Lecture 32

MLE for state-space models

Arnab Hazra

DLM with covariates (recap)

- 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.
- In the DLM, we assume the process starts with a normal vector $\boldsymbol{X}_0$, such that $\boldsymbol{X}_0 \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma_0)$.
- Here $\boldsymbol{W}_t \overset{IID}{\sim} \mathcal{N}_p(\mathbf{0}, \boldsymbol{Q})$ and the additive observation noise is $\boldsymbol{V}_t \overset{IID}{\sim} \mathcal{N}_q(\mathbf{0}, \boldsymbol{R})$.

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

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

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

ML estimation problem

- We represent the vector of unknown parameters as

$$\boldsymbol{\Theta} = \{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\Phi}, \gamma, \boldsymbol{\Gamma}, \boldsymbol{Q}, \boldsymbol{R}\}.$$

- Note that the innovations are independent Gaussian random vectors with zero means and covariance matrices $\boldsymbol{\Sigma}_t = \boldsymbol{A}_t \boldsymbol{P}_t^{t-1} \boldsymbol{A}_t' + \boldsymbol{R}$.
- The likelihood is computed using the innovations ε_t, $t = 1, \ldots, T$, defined by

$$-\log L_{\mathcal{Y}}(\boldsymbol{\Theta}) = \frac{1}{2} \sum_{t=1}^{T} \log |\boldsymbol{\Sigma}_t(\boldsymbol{\Theta})| + \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t(\boldsymbol{\Theta})' \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\Theta}) \varepsilon_t(\boldsymbol{\Theta}).$$

- $-\log L_{\mathcal{Y}}(\boldsymbol{\Theta})$ is a highly nonlinear and complicated function of $\boldsymbol{\Theta}$.
- A Newton–Raphson algorithm can be used successively to update the parameter values until the negative log-likelihood is minimized.

Newton–Raphson steps

1. Select initial values for the parameters, say, $\boldsymbol{\Theta}^{(0)}$.
2. Run the Kalman filter using the initial parameter values $\boldsymbol{\Theta}^{(0)}$ to obtain a set of innovations and error covariances, say, $\{\varepsilon_t^{(0)}; t = 1, \ldots, T\}$ and $\{\boldsymbol{\Sigma}_t^{(0)}; t = 1, \ldots, T\}$.
3. Run one iteration of a Newton–Raphson procedure with $-\log L_{\mathcal{Y}}(\boldsymbol{\Theta})$ as the objective function, to obtain a new set of estimates, say $\boldsymbol{\Theta}^{(1)}$.
4. At iteration j, repeat step 2 using $\boldsymbol{\Theta}^{(j)}$ in place of $\boldsymbol{\Theta}^{(j-1)}$ to obtain a new set of innovation values $\{\varepsilon_t; t = 1, \ldots, T\}$ and $\{\boldsymbol{\Sigma}_t; t = 1, \ldots, T\}$
5. Then repeat step 3 to obtain a new estimate $\boldsymbol{\Theta}^{(j+1)}$. Stop when the values of $\boldsymbol{\Theta}^{(j+1)}$ differ from $\boldsymbol{\Theta}^{(j)}$, or when $L_{\mathcal{Y}}(\boldsymbol{\Theta}^{(j+1)})$ differs from $L_{\mathcal{Y}}(\boldsymbol{\Theta}^{(j)})$, by some predetermined, but small amount.

Example of DLM (Global temperature data, recap)

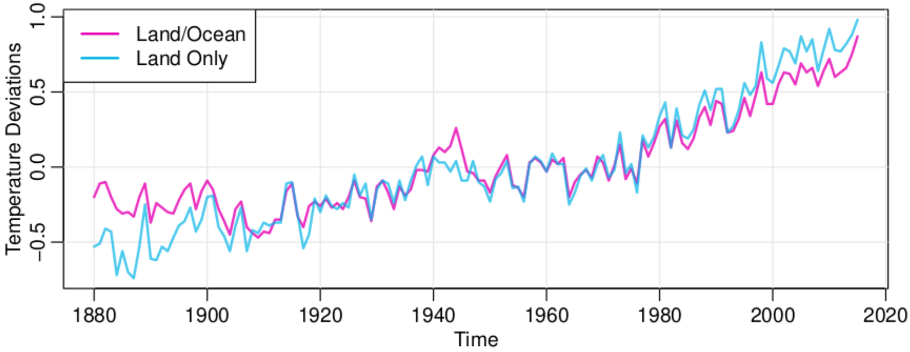

Fig. 6.3. *Annual global temperature deviation series, measured in degrees centigrade, 1880–2015. The series differ by whether or not ocean data is included.*

Example of DLM (Global temperature data, recap)

- They show two different estimators for the global temperature series from 1880 to 2015.
- First series are the global mean land-ocean temperature index data.
- The second series are the surface air temperature index data using only meteorological station data.
- Conceptually, both series should be measuring the same underlying climatic signal, and we may consider the problem of extracting this underlying signal.

Example of DLM (Global temperature data, recap)

- We suppose both series are observing the same signal with different noises; that is, $Y_{t1} = X_t + V_{t1}$ and $Y_{t2} = X_t + V_{t2}$ or more compactly as

$$(Y_{t1}, Y_{t2})' = (1, 1)' X_t + (V_{t1}, V_{t2})',$$

  where $\boldsymbol{R} = \mathrm{Var}[(V_{t1}, V_{t2})']$.

- It is reasonable to suppose that the unknown common signal X_t can be modeled as a random walk with drift of the form

$$X_t = \delta + X_{t-1} + W_t,$$

  with $Q = \mathrm{Var}(W_t)$.

- In this example, $p = 1$, $q = 2$, $\Phi = 1$, and $\gamma = \delta$ with $u_t = 1$.

Example of DLM (Global temperature data, recap)

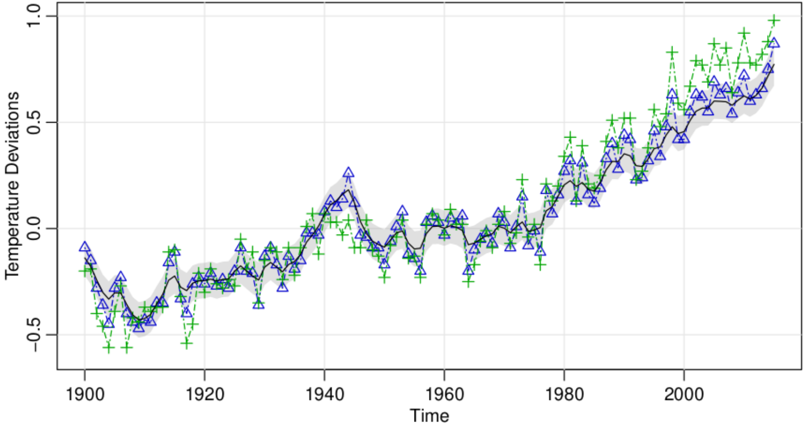

Fig. 6.5. *Plot for Example 6.7. The dashed lines with points (+ and △) are the two average global temperature deviations shown in Figure 6.3. The solid line is the estimated smoother* $\hat{x}_t^n$*, and the corresponding two root mean square error bound is the gray swatch. Only the values later than 1900 are shown.*

Example: An AR(1) Process with Observational Noise (recap)

- Consider a univariate state-space model where the observations are noisy,

$$Y_t = X_t + V_t$$

- The signal (state) is an AR(1) process,

$$X_t = \phi X_{t-1} + W_t$$

- Here $V_t \overset{IID}{\sim} \mathcal{N}(0, \sigma_V^2)$, $W_t \overset{IID}{\sim} \mathcal{N}(0, \sigma_W^2)$, and $X_0 \sim \mathcal{N}(0, (1-\phi^2)^{-1}\sigma_W^2)$

- Besides, X_0, $\{W_t\}$, and $\{V_t\}$ are independent.

Example: An AR(1) Process with Observational Noise (contd.)

- The autocovariance function of X_t is

$$\gamma_X(h) = (1-\phi^2)^{-1}\sigma_W^2\phi^h, \quad h = 0, 1, 2, \ldots$$

- The marginal variance of Y_t is

$$\gamma_Y(0) = \mathrm{Var}(X_t + V_t) = \mathrm{Var}(X_t) + \mathrm{Var}(V_t) = (1-\phi^2)^{-1}\sigma_W^2 + \sigma_V^2$$

- The autocovariance function of Y_t is

$$\gamma_Y(h) = \mathrm{Cov}(X_{t+h}+V_{t+h}, X_t+V_t) = \mathrm{Cov}(X_{t+h}, X_t) = (1-\phi^2)^{-1}\sigma_W^2\phi^h, \; h = 1, 2, \ldots$$

- The ACF of Y_t is

$$\rho_Y(h) = \gamma_Y(h)/\gamma_Y(0) = (1+\sigma_V^2/\sigma_W^2(1-\phi^2))^{-1}\phi^h, \; h = 1, 2, \ldots$$

- The ACF of Y_t is identical to the ACF of an ARMA(1,1) process.

Example: Initial parameter value selection

- Here $\rho_Y(2)/\rho_Y(1) = \phi$.
- Hence, we can calculate the sample ACF and choose $\phi^{(0)} = \hat{\rho}_Y(2)/\hat{\rho}_Y(1)$
- Next we have $\gamma_Y(1) = (1-\phi^2)^{-1}\sigma_W^2\phi$.
- Hence, we can calculate sample autocovariance $\hat{\gamma}_Y(1)$ and obtain $\sigma_W^{2(0)}$ by the equation $\hat{\gamma}_Y(1) = (1-[\phi^{(0)}]^2)^{-1}\sigma_W^{2(0)}\phi^{(0)}$.
- Next we have $\gamma_Y(0) = (1-\phi^2)^{-1}\sigma_W^2 + \sigma_V^2$.
- Hence, we can calculate $\hat{\gamma}_Y(0)$ and obtain $\sigma_V^{2(0)}$ by the equation $\hat{\gamma}_Y(0) = (1-[\phi^{(0)}]^2)^{-1}\sigma_W^{2(0)} + \sigma_V^{2(0)}$.

Thank you!