The Kalman filter and the Hamilton Model

Richard G. Pierse

1 Introduction

The Kalman filter of Kalman (1960) and Kalman and Bucy (1961) is an algorithm for generating minimum mean square error forecasts in a state space model. Since the state space model is a very general formulation for linear models in which time varying parameters, measurement errors and missing observations can all be dealt with very easily, the Kalman filter is a very useful tool. As a by-product, if Gaussian errors are assumed, the filter allows the computation of the log-likelihood function of the state space model. This allows the model parameters to be easily estimated by maximum likelihood methods.

2 State Space Models

The general state space model can be written as two equations. The first is a *measurement equation*

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{X}_t\mathbf{d} + \boldsymbol{\varepsilon}_t \tag{1}$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of observable variables, $\mathbf{Z}_t$ is an $n \times m$ matrix of variables, and $\mathbf{X}_t$ is an $n \times k$ matrix of exogenous variables. $\boldsymbol{\alpha}_t$ is an $m \times 1$ vector of possibly unobservable state variables, $\mathbf{d}$ is a $k \times 1$ vector of parameters and $\boldsymbol{\varepsilon}_t$ is an observational error with $E(\boldsymbol{\varepsilon}_t) = 0$ and

$$\text{var}(\boldsymbol{\varepsilon}_t) = \sigma^2\mathbf{H}_t$$

where $\mathbf{H}_t$ is a known $n \times n$ matrix and σ^2 is a scaling variance.

The state variables $\boldsymbol{\alpha}_t$ are generated by first order Markov process defined by the *transition equation*

$$\boldsymbol{\alpha}_t = \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t\boldsymbol{\eta}_t \tag{2}$$

where $\mathbf{T}_t$ is an $m \times m$ matrix, $\mathbf{c}_t$ is an $m \times 1$ vector, $\mathbf{R}_t$ is an $m \times g$ matrix and $\boldsymbol{\eta}_t$ is a $g \times 1$ vector of serially uncorrelated disturbances with $E(\boldsymbol{\eta}_t) = 0$ and

$$\text{var}(\boldsymbol{\eta}_t) = \sigma^2 \mathbf{Q}_t$$

where $\mathbf{Q}_t$ is a known $g \times g$ matrix.

The two disturbances $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ may be contemporaneously correlated but are assumed to be uncorrelated in all other periods so that

$$\begin{aligned} E(\boldsymbol{\eta}_t \boldsymbol{\varepsilon}_s') &= \sigma^2 \mathbf{G}_t \quad , \quad t = s \\ &= \mathbf{0} \quad , \quad t \neq s \, . \end{aligned}$$

In most cases we have $\mathbf{G}_t = \mathbf{0}$ so that the two disturbances are independent.

The matrices $\mathbf{Z}_t$, $\mathbf{H}_t$, $\mathbf{T}_t$, $\mathbf{R}_t$, $\mathbf{Q}_t$ and $\mathbf{G}_t$ are known as the *system matrices*. Most of the elements of these matrices will be fixed elements, mainly ones and zeros. However, they will also contain elements corresponding to the true underlying parameters of the system. These underlying parameters, denoted by the vector $\boldsymbol{\theta}$ are known as the system *hyperparameters*. The vectors $\mathbf{c}_t$ and $\mathbf{d}_t$ may also contain parameters but these do not affect the stochastic properties of the model but only enter the model in a determininistic way.

The disturbance ε_t in the measurement equation (1) is an error in measurement. The state space form thus naturally lends itself to modelling systems with measurement error. In this case the transition equation defines the signal $\boldsymbol{\alpha}_t$ which is unobservable and only measured with error (noise). In the univariate case where $n = m = 1$, the ratio of the variances q/h represents the signal to noise ratio.

3 Applications of State Space Models

The state space form is a very flexible specification for linear time series models. Often there will be more than one way to write down a model in state space form. This section presents a few examples of models that can be cast in state space form.

3.1 The AR(p) model

For example the pth order autoregressive process

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

can be set up in state space form with transition equation given by

$$\boldsymbol{\alpha}_t \equiv \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+2} \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-p+1} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_t \,.$$

The first equation of this system of $p-1$ equations represents the autoregressive process while the remaining $p-2$ equations are identities defining lags of y_t. These latter equations have no error so that the matrix $\mathbf{R}$ is $p-1 \times 1$ and the error process is a scalar. The measurement equation is given by

$$y_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{\alpha}_t.$$

Note that there is no error on this measurement equation because the state variables $\boldsymbol{\alpha}_t$ are observed without error. The hyperparameters are the p parameters $\phi_1, \cdots, \phi_p$ and the error variance σ_ε^2.

3.2 The ARIMA(p,d,q) model

Harvey and Pierse (1984) derive the state space representation of a general *ARIMA(p,d,q)* model

$$\phi(L)\Delta^d y_t = \psi(L)\varepsilon_t \,.$$

Defining $m = \max(p, q+1)$, the transition equation can be written as the $(m+d) \times 1$ system

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0}_{m \times d} \\ 10 \cdots 0 & \delta_1 \cdots \delta_d \\ \mathbf{0}_{d-1 \times m} & \mathbf{I}_{d-1} : 0 \end{bmatrix} \boldsymbol{\alpha}_{t-1} + \begin{bmatrix} \boldsymbol{\psi} \\ \mathbf{0}_{d \times 1} \end{bmatrix} \varepsilon_t \,.$$

where

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ \phi_{m-1} & 0 & \cdots & 0 & 1 \\ \phi_m & 0 & \cdots & 0 & 0 \end{bmatrix} \quad , \qquad \boldsymbol{\psi} = \begin{bmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-2} \\ \psi_{m-1} \end{bmatrix}$$

and $-\delta_j$ is the coefficient on L^j in the expansion of $\Delta^d = (1-L)^d$.

This state space representation has $p+q+1$ hyperparameters and a measurement equation given by

$$y_t = \begin{bmatrix} 1 & \mathbf{0}_{1 \times m-1} & \delta_1 \cdots \delta_d \end{bmatrix} \boldsymbol{\alpha}_t \,.$$

3.3 The Hodrick-Prescott filter

The Hodrick and Prescott (1980) filter generates a smoothed trend y_t^* from a time series y_t by solving the problem

$$\min_{\mathbf{y}^*} \left\{ \sum_{t=1}^{T} (y_t - y_t^*)^2 + \lambda \sum_{t=3}^{T} (\Delta^2 y_t^*)^2 \right\}$$

for a fixed value of the smoothing parameter λ. Koopman *et al.* (1995) show that this problem can be set up as the structural time series model defined by the equations

$$y_t = y_t^* + \varepsilon_t$$
$$y_t^* = y_{t-1}^* + \beta_{t-1}$$

and

$$\beta_t = \beta_{t-1} + \xi_t$$

with the special restriction that $\varepsilon_t = \sqrt{\lambda}\xi_t$. This model can be formulated in state space form with transition equation

$$\boldsymbol{\alpha}_t \equiv \begin{bmatrix} \beta_t \\ y_t^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_{t-1} \\ y_{t-1}^* \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \xi_t$$

and measurement equation

$$y_t = \begin{bmatrix} 0 & 1 \end{bmatrix} \boldsymbol{\alpha}_t + \varepsilon_t \,.$$

Note that here the disturbance terms in transition and measurement error are perfectly correlated with the 1×1 matrix $\mathbf{G}_t$ equal to the scalar $\lambda^{-\frac{1}{2}}$. The only hyperparameter in the model is the scaling variance σ^2.

4 The Kalman Filter

Let $\mathbf{a}_{t-1}$ be the *minimum mean square linear estimator* (*MMSLE*) of $\boldsymbol{\alpha}_{t-1}$ based on information available at time $t-1$. Let $\mathbf{P}_{t-1}$ be the $m \times m$ covariance matrix of the estimation error defined by

$$\mathbf{P}_{t-1} = E(\mathbf{a}_{t-1} - \boldsymbol{\alpha}_{t-1})(\mathbf{a}_{t-1} - \boldsymbol{\alpha}_{t-1})' \,.$$

The Kalman Filter comprises two sets of recursive equations:

4.1 Prediction equations

$$\mathbf{a}_{t|t-1} = \mathbf{T}_t \mathbf{a}_{t-1} + \mathbf{c}_t \tag{3}$$
$$\mathbf{P}_{t|t-1} = \mathbf{T}_t \mathbf{P}_{t-1} \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \tag{4}$$

4.2 Updating equations

When an observation arrives, the estimators $\mathbf{a}_t$ and $\mathbf{P}_t$ can be updated by the one-step ahead prediction error defined by

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1} - \mathbf{X}_t\mathbf{d}\,. \tag{5}$$

This prediction error has estimated variance given by

$$\mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t' + \mathbf{Z}_t\mathbf{R}_t\mathbf{G}_t + \mathbf{G}_t'\mathbf{R}_t'\mathbf{Z}_t' + \mathbf{H}_t\,. \tag{6}$$

The updating equations for $\mathbf{a}_t$ and $\mathbf{P}_t$ are then given by

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + (\mathbf{P}_{t|t-1}\mathbf{Z}_t' + \mathbf{R}_t\mathbf{G}_t)\mathbf{F}_t^{-1}\mathbf{v}_t \tag{7}$$

and

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - (\mathbf{P}_{t|t-1}\mathbf{Z}_t' + \mathbf{R}_t\mathbf{G}_t)\mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} + \mathbf{G}_t'\mathbf{R}_t'). \tag{8}$$

4.3 Dealing with missing observations

The Kalman filter deals very naturally with missing observations. If an observation is missing, the updating equations are simply skipped for that observation. Then the prediction equations operate on $\mathbf{a}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ to compute the *MMSLE* $\mathbf{a}_{t+1|t-1}$ and $\mathbf{P}_{t+1|t-1}$. This continues until the next non-missing observation at which stage the prediction can be updated. Harvey and Pierse (1984) use the Kalman filter to estimate *ARIMA(p,d,q)* models with missing observations.

4.4 The log-likelihood function

Assuming that the disturbances are normally distributed, the log-likelihood function for the model can be computed from the prediction errors $\mathbf{v}_t$ and associated variances $\mathbf{F}_t$ and is defined by

$$L = -\frac{nT}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{F}_t| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t\,. \tag{9}$$

The scalar parameter σ^2can be concentrated out of this log-likelihood function giving

$$\frac{\partial L}{\partial\sigma^2} = -\frac{nT}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t = 0 \Rightarrow \widetilde{\sigma}^2 = \sum_{t=1}^{T}\frac{\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t}{nT}$$

and the concentrated log-likelihood function can be written as

$$L^* = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2} - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{F}_t| - \frac{nT}{2}\log\left(\sum_{t=1}^{T}\frac{\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t}{nT}\right) \quad (10)$$

This log-likelihood can be maximised numerically to find maximum likelihood estimates of the hyperparameter vector $\boldsymbol{\theta}$.

4.5 Initialising the Filter

In order to start the Kalman filter we need initial values $\mathbf{a}_0$ and $\mathbf{P}_0$. The initial values to be used will depend on whether or not the system matrices in the state space form are time invariant.

Firstly, suppose that the system matrices $\mathbf{Z}_t$, $\mathbf{X}_t$, $\mathbf{H}_t$, $\mathbf{T}_t$, $\mathbf{c}_t$, $\mathbf{R}_t$, $\mathbf{Q}_t$ and $\mathbf{G}_t$ are all invariant over time. Then initial values for $\mathbf{a}_0$ can be given by the equation

$$\mathbf{a}_0 = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{c}$$

and initial values for $\mathbf{P}_0$ can be derived by solving the system of equations

$$\mathbf{P}_0 = \mathbf{T}\mathbf{P}_0\mathbf{T}' + \mathbf{R}\mathbf{Q}\mathbf{R}'$$

giving the solution

$$\mathrm{vec}(\mathbf{P}_0) = (\mathbf{I} - \mathbf{T}\otimes\mathbf{T})^{-1}\,\mathrm{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}')\,.$$

When the system matrices are not time-invariant, then it is conventional to set

$$\mathbf{a}_0 = \mathbf{0} \quad \text{and} \quad \mathbf{P}_0 = \kappa\mathbf{I}_m$$

where κ is a large value. This corresponds to having a diffuse prior on $\boldsymbol{\alpha}_0$ or in other words to saying we have no prior information.

4.6 Smoothing

The Kalman filter predictors $\mathbf{a}_{t|t-1}$and $\mathbf{P}_{t|t-1}$ give the optimal predictors of $\boldsymbol{\alpha}_t$ and $\mathrm{var}(\boldsymbol{\alpha}_t)$ based on information available at time $t-1$. The smoothed estimators $\mathbf{a}_{t|T}$ and $\mathbf{P}_{t|T}$ give the optimal predictors of $\boldsymbol{\alpha}_t$ and $\mathrm{var}(\boldsymbol{\alpha}_t)$ based on all the information in the sample. These smoothed estimators can be generated from the recursions:

$$\mathbf{a}_{t|T} = \mathbf{a}_t + \mathbf{P}_t^*(\mathbf{a}_{t+1|T} - \mathbf{T}_{t+1}\mathbf{a}_t) \quad (11)$$

and

$$\mathbf{P}_{t|T} = \mathbf{P}_t + \mathbf{P}_t^*(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{P}_t^{*\prime} \tag{12}$$

where

$$\mathbf{P}_t^* = \mathbf{P}_t\mathbf{T}_{t+1}\mathbf{P}_{t+1|t}^{-1}\,. \tag{13}$$

These recursions work backwards in time, $t = T-1, \cdots, 1$ from the initial conditions $\mathbf{a}_{T|T} = \mathbf{a}_T$ and $\mathbf{P}_{T|T} = \mathbf{P}_T$. Note that the matrix $\mathbf{P}_{t+1|t}$ is not guaranteed to be nonsingular. In case the matrix is singular its inverse $\mathbf{P}_{t+1|t}^{-1}$can be replaced by a generalised inverse $\mathbf{P}_{t+1|t}^{-}$.

5 The Hamilton Regime Switching Model

Hamilton (1989) presents a model of regime switching. A variable y_t is assumed to be a linear function of a vector of variables $\mathbf{x}_t$ with coefficients that depend on the state or regime in period t. There are a discrete number of states, n. Formally

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_{s_t} + \varepsilon_t \tag{14}$$

where s_t is the state in period t which can take one of n possible values, $1, \cdots, n$. Defining $\boldsymbol{\alpha}_t$ as the $n \times 1$ vector with ith element equal to one when $s_t = i$ and all other elements equal to zero, (14) can be rewritten as

$$y_t = \mathbf{x}_t'\mathbf{B}\boldsymbol{\alpha}_t + \varepsilon_t \tag{15}$$

where $\mathbf{B} = [\boldsymbol{\beta}_1 : \cdots : \boldsymbol{\beta}_n]$ and $\text{var}(\varepsilon_t) = \sigma^2$.

The state is s_t is unobserved but is assumed to follow a first order Markov process so that

$$\Pr(s_t = i|\boldsymbol{\xi}_{t-1}) = \Pr(s_t = i|s_{t-1} = j) = \pi_{ij}\,. \tag{16}$$

where $\boldsymbol{\xi}_{t-1}$ is a vector representing all the information available at time $t-1$ which includes lagged values of y_t and $\mathbf{x}_t$. The Markov assumption therefore states that all the information in $\boldsymbol{\xi}_{t-1}$ is encapsulated in the previous state s_{t-1}. The value π_{ij} is known as the *transition probability* of moving to state i from state j and is independent of time.

Representing the transition probabilities in a matrix

$$\boldsymbol{\Pi} = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1n} \\ \vdots & \ddots & \vdots \\ \pi_{n1} & \cdots & \pi_{nn} \end{bmatrix}$$

then (16) implies that

$$E(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}) = \boldsymbol{\Pi}\boldsymbol{\alpha}_{t-1} \tag{17}$$

or

$$\boldsymbol{\alpha}_t = \boldsymbol{\Pi}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \tag{18}$$

where $\boldsymbol{\eta}_t$ is a disturbance uncorrelated with $\boldsymbol{\alpha}_{t-1}$ or $\boldsymbol{\xi}_{t-1}$. Note that the transition probabilities satisfy the condition that $\sum_i \pi_{ij} = 1, \forall j$. This implies that the columns of $\boldsymbol{\Pi}$ must all sum to one or

$$\boldsymbol{\iota}'\boldsymbol{\Pi} = \boldsymbol{\iota}'$$

where $\boldsymbol{\iota}$ is an $n \times 1$ vector of ones, so that the matrix $\boldsymbol{\Pi}$ is *singular*. The disturbance vector $\boldsymbol{\eta}_t$ can take one of a possible set of n^2 discrete values and so is not normally distributed.

Equations (15) and (18) are a measurement and transition equation respectively. However, the problem differs from the Kalman filter framework in that the state vector $\boldsymbol{\alpha}_t$ is discrete rather than continuous, and there is a nonlinearity in (15) so that the state space form is nonlinear.

The Hamilton filter is an iterative algorithm for calculating the distribution of the discrete state variable $\boldsymbol{\alpha}_t$. Let $\mathbf{a}_t$ be $E(\boldsymbol{\alpha}_t|\boldsymbol{\xi}_t)$ with ith element given by

$$\Pr(s_t = i|\boldsymbol{\xi}_t)$$

and $\mathbf{a}_{t|t-1}$ be $E(\boldsymbol{\alpha}_t|\boldsymbol{\xi}_{t-1})$ with ith element given by

$$\Pr(s_t = i|\boldsymbol{\xi}_{t-1})\,.$$

Then the Hamilton filter comprises two recursive equations: one the prediction equation, defining $\mathbf{a}_{t|t-1} = \mathbf{g}(\boldsymbol{\alpha}_{t-1})$ and the other the updating equation defining $\boldsymbol{\alpha}_t = \mathbf{h}(\mathbf{a}_{t|t-1})$.

5.1 Prediction equation

The Hamilton filter prediction equation follows from (17) and is simply

$$\mathbf{a}_{t|t-1} = \boldsymbol{\Pi}\mathbf{a}_{t-1}\,. \tag{19}$$

5.2 Updating equation

The log-likelihood function of the observations y_t is given by

$$L = \sum_{t=1}^{T} \log f(y_t|\mathbf{x}_t, \boldsymbol{\xi}_{t-1})$$

where

$$f(y_t|\mathbf{x}_t, \boldsymbol{\xi}_{t-1}) = \sum_{i=1}^{n} f(y_t|s_t = i, \mathbf{x}_t, \boldsymbol{\xi}_{t-1}) \Pr(s_t = i|\boldsymbol{\xi}_{t-1}) \tag{20}$$

and the conditional distribution of y_t given knowledge of the state s_t has the standard normal density

$$f(y_t|s_t = i, \mathbf{x}_t, \boldsymbol{\xi}_{t-1}) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(y_t - \mathbf{x}_t'\boldsymbol{\beta}_i)^2\} \ .$$

Note that, by definition, we have the joint distribution

$$f(y_t, s_t = i|\mathbf{x}_t, \boldsymbol{\xi}_{t-1}) = f(y_t|s_t = i, \mathbf{x}_t, \boldsymbol{\xi}_{t-1}) \Pr(s_t = i|\boldsymbol{\xi}_{t-1})$$

and the marginal distribution

$$f(y_t|\mathbf{x}_t, \boldsymbol{\xi}_{t-1}) = \sum_{i=1}^{n} f(y_t, s_t = i|\mathbf{x}_t, \boldsymbol{\xi}_{t-1}) \, .$$

But $\boldsymbol{\xi}_t \equiv \{y_t, \mathbf{x}_t, \boldsymbol{\xi}_{t-1}\}$ and so the ratio of the two represents the optimal inference on s_t based on $\boldsymbol{\xi}_t$:

$$\Pr(s_t = i|\boldsymbol{\xi}_t) = \frac{f(y_t, s_t = i|\mathbf{x}_t, \boldsymbol{\xi}_{t-1})}{f(y_t|\mathbf{x}_t, \boldsymbol{\xi}_{t-1})} \, . \tag{21}$$

Define $\mathbf{v}_t$ to be the $n \times 1$ vector with ith element given by $f(y_t|s_t = i, \mathbf{x}_t, \boldsymbol{\xi}_{t-1})$. Then from (20)

$$f(y_t|\mathbf{x}_t, \boldsymbol{\xi}_{t-1}) = \mathbf{v}_t'\mathbf{a}_{t|t-1}$$

and $f(y_t, s_t = i|\mathbf{x}_t, \boldsymbol{\xi}_{t-1})$ is the ith element of the $n \times 1$ vector

$$\mathbf{v}_t \odot \mathbf{a}_{t|t-1}$$

where $\odot$ is the element by element multiplication operator.

Thus (21) can be written as the (nonlinear) updating equation for $\mathbf{a}_t$

$$\mathbf{a}_t = \frac{\mathbf{v}_t \odot \mathbf{a}_{t|t-1}}{\mathbf{v}_t'\mathbf{a}_{t|t-1}} \, . \tag{22}$$

5.3 Initialising the filter

If the Markov process is *stationary and ergodic*, then $E(\boldsymbol{\alpha}_t) = E(\boldsymbol{\alpha}_{t-1})$ and from the transition equation (18)

$$E(\boldsymbol{\alpha}_t) = \boldsymbol{\Pi} E(\boldsymbol{\alpha}_t)\,.$$

Note that the expectations here are *unconditional*. This can be used to define an initial value $\mathbf{a}_0$ for starting the Hamilton filter by solving

$$\mathbf{a}_0 = \boldsymbol{\Pi}\mathbf{a}_0\,.$$

This equation system cannot be solved as it stands because of the singularity of $\boldsymbol{\Pi}$. However, we know that $\boldsymbol{\iota}'\mathbf{a}_0 = 1$ and the solution for $\mathbf{a}_0$ can be derived from the $n+1$th column of the generalised inverse $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ of the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} - \boldsymbol{\Pi} \\ \boldsymbol{\iota}' \end{bmatrix}\,.$$

5.4 Smoothing

Once the unknown parameters of the model $\mathbf{B}$, $\boldsymbol{\Pi}$, and σ^2 have been estimated (by maximum likelihood), it is possible to derive estimates of the state vector based on all the sample information. These are given by $\mathbf{a}_{t|T} = E(\boldsymbol{\alpha}_t|\boldsymbol{\xi}_T)$. In the context of the Hamilton model these *smoothed* estimators represent the best estimate of the probability that the model was in state s_i in period t.

The smoothed predictors are defined by the backward recursion

$$\mathbf{a}_{t|T} = \mathbf{a}_{t|t} \odot \{\boldsymbol{\Pi}'(\mathbf{a}_{t+1|T} \oslash \mathbf{a}_{t+1|t})\}$$

where $\oslash$ is the element by element division operator.

5.5 Generalisations

The Hamilton model can be generalised in several ways. Hamilton (1989) considers the autoregressive model

$$\sum_{j=0}^{p} \rho_j (y_{t-j} - \mathbf{x}'_{t-j}\boldsymbol{\beta}_{s_{t-j}}) = \varepsilon_t \tag{23}$$

where ρ_j represents the jth order autoregressive coefficient with $\rho_0 = 1$, and where $\mathbf{x}_t$ is simply an intercept. This model can be redefined as a first order Markov process with $n^* = n^{p+1}$ states, each state $s_t^* = 1, \cdots, n^*$ representing one permutation of the history of the original states s_t, s_{t-1}, $\cdots$, s_{t-p}.

A simpler specification to (23) is

$$\sum_{j=0}^{p} \rho_j (y_{t-j} - \mathbf{x}'_{t-j} \boldsymbol{\beta}_{s_t}) = \varepsilon_t$$

which has persistence that depends on the current state only. A comparison of the implications of this model with those of (23) is given in Hamilton (1993).

Both the error variance σ^2 in (15) and the autoregressive parameters ρ_j in (23) can be allowed to be state dependent. Finally, the transition probabilities in (16) can be allowed to be endogenously determined. Hamilton (1996) presents various diagnostic tests which can be used to detect misspecification in the estimated model.

6 Further reading

Harvey (1993) is a very clear treatment of state space models and the Kalman filter, concentrating on the formulation of ARIMA models. Chapter 50 in Hamilton (1994b) treats both general state space models, and his own switching regimes model in a unified way. Hamilton (1994a) has a good chapter on regime switching models and Hamilton (1993) is also a good survey.. Harvey (1989) and Koopman *et al.* (1995) develop structural time series models using the state space form. Rigorous technical treatments of the Kalman filter can be found in the texts by Anderson and Moore (1979) or Jazwinski (1970).

References

[1] Anderson, B. D. O. and J. B. Moore (1979), *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ.

[2] Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57, 357–384.

[3] Hamilton, J. D. (1993), 'Estimation, inference and forecasting of time series subject to changes in regime', Chapter 9 in G. S. Maddala, C. R. Rao and H. D. Vinod (eds) *Handbook of Statistics, Volume 11*, Elsevier Science Publishers, Amsterdam, Netherlands.

[4] Hamilton, J. D. (1994a), *Time Series Analysis*, Princeton University Press, Princeton, NJ.

[5] Hamilton, J. D. (1994b), 'State space models', Chapter 50 in R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics, Volume IV*, Elsevier Science, Amsterdam, Netherlands.

[6] Hamilton, J. D. (1996), 'Specification testing in Markov-switching time-series models', *Journal of Econometrics*, 70, 127–157.

[7] Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

[8] Harvey, A. C. (1993), *Time Series Models*, (2nd. ed.), Harvester Wheatsheaf, Hemel Hempstead.

[9] Harvey, A. C. and R. G. Pierse, 1984, 'Estimating missing observations in economic time series', *Journal of the American Statistical Association*, 79, 125–131.

[10] Hodrick, R. and E. Prescott (1980), 'Post-war U.S. business cycles: an empirical investigation', unpublished manuscript, Carnegie-Mellon University.

[11] Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York, NY.

[12] Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering, Transactions ASME*, Series D, 82, 35–45.

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

[14] Koopman, S., A. C. Harvey, J. A. Doornik and N. Shephard (1995), *Stamp 5.0: Structural Time Series Analyser, Modeller and Predictor*, Chapman & Hall, London.