Multivariate time series

State space models and Kalman filter

V. Lefieux

State space model
Kalman filter
Estimation
References

Table of contents

State space model

Kalman filter

Estimation

Plan

State space model

Kalman filter

Estimation

Definition

Let $(Y_t)_{t\in\mathbb{N}} \in \mathbb{R}^k$ be an observed process and $(X_t)_{t\in\mathbb{N}} \in \mathbb{R}^m$ be observed (considered as deterministic here).
Let $(S_t)_{t\in\mathbb{N}} \in \mathbb{R}^n$ be an unobserved process (state process).
Let $(u_t)_{t\in\mathbb{N}} \in \mathbb{R}^k$ and $(v_t)_{t\in\mathbb{N}^*} \in \mathbb{R}^n$ be two white noises.
Let $(F_t)_{t\in\mathbb{N}}$, $(G_t)_{t\in\mathbb{N}}$ and $(H_t)_{t\in\mathbb{N}}$ be three sequences of real matrixes with respective dimensions $m \times m$, $m \times n$ and $d \times m$.
Let $(P_t)_{t\in\mathbb{N}}$ a sequence of (transition) matrixes.
A state space model is:

$$\forall t \in \mathbb{N}^* : S_t = F_{t-1}X_{t-1} + P_{t-1}S_{t-1} + v_{t-1},$$
$$Y_t = G_tX_t + H_tS_t + u_t.$$

The first equation is called output equation and the second one state equation.
Matrixes F_t, G_t, H_t and P_t are known at time t.

State space model
Kalman filter
Estimation
References

Definition

We initialize the process with X_0 and S_0.
One assume that S_0 is uncorrelated with $(u_t)_{t\in\mathbb{N}}$ and $(v_t)_{t\in\mathbb{N}}$ and we use the notation:

$$\mathbb{E}(S_0) = \mu_0,$$
$$\mathrm{Var}(S_0) = \Sigma_0.$$

State space model can be seen as a generalization of many models, it can be estimated with a powerful tool: the Kalman filter.

Remarks

- Sometimes one can find $R_{t-1}v_{t-1}$ in place of v_{t-1} where R_{t-1} is a control term (used in control theory).
- State space representation isn't unique.
  The representation is said to be minimal when the dimension k of S_t is minimal.
- The model is said to be time invariant if the matrixes F_t, G_t, H_t and P_t don't depend on t.

Model

We consider the simplified model: invariant with uncorrelated white noises $(u_t)_{t\in\mathbb{N}}$ and $(v_t)_{t\in\mathbb{N}^*}$:

$$\forall t \in \mathbb{N}^* : S_t = FX_{t-1} + PS_{t-1} + v_{t-1},$$
$$Y_t = GX_t + HS_t + u_t.$$

We can write:

$$S_t = P^t S_0 + \sum_{i=1}^{t} P^{i-1} \left(FX_{t-i} + v_{t-i} \right).$$

Stationarity

We have:

$$\mathbb{E}\left(S_t\right) = P^t \mu_0 + \sum_{i=1}^{t} P^{i-1} F X_{t-i},$$

$$\begin{aligned}
\operatorname{Cov}\left(S_t, S_{t-h}\right) &= \mathbb{E}\left[\left(X_t - \mu\right)\left(X_{t-h} - \mu\right)^\top\right] \\
&= P^t \Sigma_0 \left(P^{t-h}\right)^\top + \sum_{i=1}^{t} P^{i-1} \Sigma_v \left(P^{i-h-1}\right)^\top.
\end{aligned}$$

$$\mathbb{E}\left(Y_t\right) = G X_t + H \mathbb{E}\left(S_t\right),$$

$$\operatorname{Cov}\left(Y_t, Y_{t-h}\right) = H \operatorname{Cov}\left(S_t, S_{t-h}\right) H^\top.$$

Thus the process $\left(Y_t\right)_{t \in \mathbb{Z}}$ isn't stationary in general.

Example: MA(1)

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$be a white noise with variance σ^2.
Let $(X_t)_{t\in\mathbb{Z}}$ be a MA (1) process:

$$X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}.$$

One can write:

- State equation ($n = 2$):

$$\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}.$$

- Output equation ($k = 1$):

$$X_t = \mu + \begin{bmatrix} 1 & -\theta \end{bmatrix} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}.$$

Example: AR(p) 1/3

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$be a white noise with variance σ^2.
Soit $(X_t)_{t\in\mathbb{Z}}$ un processus $AR(p)$:

$$\forall t \in \mathbb{Z} : X_t = \nu + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t,$$

soit encore:

$$\forall t \in \mathbb{Z} : X_t - \mu = \sum_{i=1}^{p} \varphi_i \left(X_{t-i} - \mu\right) + \varepsilon_t.$$

Example: AR(p) 2/3

State equation ($n = p$):

$$\begin{pmatrix} X_t - \mu \\ X_{t-1} - \mu \\ \vdots \\ X_{t-p+1} - \mu \end{pmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \varphi_3 & \cdots & \varphi_{p-1} & \varphi_p \\ 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \ddots & & 0 \\ \vdots & 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_{t-1} - \mu \\ X_{t-2} - \mu \\ \vdots \\ X_{t-p} - \mu \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Example: AR(p) 3/3

Output equation ($k = 1$):

$$X_t - \mu = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \begin{pmatrix} X_t - \mu \\ X_{t-1} - \mu \\ \vdots \\ X_{t-p+1} - \mu \end{pmatrix}.$$

Example: ARMA(p, q) 1/4

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$be a white noise with variance σ^2.
Let $(X_t)_{t\in\mathbb{Z}}$ be an ARMA(p, q) process:

$$\forall t \in \mathbb{Z} : X_t = \nu + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

that is:

$$\forall t \in \mathbb{Z} : X_t - \mu = \sum_{i=1}^{p} \varphi_i \left(X_{t-i} - \mu\right) + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

We can write:

$$\Phi\left(B\right)\left(X_t - \mu\right) = \Theta\left(B\right)\varepsilon_t$$

with:

$$\Phi\left(B\right) = \sum_{i=1}^{p} \varphi_i B^i, \Theta\left(B\right) = \sum_{j=1}^{q} \theta_j B^j.$$

Example: ARMA(p, q) 2/4

Consider the process $(Z_t)_{t\in\mathbb{Z}}$ such that:

$$\Phi(B) Z_t = \varepsilon_t.$$

We have:

$$X_t - \mu = \Theta(B) Z_t.$$

We define $r = \max(p, q+1)$ and:

$$\forall i \in \mathbb{N}, i > p : \varphi_i = 0,$$
$$\forall j \in \mathbb{N}, j > q : \theta_j = 0.$$

Example: ARMA(p, q) 3/4

State equation $(n = r)$:

$$\begin{pmatrix} U_t \\ U_{t-1} \\ \vdots \\ U_{t-r+1} \end{pmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \varphi_3 & \cdots & \varphi_{r-1} & \varphi_r \\ 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \ddots & & 0 \\ \vdots & 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} U_{t-1} \\ U_{t-2} \\ \vdots \\ U_{t-r} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Example: ARMA(p, q) 4/4

Output equation ($k = 1$):

$$X_t - \mu = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_{r-1} \end{bmatrix} \begin{pmatrix} U_t \\ U_{t-1} \\ \vdots \\ U_{t-r+1} \end{pmatrix}.$$

Remark

Many models can be written as state space models:

- ARMA processes with missing values,
- VAR, VARMA and VARMAX models,
- etc.

Plan

State space model

Kalman filter

Estimation

Kalman filter 1/5

We want to estimate the states of a process given observations.
Under normality hypothesis estimator is:

$$\mathbb{E}\left(S_t \,/\mathcal{F}_1^t\left(Y\right)\right).$$

Generally one calculate:

$$\mathrm{Var}\left(S_t \,/\mathcal{F}_1^t\left(Y\right)\right).$$

It is the filtering step (to distinguish to the foreacsting step).
We assume that:

$$\begin{aligned} u_t &\sim \mathcal{N}\left(0, \Sigma_u\right), \\ v_t &\sim \mathcal{N}\left(0, \Sigma_v\right), \\ S_0 &\sim \mathcal{N}\left(\mu_0, \Sigma_0\right). \end{aligned}$$

and that S_0 is independant of $\left(u_t\right)_{t\in\mathbb{N}}$ and $\left(v_t\right)_{t\in\mathbb{N}^*}$.

Kalman filter 2/5

We use the notations:

$$
\begin{aligned}
S_{t/s} &= \mathbb{E}\left(S_t \,/\mathcal{F}_1^s\left(Y\right)\right), \\
Y_{t/s} &= \mathbb{E}\left(Y_t \,/\mathcal{F}_1^s\left(Y\right)\right), \\
\Sigma_{t/s}^S &= \mathrm{Var}\left(S_t \,/\mathcal{F}_1^s\left(Y\right)\right), \\
\Sigma_{t/s}^Y &= \mathrm{Var}\left(Y_t \,/\mathcal{F}_1^s\left(Y\right)\right).
\end{aligned}
$$

We have:

$$
\begin{aligned}
&\forall t \in \{2, \ldots, T\} : \mathcal{L}\left(S_t \,/\mathcal{F}_1^{t-1}\left(Y\right)\right) = \mathcal{N}\left(S_{t/t-1}, \Sigma_{t/t-1}^S\right), \\
&\forall t \in \{1, \ldots, T\} : \mathcal{L}\left(S_t \,/\mathcal{F}_1^{t-1}\left(Y\right)\right) = \mathcal{N}\left(S_{t/t-1}, \Sigma_{t/t-1}^S\right), \\
&\forall t \in \{1, \ldots, T\} : \mathcal{L}\left(Y_t \,/\mathcal{F}_1^{t-1}\left(Y\right)\right) = \mathcal{N}\left(Y_{t/t-1}, \Sigma_{t/t-1}^Y\right), \\
&\forall t \in \mathbb{N}^*, t > T : \mathcal{L}\left(S_t \,\Big/\mathcal{F}_1^{T}\left(Y\right)\right) = \mathcal{N}\left(S_{t/T}, \Sigma_{t/T}^S\right), \\
&\forall t \in \mathbb{N}^*, t > T : \mathcal{L}\left(Y_t \,/\mathcal{F}_1^{t-1}\left(Y\right)\right) = \mathcal{N}\left(Y_{t/T}, \Sigma_{t/T}^Y\right).
\end{aligned}
$$

Kalman filter 3/5

We assume here that the state space parameters F, G, H, P, μ_0, Σ_0, Σ_u and Σ_v are known.
The Kalman filter is an multi steps algorithm:

- Initialization

$$S_{0/0} = \mu_0,$$
$$\Sigma^S_{0/0} = \Sigma_0.$$

- Update

$$\forall t \in \{1, \ldots, T\} : S_{t/t-1} = FX_{t-1} + PS_{t-1/t-1},$$
$$\Sigma^S_{t/t-1} = P\Sigma^S_{t-1/t-1}P^\top + \Sigma_v,$$
$$Y_{t/t-1} = GX_t + HS_{t/t-1},$$
$$\Sigma^Y_{t/t-1} = H\Sigma^S_{t/t-1}H^\top + \Sigma_u.$$

Kalman filter 4/5

- Correction

$$\forall t \in \{1, \ldots, T\} : S_{t/t} = S_{t/t-1} + Q_t \left(Y_t - Y_{t/t-1}\right),$$
$$\Sigma^S_{t/t} = \Sigma^S_{t/t-1} - Q_t \Sigma^Y_{t/t-1} Q_t^\top.$$

where Q_t is the gain matrix a time t defined by:

$$Q_t = \Sigma^S_{t/t-1} H \left(\Sigma^Y_{t/t-1}\right)^{-1}.$$

- Forecasting

$$\begin{aligned}
\forall t \in \mathbb{N}^*, t > T : S_{t/T} &= FX_{t-1} + PS_{t-1/T}, \\
\Sigma^S_{t/T} &= P\Sigma^S_{t-1/T} P^\top + \Sigma_v, \\
Y_{t/T} &= GX_t + HS_{t/T}, \\
\Sigma^Y_{t/T} &= H\Sigma^S_{t/T} H^\top + \Sigma_u.
\end{aligned}$$

Forecast are iteratively calculated.

Kalman filter 5/5

It's also possible to smooth a process with Kalman filter:

$$\forall t \in \{1, \ldots, T-1\} : S_{t/T} = S_{t/t} + Q_t \left(S_{t+1/T} - S_{t+1/t}\right),$$
$$\Sigma^S_{t/T} = \Sigma^S_{t/t} - Q_t \left(\Sigma^S_{t+1/t} - \Sigma^S_{t+1/T}\right) Q_t^\top$$

where:

$$Q_t = \Sigma^S_{t/t} P^\top \left(\Sigma^S_{t+1/t}\right)^{-1}.$$

Plan

State space model

Kalman filter

Estimation

Estimation 1/2

θ is the vector containing all the parameters F, G, H, P, μ_0, Σ_0, Σ_u and Σ_v.
Based on $(Y_1, \ldots, Y_T)$, under normality assumption, likelihood is:

$$
\begin{aligned}
& p\left(y_T, \ldots, y_1; \theta\right) \\
=& f_\theta\left(y_1\right) \prod_{t=2}^{T} f_\theta\left(y_t \middle/ \mathcal{F}_1^{t-1}(Y)\right) \\
=& f_\theta\left(y_1\right) \prod_{t=2}^{T} \frac{1}{\sqrt{(2\pi)^k \det\left(\Sigma_{t/t-1}^{Y}\right)}} \exp\left[-\frac{1}{2}\left(y_t - y_{t/t-1}\right)^\top \left(\Sigma_{t/t-1}^{Y}\right)^{-1}\left(y_t - y_{t/t-1}\right)\right].
\end{aligned}
$$

Estimation 2/2

Using the notations:

$$Y_{1/0} = \mathbb{E}\left(Y_1\right),$$
$$\Sigma_{1/0} = \mathrm{Var}\left(Y_1\right),$$

we write:

$$p\left(y_T, \ldots, y_1; \theta\right)$$
$$= \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^k \det\left(\Sigma^Y_{t/t-1}\right)}} \exp\left[-\frac{1}{2}\left(y_t - y_{t/t-1}\right)^\top \left(\Sigma^Y_{t/t-1}\right)^{-1}\left(y_t - y_{t/t-1}\right)\right].$$

For a given θ value, one can calculate the likelihood with the Kalman filter.
We obtain the maximum likelihood estimator by using a numerical optimization algorithm.

References

[1] James Douglas Hamilton. *Time series analysis*. Princeton University Press, 1994.

[2] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer, 2005.

[3] Ruey S. Tsay. *Multivariate time series analysis with R and financial applications*. Wiley, 2014.