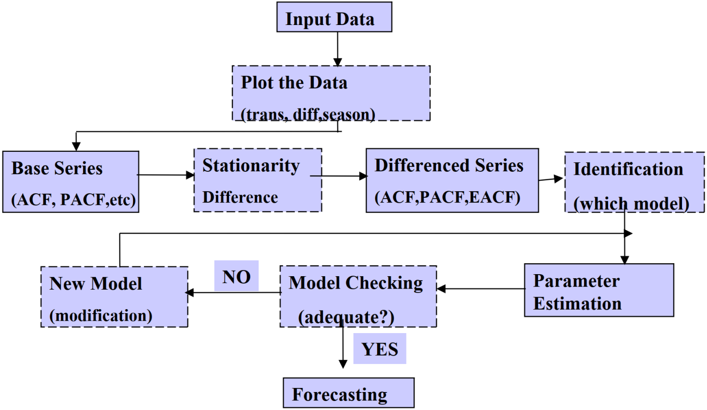
Time Series Analysis
Lecture 5 Forecasting (Out-of-Sample Validation)

Muyi Li

Wang ya'nan Institute for Studies in Economics (WISE)
Department of Statistics, School of Economics
Xiamen University

Modeling Steps



- In this chapter, we assume that the model is known completely, including all the parameters. Although such an assumption is never true, it simplifies the derivation substantially. In practice unknown parameters are replaced by their estimates.

Model Forecasting

- Based on data to the present $I_t = \{y_t, y_{t-1}, \ldots\}$, predict future values of a time series, y_{t+h}, $h = 1, 2, \ldots$.
- Let $y_{t+h|t}$ denote a forecast of y_{t+h} made at time t, define *forecast error* or *prediction error* $\varepsilon_{t+h|t}$,

$$\varepsilon_{t+h|t} = y_{t+h} - y_{t+h|t}.$$

- Find the *optimal forecast* $y_{t+h|t}$ to minimizes the mean squared error of the forecast,

$$\mathrm{MSE}(\varepsilon_{t+h|t}) \equiv E[\varepsilon^2_{t+h|t}] = E[(y_{t+h} - y_{t+h|t})^2].$$

- The *minimum MSE forecast* (best forecast) of y_{t+h} based on I_t is $E[y_{t+h}|I_t]$.

Model Forecasting

Proof:

$$
\begin{aligned}
E[(y_{t+h} - y_{t+h|t})^2] &= E\left\{[y_{t+h} - E(y_{t+h}|I_t) + E(y_{t+h}|I_t) - y_{t+h|t}]^2\right\} \\
&= E\left\{[y_{t+h} - E(y_{t+h}|I_t)]^2\right\} + 2E\left\{[y_{t+h} - E(y_{t+h}|I_t)][E(y_{t+h}|I_t) - y_{t+h|t}]\right\} \\
&\quad + E\left\{[E(y_{t+h}|I_t) - y_{t+h|t}]^2\right\} \\
&= E\left\{[y_{t+h} - E(y_{t+h}|I_t)]^2\right\} + E\left\{[E(y_{t+h}|I_t) - y_{t+h|t}]^2\right\}
\end{aligned}
$$

Taking the minimization with different values of $y_{t+h|t}$, we have

$$y_{t+h|t} = E(y_{t+h}|I_t).$$

Thus, the minimum MSE is $E[(y_{t+h} - y_{t+h|t})^2] = E[(y_{t+h} - E(y_{t+h}|I_t))^2]$.

Best forecast

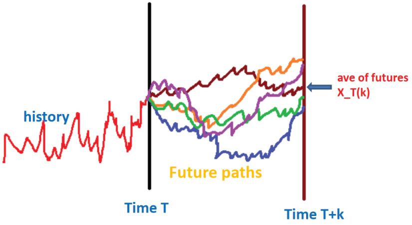


Figure 1: Illustration of best predictor of X_{T+k} based on the historical information

Forecasting AR(p) model

1-Step-Ahead Forecast: From the AR(p) model, we have

$$y_{t+1} = c + \phi_1 y_t + \cdots + \phi_p y_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2).$$

- Under the *MSE loss function*, the *point forecast* of y_{t+1} given I_t is

$$y_{t+1|t} = E[y_{t+1}|I_t] = c + \phi_1 y_t + \cdots + \phi_p y_{t+1-p},$$

- The *1-step-ahead forecast error* is

$$\varepsilon_{t+1|t} = y_{t+1} - y_{t+1|t} = \varepsilon_{t+1}.$$

- The variance of *1-step-ahead forecast error* is $Var[\varepsilon_{t+1|t}] = Var(\varepsilon_{t+1}) = \sigma^2$.
- If at is *normally distributed*, then a *95% 1-step-ahead interval forecast* of y_{t+1} is $y_{t+1|t} \pm 1.96\sigma$.

Forecasting AR(p) model

2-Step-Ahead Forecast: From the AR(p) model, we have

$$y_{t+2} = c + \phi_1 y_{t+1} + \cdots + \phi_p y_{t+2-p} + \varepsilon_{t+2}. \tag{1}$$

- Taking conditional expectation, we have

$$y_{t+2|t} = E[y_{t+2}|I_t] = c + \phi_1 y_{t+1|t} + \phi_2 y_t + \cdots + \phi_p y_{t+2-p}, \tag{2}$$

- The associated forecast error

$$\varepsilon_{t+2|t} = y_{t+2} - y_{t+2|t} = \phi_1(y_{t+1} - y_{t+1|t}) + \varepsilon_{t+2} = \varepsilon_{t+2} + \phi_1 \varepsilon_{t+1}.$$

- The variance of the forecast error is $Var[\varepsilon_{t+2|t}] = (1 + \phi_1^2)\sigma^2$.
- If ε_t is normally distributed, then a 95% 1-step-ahead interval forecast of y_{t+1} is $y_{t+2|t} \pm 1.96\sigma\sqrt{1 + \phi_1^2}$.

Multistep-Ahead Forecast: In general, we have

$$y_{t+h} = c + \phi_1 y_{t+h-1} + \cdots + \phi_p y_{t+h-p} + \varepsilon_{t+h}. \tag{3}$$

- Taking conditional expectation, we have

$$y_{t+h|t} = E[y_{t+h}|I_t] = c + \sum_{i=1}^{p} \phi_i y_{t+h-i|t}, \tag{4}$$

  where $y_{t+\ell|t} = y_{t+\ell}$ if $\ell \leq 0$. This forecast can be computed recursively using forecasts $y_{t+i|t}$ for $i = 1, \ldots, h-1$.
- The h-step-ahead forecast error is

$$\varepsilon_{t+h|t} = y_{t+h} - y_{t+h|t}.$$

- How to obtain the forecasting interval for the AR(p) model?

Forecasting MA(q) models

Consider the MA(q) model:

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ &= \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}, \quad \text{with} \theta_0 \equiv 1. \end{aligned}$$

Given the *i.i.d.* properties of ε_t, the *optimal forecast* is

$$y_{t+h|t} = \begin{cases} \sum_{i=h}^{q} \theta_i \varepsilon_{t+h-i}, & \text{for } h = 1, \ldots, q, \\ 0, & \text{for } h > q, \end{cases}$$

whereas the corresponding forecast error follows as

$$\varepsilon_{t+h|t} = \begin{cases} \sum_{i=0}^{h-1} \theta_i \varepsilon_{t+h-i}, & \text{for } h = 1, \ldots, q, \\ \sum_{i=0}^{q} \theta_i \varepsilon_{t+h-i}, & \text{for } h > q, \end{cases}$$

which can be simplified to $\varepsilon_{t+h|t} = \sum_{i=0}^{h-1} \theta_i \varepsilon_{t+h-i}$ by defining $\theta_i \equiv 0$ for $h > q$.

Forecasting MA(q) models

Given the assumptions on ε_t, it follows that

$$E[\varepsilon_{t+h|t}] = 0,$$

and for *the mean squared error (MSE) of the forecast*

$$\mathrm{MSE}(\varepsilon_{t+h|t}) = E[\varepsilon_{t+h|t}^2] = \sigma^2 \sum_{i=0}^{h-1} \theta_i^2.$$

Assuming normality, a 95% *forecasting interval* for y_{t+h} is bounded by

$$\left(y_{t+h|t} - 1.96 \cdot \mathrm{RMSE}(\varepsilon_{t+h|t}), \quad y_{t+h|t} + 1.96 \cdot \mathrm{RMSE}(\varepsilon_{t+h|t})\right),$$

where $\mathrm{RMSE}(\varepsilon_{t+h|t})$ denotes the *square root* of $\mathrm{MSE}(\varepsilon_{t+h|t})$.

Forecasting ARMA(p, q) model:

The ARMA(p, q) model for y_t is

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

or, in lag operator form $\phi_p(B) y_t = \theta_q(B) \varepsilon_t$.

- The true forecast value is

$$y_{t+h} = \phi_1 y_{t+h-1} + \cdots + \phi_p y_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \cdots + \theta_q \varepsilon_{t+h-q}.$$

- The minimum MSE forecast for y_{t+h} is

$$y_{t+h|t} = \phi_1 y_{t+h-1|t} + \cdots + \phi_p y_{t+h-p|t} + \theta_1 \varepsilon_{t+h-1|t} + \cdots + \theta_q \varepsilon_{t+h-q|t},$$

  where $y_{t+\ell|t} = y_{t+\ell}$ if $\ell \leq 0$; and $\varepsilon_{t+\ell|t} = 0$ if $\ell > 0$, otherwise $\varepsilon_{t+\ell|t} = \varepsilon_{t+\ell}$.

Forecasting the ARMA(p, q) models by MA(∞) representation

- It is convenient to rewrite the model as an MA(∞) model, that is, $y_t = \phi_p(L)^{-1}\theta_q(L)\varepsilon_t$ or

$$y_t = \varepsilon_t + \eta_1\varepsilon_{t-1} + \eta_2\varepsilon_{t-2} + \eta_3\varepsilon_{t-3} + \cdots .$$

- The minimum MSE linear forecast (best linear predictor) of y_{t+h} based on I_t is

$$y_{t+h|t} = \eta_h\varepsilon_t + \eta_{h+1}\varepsilon_{t-1} + \cdots , \quad \text{with } \eta_0 \equiv 1.$$

- From which it follows that the *h-step-ahead prediction error* is given by

$$\varepsilon_{t+h|t} = \varepsilon_{t+h} + \eta_1\varepsilon_{t+h-1} + \cdots + \eta_{h-1}\varepsilon_{t+1} = \sum_{i=0}^{h-1} \eta_i\varepsilon_{t+h-i},$$

and the MSE of the forecast error is

$$\text{MSE}(\varepsilon_{t+h|t}) = \sigma^2 \sum_{i=0}^{h-1} \eta_i^2,$$

Homework

Consider the AR(1) model

$$(1 - \phi B)(y_t - \mu) = e_t, \quad e_t \sim i.i.d. N(0, 0.1)$$

with $\phi = 0.6$, $\mu = 9$. We have $y_{97} = 9.6$, $y_{98} = 9$, $y_{99} = 9$, and $y_{100} = 8.9$.

- Forecast y_{101}, y_{102}, y_{103}, y_{104} with their associated 95% forecast limits.
- If the observation y_{101} turned to be 8.8, update the forecasts of y_{102}, y_{103}, y_{104} .