6.3   Forecasting ARMA processes

The purpose of forecasting is to predict future values of a TS based on the data collected to the present. In this section we will discuss a linear function of $\boldsymbol{X} = (X_n, X_{n-1}, \ldots, X_1)^{\mathrm{T}}$ predicting a future value of X_{n+m} for $m = 1, 2, \ldots$.

We call a function

$$f_{(n)}(\boldsymbol{X}) = \beta_0 + \beta_1 X_n + \ldots + \beta_n X_1 = \beta_0 + \sum_{i=1}^{n} \beta_i X_{n+1-i}$$

the **best linear predictor (BLP)** of X_{n+m} if it minimizes the prediction error

$$S(\boldsymbol{\beta}) = \mathrm{E}[X_{n+m} - f_{(n)}(\boldsymbol{X})]^2,$$

where $\boldsymbol{\beta}$ is the vector of the coefficients β_i and $\boldsymbol{X}$ is the vector of variables X_{n+1-i}.

Since $S(\boldsymbol{\beta})$ is a quadratic function of $\boldsymbol{\beta}$ and is bounded below by zero there is at least one value of $\boldsymbol{\beta}$ that minimizes $S(\boldsymbol{\beta})$. It satisfies the equations

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_i} = 0, \quad i = 0, 1, \ldots, n.$$

Evaluation of the derivatives gives so called **prediction equations**

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_0} = \mathrm{E}[X_{n+m} - \beta_0 - \sum_{i=1}^{n} \beta_i X_{n+1-i}] = 0 \qquad (6.18)$$

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_j} = \mathrm{E}[(X_{n+m} - \beta_0 - \sum_{i=1}^{n} \beta_i X_{n+1-i}) X_{n+1-j}] = 0 \qquad (6.19)$$

Assuming that $\mathrm{E}(X_t) = \mu$ the first equation can be written as

$$\mu - \beta_0 - \sum_{i=1}^{n} \beta_i \mu = 0,$$

which gives

$$\beta_0 = \mu(1 - \sum_{i=1}^{n} \beta_i).$$

The set of equations (6.20) gives

$$
\begin{aligned}
0 &= \mathrm{E}(X_{n+m}X_{n+1-j}) - \beta_0\mu - \sum_{i=1}^{n}\beta_i\,\mathrm{E}(X_{n+1-i}X_{n+1-j}) \\
&= \mathrm{E}(X_{n+m}X_{n+1-j}) - \mu^2\Big(1 - \sum_{i=1}^{n}\beta_i\Big) - \sum_{i=1}^{n}\beta_i\,\mathrm{E}(X_{n+1-i}X_{n+1-j}) \\
&= \gamma(m - (1-j)) - \sum_{i=1}^{n}\beta_i\gamma(i-j)
\end{aligned}
$$

That is we obtain the following form of the prediction equations (6.20).

$$
\gamma(m-1+j) = \sum_{i=1}^{n}\beta_i\gamma(i-j), \quad j = 1, \ldots, n. \tag{6.20}
$$

We obtain the same set of equations when $\mathrm{E}(X_t) = 0$. Hence, we assume further that the TS is a zero-mean stationary process. Then $\beta_0 = 0$ too.

6.3.1 One-step-ahead Prediction

Given $\{X_1, \ldots, X_n\}$ we want to forecast the value of X_{n+1}. The BLP of X_{n+1} is

$$
f_{(n)} = \sum_{i=1}^{n}\beta_i X_{n+1-i}.
$$

The coefficients β_i satisfy (6.21), where $m = 1$, that is

$$
\sum_{i=1}^{n}\beta_i\gamma(i-j) = \gamma(j), \quad j = 1, 2, \ldots, n.
$$

A convenient way of writing these equations is using matrix notation. We have

$$
\mathbf{\Gamma}_n\boldsymbol{\beta}_n = \boldsymbol{\gamma}_n, \tag{6.21}
$$

where

$$
\begin{aligned}
\mathbf{\Gamma}_n &= \{\gamma(i-j)\}_{j,i=1,2,\ldots,n} \\
\boldsymbol{\beta}_n &= (\beta_1, \ldots, \beta_n)^{\mathrm{T}} \\
\boldsymbol{\gamma}_n &= (\gamma(1), \ldots, \gamma(n))^{\mathrm{T}}.
\end{aligned}
$$

If $\mathbf{\Gamma}_n$ is nonsingular than the unique solution to (6.22) exists and is equal to

$$
\boldsymbol{\beta}_n = \mathbf{\Gamma}_n^{-1}\boldsymbol{\gamma}_n. \tag{6.22}
$$

Then the forecast of X_{n+1} based on $\boldsymbol{X} = (X_n, \ldots, X_1)^{\mathrm{T}}$ can be written as

$$X_{n+1}^{(n)} = (\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n)^{\mathrm{T}}\boldsymbol{X}. \tag{6.23}$$

The mean square one-step-ahead prediction error denoted by $P_{n+1}^{(n)}$ is

$$\begin{aligned}
P_{n+1}^{(n)} &= \mathrm{E}(X_{n+1} - X_{n+1}^{(n)})^2 \\
&= \mathrm{E}(X_{n+1} - \boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X})^2 \\
&= \mathrm{E}(X_{n+1}^2 - 2\boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X}X_{n+1} + \boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n) \\
&= \gamma(0) - 2\boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\Gamma}_n\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n \\
&= \gamma(0) - \boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n.
\end{aligned} \tag{6.24}$$

Example 6.6. **Prediction for an AR(2)**

Let

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

be a causal AR(2) process. Suppose we have one observation of X_1. Then one-step-ahead prediction function is

$$f_{(1)} = \beta_1 X_1,$$

where

$$\beta_1 = \Gamma_1^{-1}\gamma_1 = \frac{\gamma(1)}{\gamma(0)} = \rho(1) = \phi_{11}$$

and we obtain

$$X_2^{(1)} = \rho(1)X_1 = \phi_{11}X_1.$$

To predict X_3 based on X_2 and X_1 we need to calculate β_1 and β_2 in the prediction function

$$f_{(2)} = \beta_1 X_2 + \beta_2 X_1.$$

These can be obtained from (6.23) as

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(1) \\ \gamma(2) \end{pmatrix}$$

$$= \frac{1}{\gamma^2(0) - \gamma^2(1)} \begin{pmatrix} \gamma(0) & -\gamma(1) \\ -\gamma(1) & \gamma(0) \end{pmatrix} \begin{pmatrix} \gamma(1) \\ \gamma(2) \end{pmatrix}$$

$$= \frac{1}{\gamma^2(0) - \gamma^2(1)} \begin{pmatrix} \gamma(0)\gamma(1) - \gamma(1)\gamma(2) \\ -\gamma^2(1) + \gamma(0)\gamma(2) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\gamma(1)(\gamma(0)-\gamma(2))}{\gamma^2(0)-\gamma^2(1)} \\ \frac{\gamma(0)\gamma(2)-\gamma^2(1)}{\gamma^2(0)-\gamma^2(1)} \end{pmatrix} = \begin{pmatrix} \frac{\rho(1)(1-\rho(2))}{1-\rho^2(1)} \\ \frac{\rho(2)-\rho^2(1)}{1-\rho^2(1)} \end{pmatrix}.$$

From the difference equations (6.17) calculated in Example 6.4 we know that

$$\rho(1) = \frac{\phi_1}{1 - \phi_2}$$
$$\rho(2) - \phi_1\rho(1) - \phi_2\rho(0) = 0$$

That is

$$\rho(2) = \phi_1\rho(1) + \phi_2.$$

It finally gives

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}.$$

In fact, we can obtain this result directly from the model taking

$$X_3^{(2)} = \phi_1 X_2 + \phi_2 X_1$$

which satisfies the prediction equations, namely

$$\mathrm{E}[(X_3 - \phi_1 X_2 - \phi_2 X_1)X_1] = \mathrm{E}[Z_3 X_1] = 0$$
$$\mathrm{E}[(X_3 - \phi_1 X_2 - \phi_2 X_1)X_2] = \mathrm{E}[Z_3 X_2] = 0.$$

In general, for $n \geq 2$, we have

$$X_{n+1}^{(n)} = \phi_1 X_n + \phi_2 X_{n-1}, \tag{6.25}$$

i.e., $\beta_j = 0$ for $j = 3, \ldots, n$.

Similarly, it can be shown that a one-step-ahead prediction for AR(p) is

$$X_{n+1}^{(n)} = \phi_1 X_n + \phi_2 X_{n-1} + \ldots + \phi_p X_{n-p+1}, \quad \text{for } n \geq p. \qquad (6.26)$$

Remark 6.8. An interesting connection between the PACF and vector $\boldsymbol{\beta}_n$ is that in fact $\phi_{nn} = \beta_n$, the last element of the vector. For this reason, the vector $\boldsymbol{\beta}_n$ is usually denoted by $\boldsymbol{\phi}_n$ in the following way

$$\boldsymbol{\beta}_n = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \boldsymbol{\phi}_n.$$

The prediction equation (6.22) for a general ARMA(p,q) model is more difficult to calculate, particularly for large values of n when we would have to calculate an inverse of matrix $\boldsymbol{\Gamma}_n$ of large dimension. Hence some recursive solutions to calculate the predictor (6.24) and the mean square error (6.25) were proposed, one of them by Levinson in 1947 and by Durbin in 1960.

The method is known as the **Durbin-Levinson Algorithm**. Its steps are following:

Step 1 Put $\phi_{00} = 0$, $P_1^{(0)} = \gamma(0)$.

Step 2 For $n \geq 1$ calculate

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(k)} \qquad (6.27)$$

where, for $n \geq 2$

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, \ldots, n-1.$$

Step 3 For $n \geq 1$ calculate

$$P_{n+1}^{(n)} = P_n^{(n-1)}(1 - \phi_{nn}^2). \qquad (6.28)$$

Remark 6.9. Note, that the Durbin-Levinson algorithm gives an iterative method to calculate the PACF of a stationary process.

Remark 6.10. When we predict a value of the TS based only on one preceding datum, that is $n = 1$, we obtain

$$\phi_{11} = \rho(1),$$

and hence the predictor $X_2^{(1)} = \rho(1)X_1$, or in general

$$X_{n+1}^{(1)} = \rho(1)X_n$$

and its mean square error

$$P_2^{(1)} = \gamma(0)(1 - \phi_{11}^2).$$

When we predict X_{n+1} based on two preceding values, that is $n = 2$, we obtain

$$\phi_{22} = \frac{\rho(2) - \phi_{11}\rho(1)}{1 - \phi_{11}\rho(1)} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}$$

which we have also obtained solving the matrix equation (6.22) for β_2,

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \rho(1)(1 - \phi_{22}).$$

Then the predictor is

$$X_{n+1}^{(2)} = \phi_{21}X_n + \phi_{22}X_{n-1}$$

and its mean square error

$$P_3^{(2)} = \gamma(0)(1 - \phi_{11}^2)(1 - \phi_{22}^2).$$

We could continue these steps for $n = 3, 4, \ldots$.

Example 6.7. **Prediction for an AR(2), continued**

Using the Durbin-Levinson algorithm for AR(2) we obtain

$$\phi_{11} = \rho(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\phi_{22} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)} = \phi_2$$

$$\phi_{21} = \rho(1)(1 - \phi_{22}) = \phi_1$$

$$\phi_{33} = \frac{\rho(3) - \phi_1\rho(2) - \phi_2\rho(1)}{1 - \phi_1\rho(1) - \phi_2\rho(2)} = 0$$

$$\phi_{31} = \phi_{21} - \phi_{33}\phi_{22} = \phi_1$$

$$\phi_{32} = \phi_{22} - \phi_{33}\phi_{21} = \phi_2$$

$$\phi_{44} = \frac{\rho(4) - \phi_1\rho(3) - \phi_2\rho(2)}{1 - \phi_1\rho(1) - \phi_2\rho(2)} = 0$$

The results for ϕ_{33} and ϕ_{44} come from the fact that in the numerator we have the difference which is zero (difference equation).

Hence, one-step-ahead predictor for AR(2) is based only on two preceding values, as there are only two nonzero coefficients in the prediction function. As before, we obtain the result

$$X_{n+1}^{(2)} = \phi_1 X_n + \phi_2 X_{n-1}.$$

Remark 6.11. The PACF for AR(2) is

$$\begin{aligned}
\phi_{11} &= \frac{\phi_1}{1 - \phi_2} \\
\phi_{22} &= \phi_2 \\
\phi_{\tau\tau} &= 0 \quad \text{for } \tau \geq 3.
\end{aligned} \tag{6.29}$$

6.3.2   m-step-ahead Prediction

Given values of variables $\{X_1, \ldots, X_n\}$ the **m-steps-ahead predictor** is

$$X_{n+m}^{(n)} = \phi_{n1}^{(m)} X_n + \phi_{n2}^{(m)} X_{n-1} + \ldots + \phi_{nn}^{(m)} X_1, \tag{6.30}$$

where $\phi_{nj}^{(m)} = \beta_j$ satisfy the prediction equations (6.21). In matrix notation the prediction equations are

$$\mathbf{\Gamma}_n \boldsymbol{\phi}_n^{(m)} = \boldsymbol{\gamma}_n^{(m)}, \tag{6.31}$$

where

$$\boldsymbol{\gamma}_n^{(m)} = (\gamma(m), \gamma(m+1), \dots, \gamma(m+n-1))^{\mathrm{T}}$$

and

$$\boldsymbol{\phi}_n^{(m)} = (\phi_{n1}^{(m)}, \phi_{n2}^{(m)}, \dots, \phi_{nn}^{(m)})^{\mathrm{T}}.$$

The mean square m-step-ahead prediction error is

$$P_{n+m}^{(n)} = \mathrm{E}[X_{n+m} - X_{n+m}^{(n)}]^2 = \gamma(0) - (\boldsymbol{\gamma}_n^{(m)})^{\mathrm{T}} \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_n^{(m)}. \tag{6.32}$$

The mean square prediction error assesses the precision of the forecast and it is used to calculate so called **prediction interval (PI)**. When the process is Gaussian the the PI is

$$\widehat{X}_{n+m}^{(n)} \pm u_\alpha \sqrt{\widehat{P}_{n+m}^{(n)}}, \tag{6.33}$$

where u_α is such that $P(|U| < u_\alpha) = 1 - \alpha$, where U is a standard normal r.v. For $\alpha = 0.5$ we have $u_\alpha \approx 1.96$ and the 95% prediction interval boundaries are

$$\left(\widehat{X}_{n+m}^{(n)} - 1.96\sqrt{\widehat{P}_{n+m}^{(n)}}, \ \widehat{X}_{n+m}^{(n)} + 1.96\sqrt{\widehat{P}_{n+m}^{(n)}} \right).$$

Here we have used the "hat" notation as usually we do not know the values of the model parameters and we have to use their estimators. We will discuss the model parameter estimation in the next section.