Chapter 5: Forecasting

Tuan A. Luong

De Montfort University - VCREME - IREEDS

June 2018

Table of contents

Forecast

- We are interested in forecasting the value of a variable Y_{t+1} based on a set of variables $\mathbf{X}_t$ observed at time t.
- For instance, we want to forecast Y_{t+1} based on its m most recent values.
- In this case we have the expectation of
- How do we know whether our forecast is good?

The loss function

- We need to establish a loss function: it summarizes how concerned we are if our forecast is off by a particular amount.
- Usually we use the quadratic loss function

$$E(Y_{t+1} - Y^*_{t+1|t})^2 \tag{1}$$

- We choose $Y^*_{t+1|t}$ to minimize the loss function above: it must be as close to Y_{t+1} as possible.
- The quadratic loss function is also called the *mean squared error*.
- Is it better to choose $Y^*_{t+1|t}$ **higher** than Y_{t+1} or is it better to choose $Y^*_{t+1|t}$ **lower** than Y_{t+1}?

Conditional expectation

- The forecast with the smallest mean squared error is the expectation of Y_{t+1} conditional on $\mathbf{X}_t$:

$$Y^*_{t+1|t} = E(Y_{t+1}|\mathbf{X}_t) \qquad (2)$$

- If we have the $AR(1)$ process:

$$Y_t = c + \phi Y_{t-1} + \epsilon_t$$

- What is the conditional expectation:

$$E(Y_{t+1}|Y_t) = ?$$

Linear projection

- We now restrict the class of forecast to the linear function of $\mathbf{X}_t$:

$$Y^*_{t+1|t} = \alpha' \mathbf{X}_t \tag{3}$$

- Why α'?
- A linear forecast is represented by the value of α.
- A **linear projection** is the linear forecast such that the forecast error is uncorrelated with $\mathbf{X}_t$:

$$E[(Y_{t+1} - \alpha' \mathbf{X}_t)\mathbf{X}_t'] = \mathbf{0}'$$

- What is $\mathbf{0}'$?

Properties of Linear projection

- **The optimal linear forecast**: The linear projection turns out to produce the smallest mean squared error among the class of linear forecasting rules.
- What is the value of α?
- If $E(\mathbf{X}_t\mathbf{X}_t')$ is a non-singular matrix, i.e. it is reversible:

$$\alpha' = E(Y_{y+1}\mathbf{X}_t')[E(\mathbf{X}_t\mathbf{X}_t')]^{-1} \tag{4}$$

- Do you recognize this rule?

The problem

- Consider the $MA(\infty)$ process:

$$(Y_t - \mu) = \psi(L)\epsilon_t$$

with

$$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$$

$$\psi_0 = 1$$

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

- We want to forecast the value of $Y_{t+s}(s > 0)$ based on the values of ϵ through date t: $\{\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, ...\}$ and the values of $\mu : \{\mu_1, \mu_2, ...\}$

The problem(cont')

- We can rewrite the $MA(\infty)$ process:

$$Y_{t+s} = \mu + \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + ... + \psi_{s-1} \epsilon_{t+1} + \psi_s \epsilon_t + ...$$

- What is the optimal forecast?

Optimal forecast

- Recall that the ϵ are white noise.
- The forecast that minimizes the mean squared error is the conditional expectation:

$$\hat{E}(Y_{t+s}|\epsilon_t, \epsilon_{t-1}, ...) = \mu + \psi_s \epsilon_t + \psi_{s+1}\epsilon_{t-1} + ... \qquad (5)$$

Checks

- Is $\hat{E}(Y_{t+s}|\epsilon_t, \epsilon_{t-1}, ...$ a linear forecast?
- Is it a linear projection?
- Is it optimal forecast?
- What is the mean squared error?

Another example

- Consider the $MA(q)$ process:

$$(Y_t - \mu) = \psi(L)\epsilon_t$$

with

$$\psi(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$$

- What is the optimal forecast of Y_{t+s}?
- What is the mean squared error?

The problem

- Consider the $AR(p)$ process:

$$\eta(L)(Y_t - \mu) = \epsilon_t$$

with

$$\eta(L) = \sum_{j=0}^{p} \eta_j L^j$$

$$\eta_0 = 1$$

$$\sum_{j=0}^{p} |\eta_j| < \infty$$

- The problem: How do we forecast the value of $Y_{t+s}(s > 0)$ based on its past values of $Y_t, Y_{t-1}, \ldots$ and the values of η?

Example: AR(1) process

- Consider the $AR(1)$ process:

$$(1 - \phi L)(Y_t - \mu) = \epsilon_t \tag{6}$$

- We can generate a $MA(\infty)$ process:

$$\begin{aligned} \psi(L) &= (1 - \phi L)^{-1} \\ &= 1 + \phi L + \phi^2 L^2 + \ldots \\ &= \sum_{j=0}^{\infty} \phi^j L^j \end{aligned}$$

- We then can rewrite the $AR(1)$ process:

$$\begin{aligned} Y_t &= \mu + (1 - \phi L)^{-1} \epsilon_t \\ &= \mu + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \ldots \end{aligned} \tag{7}$$

Example: AR(1) process (cont')

- We can rewrite Y_{t+s} as:

$$Y_{t+s} = \mu + \epsilon_{t+s} + \phi\epsilon_{t+s-1} + \ldots + \phi^s\epsilon_t + \phi^{s+1}\epsilon_{t-1} + \ldots \quad (8)$$

- If we forecasted Y_{t+s} based on the innovation ϵ:

$$\begin{aligned} \hat{E}[Y_{t+s}|\epsilon_t, \epsilon_{t-1}, \ldots] &= \mu + \phi^s\epsilon_t + \phi^{s+1}\epsilon_{t-1} + \ldots \\ &= \mu + \phi^s \sum_{j=0}^{\infty} \phi^j L^j \epsilon_t \\ &= \mu + \phi^s(1 - \phi L)^{-1}\epsilon_t \end{aligned} \quad (9)$$

- Replace ϵ_t by Equation 6 we have:

$$\begin{aligned} \hat{E}[Y_{t+s}|Y_t, Y_{t-1}, \ldots] &= \mu + \phi^s(1 - \phi L)^{-1}(1 - \phi L)(Y_t - \mu) \\ &= \mu + \phi^s(Y_t - \mu) \end{aligned} \quad (10)$$

The Wiener-Kolmogorov prediction formula

- Consider the $AR(p)$ process:

$$\eta(L)(Y_t - \mu) = \epsilon_t \tag{11}$$

with

$$\eta(L) = \sum_{j=0}^{p} \eta_j L^j$$
$$\eta_0 = 1$$
$$\sum_{j=0}^{p} |\eta_j| < \infty$$

- These conditions ensure that we can invert $\eta(L)$:

$$\eta(L) = \psi(L)^{-1} \tag{12}$$

The Wiener-Kolmogorov prediction formula (cont')

- We can predict Y_{t+s} based on its lagged value using the Wiener-Kolmogorov prediction formula:

$$\hat{E}[Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + [\frac{\psi(L)}{L^s}]_+ \eta(L)(Y_t - \mu) \tag{13}$$

where

$$[\frac{\psi(L)}{L^s}]_+ = \psi_s + \psi_{s+1} L + ... \tag{14}$$

is called the **annihilation operator**.

Example 1: the AR(1) process

- Consider the $AR(1)$ process:

$$(1 - \phi L)(Y_t - \mu) = \epsilon_t$$

- With $\eta(L) = 1 - \phi L$ we have:

$$\begin{aligned} \psi(L) &= \eta(L)^{-1} \\ &= 1 + \phi L + \phi^2 L^2 + ... \\ &= \sum_{j=0}^{\infty} \phi^j L^j \end{aligned}$$

- Here we have $\psi_j = \phi_j$.

Example 1: the AR(1) process(cont')

- The annihilation operator is then:

$$\left[\frac{\psi(L)}{L^s}\right]_+ = \psi_s + \psi_{s+1}L + \dots$$
$$= \phi^s/(1-\phi L)$$

- Apply the Wiener-Kolmogorov prediction formula:

$$\begin{aligned}\hat{E}[Y_{t+s}|Y_t, Y_{t-1}, \dots] &= \mu + \left[\frac{\psi(L)}{L^s}\right]_+ \eta(L)(Y_t - \mu) \\ &= \mu + \phi^s/(1-\phi L)(1-\phi L)(Y_t-\mu) \\ &= \mu + \phi^s(Y_t - \mu)\end{aligned} \tag{15}$$

Example 2: the MA(1) process

- Consider the $MA(1)$ process:

$$(Y_t - \mu) = (1 + \theta L)\epsilon_t$$

  with $|\theta| < 1$
- It can be rewritten as an $AR(\infty)$ process:

$$\eta(L)(Y_t - \mu) = \epsilon_t$$

  where $\eta(L) = (1 + \theta L)^{-1}$.
- Then $\psi(L) = \eta(L)^{-1} = 1 + \theta L$.
- What is $\hat{E}[Y_{t+1}|Y_t, Y_{t-1}, ...]$?
- What are $\hat{E}[Y_{t+2}|Y_t, Y_{t-1}, ...], \hat{E}[Y_{t+3}|Y_t, Y_{t-1}, ...], ...$?

Example 3: the ARMA(1,1) process

- Consider the $ARMA(1,1)$ process:

$$(1-\phi L)(Y_t-\mu)=(1+\theta L)\epsilon_t$$

  with $|\theta|<1$ and $|\phi|<1$
- What do these conditions buy us?
- What is $\eta(L)$?
- What is ψL?
- What is $\hat{E}[Y_{t+1}|Y_t, Y_{t-1}, ...]$?