STAT 153 & 248 - Time Series
Lecture Nineteen

Spring 2025, UC Berkeley

Aditya Guntuboyina

April 01, 2025

1 Point Predictions from $AR(p)$ models

If the parameters $\phi_0, \ldots, \phi_p$ and σ (collectively denoted by θ) of the AR(p) model are fixed, then the prediction for a future value y_{n+i} is given by

$$\hat{y}_{n+i}(\theta) := \mathbb{E}\left(y_{n+i} \mid y_1, \ldots, y_n, \theta\right)$$

These values are calculated recursively for $i = 1, 2, \ldots$ using the following recursion:

$$\hat{y}_{n+i}(\theta) = \phi_0 + \phi_1 \hat{y}_{n+i-1}(\theta) + \phi_2 \hat{y}_{n+i-2}(\theta) + \cdots + \phi_p \hat{y}_{n+i-p}(\theta) \qquad \text{for } i = 1, 2, \ldots. \tag{1}$$

which is initialized by

$$\hat{y}_j(\theta) = y_j \qquad \text{for } j = n, n-1, \ldots, n+1-p, \tag{2}$$

Since θ is unknown, we can replace it by the conditional MLE $\hat{\theta}$. Alternatively, one can try to compute:

$$\mathbb{E}(y_{n+i} \mid y_1, \ldots, y_n) = \int \hat{y}_{n+i}(\theta) f_{\theta|y_1,\ldots,y_n}(\theta) d\theta \tag{3}$$

numerically using posterior samples of θ. This method is complicated and the common procedure is simply to replace θ by the conditional MLE $\hat{\theta}$.

2 Prediction Standard Errors

To compute prediction standard errors, we can again fix the parameter values θ, and then attempt to calculate:

$$V_i(\theta) := \text{var}\left(y_{n+i} \mid y_1, \ldots, y_n, \theta\right) \qquad \text{for } i = 1, 2, \ldots.$$

The prediction standard error corresponding to the predicted value for y_{n+i} can then be taken to be $\sqrt{V_i(\hat{\theta})}$ (note that θ is replaced by the conditional MLE $\hat{\theta}$).

It turns out that it is difficult to directly setup a recursion for $V_i(\theta)$. Instead, we will get the recursion by working with the conditional **covariance matrices** of $y_{n+1}, \ldots, y_{n+k}$ (given θ and the data) for $k = 1, 2, \ldots$.

Below we review some basic facts about covariance matrices.

2.1 Covariance Matrices

A finite number of random variables can be viewed together as a random vector. More precisely, a random vector is a vector whose entries are random variables. Let $Y = (Y_1, \ldots, Y_n)^T$ be an $n \times 1$ random vector. Its Expectation $\mathbb{E}Y$ is defined as a vector whose ith entry is the expectation of Y_i i.e., $\mathbb{E}Y = (\mathbb{E}Y_1, \mathbb{E}Y_2, \ldots, \mathbb{E}Y_n)^T$. The covariance matrix of Y, denoted by $\text{Cov}(Y)$, is an $n \times n$ matrix whose (i,j)th entry is the covariance between Y_i and Y_j. Two important but easy facts about $\text{Cov}(Y)$ are:

1. The diagonal entries of $\text{Cov}(Y)$ are the variances of $Y_1, \ldots, Y_n$. More specifically the (i,i)th entry of the matrix $\text{Cov}(Y)$ equals $var(Y_i)$.
2. $\text{Cov}(Y)$ is a symmetric matrix i.e., the (i,j)th entry of $\text{Cov}(Y)$ equals the (j,i) entry. This follows because $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$.

The following formulae are very important:

1. $\mathbb{E}(AY+c) = A\mathbb{E}(Y)+c$ for every deterministic matrix A and every deterministic vector c.
2. $\text{Cov}(AY + c) = A\text{Cov}(Y)A^T$ for every deterministic matrix A and every deterministic vector c.

As a consequence of the second formula above, we get

$$\text{var}(a^T Y) = a^T \text{Cov}(Y) a = \sum_{i,j} a_i a_j \text{Cov}(Y_i, Y_j) \qquad \text{for every } p \times 1 \text{ vector } a.$$

Given two random vectors Y $(p \times 1)$ and W $(q \times 1)$, we use $\text{Cov}(Y, W)$ to denote the $p \times q$ matrix whose $(i,j)^{th}$ entry equals the covariance $\text{Cov}(Y_i, W_j)$ between Y_i and W_j. With this definition, the previous notion of $\text{Cov}(Y)$ equals simply $\text{Cov}(Y, Y)$. It can be checked that

$$\text{Cov}(AY + c, BW + d) = A\text{Cov}(Y, W)B^T.$$

2.2 Covariance Recursion in $AR(p)$

We shall set up a recursion for the covariance matrices:

$$\Gamma_k(\theta) := \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k} \end{pmatrix} \mid \theta, y_1, \ldots, y_n \right)$$

The $(i,j)^{th}$ entry of $\Gamma_k(\theta)$ is

$$\text{Cov}\left(y_{n+i}, y_{n+j} \mid y_1, \ldots, y_n, \theta\right).$$

The diagonal entries of $\Gamma_k(\theta)$ equal $V_1(\theta), \ldots, V_k(\theta)$.

To initialize the recursion for $\Gamma_k(\theta)$, note that

$$\begin{aligned} \Gamma_1(\theta) &= \text{var}\left(y_{n+1} \mid y_1, \ldots, y_n, \theta\right) \\ &= \text{var}\left(\phi_0 + \phi_1 y_n + \ldots \phi_p y_{n+1-p} + \epsilon_{n+1} \mid y_1, \ldots, y_n, \theta\right) \\ &= \text{var}\left(\epsilon_{n+1} \mid y_1, \ldots, y_n, \theta\right) = \sigma^2 \end{aligned}$$

We now relate $\Gamma_{k+1}(\theta)$ to $\Gamma_k(\theta)$ to establish the recursion. We can write

$$\Gamma_k(\theta) = \begin{pmatrix} \Gamma_{k-1}(\theta) & \gamma_{k1}(\theta) \\ \gamma_{k1}^T(\theta) & V_k(\theta) \end{pmatrix}$$

where

$$\gamma_{k1}(\theta) := \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix}, y_{n+k} \mid \theta, y_1, \ldots, y_n \right)$$

and, as before,

$$V_k(\theta) = \text{var}\left(y_{n+k} \mid \theta, y_1, \ldots, y_n\right).$$

We compute $\hat{\gamma}_{k1}$ as

$$\begin{aligned}
\gamma_{k1}(\theta) &:= \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix}, y_{n+k} \mid \theta, y_1, \ldots, y_n \right) \\
&= \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix}, \phi_0 + \phi_1 y_{n+k-1} + \phi_2 y_{n+k-2} + \cdots + \phi_p y_{n+k-p} + \epsilon_{n+k} \mid \theta, \text{data} \right) \\
&= \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix}, \phi_1 y_{n+k-1} + \phi_2 y_{n+k-2} + \cdots + \phi_p y_{n+k-p} \mid \theta, \text{data} \right) \\
&= \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix}, \sum_{i=1}^{k-1} a_i y_{n+i} \mid \theta, \text{data} \right)
\end{aligned}$$

where, for $i = 1, \ldots, k-1$,

$$a_i = \begin{cases} \phi_{k-i} & \text{provided } k-p \le i \le k-1 \\ 0 & \text{otherwise} \end{cases}$$

Thus if a is the $(k-1) \times 1$ vector with entries $a_1, \ldots, a_{k-1}$, we have

$$\gamma_{k1}(\theta) = \text{Cov}\left(\begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix}, a^T \begin{pmatrix} y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{n+k-1} \end{pmatrix} \mid \theta, \text{data} \right) = \Gamma_{k-1}(\theta) a.$$

Further

$$\begin{aligned} V_k(\theta) &= \text{var}\left(y_{n+k} \mid \theta, \text{data}\right) \\ &= \text{var}\left(\phi_0 + \phi_1 y_{n+k-1} + \phi_2 y_{n+k-2} + \cdots + \phi_p y_{n+k-p} + \epsilon_{n+k} \mid \theta, \text{data}\right) \\ &= \text{var}\left(\phi_0 + \phi_1 y_{n+k-1} + \phi_2 y_{n+k-2} + \cdots + \phi_p y_{n+k-p} \mid \theta, \text{data}\right) + \sigma^2 \\ &= \text{var}\left(\sum_{i=1}^{k-1} a_i y_{n+i} \mid \theta, \text{data}\right) + \sigma^2 = a^T \Gamma_{k-1}(\theta) a + \sigma^2. \end{aligned}$$

The equation for obtaining $\Gamma_k(\theta)$ from $\Gamma_{k-1}(\theta)$ is therefore

$$\Gamma_k(\theta) = \begin{pmatrix} \Gamma_{k-1}(\theta) & \gamma_{k1}(\theta) \\ \gamma_{k1}^T(\theta) & V_k(\theta) \end{pmatrix} = \begin{pmatrix} \Gamma_{k-1}(\theta) & \Gamma_{k-1}(\theta)a \\ a^T\Gamma_{k-1}(\theta) & a^T\Gamma_{k-1}(\theta)a + \sigma^2 \end{pmatrix}$$

The algorithm for calculating the variances $V_i(\theta)$ for $i = 1, 2, \ldots, K$ is thus given by

1. Initialize with $\Gamma_1(\theta) = V_1(\theta) = \sigma^2$.
2. For $k = 2, 3, \ldots, K$, repeat the following
   a) Form the $(k-1) \times 1$ vector a whose i^{th} entry is ϕ_{k-i} if $k - p \leq i \leq k-1$ and 0 otherwise.
   b) Calculate $\Gamma_k(\theta)$ using $\Gamma_{k-1}(\theta)$ and a by the formula given above.
3. The variances $V_i(\theta), i = 1, 2, \ldots, K$ are given by the diagonal entries of the matrix $\Gamma_K(\theta)$.

Because θ is unknown, in practice, we run this recursion with θ replaced by its conditional MLE $\hat{\theta}$. The prediction standard errors are then the square roots of $V_i(\hat{\theta})$.

3 Time Series Models

Before proceeding further, let us take a general look at time series models (especially the models that we have already considered), and introduce the concept of "Stationarity". Stationarity is an important property that some time series models satisfy while others do not.

We have already seen many models for observed time series data $y_1, \ldots, y_n$. These models describe the distribution of $y_1, \ldots, y_n$ in terms of various parameters. Even though, the observed data only corresponds to times $t = 1, \ldots, n$, it makes sense for the model to describe the distribution of y_t for all t past and present i.e., $t = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$. This is because one may be interested in predicting the values of y_t for unobserved times.

All of our models have error terms ϵ_t that we assume are i.i.d Gaussian $N(0, \sigma^2)$. This ensures that the whole time series $\{y_t\}$ is jointly Gaussian. Gaussianity ensures that the behavior of the time series model is characterized by means (expectations) and covariances.

Example 3.1. *$y_t = \beta_0 + \beta_1 t + \epsilon_t$. The means are given by:*

$$\mathbb{E}y_t = \beta_0 + \beta_1 t$$

and covariances are:

$$var(y_t) = \sigma^2 \quad \text{and} \quad cov(y_{t_1}, y_{t_2}) = 0 \ \text{ for } t_1 \neq t_2.$$

So the mean changes with t, variance is constant and there is no correlation between different time points.

Example 3.2. *$y_t = \beta_0 + \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft) + \epsilon_t$ The means are given by:*

$$\mathbb{E}y_t = \beta_0 + \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft)$$

and covariances are:

$$var(y_t) = \sigma^2 \quad \text{and} \quad cov(y_{t_1}, y_{t_2}) = 0 \ \text{ for } t_1 \neq t_2.$$

Again the mean changes with t, variance is constant and there is no correlation between different time points.

Example 3.3. *Consider the Spectrum model:*

$$y_t = \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi jt}{n} + \beta_{2j} \sin \frac{2\pi jt}{n} \right)$$

with $\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \ldots, \beta_{1m}, \beta_{2m}$ all independent with

$$\beta_{1j}, \beta_{2j} \overset{i.i.d}{\sim} N(0, \tau_j^2).$$

For this model, $\mathbb{E}y_t = \beta_0$ so that the mean is constant over time (unlike the previous two models). The covariance is given by

$$\begin{aligned}
&cov(y_{t_1}, y_{t_2}) \\
&= cov\left(\beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi jt_1}{n} + \beta_{2j} \sin \frac{2\pi jt_1}{n} \right), \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi jt_2}{n} + \beta_{2j} \sin \frac{2\pi jt_2}{n} \right) \right) \\
&= \sum_{j=1}^{m} \left\{ cov\left(\beta_{1j} \cos \frac{2\pi jt_1}{n}, \beta_{1j} \cos \frac{2\pi jt_2}{n} \right) + cov\left(\beta_{2j} \sin \frac{2\pi jt_1}{n}, \beta_{2j} \sin \frac{2\pi jt_2}{n} \right) \right\} \\
&= \sum_{j=1}^{m} \left\{ \tau_j^2 \cos \frac{2\pi jt_1}{n} \cos \frac{2\pi jt_2}{n} + \tau_j^2 \sin \frac{2\pi jt_1}{n} \sin \frac{2\pi jt_2}{n} \right\} \\
&= \sum_{j=1}^{m} \tau_j^2 \cos \frac{2\pi j(t_1 - t_2)}{n} = \sum_{j=1}^{m} \tau_j^2 \cos \frac{2\pi j|t_1 - t_2|}{n}
\end{aligned}$$

This model incorporates dependence between y_t at different time points t (unlike the previous two models). Further, the covariance between y_{t_1} and y_{t_2} only depends on the distance $|t_1 - t_2|$ between the two time points.

4 Stationarity

Definition 4.1 (Stationarity)**.** *A doubly infinite sequence of random variables y_t is stationary*

1. *The mean of y_t (denoted by $\mathbb{E}y_t$) is the same for all times t*
2. *The covariance between y_{t_1} and y_{t_2} only depends on the distance $|t_1 - t_2|$ between t_1 and t_2.*

The third example of the previous section is a stationary model while the other two models are not stationary because the means therein change with time. We shall see more examples of stationary models next week.

4.1 Additional Optional Reading

1. For more on predictions, see Section 3.4 of Shumway-Stoffer 4th edition.
2. For more on stationarity, see Chapter 1 of Shumway-Stoffer 4th edition.