

Unit 12: Forecasting

Jeffrey Woo

Department of Statistics, University of Virginia

Spring 2016

Readings for Unit 12

Textbook chapter 3.5.

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 a forecasting: predicting future values of a time series based on observed data.

1 Best Linear Predictor

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 $\mathbf{x} = \{x_n, x_{n-1}, \dots, x_1\}$. In this unit, we assume $\{x_t\}$ is stationary.

Forecasting

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

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

as the conditional expectation minimizes the mean square error $E[x_{n+m} - g(\mathbf{x})]^2$, 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.

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

where $\alpha_0, \alpha_1, \dots, \alpha_n \in \mathbb{R}$. 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*
- $\langle z, w \rangle = 0$.*

Projection Theorem

Projection Theorem: Linear Prediction

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

Note that:

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

Projection Theorem: Linear Prediction

Let x_{n+m} 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 $(1, x_1, \dots, x_n)$.
- 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, \dots, 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

$$E [(x_{n+m} - x_{n+m}^n)x_k] = 0 \text{ for } k = 0, 1, \dots, n, \quad (3)$$

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

One-Step-Ahead Linear Prediction

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

$$x_{n+1}^n = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \dots + \phi_{nn}x_1. \quad (4)$$

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

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

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 $N(0, P_{n+1}^n)$.

1 Best Linear Predictor

2 ARMA Forecasting

ARMA Forecasting

The prediction equations (3) for stationary processes provide little insight into forecasting for ARMA models. 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

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

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

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

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

ARMA Forecasting

Note also that

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

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

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

ARMA Forecasting

So

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

ARMA Forecasting

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

ARMA Forecasting

This leads to

$$\tilde{x}_{n+m} = - \sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}. \quad (9)$$

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

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

So, start by finding \tilde{x}_{n+1} and then recursively use (9) to find the later \tilde{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 n th term, i.e. x_n, x_{n-1}, \dots . 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

$$x_{n+1} = \phi_1 x_n + \phi_2 x_{n-1} + w_{n+1},$$

$$x_{n+2} = \phi_1 x_{n+1} + \phi_2 x_n + w_{n+2}.$$

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

$$P_{n+m}^n = E(x_{n+m} - \tilde{x}_{n+m})^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2, \quad (10)$$

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

$$\sqrt{\hat{\sigma}_w^2 \sum_{j=0}^{m-1} \psi_j^2}. \quad (11)$$

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 farther 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 $h \geq 1$,

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

Prediction Interval

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

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

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