Lecture 13

State-Space Models I

Readings: SS17 Chapter 6.1-6.2; BD Chapter 9.1-9.3

MATH 8090 Time Series Analysis

Week 13

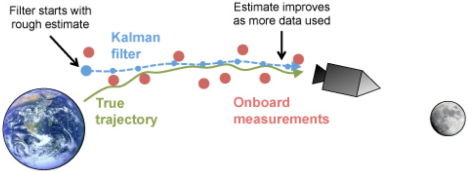
Whitney Huang
Clemson University

Agenda

1. **Background**
2. **Multivariate Gaussian and Regression Lemmas**
3. **Forecasting, Filtering, and Smoothing**

Historical Background

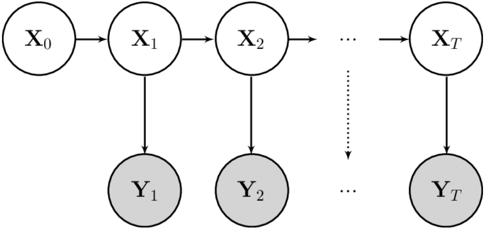
- The original model emerged in the context of space tracking [Kalman, 1960, Kalman and Bucy, 1961]
- The "state equation" defines the motion equations for the position of a spacecraft with location $\boldsymbol{x}_t$



- The data $\boldsymbol{y}_t$ reflect information that can be observed from a tracking device, such as velocity and azimuth

> The main goal was to retrieve the underling state $\{x_t\}$ based on observed data $\{y_t\}$

State-Space Model



$$\text{State: } \boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t, \quad \boldsymbol{V}_t \overset{i.i.d.}{\sim} \mathrm{WN}(\mathbf{0}, Q_t), \quad t = 1, 2, \cdots$$

$$\text{Observation: } \boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t, \quad \boldsymbol{W}_t \overset{i.i.d.}{\sim} \mathrm{WN}(\mathbf{0}, R_t), \quad t = 1, 2, \cdots$$

- $\boldsymbol{X}_{\boldsymbol{t}} \in \mathbb{R}^p$ and $\boldsymbol{Y}_{\boldsymbol{t}} \in \mathbb{R}^q$ are the state vector and the observation vector at time t
- $\boldsymbol{M}_t$ is the $p \times p$ transition matrix, and $\boldsymbol{H}_t$ is the $q \times p$ observation matrix
- $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ are the state and observation noises

Additional Assumptions of State-Space Models

State equation:

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t, \quad t = 1, 2 \cdots$$

Observation equation:

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t, \quad t = 1, 2, \cdots$$

- $\mathrm{E}(\boldsymbol{W}_s \boldsymbol{V}_t^T) = 0$ for all s and t, that is, every observation noise is uncorrelated with every state-transition noise

Additional Assumptions of State-Space Models

State equation:

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t, \quad t = 1, 2 \cdots$$

Observation equation:

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t, \quad t = 1, 2, \cdots$$

- $\mathrm{E}(\boldsymbol{W}_s \boldsymbol{V}_t^T) = 0$ for all s and t, that is, every observation noise is uncorrelated with every state-transition noise

- Assuming $\mathrm{E}(\boldsymbol{X}_0) = \boldsymbol{\mu}_0$, $\mathrm{E}(\boldsymbol{X}_0 \boldsymbol{W}_t^T) = 0$ and $\mathrm{E}(\boldsymbol{X}_0 \boldsymbol{V}_t^T) = 0$ for all t, that is, initial state vector are uncorrelated with both observation and state transition noises

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The Kalman recursions for state-space models provide elegant solution for forecasting, filtering, and smoothing

  To estimate $\boldsymbol{X}_t$ with $\boldsymbol{Y}_{1:s} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_s\}$:

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The Kalman recursions for state-space models provide elegant solution for forecasting, filtering, and smoothing

  To estimate $\boldsymbol{X}_t$ with $\boldsymbol{Y}_{1:s} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_s\}$:
  - When $s < t \Rightarrow$ forecasting

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The Kalman recursions for state-space models provide elegant solution for forecasting, filtering, and smoothing

  To estimate $\boldsymbol{X}_t$ with $\boldsymbol{Y}_{1:s} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_s\}$:
  - When $s < t \Rightarrow$ forecasting
  - When $s = t \Rightarrow$ filtering

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The Kalman recursions for state-space models provide elegant solution for forecasting, filtering, and smoothing

  To estimate $\boldsymbol{X}_t$ with $\boldsymbol{Y}_{1:s} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_s\}$:
  - When $s < t \Rightarrow$ forecasting
  - When $s = t \Rightarrow$ filtering
  - When $s > t \Rightarrow$smoothing

Applications of State-Space Models

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The Kalman recursions for state-space models provide elegant solution for forecasting, filtering, and smoothing

  To estimate $\boldsymbol{X}_t$ with $\boldsymbol{Y}_{1:s} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_s\}$:
  - When $s < t \Rightarrow$ forecasting
  - When $s = t \Rightarrow$ filtering
  - When $s > t \Rightarrow$smoothing
- State-space models and Kalman recursions can be readily adapted to handle time series with missing values

AR(1) Process as a State-Space Model: I

- State-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $|\phi| < 1$

AR(1) Process as a State-Space Model: I

- State-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

  is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

  with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting

AR(1) Process as a State-Space Model: I

- State-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\boldsymbol{X}_t = Y_t$; $M_t = \phi$

AR(1) Process as a State-Space Model: I

- State-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

  is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

  with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\boldsymbol{X}_t = Y_t$; $M_t = \phi$
  - $\boldsymbol{V}_t = Z_t$ along with $Q_t \stackrel{\text{def}}{=} \mathrm{E}(\boldsymbol{V}_t \boldsymbol{V}_t^T) = \mathrm{E}(Z_t^2) = \sigma_Z^2$

AR(1) Process as a State-Space Model: I

- State-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

  is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

  with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\boldsymbol{X}_t = Y_t$; $M_t = \phi$
  - $\boldsymbol{V}_t = Z_t$ along with $Q_t \stackrel{\text{def}}{=} \mathrm{E}(\boldsymbol{V}_t \boldsymbol{V}_t^T) = \mathrm{E}(Z_t^2) = \sigma_Z^2$

AR(1) Process as a State-Space Model: I

- State-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\boldsymbol{X}_t = Y_t$; $M_t = \phi$
  - $\boldsymbol{V}_t = Z_t$ along with $Q_t \stackrel{\text{def}}{=} \mathrm{E}(\boldsymbol{V}_t \boldsymbol{V}_t^T) = \mathrm{E}(Z_t^2) = \sigma_Z^2$

and by using a degenerate form of the observation equation: $\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t$ in which $H_t = 1$ and $\boldsymbol{W}_t = 0$ so that $\boldsymbol{Y}_t = X_t$

AR(1) Process as a State-Space Model: II

Need to define the initial state X_0 in order to complete the model:

- A natural choice is

$$X_0 = \sum_{j=1}^{\infty} \phi^j Z_{1-j}, \quad \text{for which } \mathrm{Var}(X_0) = \frac{\sigma^2}{1-\phi^2}$$

AR(1) Process as a State-Space Model: II

Need to define the initial state X_0 in order to complete the model:

- A natural choice is

$$X_0 = \sum_{j=1}^{\infty} \phi^j Z_{1-j}, \quad \text{for which } \mathrm{Var}(X_0) = \frac{\sigma^2}{1-\phi^2}$$

- With this choice, the required conditions, namely, $\mathrm{E}(X_0 \boldsymbol{W}_t^T) = 0$ and $\mathrm{E}(X_0 \boldsymbol{V}_t^T) = 0$ hold

AR(1) Process as a State-Space Model: II

Need to define the initial state X_0 in order to complete the model:

- A natural choice is

$$X_0 = \sum_{j=1}^{\infty} \phi^j Z_{1-j}, \quad \text{for which } \mathrm{Var}(X_0) = \frac{\sigma^2}{1-\phi^2}$$

- With this choice, the required conditions, namely, $\mathrm{E}(X_0 \boldsymbol{W}_t^T) = 0$ and $\mathrm{E}(X_0 \boldsymbol{V}_t^T) = 0$ hold

- Could also set $X_0 = Z_0 \frac{\sigma}{\sqrt{1-\phi^2}}$ to get a AR(1) process, but using $X_0 = Z_0$ would lead to a valid state-space model that is **not** a true AR(1) model

AR(1) Process as a State-Space Model: III

AR(1) process with $0 < \phi < 1$ is known as "red noise", red noise is related to a 1st order stochastic differential equation, rendering it a model for various geophysical processes:

- Typically only observe red noise process of interest in presence of observational noise (often taken to be white noise)

AR(1) Process as a State-Space Model: III

AR(1) process with $0 < \phi < 1$ is known as “red noise”, red noise is related to a 1st order stochastic differential equation, rendering it a model for various geophysical processes:

- Typically only observe red noise process of interest in presence of observational noise (often taken to be white noise)
- Can modify this setup by changing observational noise from $\boldsymbol{W}_t = 0$ to $\boldsymbol{W}_t = W_t \sim \mathrm{WN}(0, \sigma_W^2)$, where W_t is uncorrelated with Z_t's

AR(1) Process as a State-Space Model: III

AR(1) process with $0 < \phi < 1$ is known as "red noise", red noise is related to a 1st order stochastic differential equation, rendering it a model for various geophysical processes:

- Typically only observe red noise process of interest in presence of observational noise (often taken to be white noise)
- Can modify this setup by changing observational noise from $\boldsymbol{W}_t = 0$ to $\boldsymbol{W}_t = W_t \sim \mathrm{WN}(0, \sigma_W^2)$, where W_t is uncorrelated with Z_t's
- The observation and state-transition equations become

$$Y_t = X_t + W_t \text{ and } X_t = \phi X_{t-1} + Z_t$$

ARMA(1,1) Process as a State-Space Model: I

Recall ARMA(1,1) process $Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}$

- Expressing ARMA(1,1) as $\phi(B)Y_t = \theta(B)Z_t$, note that one can create Y_t by taking causal AR(1) process $X_t = \phi^{-1}(B)Z_t$ and subjecting it to a $\theta(B)$ filter to obtain output $Y_t = \theta(B)X_t = \theta(B)\phi^{-1}(B)Z_t$

ARMA(1,1) Process as a State-Space Model: I

Recall ARMA(1,1) process $Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}$

- Expressing ARMA(1,1) as $\phi(B)Y_t = \theta(B)Z_t$, note that one can create Y_t by taking causal AR(1) process $X_t = \phi^{-1}(B)Z_t$ and subjecting it to a $\theta(B)$ filter to obtain output $Y_t = \theta(B)X_t = \theta(B)\phi^{-1}(B)Z_t$

- Can express filtering of AR(1) process by

$$Y_t = [1 \quad \theta]\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix},$$

which matches up with observation equation

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t$$

if $\boldsymbol{Y}_t = Y_t$, $H_t = [1 \quad \theta]$, $\boldsymbol{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix}$ and $\boldsymbol{W}_t = 0$

ARMA(1,1) Process as a State-Space Model: II

- Given $\boldsymbol{X}_t = \begin{bmatrix} X_t & X_{t-1} \end{bmatrix}^T$, can express $X_t = \phi X_{t-1} + Z_t$ in the 1st row of matrix equation

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} Z_t \\ 0 \end{bmatrix},$$

which matches up with state-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

if $M_t = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix}$ and $\boldsymbol{V}_t = \begin{bmatrix} Z_t \\ 0 \end{bmatrix}$ with

$Q_t \stackrel{\text{def}}{=} \mathrm{E}(\boldsymbol{V}_t \boldsymbol{V}_t^T) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$

ARMA(1,1) Process as a State-Space Model: II

- Given $\boldsymbol{X}_t = \begin{bmatrix} X_t & X_{t-1} \end{bmatrix}^T$, can express $X_t = \phi X_{t-1} + Z_t$ in the 1st row of matrix equation

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} Z_t \\ 0 \end{bmatrix},$$

which matches up with state-transition equation

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t$$

if $M_t = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix}$ and $\boldsymbol{V}_t = \begin{bmatrix} Z_t \\ 0 \end{bmatrix}$ with $Q_t \stackrel{\text{def}}{=} \mathrm{E}(\boldsymbol{V}_t \boldsymbol{V}_t^T) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$

- to complete the model, let

$$\boldsymbol{X}_0 = \begin{bmatrix} X_0 \\ X_{-1} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{\infty} \phi^j Z_{1-j} \\ \sum_{j=1}^{\infty} \phi^j Z_{-j} \end{bmatrix},$$

noting that $\boldsymbol{X}_0$ and $\boldsymbol{V}_t$ for $t \geq 1$ are uncorrelated, as required

ARMA(1,1) Process as a State-Space Model: III

Since

$$\mathrm{E}(\boldsymbol{X}_0\boldsymbol{X}_0^T) = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} = \frac{\sigma^2}{1-\phi^2}\begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix},$$

can alternatively stipulate

$$\boldsymbol{X}_0 = \begin{bmatrix} 1 & \frac{\phi}{\sqrt{1-\phi^2}} \\ 0 & \frac{\phi}{\sqrt{1-\phi^2}} \end{bmatrix}\begin{bmatrix} Z_0 \\ Z_{-1} \end{bmatrix},$$

yielding

$$\begin{aligned}\mathrm{E}(\boldsymbol{X}_0\boldsymbol{X}_0^T) &= \begin{bmatrix} 1 & \frac{\phi}{\sqrt{1-\phi^2}} \\ 0 & \frac{\phi}{\sqrt{1-\phi^2}} \end{bmatrix}\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \frac{\phi}{\sqrt{1-\phi^2}} & \frac{1}{\sqrt{1-\phi^2}} \end{bmatrix} \\ &= \frac{\sigma^2}{1-\phi^2}\begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}\end{aligned}$$

as required

The Linear Gaussian State-Space Model

- State equation:

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t,$$

where $\boldsymbol{V}_t \overset{iid}{\sim} \mathrm{N}(\boldsymbol{0}, Q_t)$ with $\boldsymbol{X}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0)$

- Observation equation:

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t,$$

where $\boldsymbol{W}_t \overset{iid}{\sim} \mathrm{N}(\boldsymbol{0}, R_t)$

- Additional assumptions: $\boldsymbol{X}_0$, $\{\boldsymbol{V}_t\}$, and $\{\boldsymbol{W}_t\}$ are uncorrelated

Forecasting, Filtering, and Smoothing

Goal: To estimate the underlying unobserved signal $\boldsymbol{X}_t$, given the data $\boldsymbol{y}_{1:s} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_s\}$:

- 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

In addition to these estimates, we would also want to measure their precision. The solution to these problems is accomplished via the Kalman filter and Kalman smoother

CLEMSON UNIVERSITY

The Kalman Filter: General Results

Assume the filtering distribution at time $t-1$ is

$$[\boldsymbol{X}_{t-1}|\boldsymbol{y}_{1:t-1}] \sim \mathrm{N}\left(\boldsymbol{\mu}^a_{t-1}, \Sigma^a_{t-1}\right)$$

- **Forecast Step:** Gives the forecast distribution at time t:

$$[\boldsymbol{X}_{t}|\boldsymbol{y}_{1:t-1}] \sim \mathrm{N}\left(\boldsymbol{\mu}^f_{t}, \Sigma^f_{t}\right),$$

where $\boldsymbol{\mu}^f_t = M_t \boldsymbol{\mu}^a_{t-1}$, and $\Sigma^f_t = M_t \Sigma^a_{t-1} M_t^T + Q_t$.

The Kalman Filter: General Results

Assume the filtering distribution at time $t-1$ is

$$[\boldsymbol{X}_{t-1}|\boldsymbol{y}_{1:t-1}] \sim \mathrm{N}\left(\boldsymbol{\mu}_{t-1}^{a}, \Sigma_{t-1}^{a}\right)$$

- **Forecast Step:** Gives the forecast distribution at time t:

$$[\boldsymbol{X}_{t}|\boldsymbol{y}_{1:t-1}] \sim \mathrm{N}\left(\boldsymbol{\mu}_{t}^{f}, \Sigma_{t}^{f}\right),$$

  where $\boldsymbol{\mu}_t^f = M_t \boldsymbol{\mu}_{t-1}^a$, and $\Sigma_t^f = M_t \Sigma_{t-1}^a M_t^T + Q_t$.

- **Update Step:** updates the forecast distribution using new data $\boldsymbol{y}_t$

$$[\boldsymbol{X}_{t}|\boldsymbol{y}_{1:t}] \sim \mathrm{N}\left(\boldsymbol{\mu}_{t}^{a}, \Sigma_{t}^{a}\right),$$

  where $\boldsymbol{\mu}_t^a = \boldsymbol{\mu}_t^f + K_t\left(\boldsymbol{y}_t - H_t \boldsymbol{\mu}_t^f\right)$, and $\Sigma_t^a = \left(I - K_t H_t^T\right)\Sigma_t^f$, and

$$K_t = \Sigma_t^f H_t^T \left(H_t \Sigma_t^f H_t^T + R_t\right)^{-1}$$

  is the Kalman gain matrix

Local Level Model: Part I

Let's begin with a particularly simple example of a state space model: the local level model. We will develop the basic state space techniques for this model.

- **Observation equation**:

$$Y_t = X_t + W_t, \quad \{W_t\} \overset{iid}{\sim} \mathrm{N}(0, \sigma_W^2)$$

Local Level Model: Part I

Let's begin with a particularly simple example of a state space model: the local level model. We will develop the basic state space techniques for this model.

- **Observation equation**:

$$Y_t = X_t + W_t, \quad \{W_t\} \overset{iid}{\sim} \mathrm{N}(0, \sigma_W^2)$$

- **State equation**:

$$X_t = X_{t-1} + V_t, \quad \{V_t\} \overset{iid}{\sim} \mathrm{N}(0, \sigma_V^2)$$

Local Level Model: Part I

Let's begin with a particularly simple example of a state space model: the local level model. We will develop the basic state space techniques for this model.

- **Observation equation**:

$$Y_t = X_t + W_t, \quad \{W_t\} \overset{iid}{\sim} \mathrm{N}(0, \sigma_W^2)$$

- **State equation**:

$$X_t = X_{t-1} + V_t, \quad \{V_t\} \overset{iid}{\sim} \mathrm{N}(0, \sigma_V^2)$$

- Assume $\mathrm{E}(X_0) = \mu_0$ and $\mathrm{Var}(X_0) = \sigma_0^2$ and X_0 is uncorrected with W_t's and V_t's

Local Level Model: Part II

- Since $X_t = X_{t-1} + V_t$, state variable X_t is a random walk starting from $\boldsymbol{\mu}_0$ (intended to model a slowly varying trend)

Local Level Model: Part II

- Since $X_t = X_{t-1} + V_t$, state variable X_t is a random walk starting from $\boldsymbol{\mu}_0$ (intended to model a slowly varying trend)
- Since V_t and X_t are uncorrelated,

$$\mathrm{E}(X_{t+1}|X_t) = \mathrm{E}(X_t + V_t|X_t) = X_t + \mathrm{E}(V_t) = X_t;$$

i.e., if state variable is at a certain 'level' at time t, we can expect no change in its level at time $t + 1$

Local Level Model: Part II

- Since $X_t = X_{t-1} + V_t$, state variable X_t is a random walk starting from $\boldsymbol{\mu}_0$ (intended to model a slowly varying trend)

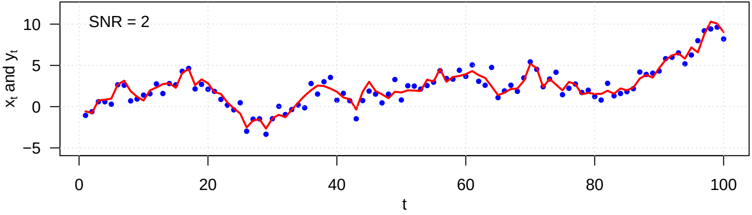
- Since V_t and X_t are uncorrelated,

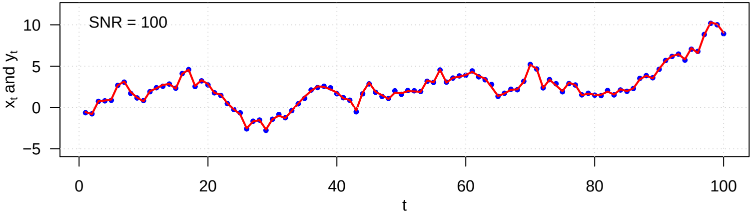
$$\mathrm{E}(X_{t+1}|X_t) = \mathrm{E}(X_t + V_t|X_t) = X_t + \mathrm{E}(V_t) = X_t;$$

  i.e., if state variable is at a certain 'level' at time t, we can expect no change in its level at time $t+1$

- When $\sigma_W^2 > 0$, trend is corrupted by noise, so ability to pick out trend depends upon "signal to noise" ratio (SNR) $\frac{\sigma_V^2}{\sigma_W^2}$

Local Level Model: Examples of Different SNR





Four Problems in State-Space Models

Given observations $\{Y_i\}_{i=1}^t$ of a local level process,

1. Filtering: what is best predictor of state X_t?
2. Forecasting: what is best predictor of state X_{t+1}?
3. Smoothing: what is best predictor of state X_s for $s < t$?
4. Estimation: what are best estimates of model parameters $\sigma_W^2, \sigma_V^2, \mu_0, \sigma_0^2$?

Four Problems in State-Space Models

Given observations $\{Y_i\}_{i=1}^{t}$ of a local level process,

1. Filtering: what is best predictor of state X_t?
2. Forecasting: what is best predictor of state X_{t+1}?
3. Smoothing: what is best predictor of state X_s for $s < t$?
4. Estimation: what are best estimates of model parameters $\sigma_W^2, \sigma_V^2, \mu_0, \sigma_0^2$?

First, we will focus on filtering and forecasting problems, with ‘best‘ defined as the minimum mean square error (MSE).

Four Problems in State-Space Models

Given observations $\{Y_i\}_{i=1}^{t}$ of a local level process,

1. Filtering: what is best predictor of state X_t?
2. Forecasting: what is best predictor of state X_{t+1}?
3. Smoothing: what is best predictor of state X_s for $s < t$?
4. Estimation: what are best estimates of model parameters $\sigma_W^2, \sigma_V^2, \mu_0, \sigma_0^2$?

First, we will focus on filtering and forecasting problems, with 'best' defined as the minimum mean square error (MSE).

To facilitate discussion, let's assume that X_0, V_t's, and W_t are normals, implying that Y_t and the remaining X_t's share this property.

Regression Lemma I

- Suppose random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are jointly normal with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, to be denoted by

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$$

Regression Lemma I

- Suppose random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are jointly normal with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, to be denoted by

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$$

- Can partition both $\boldsymbol{\mu}$ and Σ:

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\mu}_{\boldsymbol{Y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\boldsymbol{XX}} & \Sigma_{\boldsymbol{XY}} \\ \Sigma_{\boldsymbol{YX}} & \Sigma_{\boldsymbol{YY}} \end{bmatrix} \right),$$

where $\boldsymbol{\mu}_{\boldsymbol{X}}$ ($\boldsymbol{\mu}_{\boldsymbol{Y}}$) and $\Sigma_{\boldsymbol{XX}}$ ($\Sigma_{\boldsymbol{YY}}$) are mean and covariance matrix for $\boldsymbol{X}$ ($\boldsymbol{Y}$); $\Sigma_{\boldsymbol{XY}}$ is the cross-covariance matrix between $\boldsymbol{X}$ and $\boldsymbol{Y}$

CLEMSON UNIVERSITY

Regression Lemma II

- Conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is multivariate normal with mean vector

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}} = \boldsymbol{\mu}_{\boldsymbol{X}} + \Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{YY}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

and covariance matrix

$$\Sigma_{\boldsymbol{X}|\boldsymbol{y}} = \Sigma_{\boldsymbol{XX}} - \Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{YY}}^{-1}\Sigma_{\boldsymbol{XY}}^{T}$$

Regression Lemma II

- Conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is multivariate normal with mean vector

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}} = \boldsymbol{\mu}_{\boldsymbol{X}} + \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{YY}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

  and covariance matrix

$$\Sigma_{\boldsymbol{X}|\boldsymbol{y}} = \Sigma_{\boldsymbol{XX}} - \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{YY}}^{-1} \Sigma_{\boldsymbol{XY}}^{T}$$

- Best (under MSE) predictor of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y}} = \boldsymbol{\mu}_{\boldsymbol{X}} + \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{YY}}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

Regression Lemma III

- Recall that, if random vector $\boldsymbol{U}$ has covariance matrix $\Sigma_{\boldsymbol{U}}$, then covariance matrix for $A\boldsymbol{U}$ is $A\Sigma_{\boldsymbol{U}}A^T$

  $\Rightarrow$ covariance matrix of $c + A(\boldsymbol{U} - \mu_{\boldsymbol{U}})$ is also $A\Sigma_{\boldsymbol{U}}A^T$

Regression Lemma III

- Recall that, if random vector $\boldsymbol{U}$ has covariance matrix $\Sigma_{\boldsymbol{U}}$, then covariance matrix for $A\boldsymbol{U}$ is $A\Sigma_{\boldsymbol{U}}A^T$

  $\Rightarrow$ covariance matrix of $c + A(\boldsymbol{U} - \mu_{\boldsymbol{U}})$ is also $A\Sigma_{\boldsymbol{U}}A^T$

- Covariance matrix for

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y}} = \boldsymbol{\mu}_{\boldsymbol{X}} + \Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{YY}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

  is thus

$$\Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{YY}}^{-1}\Sigma_{\boldsymbol{YY}}\Sigma_{\boldsymbol{YY}}^{-1}\Sigma_{\boldsymbol{XY}}^{T} = \Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{YY}}^{-1}\Sigma_{\boldsymbol{XY}}^{T}$$

Note: it is not the same as $\Sigma_{\boldsymbol{X}|\boldsymbol{y}} = \Sigma_{\boldsymbol{XX}} - \Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{YY}}^{-1}\Sigma_{\boldsymbol{XY}}^{T}$

Regression Lemma IV

Consider prediction error $\boldsymbol{U}$ associated with best linear predictor of $\boldsymbol{X}$:

$$\boldsymbol{U} = \boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})$$

- Since $\mathrm{E}\left[\mathrm{E}\left(\boldsymbol{X}|\boldsymbol{Y}\right)\right] = \boldsymbol{\mu}_{\boldsymbol{X}} \Rightarrow \mathrm{E}(\boldsymbol{U}) = \boldsymbol{0}$

Regression Lemma IV

Consider prediction error $\boldsymbol{U}$ associated with best linear predictor of $\boldsymbol{X}$:

$$\boldsymbol{U} = \boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})$$

- Since $\mathrm{E}\left[\mathrm{E}\left(\boldsymbol{X}|\boldsymbol{Y}\right)\right] = \boldsymbol{\mu}_{\boldsymbol{X}} \Rightarrow \mathrm{E}(\boldsymbol{U}) = \mathbf{0}$

- Covariance matrix for $\boldsymbol{U}$ is given by

$$\begin{aligned}\mathrm{E}(\boldsymbol{U}\boldsymbol{U}^T) =& \mathrm{E}\left([\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})]\,[\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})]^T\right) \\ =& \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^T) + \mathrm{E}[\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})^T] \\ &- \mathrm{E}[\boldsymbol{X}\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})^T] - \mathrm{E}[\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})\boldsymbol{X}^T] \\ =& \Sigma_{\boldsymbol{X}\boldsymbol{X}} - \Sigma_{\boldsymbol{X}\boldsymbol{Y}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}\Sigma_{\boldsymbol{X}\boldsymbol{Y}}^T,\end{aligned}$$

  which is equal to $\Sigma_{\boldsymbol{X}|\boldsymbol{y}}$, the conditional covariance matrix

Regression Corollary

Now specialize to the case where $\boldsymbol{X}$ has just one element, say, X

- Corollary: conditional distribution of X given $\boldsymbol{Y} = \boldsymbol{y}$ is normal with mean

$$\mu_X + \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

and conditional variance

$$\Sigma_{X|\boldsymbol{y}} = \sigma_X^2 - \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} \Sigma_{X\boldsymbol{Y}},$$

where $\sigma_X^2 = \mathrm{Var}(X)$ and $\Sigma_{X\boldsymbol{Y}}$ is a column vector containing covariance between X and $\boldsymbol{Y}$

Regression Corollary

Now specialize to the case where $\boldsymbol{X}$ has just one element, say, X

- Corollary: conditional distribution of X given $\boldsymbol{Y} = \boldsymbol{y}$ is normal with mean

$$\mu_X + \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

and conditional variance

$$\Sigma_{X|\boldsymbol{y}} = \sigma_X^2 - \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} \Sigma_{X\boldsymbol{Y}},$$

where $\sigma_X^2 = \mathrm{Var}(X)$ and $\Sigma_{X\boldsymbol{Y}}$ is a column vector containing covariance between X and $\boldsymbol{Y}$

- Since conditional variance is same as MSE for X, will refer to $\Sigma_{X|\boldsymbol{y}}$ as MSE

Aside – Revisiting Time Series Prediction: I

Suppose $\{X_t\}$ is zero mean stationary process with ACF $\gamma(h)$

- Set X to X_{n+1} and put $X_1, \cdots, X_n$ into $\boldsymbol{Y}$
- Corollary says best linear predictor $\hat{X}_{n+1}$ of X_{n+1} given $X_1, \cdots, X_n$ is

$$\hat{X}_{n+1} = \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} \boldsymbol{Y} = \gamma_n^T \Gamma_n^{-1} \boldsymbol{Y} \stackrel{\text{def}}{=} \boldsymbol{c}_n^T \boldsymbol{Y},$$

where

1. $\gamma_n = [\gamma(1), \gamma(2), \cdots, \gamma(n)]^T = \Sigma_{X\boldsymbol{Y}}$
2. (i, j)th entry of matrix $\Gamma_n = \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}$ is $\gamma(i - j)$
3. $\boldsymbol{c}_n^T \stackrel{\text{def}}{=} \gamma_n^T \Gamma_n^{-1}$ and hence $\boldsymbol{c}_n = \Gamma_n^{-1} \gamma_n$

Aside – Revisiting Time Series Prediction: II

Recall that MSE for $\hat{X}_{n+1}$ is

$$\begin{aligned} v_n &= \mathrm{Var}(X_{n+1}) - \boldsymbol{c}_n^T \boldsymbol{\gamma}_n \\ &= \sigma_X^2 - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n \\ &= \sigma_X^2 - \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} \Sigma_{X\boldsymbol{Y}} \\ &= \Sigma_{X|\boldsymbol{y}} \end{aligned}$$

This is a special case of regression corollary

Filtering for Local Level Model: I

Let's begin with a particularly simple example of a state space model: the local level model

- Local level model:

$$Y_t = X_t + W_t, \quad \{W_t\} \sim \mathrm{N}(0, \sigma_W^2)$$
$$X_t = X_{t-1} + V_t, \quad \{V_t\} \sim \mathrm{N}(0, \sigma_V^2)$$

  and X_0 is a R.V. that

  - is uncorrelated with W_t's and V_t's
  - has $\mathrm{E}(X_0) = \mu_0$ and $\mathrm{Var}(X_0) = \sigma_0^2$

- Filtering problem is to predict unknown state X_t based on data up to time t, i.e., $\boldsymbol{Y}_{1:t} = (y_1, \cdots, y_t)^T$

Filtering for Local Level Model: II

Best linear predictor of X_t given $\boldsymbol{Y}_{1:t}$ is

$$\mu_t^a \stackrel{\text{def}}{=} \mathrm{E}(X_t|\boldsymbol{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = \mathrm{E}(X_t)$, $\boldsymbol{\mu}_{1:t}$ is a vector containing, for $j = 1, \cdots, t$,

$$\mu_j \stackrel{\text{def}}{=} \mathrm{E}(X_j) = \mathrm{E}(X_j + W_j) = \mathrm{E}(Y_j)$$

Filtering for Local Level Model: II

Best linear predictor of X_t given $\boldsymbol{Y}_{1:t}$ is

$$\mu_t^a \stackrel{\text{def}}{=} \mathrm{E}(X_t|\boldsymbol{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = \mathrm{E}(X_t)$, $\boldsymbol{\mu}_{1:t}$ is a vector containing, for $j = 1, \cdots, t$,

$$\mu_j \stackrel{\text{def}}{=} \mathrm{E}(X_j) = \mathrm{E}(X_j + W_j) = \mathrm{E}(Y_j)$$

- Vector $\Sigma_{t,t}$ contains covarinces between X_t and $\boldsymbol{Y}_{1:t}$

Filtering for Local Level Model: II

Best linear predictor of X_t given $\boldsymbol{Y}_{1:t}$ is

$$\mu_t^a \stackrel{\text{def}}{=} \mathrm{E}(X_t|\boldsymbol{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = \mathrm{E}(X_t)$, $\boldsymbol{\mu}_{1:t}$ is a vector containing, for $j = 1, \cdots, t$,

$$\mu_j \stackrel{\text{def}}{=} \mathrm{E}(X_j) = \mathrm{E}(X_j + W_j) = \mathrm{E}(Y_j)$$

- Vector $\Sigma_{t,t}$ contains covarinces between X_t and $\boldsymbol{Y}_{1:t}$
- (i, j)th element of matrix $\Sigma_{Y,t}$ is covariance between Y_i and Y_j

Filtering for Local Level Model: II

Best linear predictor of X_t given $\boldsymbol{Y}_{1:t}$ is

$$\mu_t^a \stackrel{\text{def}}{=} \mathrm{E}(X_t|\boldsymbol{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = \mathrm{E}(X_t)$, $\boldsymbol{\mu}_{1:t}$ is a vector containing, for $j = 1, \cdots, t$,

$$\mu_j \stackrel{\text{def}}{=} \mathrm{E}(X_j) = \mathrm{E}(X_j + W_j) = \mathrm{E}(Y_j)$$

- Vector $\Sigma_{t,t}$ contains covarinces between X_t and $\boldsymbol{Y}_{1:t}$
- (i, j)th element of matrix $\Sigma_{Y,t}$ is covariance between Y_i and Y_j
- Note: $\mathrm{E}(\mu_t^a) = \mathrm{E}[\mathrm{E}(X_t|\boldsymbol{Y}_{1:t})] = \mathrm{E}(X_t) = \mu_t$

Filtering for Local Level Model: II

Best linear predictor of X_t given $\boldsymbol{Y}_{1:t}$ is

$$\mu_t^a \stackrel{\text{def}}{=} \mathrm{E}(X_t|\boldsymbol{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = \mathrm{E}(X_t)$, $\boldsymbol{\mu}_{1:t}$ is a vector containing, for $j = 1, \cdots, t$,

$$\mu_j \stackrel{\text{def}}{=} \mathrm{E}(X_j) = \mathrm{E}(X_j + W_j) = \mathrm{E}(Y_j)$$

- Vector $\Sigma_{t,t}$ contains covarinces between X_t and $\boldsymbol{Y}_{1:t}$
- (i,j)th element of matrix $\Sigma_{Y,t}$ is covariance between Y_i and Y_j
- Note: $\mathrm{E}(\mu_t^a) = \mathrm{E}[\mathrm{E}(X_t|\boldsymbol{Y}_{1:t})] = \mathrm{E}(X_t) = \mu_t$
- With $\sigma_t^2 \stackrel{\text{def}}{=} \mathrm{Var}(X_t)$, MSE for predictor is

$$\mathrm{E}[(X_t - \mu_t^a)^2] = \sigma_t^2 - \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} \Sigma_{t,t} \stackrel{\text{def}}{=} \Sigma_t^a$$

Forecasting for Local Level Model: I

Forecasting: estimate X_{t+1} given $\boldsymbol{Y}_{1:t}$

- Best linear predictor of X_{t+1} given $\boldsymbol{Y}_{1:t}$ is

$$\mu^f_{t+1} \stackrel{\text{def}}{=} \mathrm{E}(X_{t+1}|\boldsymbol{Y}_{1:t}) = \mu_{t+1} + \Sigma^T_{t+1,t}\Sigma^{-1}_{Y,t}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where vector $\Sigma_{t+1,t}$ has covaraince between X_{t+1} and $\boldsymbol{Y}_{1:t}$

Forecasting for Local Level Model: I

Forecasting: estimate X_{t+1} given $\boldsymbol{Y}_{1:t}$

- Best linear predictor of X_{t+1} given $\boldsymbol{Y}_{1:t}$ is

$$\mu_{t+1}^{f} \stackrel{\text{def}}{=} \mathrm{E}(X_{t+1}|\boldsymbol{Y}_{1:t}) = \mu_{t+1} + \Sigma_{t+1,t}^{T}\Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

  where vector $\Sigma_{t+1,t}$ has covaraince between X_{t+1} and $\boldsymbol{Y}_{1:t}$

- Note: $\mathrm{E}(\mu_{t+1}^{f}) = \mathrm{E}[\mathrm{E}(X_{t+1}|\boldsymbol{Y}_{1:t})] = \mathrm{E}(X_{t+1}) = \mu_{t+1}$

Forecasting for Local Level Model: I

Forecasting: estimate X_{t+1} given $\boldsymbol{Y}_{1:t}$

- Best linear predictor of X_{t+1} given $\boldsymbol{Y}_{1:t}$ is

$$\mu_{t+1}^{f} \stackrel{\text{def}}{=} \mathrm{E}(X_{t+1}|\boldsymbol{Y}_{1:t}) = \mu_{t+1} + \Sigma_{t+1,t}^{T}\Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

  where vector $\Sigma_{t+1,t}$ has covaraince between X_{t+1} and $\boldsymbol{Y}_{1:t}$

- Note: $\mathrm{E}(\mu_{t+1}^{f}) = \mathrm{E}[\mathrm{E}(X_{t+1}|\boldsymbol{Y}_{1:t})] = \mathrm{E}(X_{t+1}) = \mu_{t+1}$

- MSE for predictor is

$$\mathrm{E}[(X_{t+1} - \mu_{t+1}^{f})^2] = \sigma_{t+1}^2 - \Sigma_{t+1,t}^{T}\Sigma_{Y,t}^{-1}\Sigma_{t+1,t} \stackrel{\text{def}}{=} \Sigma_{t+1}^{f}$$

Forecasting for Local Level Model: II

- Let's also consider best linear predictor of Y_{t+1} given $\boldsymbol{Y}_{1:t}$:

$$Y_{t+1}^{t} \stackrel{\text{def}}{=} \mathrm{E}(Y_{t+1}|\boldsymbol{Y}_{1:t}) = \mu_{Y,t+1} + \tilde{\Sigma}_{t+1,t}^{T}\Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}),$$

  where the vector $\tilde{\Sigma}_{t+1,t}$ has covarainces between Y_{t+1} and $\boldsymbol{Y}_{1:t}$

Forecasting for Local Level Model: II

- Let's also consider best linear predictor of Y_{t+1} given $\boldsymbol{Y}_{1:t}$:

$$Y_{t+1}^{t} \overset{\text{def}}{=} \mathrm{E}(Y_{t+1}|\boldsymbol{Y}_{1:t}) = \mu_{Y,t+1} + \tilde{\Sigma}_{t+1,t}^{T}\Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}),$$

  where the vector $\tilde{\Sigma}_{t+1,t}$ has covaraincés between Y_{t+1} and $\boldsymbol{Y}_{1:t}$

- However, note that, for $j = 1, \cdots, t$

$$\mathrm{Cov}(Y_{t+1}, Y_j) = \mathrm{Cov}(X_{t+1} + W_{t+1}, Y_j) = \mathrm{Cov}(X_{t+1}, Y_j)$$

Forecasting for Local Level Model: II

- Let's also consider best linear predictor of Y_{t+1} given $\boldsymbol{Y}_{1:t}$:

$$Y_{t+1}^t \stackrel{\text{def}}{=} \mathrm{E}(Y_{t+1}|\boldsymbol{Y}_{1:t}) = \mu_{Y,t+1} + \tilde{\Sigma}_{t+1,t}^T \Sigma_{Y,t}^{-1}(\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}),$$

  where the vector $\tilde{\Sigma}_{t+1,t}$ has covaraínces between Y_{t+1} and $\boldsymbol{Y}_{1:t}$

- However, note that, for $j = 1, \cdots, t$

$$\mathrm{Cov}(Y_{t+1}, Y_j) = \mathrm{Cov}(X_{t+1} + W_{t+1}, Y_j) = \mathrm{Cov}(X_{t+1}, Y_j)$$

- Thus $\tilde{\Sigma}_{t+1,t} = \Sigma_{t+1,t}$, yielding

$$Y_{t+1}^t = \mu_{Y,t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1}(\boldsymbol{y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}) = \mu_{t+1}^f$$

  $\Rightarrow$ difference between Y_{t+1} and X_{t+1} is W_{t+1}, therefore they have the same estimator, but their MSEs differ:

$$\mathrm{E}\left[(Y_{t+1} - Y_{t+1}^f)^2\right] = \Sigma_{t+1}^f + \sigma_W^2$$

Filtering for Local Level Model: III

- To implement filtering, i.e., compute μ_t^a, need to determine:

Filtering for Local Level Model: III

- To implement filtering, i.e., compute μ_t^a, need to determine:

  1. $\mu_j = \mathrm{E}(X_j),\ j = 1, \cdots, t$

Filering for Local Level Model: III

- To implement filtering, i.e., compute μ_t^a, need to determine:

  1. $\mu_j = \mathrm{E}(X_j)$, $j = 1, \cdots, t$
  2. Elements of $\Sigma_{t,t}$, i.e., covaraince between X_t and $\boldsymbol{Y}_{1:t}$

Filtering for Local Level Model: III

- To implement filtering, i.e., compute μ_t^a, need to determine:

1. $\mu_j = \mathrm{E}(X_j)$, $j = 1, \cdots, t$
2. Elements of $\Sigma_{t,t}$, i.e., covaraince between X_t and $\boldsymbol{Y}_{1:t}$
3. Elements of $\Sigma_{Y,t}$, i.e., covariances between Y_j and Y_k, $1 \leq j \leq k \leq t$

Filtering for Local Level Model: III

- To implement filtering, i.e., compute μ_t^a, need to determine:

  1. $\mu_j = \mathrm{E}(X_j)$, $j = 1, \cdots, t$
  2. Elements of $\Sigma_{t,t}$, i.e., covaraince between X_t and $\boldsymbol{Y}_{1:t}$
  3. Elements of $\Sigma_{Y,t}$, i.e., covariances between Y_j and Y_k, $1 \le j \le k \le t$

- To compute Σ_t^a, i.e., MSE for μ_t^a, need $\sigma_t^2 = \mathrm{Var}(X_t)$ in addition to 2 and 3 above

Filtering for Local Level Model: III

- To implement filtering, i.e., compute μ_t^a, need to determine:
  1. $\mu_j = \mathrm{E}(X_j)$, $j = 1, \cdots, t$
  2. Elements of $\Sigma_{t,t}$, i.e., covaraince between X_t and $\boldsymbol{Y}_{1:t}$
  3. Elements of $\Sigma_{Y,t}$, i.e., covariances between Y_j and Y_k, $1 \le j \le k \le t$
- To compute Σ_t^a, i.e., MSE for μ_t^a, need $\sigma_t^2 = \mathrm{Var}(X_t)$ in addition to 2 and 3 above
- Since $X_t = X_{t-1} + V_t$ and $Y_t = X_t + W_t$, telescoping yields $X_j = X_0 + \sum_{l=1}^{j} V_l$ and $Y_j = X_0 + \sum_{l=1}^{j} V_l + W_j$, $j = 1, \cdots, t$

Filtering for Local Level Model: IV

Using

$$X_j = X_0 + \sum_{l=1}^{j} V_l \text{ and } Y_j = X_0 + \sum_{l=1}^{j} + W_j, \quad j = 1, \cdots, t,$$

get $\mu_j = \mathrm{E}[X_j] = \mathrm{E}[X_0] = \mu_0$ and (assuming $j \le k \le t$)

$$\begin{aligned}
\mathrm{Cov}(X_t, Y_j) &= \mathrm{Cov}\left(X_0 + \sum_{l=1}^{t} V_l, X_0 + \sum_{l=1}^{j} V_l + W_j \right) \\
&= \sigma_0^2 + j\sigma_V^2 \\
\mathrm{Cov}(Y_j, Y_k) &= \mathrm{Cov}\left(X_0 + \sum_{l=1}^{j} V_l + W_j, X_0 + \sum_{l=1}^{k} V_l + W_k \right) \\
&= \sigma_0^2 + \min(j,k)\sigma_V^2 + \delta_{jk}\sigma_W^2,
\end{aligned}$$

where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \ne k$

Filtering for Local Level Model: V

Using

$$X_t = X_0 + \sum_{l=1}^{t} V_l,$$

get

$$\sigma_t^2 = \mathrm{Var}(X_t) = \sigma_0^2 + t\sigma_V^2$$

- Now we have all the pieces needed to form μ_t^a and its MSE Σ_t^a
- **Note**: similar argument leads to pieces needed to form forecast μ_{t+1}^f and its MSE Σ_{t+1}^f

Kalman Recursions for Filtering/Forecasting: I

While straightforward conceptually, forming

$$\mu_t^a = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

and

$$\mu_{t+1}^f = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

via these equations requires inversion of matrix $\Sigma_{Y,t}$ whose dimension $t \times t$ becomes problematic as t gets large ☹

Kalman Recursions for Filtering/Forecasting: I

While straightforward conceptually, forming

$$\mu_t^a = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

and

$$\mu_{t+1}^f = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\boldsymbol{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

via these equations requires inversion of matrix $\Sigma_{Y,t}$ whose dimension $t \times t$ becomes problematic as t gets large ☹

> ⇒ The celebrated Kalman recursions give a recipe that avoids explicit matrix inversion

- **Idea**: at time $t-1$, we have 4 quantities of interest: fitted value μ_{t-1}^a, and forecast μ_t^f and their associated MSEs Σ_{t-1}^a and Σ_t^f
- **Note**: $\mu_{t-1}^a = \mu_t^f$ for local level model (but not others)

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields
  - fitted values μ_t^a and forecast μ_{t+1}^f

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields
  - fitted values μ_t^a and forecast μ_{t+1}^f
  - associated MSEs Σ_t^a and Σ_{t+1}^f

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields
  - fitted values μ_t^a and forecast μ_{t+1}^f
  - associated MSEs Σ_t^a and Σ_{t+1}^f
- There are six steps in the Kalman recursion:

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields
  - fitted values μ_t^a and forecast μ_{t+1}^f
  - associated MSEs Σ_t^a and Σ_{t+1}^f
- There are six steps in the Kalman recursion:
  1. steps 1 and 2 are preparatory

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields
  - fitted values μ_t^a and forecast μ_{t+1}^f
  - associated MSEs Σ_t^a and Σ_{t+1}^f
- There are six steps in the Kalman recursion:
  1. steps 1 and 2 are preparatory
  2. steps 3 and 4 yield μ_t^a and Σ_t^a (filtering)

Kalman Recursions for Filtering/Forecasting: II

- At time t, new observation Y_t becomes available
- Kalman recursion takes μ_t^f, Σ_t^f and Y_t and yields
  - fitted values μ_t^a and forecast μ_{t+1}^f
  - associated MSEs Σ_t^a and Σ_{t+1}^f
- There are six steps in the Kalman recursion:
  1. steps 1 and 2 are preparatory
  2. steps 3 and 4 yield μ_t^a and Σ_t^a (filtering)
  3. steps 5 and 6 yield μ_{t+1}^f and Σ_{t+1}^f (forecasting)

Kalman Recursions for Filtering/Forecasting: III

1. Compute innovation:

$$U_t = Y_t - Y_t^{t-1} = Y_t - \mu_t^f$$

2. Compute MSE for Y_t^{t-1}:

$$\Sigma_t^f + \sigma_W^2 \stackrel{\text{def}}{=} F_t$$

3. Compute new filtered value:

$$\mu_t^a = \mu_t^f + K_t U_t,$$

where $K_t \stackrel{\text{def}}{=} \Sigma_t^f / F_t$ is the so-called Kalman gain

4. Compute MSE for new filtered value:

$$\Sigma_t^a = \Sigma_t^f (1 - K_t)$$

Kalman Recursions for Filtering/Forecasting: IV

5. Compute new forecast:

$$\mu_{t+1}^{f} = \mu_t^{f} + K_t U_t = \mu_t^{a}$$

6. Compute MSE for new forecast:

$$\Sigma_{t+1}^{f} = \Sigma_t (1 - K_t) + \sigma_W^2 = \Sigma_t^{a} + \sigma_W^2$$

Recursions are carried out for $t = 0, \cdots, n$ with inputs $\mathrm{E}[X_0] = \mu_0$, $\mathrm{Var}(X_0) = \sigma_0^2$ and $Y_t' s$

Kalman Recursions for Filtering/Forecasting: V

To prove validity of steps 3 and 4, need to show that $\mu_t^f + K_t U_t$ is equal to μ_t^a, and $\Sigma_t^f(1 - K_t)$ is equal to Σ_t^a

Kalman Recursions for Filtering/Forecasting: V

To prove validity of steps 3 and 4, need to show that $\mu_t^f + K_t U_t$ is equal to μ_t^a, and $\Sigma_t^f(1 - K_t)$ is equal to Σ_t^a

- **Key fact**: X_t conditioned on both $U_t = Y_t - Y_t^{t-1}$ and $\boldsymbol{Y}_{1:t-1}$ is the same as X_t conditioned on $\boldsymbol{Y}_{1:t-1}$ becasue

$$\begin{aligned}\mathrm{Cov}(X_t, U_t|\boldsymbol{Y}_{1:t-1}) &= \mathrm{Cov}(X_t, Y_t - Y_t^{t-1}|\boldsymbol{Y}_{1:t-1}) \\ &= \mathrm{Cov}(X_t, X_t + W_t|\boldsymbol{Y}_{1:t-1}) = \mathrm{Var}(X_t|\boldsymbol{Y}_{1:t-1}) \\ &= \Sigma_t^f\end{aligned}$$

We have

$$\mu_t^a = \mu_t^f + \frac{\Sigma_t^f}{F_t}U_t, \text{ and } \Sigma_t^a = \Sigma_t^f - \frac{\left(\Sigma_t^f\right)^2}{F_t}$$

since $K_t \stackrel{\text{def}}{=} \frac{\Sigma_t^f}{F_t}$, we get required

$$\mu_t^a = \mu_t^f + K_t U_t \text{ and } \Sigma_t^a = \Sigma_t^f(1 - K_t)$$