Unit 12: Forecasting

Jeffrey Woo

Department of Statistics, University of Virginia

Spring 2020

Readings for Unit 12

Textbook chapter 3.4 (skip pages 103 to 106).

Last Unit

1. ACF for MA(q)
2. ACF for Causal ARMA(p,q)
3. Partial Autocorrelation Function

This Unit

1. Best linear predictor
2. ARMA forecasting

Motivation

In this unit, we explore forecasting: predicting future values of a time series based on observed data.

1. Forecasting for Stationary Processes

2. ARMA Forecasting

Forecasting

In forecasting, the goal is to predict future values of a time series, x_{n+m}, based on the observed data $\mathbf{x} = \{x_n, x_{n-1}, \cdots, x_1\}$. In this unit, we assume $\{x_t\}$ is stationary.

Forecasting

The minimum mean square error predictor of x_{n+m} is

$$x_{n+m}^{n} = \mathrm{E}(x_{n+m} | \mathbf{x}) \tag{1}$$

as the conditional expectation minimizes the mean square error $\mathrm{E}\left[x_{n+m} - g(\mathbf{x})\right]^2$, where $g(\mathbf{x})$ is a function of the observations.

Forecasting

First, we restrict our attention to predictors that are linear functions of the observations, i.e.

$$x_{n+m}^{n} = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j \tag{2}$$

where $\alpha_0, \alpha_1, \cdots, \alpha_n \in \mathbb{R}$. Linear predictors of the form (2) that minimize the mean square prediction error are called ____________ __________________.

Linear prediction depends on on the second-order moments of the process, which can be estimated from the data.

Projection Theorem

Theorem

Let $\mathcal{M}$ be a closed subspace of the Hilbert space $\mathcal{H}$ and let y be an element in $\mathcal{H}$. Then, y can be uniquely represented as $y = \hat{y} + z$ where $\hat{y} \in \mathcal{M}$ and z is orthogonal to $\mathcal{M}$. Therefore, for any $w \in \mathcal{M}$,

- $\|y - w\| \geq \|y - \hat{y}\|$ *and*
- $< z, w >= 0$.

Projection Theorem

Projection Theorem: Linear Prediction

Given $1, x_1, x_2, \cdots, x_n \in \{X : \mathrm{E}(X^2) < \infty\}$, choose $\alpha_0, \alpha_1, \cdots, \alpha_n \in \mathbb{R}$ so that $U = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j$ minimizes $\mathrm{E}(x_{n+m} - U)^2$.

Note that:
$\mathcal{M} = \{U = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j : \alpha_j \in \mathbb{R}\} = \bar{sp}\{1, x_1, \cdots, x_n\}$ and $y = x_{n+m}$.

Projection Theorem: Linear Prediction

Let x_{n+m}^n denote the best linear predictor, i.e.

$$\|x_{n+m}^n - x_{n+m}\|^2 \leq \|U - x_{n+m}\|^2$$

for all $U \in \mathcal{M}$. The projection theorem implies

Projection Theorem: Linear Prediction

- The prediction errors $x_{n+m}^n - x_{n+m}$ are orthogonal to the prediction variables $(1, x_1, \cdots, x_n)$.
- Orthogonality of prediction error and 1 implies we can ________________ from all variables x_{n+m} and x_i.
- Therefore, we typically assume $\mu = 0$ for forecasting.

BLP for Stationary Process

Given $x_1, \cdots, x_n$, the best linear predictor for stationary processes, $x_{n+m}^n = \alpha_0 + \sum_{j=1}^n \alpha_j x_j$, of x_{n+m}, for $m \geq 1$, is found by solving

$$\mathrm{E}\left[(x_{n+m} - x_{n+m}^n)x_k\right] = 0 \text{ for } k = 0, 1, \cdots, n, \tag{3}$$

where $x_0 = 1$, for $\alpha_0, \alpha_1, \cdots, \alpha_n$. The equations (3) are called the **prediction equations**.

One-Step-Ahead Linear Prediction

Consider one-step-ahead prediction. Given $x_1, \cdots, x_n$, we want to forecast x_{n+1}. The BLP takes the form

$$x_{n+1}^n = \phi_{n1} x_n + \phi_{n2} x_{n-1} + \cdots + \phi_{nn} x_1. \tag{4}$$

Therefore, the prediction equations (3) become

One-Step-Ahead Linear Prediction

In matrix form:

One-Step-Ahead Linear Prediction

The mean square one-step-ahead prediction error is

$$
\begin{aligned}
P_{n+1}^{n} &= \mathrm{E}(x_{n+1} - x_{n+1}^{n})^2 \\
&= \\
&= \\
&= \\
&= \\
&= \qquad (5)
\end{aligned}
$$

Prediction Intervals

Construct prediction interval:

$$x_{n+1}^{n} \pm 1.96\sqrt{P_{n+1}^{n}}.$$

for Gaussian processes. The prediction error has distribution $N(0, P_{n+1}^{n})$.

ARMA Forecasting

Let's consider an ARMA model that is causal and invertible

$$\phi(B)X_t = \theta(B)w_t,$$

where

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \quad \text{and} \quad \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

ARMA Forecasting

By causality and invertibility, we have

$$x_{n+m} = \sum_{j=0}^{\infty} \psi_j w_{n+m-j}, \qquad \psi_0 = 1, \tag{6}$$

where $\psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$, and

$$w_{n+m} = \sum_{j=0}^{\infty} \pi_j x_{n+m-j}, \qquad \pi_0 = 1, \tag{7}$$

where $\pi(z) = \frac{\phi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j z^j$.

ARMA Forecasting

Given the past information $x_n, x_{n-1}, \ldots$, we are interested in predicting x_{n+m}. We use $\widetilde{x}_{n+m} = E(x_{n+m}|x_n, x_{n-1}, ...)$, the conditional expectation of x_{n+m} given all the past $x_n, x_{n-1}, \ldots$, to forecast x_{n+m}.

ARMA Forecasting

Note also that

$$E(w_t|x_n, x_{n-1}, ...) = 0$$

for $t > n$ because of causality. For $t \leq n$, w_t is determined by $x_t, x_{t-1}, \ldots$, which are included in $x_n, x_{n-1}, \ldots$. Thus,

$$E(w_t|x_n, x_{n-1}, ...) = w_t.$$

ARMA Forecasting

So

$$E(w_t | x_n, x_{n-1}, \ldots) = \begin{cases} 0, & t > n, \\ w_t, & t \leq n. \end{cases} \tag{8}$$

ARMA Forecasting

Now, we take the infinite AR representation (7) and take the conditional expectation (conditioning on $x_n, x_{n-1}, ...$) on both sides of (7) to get

ARMA Forecasting

This leads to

$$\widetilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \widetilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}. \tag{9}$$

Letting $m = 1$ in (9), we have

$$\widetilde{x}_{n+1} = -\sum_{j=1}^{\infty} \pi_j x_{n+1-j} = -\sum_{j'=0}^{\infty} \pi_{j'+1} x_{n-j'}.$$

So, start by finding $\widetilde{x}_{n+1}$ and then recursively use (9) to find the later $\widetilde{x}_{n+m}$. This is called the ____________________.

Worked Example

We have an AR(2) model $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$. Suppose we have observations up to the nth term, i.e. $x_n, x_{n-1}, \cdots$. So we wish to use the observed data and estimated AR(2) model to forecast the value of x_{n+1} and x_{n+2}. Writing out these two values, we have

$$\begin{aligned} x_{n+1} &= \phi_1 x_n + \phi_2 x_{n-1} + w_{n+1}, \\ x_{n+2} &= \phi_1 x_{n+1} + \phi_2 x_n + w_{n+2}. \end{aligned}$$

Worked Example

To forecast x_{n+1}, we use the observed values of x_n and x_{n-1} and replace w_{n+1} by its expected value of 0.

Forecasting x_{n+2} poses a challenge, since it requires the unobserved value of x_{n+1}. We use the forecasted value of x_{n+1}.

Framework in Forecasting

In general, the forecasting procedure for an ARMA(p,q) model is as follows:

- For any w_j with $1 \leq j \leq n$, use the sample residual at time j.
- For any w_j with $j > n$, use the expected value of w_j, which is 0.
- For any x_j with $1 \leq j \leq n$, use the observed value of x_j.
- For any x_j with $j > n$, use the forecasted value of x_j.

Prediction Error

We use the infinite MA representation (6) and write

Prediction Error

Therefore, the mean-squared prediction error, or variance of the difference between the forecasted value and the true value at time $n+m$ is

$$P^n_{n+m} = E(x_{n+m} - \widetilde{x}_{n+m})^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2, \tag{10}$$

and the standard error of the forecast error at time $n+m$ is

$$\sqrt{\hat{\sigma_w^2} \sum_{j=0}^{m-1} \psi_j^2}. \tag{11}$$

Prediction Error

Question: Write out the standard error of the forecast error for $m = 1$ and $m = 2$.

Prediction Error

Notice that as m gets larger–i.e. as we predict further into the future, this is __________ but essentially asymptotes. This means that you are getting essentially a constant prediction interval after a certain distance into the future, as if we do not know what was going on previously.

Prediction Error

Also, for fixed sample size n, the prediction errors are correlated. For $h \geq 1$,

$$E\left\{(x_{n+m} - \widetilde{x}_{n+m})(x_{n+m+h} - \widetilde{x}_{n+m+h})\right\} = \sigma^2 \sum_{j=0}^{m-1} \psi_j \psi_{j+k}.$$

Prediction Interval

For Gaussian processes, the 95% prediction interval for x_{n+m}, the future value of the series at time $n+m$ is

$$x_{n+m}^{n} \pm 1.96 \sqrt{\hat{\sigma}_w^2 \sum_{j=0}^{m-1} \psi_j^2}. \tag{12}$$

Worked Example

Question: We have an AR(1) model $x_t = 40 + 0.6x_{t-1} + w_t$. Suppose we have $n = 100$ observations, $\hat{\sigma_w^2} = 1$ and $x_{100} = 80$. We wish to forecast the values at times 101 and 102.

Worked Example