

# FAME: Applied Econometrics

## Lecture Note

### The Kalman Filter

Michael Rockinger

August 26, 2004

The following section is heavily inspired by Thierry Roncalli's book: "Applications à la Finance et à l'Econométrie", Volume 2", the book by Andrew Harvey: "Forecasting structural time series models and the Kalman filter", Cambridge University Press, as well as on the section concerning the Kalman filter in Gouriéroux and Monfort's book on econometrics. We will use the same notation as in Harvey. For copyright reasons this document should not circulate beyond the FAME course.

## 1 The state-space model

A state-space model consists in a measurement equation linking actual observations with latent variables. These latent variables may be seen as defining a given state in which the observations are in. It is assumed that the state variables are a first order Markov process. This is without loss of generality since a higher order Markov process may be transformed into a first order process by augmenting the parameters characterizing a state. We denote this state vector by  $\alpha_t$ . We have

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \quad t = 1, \dots, \tau,$$

where  $\alpha_t$  is the  $m$  dimensional state vector.  $T_t$  is an  $m \times m$  matrix.  $c_t$  is an  $m \times 1$  vector and  $R_t$  is  $m \times g$ .

The measurement equation is given by

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad t = 1, \dots, T,$$

where  $y_t$  is a given time-series with  $N \times 1$  elements.  $Z_t$  is an  $n \times m$  matrix,  $d_t$  an  $N \times 1$  vector.  $\eta_t$  and  $\varepsilon_t$  are supposed to be white noise with dimensions  $g \times 1$  and  $N \times 1$ . Both  $\eta_t$  and  $\varepsilon_t$  are supposed to be normally distributed with 0 correlation which is equivalent to assuming independance of the two. The first two moments are given as follows:

$$\begin{aligned} E[\eta_t] &= 0 \text{ and } V[\eta_t] = Q_t, \\ E[\varepsilon_t] &= 0 \text{ and } V[\varepsilon_t] = H_t. \end{aligned}$$

We also assume that, initially, the state vector follows a Gaussian distribution with mean  $E[\alpha_0] = a_0$  and  $V[\alpha_0] = P_0$

Consider  $a_t$  the best estimate of  $\alpha_t$  given all available information up to time  $t$ , meaning that

$$a_t = E_t[\alpha_t],$$

and the variance-covariance matrix associated to  $a_t$  is given by

$$P_t = E_t[(a_t - \alpha_t)(a_t - \alpha_t)']$$

## 2 The filter

### 2.1 Mathematical analysis.

The Kalman filter is given by the following recursive equations:

$$\begin{aligned} a_{t|t-1} &= T_t a_{t-1} + c_t, \\ P_{t|t-1} &= T_t P_{t-1} T_t' + R_t Q_t R_t', \\ \tilde{y}_{t|t-1} &= Z_t a_{t|t-1} + d_t, \\ v_t &= y_t - \tilde{y}_{t|t-1}, \\ F_t &= Z_t P_{t|t-1} Z_t' + H_t, \\ a_t &= a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t, \\ P_t &= (I_m - P_{t|t-1} Z_t' F_t^{-1} Z_t) P_{t|t-1}. \end{aligned}$$

where  $a_{t|t-1}$  and  $P_{t|t-1}$  are the best estimates of  $\alpha_t$  and  $P_t$  conditionally on all the information available at time  $t - 1$ . The process  $v_t$  is the difference between the actual observation and its best predictor. This process is also called the innovation.

The log-likelihood of the observation at time  $t$  corresponds to

$$l_t = -N/2\ln(2\pi) - 1/2\ln |F_t| - 1/2v_t'F_t^{-1}v_t$$

where  $|F_t|$  is the determinant of  $F_t$ .

## 2.2 Implementation.

For the case where all the matrices and vectors  $T_t$ ,  $l_t$ ,  $R_t$ ,  $Z_t$ ,  $d_t$ ,  $Q_t$  and  $H_t$  are time-invariant, then the implementation of the formulas is straightforward. One only needs to store the matrices as is. Whenever the matrices and vectors are variable through time, then the implementation is slightly more involved because of a storage problem. For instance, to define  $T_t$  for  $t = 1, \dots, T$  we need to store  $T_t$  for each of the time periods. Roncalli solves the problem by stocking vertically the various matrices in a new matrix, say  $M$  of dimensions  $T \times (m \times m)$  where the  $i$ -th line of  $M$  contains the various rows of  $T_t$ . The instruction defining line  $i$  of  $M$  is

$$M(i,:) = \text{vecr}(T_t)'.$$

The `vecr` operator is implemented in LeSage's toolbox. You may wish to download this toolbox if you have not done it already. A parameter `timevar` should indicate ( $= 1$ ) if the matrices are time varying or not.

The implementation also adjusts  $F_t$  in the case that this matrix is zero.

The program `Kalman_filter.m` implements the filter. For larger applications, it is recommended to run some compiled code. A few years ago, it was nearly impossible to estimate univariate models with more than a few hundred observations. In 2004 I estimated a model with some 10'000 observations in 5 minutes.

## 3 Forecasting in a state-space model.

The forecast of  $y_t$ , for  $t \geq t_0$  is given by the following recursive equations

$$\begin{aligned} a_{t_0+l|t_0} &= T_{t_0+l}a_{t_0+l-1|t_0} + c_{t_0+l}, \\ P_{t_0+l|t_0} &= T_{t_0+l}P_{t_0+l-1|t_0}T_{t_0+l}' + R_{t_0+l}Q_{t_0+l}R_{t_0+l}^\top, \\ \hat{y}_{t_0+l|t_0} &= Z_{t_0+l}a_{t_0+l|t_0} + d_{t_0+l}. \end{aligned}$$

And the mean square error of the forecast on  $\hat{y}_{t_0+l|t_0}$  is given by

$$MSE(\hat{y}_{t_0+l|t_0}) = Z_{t_0+l}P_{t_0+l|t_0}Z'_{t_0+l} + H_{t_0+l}.$$

The implementation of this problem is straightforward and may be found in the Matlab module `Kalman_forecasting.m`.

## 4 Smoothing

By smoothing is meant the estimation of past values given all the available information. The name *smoother* comes from the fact that the estimated series look much smoother than the just filtered series... You will see this in the empirical implementation.

The smoothing equations are given by

$$\begin{aligned} P_t^* &= P_t T'_{t+1} P_{t+1|t}^{-1}, \\ a_{t|\tau} &= a_t + P_t^* (a_{t+1|\tau} - a_{t+1|t}), \\ P_{t|\tau} &= P_t + P_t^* (P_{t+1|\tau} - P_{t+1|t}) P_t^{*'} \end{aligned}$$

Notice that the smoother actually works backwards, starting from  $\tau$  and going back to 1.

Again, the implementation is easy. See also `Kalman_Smoothing.m`.

## 5 Simulation

The simulation of the Kalman filter is very easy. It boils down to a recursive series simulation. In certain situations such as the estimation of stochastic volatility models, the estimation may be more involved since the normality of the innovations may be challenged. Stochastic volatility models also involve more complex estimation techniques, based on a Bayesian paradigm. These techniques usually involve the Kalman filter. See `Kalman_Simulation.m` for a simulation code.

## 6 Uses of the Kalman filter

In this section we follow Harvey (1989) for the models and we follow Roncalli for the programs. Once these examples are well understood, the reader may adapt the programs for his own purposes and use data that is specific to his own research.

## 6.1 Local level / Random walk with noise

This model is written as

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & E[\varepsilon_t] &= 0, & V[\varepsilon_t] &= \sigma_\varepsilon^2, \\ \mu_t &= \mu_{t-1} + \eta_t, & E[\eta_t] &= 0, & V[\eta_t] &= \sigma_\eta^2. \end{aligned}$$

Both  $\varepsilon_t$  and  $\eta_t$  are supposed to be independent.

It is very easy to write this problem as a state-space model:

$$\begin{aligned} y_t &= 1 \times \mu_t + 0 + \varepsilon_t, \\ \mu_t &= 1 \times \mu_{t-1} + 0 + 1 \times \eta_t. \end{aligned}$$

Hence,  $y_t = y_t$ ,  $Z_t = 1$ ,  $d_t = 0$ ,  $H_t = \sigma_\varepsilon^2$ ,  $\alpha_t = \mu_t$ ,  $T_t = 1$ ,  $c_t = 0$ ,  $R_t = 1$ ,  $Q_t = \sigma_\eta^2$ .

Obviously, all the vectors and matrices are constant, hence, there is no time variation. The parameters that need to be estimated are  $\sigma_\varepsilon^2$  and  $\sigma_\eta^2$ . Various possibilities exist to estimate  $a_0$  and  $P_0$ . We set  $a_0 = y_1$  and  $P_0 = 0$ . It is easy to estimate the parameters with maximum likelihood.

An implementation of the problem may be found in `KLocal.m`.

Notice that  $(y_t - \hat{\mu}_t)/\hat{\sigma}_\varepsilon$  should be distributed iid normal with zero mean and unit variance. A test of goodness of fit can be built on this.

## 6.2 Local linear trend

This model is written as

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & E[\varepsilon_t] &= 0, & V[\varepsilon_t] &= \sigma_\varepsilon^2, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \zeta_{1t}, & E[\zeta_{1t}] &= 0, & V[\zeta_{1t}] &= \sigma_{\zeta_1}^2, \\ \beta_t &= \beta_{t-1} + \zeta_{2t}, & E[\zeta_{2t}] &= 0, & V[\zeta_{2t}] &= \sigma_{\zeta_2}^2. \end{aligned}$$

All the sources of uncertainty  $\varepsilon_t$ ,  $\zeta_{1t}$ , and  $\zeta_{2t}$  are assumed to be independent. Now, we may cast the model in the state-space representation. This gives

$$\begin{aligned} y_t &= [1 \quad 0] \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + 0 + \varepsilon_t, \\ \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + 0 + \eta_t \end{aligned}$$

Hence,

$$\begin{aligned}
y_t &= y_t, \\
Z_t &= [1 \ 0], \\
d_t &= 0, \\
H_t &= \sigma_\varepsilon^2, \\
\alpha_t &= [\mu_t \ \beta_t]', \\
T_t &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\
c_t &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\
R_t &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\
Q_t &= \begin{bmatrix} \sigma_{\zeta_1}^2 & 0 \\ 0 & \sigma_{\zeta_2}^2 \end{bmatrix}.
\end{aligned}$$

The parameters to be estimated are  $\sigma_\varepsilon^2$ ,  $\sigma_{\zeta_1}^2$ , and  $\sigma_{\zeta_2}^2$ . Again, the estimation may be easily done with maximum likelihood. One may use as initial values

$$a_0 = \begin{bmatrix} y(1) \\ 0 \end{bmatrix} \text{ and } P_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which boils down to using a non-informative prior, in bayesian-speak.

### 6.3 Linear model with time varying parameters.

This model may be written as

$$\begin{aligned}
y_t &= x_t' \beta_t + \varepsilon_t, & E[\varepsilon_t] &= 0, & V[\varepsilon_t] &= \sigma_\varepsilon^2, \\
\beta_t &= \beta_{t-1} + \eta_t, & E[\eta_t] &= 0, & V[\eta_t] &= \begin{bmatrix} \sigma_{\zeta_1}^2 & 0 \\ 0 & \sigma_{\zeta_2}^2 \end{bmatrix}.
\end{aligned}$$

Here, we have set up the model for the case where we have 2 parameters. Obviously, the model may be extended to the situation where there are more parameters and where there is some dependency between the parameters.

In this model, the parameters are integrated processes which means that they will be “smooth” series. The expected value of  $\beta_t$  is affected by the initial values  $\beta_0$ . Hence, for this model, the starting values for the drift are very important and need to be estimated.

The model is again easy to be cast into the setting of the state-space model. Here

$$\begin{aligned}
y_t &= y_t, \\
Z_t &= x'_t, \\
\alpha_t &= \beta_t, \\
d_t &= 0, \\
T_t &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\
Q_t &= \begin{bmatrix} \sigma_{\zeta_1}^2 & 0 \\ 0 & \sigma_{\zeta_2}^2 \end{bmatrix}.
\end{aligned}$$

The parameters to be estimated are  $\theta = (\sigma_\varepsilon^2, \sigma_{\zeta_1}^2, \sigma_{\zeta_2}^2, \beta_0)$ . (where  $\beta_0$  has, for the way the problem is set up, two components) Again, maximum likelihood may be used for an estimation.