Time Series Analysis

Henrik Madsen
hm@imm.dtu.dk

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby
Outline of the lecture

State space models, 2nd part:
- The Kalman filter when some observations are missing
- ARMA-models on state space form, Sec. 10.4 (not 10.4.1)
- ML-estimates of state space models, Sec. 10.6

Cursory material:
- Signal extraction, Sec. 10.4.1
- Time series with missing observations, Sec. 10.5
The linear stochastic state space model

System equation: \( X_t = AX_{t-1} + Bu_{t-1} + e_{1,t} \)
Observation equation: \( Y_t = CX_t + e_{2,t} \)

- \( X \): State vector
- \( Y \): Observation vector
- \( u \): Input vector
- \( e_1 \): System noise
- \( e_2 \): Observation noise

\[ \text{dim}(X_t) = m \] is called the order of the system
\( \{e_{1,t}\} \) and \( \{e_{2,t}\} \) mutually independent white noise
\( V[e_1] = \Sigma_1, V[e_2] = \Sigma_2 \)
\( A, B, C, \Sigma_1, \text{ and } \Sigma_2 \) are known matrices
The Kalman filter

Initialization: \( \hat{X}_{1|0} = \mu_0, \Sigma_{1|0} = V_0 \Rightarrow \Sigma_{1|0} = C\Sigma_{1|0}C^T + \Sigma_2 \)

For: \( t = 1, 2, 3, \ldots \)

\[
K_t = \Sigma_{t|t-1} \Sigma_{t|t-1} C^T \left( \Sigma_{t|t-1} C^T \right)^{-1}
\]

Reconstruction:

\[
\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t \left( Y_t - C\hat{X}_{t|t-1} \right)
\]

\[
\Sigma_{t|t} = \Sigma_{t|t-1} - K_t \Sigma_{t|t-1} K_t^T
\]

Prediction:

\[
\hat{X}_{t+1|t} = A\hat{X}_{t|t} + Bu_t
\]

\[
\Sigma_{t+1|t} = A\Sigma_{t|t} A^T + \Sigma_1
\]

\[
\Sigma_{t+1|t} = C\Sigma_{t|t} C^T + \Sigma_2
\]
The Kalman filter

Initialization: \( \hat{X}_{1|0} = \mu_0, \Sigma_{1|0}^{xx} = V_0 \Rightarrow \Sigma_{1|0}^{yy} = C \Sigma_{1|0}^{xx} C^T + \Sigma_2 \)

For: \( t = 1, 2, 3, \ldots \)

Reconstruction:
\[
K_t = \Sigma_{t|t-1}^{xx} C^T \left( \Sigma_{t|t-1}^{yy} \right)^{-1}
\]
\[
\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t \left( Y_t - C \hat{X}_{t|t-1} \right)
\]
\[
\Sigma_{t|t}^{xx} = \Sigma_{t|t-1}^{xx} - K_t \Sigma_{t|t-1}^{yy} K_t^T
\]

Prediction:
\[
\hat{X}_{t+1|t} = A \hat{X}_{t|t} + B u_t
\]
\[
\Sigma_{t+1|t}^{xx} = A \Sigma_{t|t}^{xx} A^T + \Sigma_1
\]
\[
\Sigma_{t+1|t}^{yy} = C \Sigma_{t+1|t}^{xx} C^T + \Sigma_2
\]

What happens if the observation \( Y_t \) is missing for some \( t \)?
Estimation in $ARMA(p, q)$-models using the KF

- Using the Kalman filter we can get the mean and variance of the one-step predictions of the observations:

$$\hat{Y}_{t+1|t} = C\hat{X}_{t+1|t}$$
$$\Sigma_{yy_{t+1|t}} = C\Sigma_{xx_{t+1|t}}C^T + \Sigma_2$$

- The Kalman filter can handle missing observations
- An $ARMA(p, q)$-model can be written as a state space model

- This gives us a way of calculating ML-estimates in the $ARMA(p, q)$-model even when some observations are missing.
**ARMA**\((p, q)\)-models on state space form

\[
Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_t \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}
\]

State space form:

\[
X_t = AX_{t-1} + e_{1,t}, \quad Y_t = CX_t
\]

\[
X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})^T, \quad d = \max(p, q + 1)
\]

\[
A = \begin{bmatrix}
-\phi_1 & 1 & 0 & \cdots & 0 \\
-\phi_2 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\phi_{d-1} & 0 & 0 & \cdots & 1 \\
-\phi_d & 0 & 0 & \cdots & 0 \\
\end{bmatrix}, \quad e_{1,t} = G\varepsilon_t = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{d-1} \end{bmatrix} \varepsilon_t
\]

\[
C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}
\]
ML-estimates in state space models

\[
X_t = AX_{t-1} + Ge_{1,t} \\
Y_t = CX_t + e_{2,t}
\]

- \{e_{1,t}\} and \{e_{2,t}\} are mutually uncorrelated normally distributed white noise.
- \(V(e_{1,t}) = \Sigma_1\) and \(V(e_{2,t}) = \Sigma_2\).
- For \(ARMA(p, q)\)-models we have \(A\), \(C\), and \(G\) as stated on the previous slide. Furthermore, \(e_{1,t} = \varepsilon_t\), \(\Sigma_1 = \sigma^2_{\varepsilon}\), and \(\Sigma_2 = 0\).
Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\theta$ contain the parameters of the model.

- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\theta; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*}|\theta)$$

- The ML-estimates is found by selecting $\theta$ so that the density function is as large as possible at the actual observations.

- The random variables $Y_{N^*}|\mathcal{Y}_{N^*-1}$ and $\mathcal{Y}_{N^*-1}$ are independent:

$$L(\theta; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*}|\theta) = f(Y_{N^*}|\mathcal{Y}_{N^*-1}, \theta) f(\mathcal{Y}_{N^*-1}|\theta)$$

$$= f(Y_{N^*}|\mathcal{Y}_{N^*-1}, \theta) f(Y_{N^*-1}|\mathcal{Y}_{N^*-2}, \theta) \cdots f(Y_1|\theta)$$

- The conditional densities can be found using the Kalman filter.
MLE / KF

- Assume that at time $t$ we have:
  \[
  \hat{X}_{t|t} = E[ X_t | \mathcal{Y}_t ] \quad \text{and} \quad \Sigma_{x|x|t} = V[ X_t | \mathcal{Y}_t ]
  \]

- Using the model we obtain predictions for time $t + 1$: 
  \[
  \hat{X}_{t+1|t} = A \hat{X}_{t|t}
  \]
  \[
  \Sigma_{x|x|t+1} = A \Sigma_{x|x|t} A^T + G \Sigma_1 G^T
  \]
  \[
  \hat{Y}_{t+1|t} = C \hat{X}_{t+1|t}
  \]
  \[
  \Sigma_{y|y|t+1} = C \Sigma_{x|x|t+1} C^T + \Sigma_2
  \]

- Due to the normality of the white noise process $f(Y_{t+1|\mathcal{Y}_t, \theta})$ is then the (multivariate) normal density (see Chapter 2) with mean $\hat{Y}_{t+1|t}$ and variance-covariance $\Sigma_{y|y|t+1|t} (= R_{t+1})$
MLE / KF (cont’nd)

At time \( t + 1 \) there is two possibilities:

The observation \( Y_{t+1} \) is available: We update the state estimate using the reconstruction step of the Kalman Filter:

\[
K_{t+1} = \Sigma_{x|x|t+1} C^T \left( \Sigma_{y|y|t+1} \right)^{-1}
\]

\[
\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1} \left( Y_{t+1} - \hat{Y}_{t+1|t} \right)
\]

\[
\Sigma_{x|x|t+1} = \Sigma_{x|x|t} - K_{t+1} \Sigma_{y|y|t+1} K_{t+1}^T
\]

The observation \( Y_{t+1} \) is missing: We got no new information and we use:

\[
\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t}
\]

\[
\Sigma_{x|x|t+1} = \Sigma_{x|x|t}
\]

And then we predict for time \( t + 2 \)
MLE / KF (cont’nd)

- Using the prediction errors and variances

\[ \tilde{Y}_i = Y_i - \hat{Y}_{i|i-1} \]
\[ R_i = \sum_{i|i-1}^{y y} \]

- The likelihood function can be expressed as

\[
L(\theta; \mathcal{Y}_{N^*}) = \prod_{i=1}^{N^*} \left[ (2\pi)^m \det R_i \right]^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \tilde{Y}_i^T R_i^{-1} \tilde{Y}_i \right]
\]

- In practice optimization is based on \( \log L(\theta; \mathcal{Y}_{N^*}) \) and the variance of the estimates can be approximated by the 2’nd order derivatives of log-likelihood.
The only outstanding issue is “prediction” of $Y_1$, i.e. calculation of $\hat{Y}_{1|0}$

This can be done by setting $\hat{X}_{0|0} = 0$ and $\Sigma_{0|0}^{xx} = \alpha I$, where $I$ is the identity matrix and $\alpha$ is a ‘large’ constant (we don’t know what it is)

Alternatively, we can estimate the initial state $\hat{X}_{0|0}$ and set $\Sigma_{0|0}^{xx} = 0$, whereby $\Sigma_{1|0}^{xx} = G \Sigma_1 G^T$