STAT 153 - Introduction to Time Series
Lecture Fifteen

Fall 2022, UC Berkeley

Aditya Guntuboyina

October 15, 2022

1 Last Class

In the last class, we saw how we can fit the $AR(p)$ model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t \qquad \text{with } Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2) \tag{1}$$

to observed time series data $y_1, \ldots, y_n$. Standard linear regression methodology gives estimates $\hat{\phi}_0, \ldots, \hat{\phi}_p, \hat{\sigma}$ as well as uncertainty quantification.

We saw how to predict future values of the time series $Y_{n+1}, Y_{n+2}, \ldots$ using this model. Point estimates of the predictions are given by

$$\hat{Y}_{n+i} := \mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \ldots, Y_n = y_n, \hat{\theta}\right) \qquad \text{for } i = 1, 2, \ldots$$

where $\hat{\theta} = (\hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_p)$ denote the estimates. We calculate these via the recursion

$$\hat{Y}_{n+i} = \hat{\phi}_0 + \hat{\phi}_1 \hat{Y}_{n+i-1} + \hat{\phi}_2 \hat{Y}_{n+i-2} + \cdots + \phi_p \hat{Y}_{n+i-p} \qquad \text{for } i = 1, 2, \ldots \tag{2}$$

which is initialized with

$$\hat{Y}_j = y_j \qquad \text{for } j = n, n-1, \ldots, n+1-p. \tag{3}$$

The recursion (2) is evaluated in sequence for $i = 1, 2, \ldots$ to calculate $\hat{Y}_{n+i}$ for all $i \geq 1$.

2 Difference Equations and their Solutions

To understand the behavior of predictions generated by $AR(p)$ models, we need to look at solutions to difference equations of the form (2). Consider the difference equation

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + \cdots + \alpha_p u_{k-p} \qquad \text{for } k = p, p+1, p+2, \ldots. \tag{4}$$

which is initialized by specifying the values of $u_0, u_1, \ldots, u_{p-1}$.

We looked at the solutions to the first order ($p = 1$) and second order ($p = 2$) cases in the last class.

2.1 First Order

Here $p = 1$ so the difference equation becomes:

$$u_k = \alpha_0 + \alpha_1 u_{k-1} \qquad \text{for } k = 1, 2, \ldots$$

along with an initial value specification for u_0. The solution is given by

$$u_k = \alpha_0 \sum_{0 \le j < k} \alpha_1^j + \alpha_1^k u_0 \qquad \text{for } k = 0, 1, 2, \ldots. \tag{5}$$

2.2 Second Order

Here $p = 2$ so the difference equation becomes

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + \alpha_2 u_{k-2} \qquad \text{for } k = 2, 3, \ldots$$

along with an initial value specification for u_0 and u_1.

We wrote down the solution to this equation when $\alpha_1 + \alpha_2 \neq 1$. In this case, it is easy to see that

$$v_k = u_k - \frac{\alpha_0}{1 - \alpha_1 - \alpha_2} \tag{6}$$

solves the homogeneous equation:

$$v_k = \alpha_1 v_{k-1} + \alpha_2 v_{k-2} \qquad \text{for } k = 2, 3, \ldots. \tag{7}$$

A key role in the general solution for this equation is played by the characteristic equation:

$$\phi(z) = 1 - \alpha_1 z - \alpha_2 z^2.$$

Let z_1 and z_2 denote the roots of this polynomial. We then consider the three cases:

Case 1: z_1 and z_2 are real and distinct: Here the general solution is given by

$$v_k = c_1 z_1^{-k} + c_2 z_2^{-k} \tag{8}$$

is a solution to (7) for every real c_1 and c_2. If we now choose c_1 and c_2 so as to satisfy the initial values given for v_0 and v_1, we would obtain the solution we are seeking.

Case 2: z_1 and z_2 are complex and distinct: Here we must have $z_2 = \bar{z}_1$. The general solution is again given by (8) but because v_k needs to be real, we have to take $c_2 = \bar{c_1}$. By writing $z_1 = |z_1| e^{i\theta}$ and $\bar{c_1} = a e^{ib}$, we can simplify the expression (8) for v_k as

$$v_k = |z_1|^{-k} 2a \cos(b + k\theta) \tag{9}$$

This expression has two constants a and b which need to be chosen so as to satisfy the initial conditions for v_0 and v_1. The interesting thing to note is that, in this case where the roots of $\phi(z)$ are complex, the predictions will be oscillating.

Case 3: z_1 and z_2 are equal. Here $z_1 = z_2$ must necessarily be real. The solution to (7) is now given by

$$v_k = c_1 z_1^{-k} + c_2 k z_1^{-k} = z_1^{-k} \left(c_1 + k c_2 \right).$$

In the case when $\alpha_1 + \alpha_2 = 1$, the above analysis is not applicable because we cannot define v_k as in (6). We instead do the following. As $\alpha_2 = 1 - \alpha_1$, we have

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + (1 - \alpha_1) u_{k-2}$$

which is same as

$$u_k - u_{k-2} = \alpha_0 + \alpha_1 \left(u_{k-1} - u_{k-2} \right).$$

Let $w_k := u_{k+1} - u_k$ for $k = 0, 1, 2, \dots$ so the above equation can be written as

$$w_{k-1} + w_{k-2} = \alpha_0 + \alpha_1 w_{k-2}.$$

Moving the w_{k-2} to the right hand side and switching the index $k-1$ to k, we obtain

$$w_k = \alpha_0 + (\alpha_1 - 1)\, w_{k-1} \qquad \text{for } k = 1, 2, \dots$$

This is a first order equation to (5) gives

$$w_k = \alpha_0 \sum_{0 \le j < k} (\alpha_1 - 1)^j + (\alpha_1 - 1)^k\, w_0 = \alpha_0 \sum_{0 \le j < k} (\alpha_1 - 1)^j + (\alpha_1 - 1)^k \left(u_1 - u_0 \right).$$

The solution of u_k for $k \ge 2$ is then given by (remember $w_k = u_{k+1} - u_k$)

$$u_k = u_1 + (u_2 - u_1) + \cdots + (u_k - u_{k-1}) = u_1 + w_1 + \cdots + w_{k-1} \qquad \text{for } k = 2, 3, \dots$$

2.3 General Order

Now we describe solutions for the general order difference equation (4) with initial conditions for $u_0, u_1, \dots, u_{p-1}$. When

$$\alpha_1 + \cdots + \alpha_p \neq 1,$$

we can take

$$v_k = u_k - \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_p} \tag{10}$$

which satisfies the homogeneous equation

$$v_i = \alpha_1 v_{k-1} + \alpha_2 v_{k-2} + \cdots + \alpha_p v_{k-p}. \tag{11}$$

The characteristic equation of this homogeneous equation is

$$\phi(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p.$$

Let the roots of this equation be $z_1, \dots, z_r$ with multiplicities $m_1, \dots, m_r$ (each m_j is at least 1 and $m_1 + \cdots + m_r = p$). The general solution to (11) is given by

$$v_k = z_1^{-k} P_1(k) + z_2^{-k} P_2(k) + \cdots + z_r^{-k} P_r(k)$$

where $P_j(k)$ is a polynomial of degree $m_j - 1$ for each $j = 1, \dots, r$. The total number of coefficients of these polynomials is $m_1 + \cdots + m_r = p$ which is same as the number of initial values $v_0, \dots, v_{p-1}$. These can thus be determined from the initial values.

When $\alpha_1 + \cdots + \alpha_p = 1$, we cannot use (10). Then writing $\alpha_p = 1 - \alpha_1 - \cdots - \alpha_{p-1}$, the equation (4) is equivalent to

$$u_k - u_{k-p} = \phi_0 + \phi_1 (u_{k-1} - u_{k-p}) + \cdots + \phi_{p-1} (u_{k-p+1} - u_{k-p}).$$

Letting $w_k = u_{k+1} - u_k$ so that $u_j - u_{k-p} = w_{j-1} + \cdots + w_{k-p}$, the above equation is equivalent to

$$w_{k-1} = \alpha_0 + (\alpha_1 - 1)\, w_{k-2} + (\alpha_1 + \alpha_2 - 1)\, w_{k-3} + \cdots + (\alpha_1 + \cdots + \alpha_{p-1} - 1)\, w_{k-p}$$

which is same as (increasing index by 1 so the right hand side w_k)

$$w_k = \alpha_0 + (\alpha_1 - 1)\, w_{k-1} + (\alpha_1 + \alpha_2 - 1)\, w_{k-2} + \cdots + (\alpha_1 + \cdots + \alpha_{p-1} - 1)\, w_{k-p+1}$$

for $k = p-1, p, \dots$. One can then solve this and then get the solution for u_k via

$$u_k = u_{p-1} + w_{p-1} + \dots w_{k-1} \qquad \text{for } k = p, p+1, p+2, \dots.$$

3 Prediction Uncertainty

We shall next discuss how to provide uncertainty quantification for predictions given by the $AR(p)$ model. This can be done via the variance of the future observations given the data. Specifically by

$$\operatorname{var}(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n) \qquad \text{for } i = 1, 2, \dots.$$

and the corresponding standard deviations. We approximate these conditional variances as (below we write "data" for $Y_1 = y_1, \dots, Y_n = y_n$)

$$\operatorname{var}(Y_{n+i} \mid \text{data}) = \mathbb{E}\left(\operatorname{var}(Y_{n+i} \mid \theta, \text{data}) \mid \text{data}\right) + \operatorname{var}\left(\mathbb{E}(Y_{n+i} \mid \theta, \text{data}) \mid \text{data}\right)$$

The second term above is generally small. This is because $\mathbb{E}(Y_{n+i} \mid \theta, \text{data})$ is a function of θ and then we take the expectation of θ with respect to the posterior distribution. Because the posterior distribution is usually quite concentrated, the variance will be small. We shall thus ignore the second term and write

$$\operatorname{var}(Y_{n+i} \mid \text{data}) \approx \mathbb{E}\left(\operatorname{var}(Y_{n+i} \mid \theta, \text{data}) \mid \text{data}\right)$$

Because the posterior of θ will be concentrated around $\hat{\theta}$, we get

$$\operatorname{var}(Y_{n+i} \mid \text{data}) \approx \operatorname{var}\left(Y_{n+i} \mid \theta = \hat{\theta}, \text{data}\right).$$

We thus determine uncertainty in our predictions by the variances

$$\hat{V}_i := \operatorname{var}\left(Y_{n+i} \mid \theta = \hat{\theta}, \text{data}\right)$$

and the associated standard deviations $\sqrt{\hat{V}_i}$ for $i = 1, 2, \dots$. In order to calculate, we can attempt to set up a recursion for these quantities. However, it is difficult to do this. We will get the recursion by instead working with the conditional covariance matrices of $Y_{n+1}, \dots, Y_{n+k}$ for $k = 1, 2, \dots$. Let us first review some basic formulae for covariance matrices.

4 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 $Y = (Y_1, \dots, Y_n)^T$ 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., $\mathbb{E}Y = (\mathbb{E}Y_1, \mathbb{E}Y_2, \dots, \mathbb{E}Y_n)^T$. 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

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

5 Back to Prediction Uncertainty for $AR(p)$ Models

We shall set up a recursion for the covariance matrices:

$$\hat{\Gamma}_k := \text{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k} \end{pmatrix} \mid \theta = \hat{\theta}, \text{data} \right)$$

The $(i,j)^{th}$ entry of $\hat{\Gamma}_k$ is

$$\text{Cov}\left(Y_{n+i}, Y_{n+j} \mid Y_1 = y_1, \ldots, Y_n = y_n, \theta = \hat{\theta} \right).$$

To initialize this recursion, note that

$$\begin{aligned}
\hat{\Gamma}_1 &= \text{var}\left(Y_{n+i} \mid Y_1 = y_1, \ldots, Y_n = y_n, \theta = \hat{\theta} \right) \\
&= \text{var}\left(\hat{\phi}_0 + \hat{\phi}_1 y_n + \ldots \hat{\phi}_p y_{n+1-p} + Z_{n+1} \mid Y_1 = y_1, \ldots, Y_n = y_n, \theta = \hat{\theta} \right) \\
&= \text{var}\left(Z_{n+1} \mid Y_1 = y_1, \ldots, Y_n = y_n, \theta = \hat{\theta} \right) = \hat{\sigma}^2
\end{aligned}$$

Let us now relate $\hat{\Gamma}_{k+1}$ to $\hat{\Gamma}_k$. We can write

$$\hat{\Gamma}_k = \begin{pmatrix} \hat{\Gamma}_{k-1} & \hat{\gamma}_{k1} \\ \hat{\gamma}_{k1}^T & \hat{V}_k \end{pmatrix}$$

where

$$\hat{\gamma}_{k1} := \text{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix}, Y_{n+k} \mid \theta = \hat{\theta}, \text{data} \right)$$

and, as before,

$$\hat{V}_k = \operatorname{var}\left(Y_{n+k} \mid \theta = \hat{\theta}, \text{data}\right).$$

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

$$\begin{aligned}
\hat{\gamma}_{k1} &:= \operatorname{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix}, Y_{n+k} \mid \theta = \hat{\theta}, \text{data}\right) \\
&= \operatorname{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix}, \hat{\phi}_0 + \hat{\phi}_1 Y_{n+k-1} + \hat{\phi}_2 Y_{n+k-2} + \cdots + \hat{\phi}_p Y_{n+k-p} + Z_{n+k} \mid \theta = \hat{\theta}, \text{data}\right) \\
&= \operatorname{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix}, \hat{\phi}_1 Y_{n+k-1} + \hat{\phi}_2 Y_{n+k-2} + \cdots + \hat{\phi}_p Y_{n+k-p} \mid \theta = \hat{\theta}, \text{data}\right) \\
&= \operatorname{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix}, \sum_{i=1}^{k-1} a_i Y_{n+i} \mid \theta = \hat{\theta}, \text{data}\right)
\end{aligned}$$

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

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

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

$$\hat{\gamma}_{k1} = \operatorname{Cov}\left(\begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix}, a^T \begin{pmatrix} Y_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{n+k-1} \end{pmatrix} \mid \theta = \hat{\theta}, \text{data}\right) = \hat{\Gamma}_{k-1} a.$$

Further

$$\begin{aligned}
\hat{V}_k &= \operatorname{var}\left(Y_{n+k} \mid \theta = \hat{\theta}, \text{data}\right) \\
&= \operatorname{var}\left(\hat{\phi}_0 + \hat{\phi}_1 Y_{n+k-1} + \hat{\phi}_2 Y_{n+k-2} + \cdots + \hat{\phi}_p Y_{n+k-p} + Z_{n+k} \mid \theta = \hat{\theta}, \text{data}\right) \\
&= \operatorname{var}\left(\hat{\phi}_0 + \hat{\phi}_1 Y_{n+k-1} + \hat{\phi}_2 Y_{n+k-2} + \cdots + \hat{\phi}_p Y_{n+k-p} \mid \theta = \hat{\theta}, \text{data}\right) + \hat{\sigma}^2 \\
&= \operatorname{var}\left(\sum_{i=1}^{k-1} a_i Y_{n+i} \mid \theta = \hat{\theta}, \text{data}\right) + \hat{\sigma}^2 = a^T \hat{\Gamma}_{k-1} a + \hat{\sigma}^2.
\end{aligned}$$

The equation for obtaining $\hat{\Gamma}_k$ from $\hat{\Gamma}_{k-1}$ is therefore

$$\hat{\Gamma}_k = \begin{pmatrix} \hat{\Gamma}_{k-1} & \hat{\gamma}_{k1} \\ \hat{\gamma}_{k1}^T & \hat{V}_k \end{pmatrix} = \begin{pmatrix} \hat{\Gamma}_{k-1} & \hat{\Gamma}_{k-1} a \\ a^T \hat{\Gamma}_{k-1} & a^T \hat{\Gamma}_{k-1} a + \hat{\sigma}^2 \end{pmatrix}$$

The algorithm for calculating the variances $\hat{V}_i$ for $i = 1, 2, \ldots$ is thus given by

1. Initialize with $\hat{\Gamma}_1 = \hat{V}_1 = \hat{\sigma}^2$.

2. For $k = 2, 3, \dots, K$, repeat the following

    a) Form the $(k-1) \times 1$ vector a whose i^{th} entry is $\hat{\phi}_{k-i}$ if $k - p \le i \le k-1$ and 0 otherwise.

    b) Calculate $\hat{\Gamma}_k$ using $\hat{\Gamma}_{k-1}$ and a by the formula given above.

3. The variances $\hat{V}_i, i = 1, 2, \dots, K$ are given by the diagonal entries of the matrix $\hat{\Gamma}_K$.

The square roots of $\hat{V}_i, i = 1, 2, \dots$ provide uncertainty quantification for the point predictions $\hat{Y}_{n+i}, i = 1, 2, \dots$.

6 Recommended Reading for Today

1. For more on difference equations, see Section 3.2 of the book by Shumway and Stoffer titled *Time Series Analysis and its applications* (Fourth Edition).

2. For more on fitting $AR(p)$ models to data, see Section 3.4 of the Shumway-Stoffer book.