Lecture 31

Filtering, Smoothing, & Forecasting
Part 2

Arnab Hazra

Definitions (recap)

- A primary aim of state space models is to produce estimators for the underlying unobserved signal $\boldsymbol{X}_t$, given the data $\mathcal{Y}_{1:s} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_s\}$.
- State estimation is an essential component of parameter estimation.
- In addition to these estimates, we would also want to measure their precision.
- When $s < t$, the problem is called forecasting or prediction.
- When $s = t$, the problem is called filtering.
- When $s > t$, the problem is called smoothing.

DLM with covariates (recap)

- The ARMAX model involves covariates that may enter into the states or into the observations.
- In this case, we suppose we have an $r \times 1$ vector of inputs $\boldsymbol{u}_t$, and write the model as

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \gamma \boldsymbol{u}_t + \boldsymbol{W}_t$$

$$\boldsymbol{Y}_t = \boldsymbol{A}_t \boldsymbol{X}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{V}_t$$

- Here γ is $p \times r$ and Γ is $q \times r$; either of these matrices may be the zero matrix.

Notations (recap)

We will use the following definitions:

- $$\boldsymbol{X}_t^s = E(\boldsymbol{X}_t | \mathcal{Y}_{1:s})$$

- $$\boldsymbol{P}_{t_1,t_2}^s = E[(\boldsymbol{X}_{t_1} - \boldsymbol{X}_{t_1}^s)(\boldsymbol{X}_{t_2} - \boldsymbol{X}_{t_2}^s)']$$

- $$\boldsymbol{P}_t^s = \boldsymbol{P}_{t,t}^s = \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^s)$$

- Due to Gaussian assumption,

  $$\boldsymbol{P}_{t_1,t_2}^s = E[(\boldsymbol{X}_{t_1} - \boldsymbol{X}_{t_1}^s)(\boldsymbol{X}_{t_2} - \boldsymbol{X}_{t_2}^s)' | \mathcal{Y}_{1:s}],$$

  follows from the fact that the covariance matrix between $(\boldsymbol{X}_t - \boldsymbol{X}_t^s)$ and $\mathcal{Y}_{1:s}$, for any t and s, is zero.

The Kalman Filter (recap)

- With initial conditions $\boldsymbol{X}_0^0 = \mu_0$ and $\boldsymbol{P}_0^0 = \Sigma_0$, for $t = 1, \ldots, T$,

$$\boldsymbol{X}_t^{t-1} = \Phi \boldsymbol{X}_{t-1}^{t-1} + \gamma \boldsymbol{u}_t, \quad \boldsymbol{P}_t^{t-1} = \Phi \boldsymbol{P}_{t-1}^{t-1} \Phi' + \boldsymbol{Q}$$

  with

$$\boldsymbol{X}_t^t = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{A}_t \boldsymbol{X}_t^{t-1} - \Gamma \boldsymbol{u}_t), \quad \boldsymbol{P}_t^t = [\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{A}_t] \boldsymbol{P}_t^{t-1}.$$

- Here the Kalman gain is

$$\boldsymbol{K}_t = \boldsymbol{P}_t^{t-1} \boldsymbol{A}_t' [\boldsymbol{A}_t \boldsymbol{P}_t^{t-1} \boldsymbol{A}_t' + \boldsymbol{R}]^{-1}.$$

- Prediction for $t > T$ is accomplished via $\boldsymbol{X}_t^{t-1} = \Phi \boldsymbol{X}_{t-1}^{t-1} + \gamma \boldsymbol{u}_t$ and $\boldsymbol{P}_t^{t-1} = \Phi \boldsymbol{P}_{t-1}^{t-1} \Phi' + \boldsymbol{Q}$ with initial conditions $\boldsymbol{X}_T^T$ and $\boldsymbol{P}_T^T$.

Proofs

- Show that $\boldsymbol{X}_t^{t-1} = \Phi \boldsymbol{X}_{t-1}^{t-1} + \gamma \boldsymbol{u}_t$.
-

$$
\begin{aligned}
\boldsymbol{X}_t^{t-1} &= E[\boldsymbol{X}_t | \mathcal{Y}_{1:(t-1)}] \\
&= E[\Phi \boldsymbol{X}_{t-1} + \gamma \boldsymbol{u}_t + \boldsymbol{W}_t | \mathcal{Y}_{1:(t-1)}] \\
&= \Phi E[\boldsymbol{X}_{t-1} | \mathcal{Y}_{1:(t-1)}] + \gamma \boldsymbol{u}_t + E[\boldsymbol{W}_t] \\
&= \Phi \boldsymbol{X}_{t-1}^{t-1} + \gamma \boldsymbol{u}_t
\end{aligned}
$$

Proofs (contd.)

- Show that $\boldsymbol{P}_t^{t-1} = \Phi \boldsymbol{P}_{t-1}^{t-1} \Phi' + \boldsymbol{Q}$.

-

$$
\begin{aligned}
\boldsymbol{P}_t^{t-1} &= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1} | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}([\Phi \boldsymbol{X}_{t-1} + \gamma \boldsymbol{u}_t + \boldsymbol{W}_t] - [\Phi \boldsymbol{X}_{t-1}^{t-1} + \gamma \boldsymbol{u}_t] | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}(\Phi [\boldsymbol{X}_{t-1} - \boldsymbol{X}_{t-1}^{t-1}] + \boldsymbol{W}_t | \mathcal{Y}_{1:(t-1)}) \\
&= \Phi \mathrm{Cov}[\boldsymbol{X}_{t-1} - \boldsymbol{X}_{t-1}^{t-1} | \mathcal{Y}_{1:(t-1)}] \Phi' + \mathrm{Cov}[\boldsymbol{W}_t] \\
&= \Phi \boldsymbol{P}_{t-1}^{t-1} \Phi' + \boldsymbol{Q}.
\end{aligned}
$$

The Kalman Filter (contd.)

- Important byproducts of the filter are the innovations (prediction errors)

$$\varepsilon_t = \boldsymbol{Y}_t - E(\boldsymbol{Y}_t | \mathcal{Y}_{1:(t-1)}) = \boldsymbol{Y}_t - \boldsymbol{A}_t \boldsymbol{X}_t^{t-1} - \Gamma \boldsymbol{u}_t,$$

  and the corresponding variance-covariance matrices

$$\Sigma_t \stackrel{def}{=} \mathrm{Cov}(\varepsilon_t) = \mathrm{Cov}[\boldsymbol{A}_t(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}) + \boldsymbol{V}_t] = \boldsymbol{A}_t \boldsymbol{P}_t^{t-1} \boldsymbol{A}_t' + \boldsymbol{R}$$

  for $t = 1, \ldots, T$.

- We assume that $\Sigma_t > 0$ (is positive definite), which is guaranteed, for example, if $\boldsymbol{R} > 0$. This assumption is not necessary and may be relaxed.

Proofs (contd.)

- We next show that $\boldsymbol{X}_t^t = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{A}_t\boldsymbol{X}_t^{t-1} - \Gamma \boldsymbol{u}_t) = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t\varepsilon_t$.
- We do not show the result but assume: the innovations are independent of the past observations, implying, $\mathrm{Cov}(\varepsilon_t, \boldsymbol{Y}_s) = 0$ for $s < t$. See Appendix B.1 of Shumway and Stoffer, if interested.
- Here $\boldsymbol{X}_t^{t-1}$ is a linear function of $\mathcal{Y}_{1:(t-1)}$ and hence $\boldsymbol{X}_t^{t-1} = \sum_{s=1}^{t-1} \boldsymbol{B}_s \boldsymbol{Y}_s$ for some $\boldsymbol{B}_s$.
- Hence, $\mathrm{Cov}(\boldsymbol{X}_t^{t-1}, \varepsilon_t) = \mathrm{Cov}(\sum_{s=1}^{t-1} \boldsymbol{B}_s \boldsymbol{Y}_s, \varepsilon_t) = \sum_{s=1}^{t-1} \boldsymbol{B}_s \mathrm{Cov}(\boldsymbol{Y}_s, \varepsilon_t) = 0$.
- Besides, by the previously used logic, $\mathrm{Cov}(\boldsymbol{X}_t^{t-1}, \varepsilon_t | \mathcal{Y}_{1:(t-1)}) = 0$.

Proofs (contd.)

- First, prove that $\mathrm{Cov}(\boldsymbol{X}_t, \varepsilon_t | \mathcal{Y}_{1:(t-1)}) = \boldsymbol{P}_t^{t-1} \boldsymbol{A}_t'$.
- $$\begin{aligned}
\mathrm{Cov}(\boldsymbol{X}_t, \varepsilon_t | \mathcal{Y}_{1:(t-1)}) &= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}, \varepsilon_t | \mathcal{Y}_{1:(t-1)}) + \mathrm{Cov}(\boldsymbol{X}_t^{t-1}, \varepsilon_t | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}, \varepsilon_t | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}, \boldsymbol{Y}_t - \boldsymbol{A}_t \boldsymbol{X}_t^{t-1} - \Gamma \boldsymbol{u}_t | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}, \boldsymbol{A}_t \boldsymbol{X}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{V}_t - \boldsymbol{A}_t \boldsymbol{X}_t^{t-1} - \Gamma \boldsymbol{u}_t | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}, \boldsymbol{A}_t [\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}] + \boldsymbol{V}_t | \mathcal{Y}_{1:(t-1)}) \\
&= \mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}) \boldsymbol{A}_t' \\
&= \boldsymbol{P}_t^{t-1} \boldsymbol{A}_t'.
\end{aligned}$$

Proofs (contd.)

- Now prove that $\boldsymbol{X}_t^t = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{A}_t\boldsymbol{X}_t^{t-1} - \Gamma\boldsymbol{u}_t) = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t\varepsilon_t$.

- $$\begin{bmatrix} \boldsymbol{X}_t \\ \varepsilon_t \end{bmatrix} |\mathcal{Y}_{1:(t-1)} \sim \mathcal{N}_{p+q}\left(\begin{bmatrix} \boldsymbol{X}_t^{t-1} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_t^{t-1} & \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t' \\ \boldsymbol{A}_t\boldsymbol{P}_t^{t-1} & \Sigma_t \end{bmatrix}\right)$$

- Hence,

$$\boldsymbol{X}_t^t = E[\boldsymbol{X}_t|\mathcal{Y}_{1:t}] = E[\boldsymbol{X}_t|\mathcal{Y}_{1:(t-1)}, \varepsilon_t] = \boldsymbol{X}_t^{t-1} + \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t'\Sigma_t^{-1}\varepsilon_t = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t\varepsilon_t,$$

where $\boldsymbol{K}_t = \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t'\Sigma_t^{-1}$.

Proofs (contd.)

- Now prove that $\Sigma_t \stackrel{def}{=} \mathrm{Cov}(\varepsilon_t) = \mathrm{Cov}[\boldsymbol{A}_t(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}) + \boldsymbol{V}_t] = \boldsymbol{A}_t\boldsymbol{P}_t^{t-1}\boldsymbol{A}_t' + \boldsymbol{R}$
- $$\begin{aligned}
\Sigma_t &= \mathrm{Cov}(\varepsilon_t) \\
&= \mathrm{Cov}(\boldsymbol{Y}_t - \boldsymbol{A}_t X_t^{t-1} - \Gamma \boldsymbol{u}_t) \\
&= \mathrm{Cov}(\boldsymbol{A}_t\boldsymbol{X}_t + \Gamma\boldsymbol{u}_t + \boldsymbol{V}_t - \boldsymbol{A}_t X_t^{t-1} - \Gamma\boldsymbol{u}_t) \\
&= \mathrm{Cov}(\boldsymbol{A}_t[\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1}] + \boldsymbol{V}_t) \\
&= \boldsymbol{A}_t\mathrm{Cov}(\boldsymbol{X}_t - \boldsymbol{X}_t^{t-1})\boldsymbol{A}_t' + \mathrm{Cov}(\boldsymbol{V}_t) \\
&= \boldsymbol{A}_t\boldsymbol{P}_t^{t-1}\boldsymbol{A}_t' + \boldsymbol{R}
\end{aligned}$$
- Hence,
$\boldsymbol{X}_t^t = \boldsymbol{X}_t^{t-1} + \boldsymbol{K}_t\varepsilon_t = \boldsymbol{X}_t^{t-1} + \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t'\Sigma_t^{-1}\varepsilon_t = \boldsymbol{X}_t^{t-1} + \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t'[\boldsymbol{A}_t\boldsymbol{P}_t^{t-1}\boldsymbol{A}_t' + \boldsymbol{R}]^{-1}\varepsilon_t$

Proofs (contd.)

- Now prove that $\boldsymbol{P}_t^t = [\boldsymbol{I}_p - \boldsymbol{K}_t\boldsymbol{A}_t]\boldsymbol{P}_t^{t-1}$
-

$$\begin{bmatrix} \boldsymbol{X}_t \\ \varepsilon_t \end{bmatrix} | \mathcal{Y}_{1:(t-1)} \sim \mathcal{N}_{p+q}\left(\begin{bmatrix} \boldsymbol{X}_t^{t-1} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_t^{t-1} & \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t' \\ \boldsymbol{A}_t\boldsymbol{P}_t^{t-1} & \Sigma_t \end{bmatrix} \right)$$

- Thus, we have

$$\begin{aligned} \boldsymbol{P}_t^t &= \mathrm{Cov}(\boldsymbol{X}_t | \mathcal{Y}_{1:t}) \\ &= \mathrm{Cov}(\boldsymbol{X}_t | \mathcal{Y}_{1:(t-1)}, \varepsilon_t) \\ &= \boldsymbol{P}_t^{t-1} - \boldsymbol{P}_t^{t-1}\boldsymbol{A}_t'\Sigma_t^{-1}\boldsymbol{A}_t\boldsymbol{P}_t^{t-1} \\ &= \boldsymbol{P}_t^{t-1} - [\boldsymbol{P}_t^{t-1}\boldsymbol{A}_t'\Sigma_t^{-1}]\boldsymbol{A}_t\boldsymbol{P}_t^{t-1} \\ &= \boldsymbol{P}_t^{t-1} - \boldsymbol{K}_t\boldsymbol{A}_t\boldsymbol{P}_t^{t-1} \\ &= [\boldsymbol{I}_p - \boldsymbol{K}_t\boldsymbol{A}_t]\boldsymbol{P}_t^{t-1}. \end{aligned}$$

The Kalman Smoother

For the DLM with covariates, with initial conditions $\boldsymbol{X}_T^T$ and $\boldsymbol{P}_T^T$ obtained Kalman Filter, for $t = T, T-1, \ldots, 1$,

- $$\boldsymbol{X}_{t-1}^T = \boldsymbol{X}_{t-1}^{t-1} + \boldsymbol{J}_{t-1}(\boldsymbol{X}_t^T - \boldsymbol{X}_t^{t-1}),$$

- $$\boldsymbol{P}_{t-1}^T = \boldsymbol{P}_{t-1}^{t-1} + \boldsymbol{J}_{t-1}(\boldsymbol{P}_t^T - \boldsymbol{P}_t^{t-1})\boldsymbol{J}_{t-1}',$$

  where

  $$\boldsymbol{J}_{t-1} = \boldsymbol{P}_{t-1}^{t-1}\Phi'[\boldsymbol{P}_t^{t-1}]^{-1}.$$

Proof: Read from the book

The Lag-One Covariance Smoother

For the DLM with covariates, with $\boldsymbol{K}_t$, $\boldsymbol{J}_t$ $(t = 1, \ldots, T)$, and $\boldsymbol{P}_T^T$ obtained from Kalman filter and Kalman smoother, and with initial condition

$$\boldsymbol{P}_{T,T-1}^{T} = (\boldsymbol{I}_q - \boldsymbol{K}_T \boldsymbol{A}_T)\Phi \boldsymbol{P}_{T-1}^{T-1},$$

- For $t = T, T-1, \ldots, 2,$

$$\boldsymbol{P}_{t-1,t-2}^{T} = \boldsymbol{P}_{t-1}^{t-1}\boldsymbol{J}_{t-2}' + \boldsymbol{J}_{t-1}(\boldsymbol{P}_{t,t-1}^{T} - \Phi \boldsymbol{P}_{t-1}^{t-1})\boldsymbol{J}_{t-2}'.$$

Proof: Read from the book

Thank you!