Suppose that the first five observations of the y_t series are given by:

t	1	2	3	4	5
y_t	2.0570	0.4980	1.2315	-1.5968	2.2541

- Although we do not know the initial conditions of the system, suppose that we are at the very beginning of period 1 and have not, as yet, observed y_t

A simplified Kalman filter in action (4)

- If the system was just beginning – so that $x_0 = 0$ – it might be reasonable to set $x_{0|0} = 0$ and to assign an initial value of $P_{0|0} = 0$
- We now have the initial conditions necessary for using the Kalman filter...
- ...and can consider the iterations of the Kalman filter
- Given these initial conditions, we use the prediction equations (40) and (41) to obtain $x_{1|0} = 0$ and $P_{1|0} = 1$
- Prediction:

$$x_{1|0} = 0.5x_{0|0} = 0.5(0) = 0$$

$$P_{1|0} = 0.25P_{0|0} + 1 = 0.25(0) + 1 = 1$$

A simplified Kalman filter in action (5)

- Once we observe $y_1 = 2.0570$, we use the updating equations (42), (43) and (44) to obtain:

$$K_1 = P_{0|0}/(P_{1|0} + 1) = 1/(1 + 1) = 0.5$$

$$x_{1|1} = x_{0|0} + K_1(y_1 - x_{1|0}) = 0 + 0.5(2.0570 - 0) = 1.0285$$

$$P_{1|1} = (1 - K_1)P_{1|0} = (1 - 0.5)(1) = 0.5$$

A simplified Kalman filter in action (6)

- We next use this information to predict $x_{2|1}$ and $P_{2|1}$
- From the prediction equations (40) and (41) we obtain:

$$x_{2|1} = 0.5x_{1|1} = 0.5(1.0285) = 0.5143$$

$$P_{2|1} = 0.25P_{1|1} + 1 = 0.25(0.5) + 1 = 1.1250$$

A simplified Kalman filter in action (7)

- Once we observe y_2 = 0.4980, we use the updating equations (42), (43) and (44) to obtain:

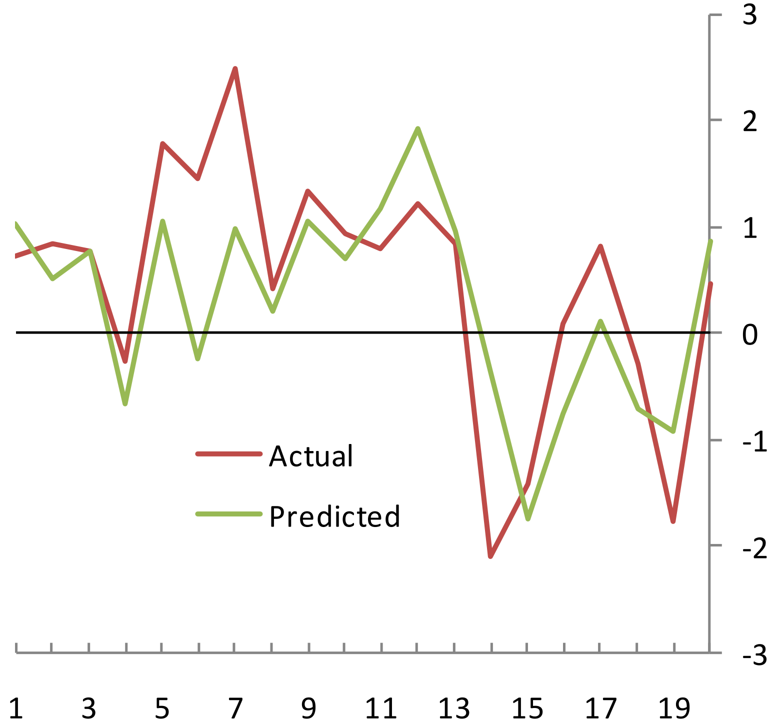
$$K_2 = P_{2|1}/(P_{2|1} + 1) = 1.1250/(1.1250 + 1) = 0.5294$$

$$x_{2|2} = x_{2|1} + K_2(y_2 – x_{2|1})$$

$$= 0.5143 + 0.5294(0.4980 – 0.5413) = 0.5056$$

$$P_{2|2} = (1 – k_2)P_{2|1} = (1 – 0.5294)(1.1250) = 0.5294$$

A simplified Kalman filter in action (8)



- Continuing in this fashion, we can obtain the complete set of forecasts for the series y_t

Entering state-space models in EViews (1)

- The time-varying parameter model:

```
param c(1) -1 c(2) -1 c(3) -1

@signal inf = sv1 + sv2*exrate + [var = exp(c(1))]
     @state sv1 = sv1(-1) + [var = exp(c(2))]
     @state sv2 = sv2(-1) + [var = exp(c(3))]
```

Entering state-space models in EViews (2)

- The trend-cycle model:

```
param c(1) 0 c(2) 0 c(3) -1 c(4) -1

@signal inf = trend + cycle
@state trend = c(1) + trend(-1) + [var = exp(c(3))]
@state cycle = c(2)*cycle(-1) + [var = exp(c(4))]
```

Summary (1)

- The Kalman filter is an algorithm for generating minimum mean square error forecasts in a state-space model
- The state-space form is a very flexible specification for linear time series models
- The state-space model and the Kalman filter go hand-in-hand: to use the Kalman filter, it is necessary to be able to write the model in state-space form
- The Kalman filter consists of two prediction and three updating equations – Kalman filtering is therefore a combination of prediction and correction
- It is important to understand that the Kalman filter is a **dynamic** process

Summary (2)

- If Gaussian errors are assumed, the Kalman filter allows the computation of the log-likelihood function of the state-space model, which allows the model parameters to be easily estimated by maximum likelihood methods
- The procedure used when estimating a model with the help of the Kalman filter is to:
  - express the model in **state-space form**;
  - generate $\eta_{t|t\text{-}1}$ and $f_{t|t\text{-}1}$ using the Kalman filter recursions;
  - use $\eta_{t|t\text{-}1}$ and $f_{t|t\text{-}1}$ to set up th e prediction error decomposition of the likelihood function; and
  - maximise the likelihood function with respect to the unknown parameters

References and further reading (1)

Commandeur, J J F and Koopman, S J (2007), *An introduction to state space time series analysis*, Oxford, Oxford University Press.

Durbin, J and Koopman, S J (2001), *Time series analysis by state space methods*, Oxford, Oxford University Press.

Hamilton, J D (1994a), *Time series analysis*, Princeton, Princeton University Press.

Hamilton, J D (1994b), 'State-space models', Chapter 50 in Engle, R F and McFadden, D L (eds), *Handbook of econometrics, volume IV*, Amsterdam, North-Holland, pages 3039-80.

Harvey, A C and Pierse, R G (1984), 'Estimating missing observations in economic time series', *Journal of the American Statistical Association*, Vol. 97, No. 385, pages 125-31.

References and further reading (2)

Kalman, R E (1960), 'A new approach to linear filtering and prediction problems', *Transactions of the ASME, Series D – Journal of Basic Engineering*, Vol. 82, No. 1, pages 34-45.

Kalman, R E (1963), 'New methods in Wiener filtering theory', in Bogdanoff, J L and Kozin, F (eds), *Proceedings of the first symposium on engineering applications of random function theory and probability*, New York, Wiley, pages 270-388.

Kalman, R E and Bucy, R S (1961), 'New results in linear filtering and prediction theory', *Transactions of the ASME, Series D – Journal of Basic Engineering*, Vol. 83, pages 95-108.

Kim, C-J and Nelson, C R (1999), *State-space models with regime switching*, Cambridge, Massachusetts, MIT Press.

References and further reading (3)

Schweppe, F C (1965), 'Evaluation of likelihood functions for Gaussian signals', IEEE *Transactions on Information Theory*, Vol. 11, No. 1, pages 61-70.