

3.4: Forecasting

Suppose that the variables X_1, \dots, X_n of a weakly stationary time series $(X_t: t \in \mathbb{Z})$ have been observed with the goal to predict or forecast the future values of X_{n+1}, X_{n+2}, \dots . The focus is here on so-called *one-step best linear predictors (BLP)*. These are, by definition, linear combinations

$$\hat{X}_{n+1} = \phi_{n0} + \phi_{n1}X_n + \dots + \phi_{nn}X_1 \quad (3.4.1)$$

of the observed variables X_1, \dots, X_n that minimize the mean-squared error

$$E[\{X_{n+1} - g(X_1, \dots, X_n)\}^2]$$

for functions g of X_1, \dots, X_n . Straightforward generalizations yield definitions for the m -step best linear predictors \hat{X}_{n+m} of X_{n+m} for arbitrary $m \in \mathbb{N}$ in the same fashion. Using Hilbert space theory, one can prove the following theorem which will be the starting point for our considerations.

Theorem 3.4.1: Best linear prediction (BLP)

Let $(X_t: t \in \mathbb{Z})$ be a weakly stationary stochastic process of which X_1, \dots, X_n are observed. Then, the one-step BLP \hat{X}_{n+1} of X_{n+1} is determined by the equations

$$E[(X_{n+1} - \hat{X}_{n+1})X_{n+1-j}] = 0$$

for all $j = 1, \dots, n+1$, where $X_0 = 1$.

The equations specified in Theorem 3.4.1 can be used to calculate the coefficients $\phi_{n0}, \dots, \phi_{nn}$ in Equation 3.4.1. It suffices to focus on mean zero processes $(X_t: t \in \mathbb{Z})$ and thus to set $\phi_{n0} = 0$ as the following calculations show. Assume that $E[X_t] = \mu$ for all $t \in \mathbb{Z}$. Then, Theorem 3.4.1 gives that $E[\hat{X}_{n+1}] = E[X_{n+1}] = \mu$ (using the equation with $j = n+1$). Consequently, it holds that

$$\mu = E[\hat{X}_{n+1}] = E\left[\phi_{n0} + \sum_{\ell=1}^n \phi_{n\ell}X_{n+1-\ell}\right] = \phi_{n0} + \sum_{\ell=1}^n \phi_{n\ell}\mu.$$

Using now that $\phi_{n0} = \mu(1 - \phi_{n1} - \dots - \phi_{nn})$, Equation 3.4.1 can be rewritten as

$$\hat{Y}_{n+1} = \phi_{n1}Y_n + \dots + \phi_{nn}Y_1,$$

where $\hat{Y}_{n+1} = \hat{X}_{n+1} - \mu$ has mean zero.

With the ACVF γ of $(X_t: t \in \mathbb{Z})$, the equations in Theorem 3.4.1 can be expressed as

$$\sum_{\ell=1}^n \phi_{n\ell}\gamma(j-\ell) = \gamma(j), \quad j = 1, \dots, n. \quad (3.4.2)$$

Note that due to the convention $\phi_{n0} = 0$, the last equation in Theorem 3.4.1 (for which $j = n+1$) is omitted. More conveniently, this is restated in matrix notation. To this end, let $\Gamma_n = (\gamma(j-\ell))_{j,\ell=1,\dots,n}$, $\phi_n = (\phi_{n1}, \dots, \phi_{nn})^T$ and $\gamma_n = (\gamma(1), \dots, \gamma(n))^T$, where T denotes the transpose. With these notations, (3.4.2) becomes

$$\Gamma_n \phi_n = \gamma_n \iff \phi_n = \Gamma_n^{-1} \gamma_n, \quad (3.4.3)$$

provided that Γ_n is nonsingular.

The determination of the coefficients $\phi_{n\ell}$ has thus been reduced to solving a linear equation system and depends only on second-order properties of $(X_t: t \in \mathbb{Z})$ which are given by the ACVF γ .

Let $X_n = (X_n, X_{n-1}, \dots, X_1)^T$. Then, $\hat{X}_{n+1} = \phi_n^T X_n$. To assess the quality of the prediction, one computes the mean-squared error with the help of Equation 3.4.3 as follows:

$$\begin{aligned}
 P_{n+1} &= E \left[(X_{n+1} - \hat{X}_{n+1})^2 \right] \\
 &= E \left[(X_{n+1} - \phi_n^T X_n)^2 \right] \\
 &= E \left[(X_{n+1} - \gamma_n^T \Gamma_n^{-1} X_n)^2 \right] \\
 &= E \left[X_{n+1}^2 - 2\gamma_n^T \Gamma_n^{-1} X_n X_{n+1} + \gamma_n^T \Gamma_n^{-1} X_n X_n^T \Gamma_n^{-1} \gamma_n \right] \\
 &= \gamma(0) - 2\gamma_n^T \Gamma_n^{-1} \gamma_n + \gamma_n^T \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \gamma_n \\
 &= \gamma(0) - \gamma_n^T \Gamma_n^{-1} \gamma_n.
 \end{aligned} \tag{3.4.4}$$

As an initial example, we explain the prediction procedure for an autoregressive process of order 2.

Example 3.4.1: Prediction of an AR(2) Process

Let $(X_t: t \in \mathbb{Z})$ be the causal AR(2) process $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$. Suppose that only an observation of X_1 is available to forecast the value of X_2 . In this simplified case, the single prediction Equation 3.4.2 is

$$\phi_{11} \gamma(0) = \gamma(1),$$

so that $\phi_{11} = \rho(1)$ and $\hat{X}_{1+1} = \rho(1)X_1$.

In the next step, assume that observed values of X_1 and X_2 are at hand to forecast the value of X_3 . Then, one similarly obtains from (3.4.2.) that the predictor can be computed from

$$\begin{aligned}
 \hat{X}_{2+1} &= \phi_{21}X_2 + \phi_{22}X_1 = \phi_2^T X_2 = (\Gamma_2^{-1} \gamma_2)^T X_2 \\
 &= (\gamma(1), \gamma(2)) \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} X_2 \\ X_1 \end{pmatrix}.
 \end{aligned}$$

However, applying the arguments leading to the definition of the PAC in Section 3.3.3., one finds that

$$\begin{aligned}
 E[\{X_3 - (\phi_1 X_2 + \phi_2 X_1)\}X_1] &= E[Z_3 X_1] = 0, \\
 E[\{X_3 - (\phi_1 X_2 + \phi_2 X_1)\}X_2] &= E[Z_3 X_2] = 0.
 \end{aligned}$$

Hence, $\hat{X}_{2+1} = \phi_1 X_2 + \phi_2 X_1$ and even $\hat{X}_{n+1} = \phi_1 X_n + \phi_2 X_{n-1}$ for all $n \geq 2$, exploiting the particular autoregressive structure.

Since similar results can be proved for general causal AR(p) processes, the one-step predictors have the form

$$\hat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n-p+1}$$

whenever the number of observed variables n is at least p .

The major drawback of this approach is immediately apparent from the previous example: For larger sample sizes n , the prediction procedure requires the calculation of the inverse matrix Γ_n^{-1} which is computationally expensive. In the remainder of this section, two recursive prediction methods are introduced that bypass the inversion altogether. They are known as *Durbin-Levinson algorithm* and *innovations algorithm*. Finally, predictors based on the *infinite past* are introduced which are often easily applicable for the class of causal and invertible ARMA processes.

Method 1: The Durbin-Levinson algorithm

If $(X_t: t \in \mathbb{Z})$ is a zero mean weakly stationary process with ACVF γ such that $\gamma(0) > 0$ and $\gamma(h) \rightarrow 0$ as $h \rightarrow \infty$, then the coefficients $\phi_{n\ell}$ in (3.4.2.) and the mean squared errors P_n in (3.4.4.) satisfy the recursions

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)}, \quad P_0 = \gamma(0),$$

and, for $n \geq 1$,

$$\phi_{nn} = \frac{1}{P_{n-1}} \left(\gamma(n) - \sum_{\ell=1}^{n-1} \phi_{n-1,\ell} \gamma(n-\ell) \right),$$

$$\begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{nn} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$

and

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

It can be shown that under the assumptions made on the process $(X_t: t \in \mathbb{Z})$, it holds indeed that ϕ_{nn} is equal to the value of the PACF of $(X_t: t \in \mathbb{Z})$ at lag n . The result is formulated as Corollary 5.2.1 in Brockwell and Davis (1991). This fact is highlighted in an example.

The PACF of an AR(2) process

Let $(X_t: t \in \mathbb{Z})$ be a causal AR(2) process. Then, $\rho(1) = \phi_1/(1 - \phi_2)$ and all other values can be computed recursively from

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0, \quad h \geq 2.$$

Note that the ACVF γ satisfies a difference equation with the same coefficients, which is seen by multiplying the latter equation with $\gamma(0)$. Applying the Durbin-Levinson algorithm gives first that

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)} = \rho(1) \quad \text{and} \quad P_1 = P_0(1 - \phi_{11}^2) = \gamma(0)(1 - \rho(1)^2).$$

Ignoring the recursion for the error terms P_n in the following, the next $\phi_{n\ell}$ values are obtained a

$$\begin{aligned} \phi_{22} &= \frac{1}{P_1} [\gamma(2) - \phi_{11}\gamma(1)] = \frac{1}{1 - \rho(1)^2} [\rho(2) - \rho(1)^2] \\ &= \frac{\phi_1^2(1 - \phi_2)^{-1} + \phi_2 - [\phi_1(1 - \phi_2)^{-1}]^2}{1 - [\phi_1(1 - \phi_2)^{-1}]^2} = \phi_2, \\ \phi_{21} &= \phi_{11} - \phi_{22}\phi_{11} = \rho(1)(1 - \phi_2) = \phi_1, \\ \phi_{33} &= \frac{1}{P_2} [\gamma(3) - \phi_{21}\gamma(2) - \phi_{22}\gamma(1)] = \frac{1}{P_2} [\gamma(3) - \phi_1\gamma(2) - \phi_2\gamma(1)] = 0. \end{aligned}$$

Now, referring to the remarks after Example 3.3.7., no further computations are necessary to determine the PACF because $\phi_{nn} = 0$ for all $n > p = 2$.

Method 2: The innovations algorithm

In contrast to the Durbin-Levinson algorithm, this method can also be applied to nonstationary processes. It should thus, in general, be preferred over Method 1. The innovations algorithm gets its name from the fact that one directly uses the form of the prediction equations in Theorem 3.4.1. which are stated in terms of the *innovations* $(X_{t+1} - \hat{X}_{t+1})_{t \in \mathbb{Z}}$. Observe that the sequence consists of uncorrelated random variables.

The one-step predictors \hat{X}_{n+1} can be calculated from the recursions

$$\hat{X}_{0+1} = 0, \quad P_1 = \gamma(0)$$

and, for $n \geq 1$,

$$\hat{X}_{n+1} = \sum_{\ell=1}^n \theta_{n\ell} (X_{n+1-\ell} - \hat{X}_{n+1-\ell})$$

$$P_{n+1} = \gamma(0) - \sum_{\ell=0}^{n-1} \theta_{n,n-\ell}^2 P_{\ell+1},$$

where the coefficients are obtained from the equations

$$\theta_{n,n-\ell} = \frac{1}{P_{\ell+1}} \left[\gamma(n-\ell) - \sum_{i=0}^{\ell-1} \theta_{\ell,\ell-i} \theta_{n,n-i} P_{i+1} \right], \quad \ell = 0, 1, \dots, n-1.$$

As example we show how the innovations algorithm is applied to a moving average time series of order 1.

Example 3.4.3: Prediction of an MA(1) Process

Let $(X_t: t \in \mathbb{Z})$ be the MA(1) process $X_t = Z_t + \theta Z_{t-1}$. Note that

$$\gamma(0) = (1 + \theta^2)\sigma^2, \quad \gamma(1) = \theta\sigma^2 \quad \text{and} \quad \gamma(h) = 0 \quad (h \geq 2).$$

Using the innovations algorithm, one can compute the one-step predictor from the values

$$\theta_{n1} = \frac{\theta\sigma^2}{P_n}, \quad \theta_{n\ell} = 0 \quad (\ell = 2, \dots, n-1),$$

and

$$P_1 = (1 + \theta^2)\sigma^2, \\ P_{n+1} = (1 + \theta^2 - \theta\theta_{n1})\sigma^2$$

as

$$\hat{X}_{n+1} = \frac{\theta\sigma^2}{P_n} (X_n - \hat{X}_n).$$

Method 3: Prediction based on the infinite past

Suppose that a causal and invertible ARMA(p, q) process is analyzed. Assume further that (unrealistically) the complete history of the process can be stored and that thus all past variables $(X_t: t \leq n)$ can be accessed. Define then

$$\tilde{X}_{n+m} = E[X_{n+m} | X_n, X_{n-1}, \dots],$$

as the m -step ahead predictor based on the infinite past. It can be shown that, for large sample sizes n , the difference between the values of \hat{X}_{n+m} and \tilde{X}_{n+m} vanishes at an exponential rate. Exploiting causality and invertibility of the ARMA process, one can transform the predictor \tilde{X}_{n+m} so that it is in a computationally more feasible form. To do so, note that by causality

$$\begin{aligned} \tilde{X}_{n+m} &= E[X_{n+m} | X_n, X_{n-1}, \dots] \\ &= E \left[\sum_{j=0}^{\infty} \psi_j Z_{n+m-j} \middle| X_n, X_{n-1}, \dots \right] \\ &= \sum_{j=m}^{\infty} \psi_j Z_{n+m-j} \end{aligned} \tag{3.4.5}$$

because $E[Z_t | X_n, X_{n-1}, \dots]$ equals zero if $t > n$ and equals Z_t if $t \leq n$ (due to invertibility!). The representation in (3.4.5.) can be used to compute the mean squared prediction error \tilde{P}_{n+m} . It follows from causality that

$$\tilde{P}_{n+m} = E[(X_{n+m} - \tilde{X}_{n+m})^2] = E \left[\left(\sum_{j=0}^{m-1} \psi_j Z_{n+m-j} \right)^2 \right] = \sigma^2 \sum_{j=0}^{m-1} \psi_j^2. \tag{3.4.6}$$

On the other hand, Equation 3.4.5 does not allow to directly calculate the forecasts because \tilde{X}_{n+m} is given in terms of the noise variables Z_{n+m-j} . Instead invertibility will be utilized. Observe first that

$$E[X_{n+m-j}|X_n, X_{n-1}, \dots] = \begin{cases} \tilde{X}_{n+m-j}, & j < m. \\ X_{n+m-j}, & j \geq m. \end{cases}$$

By invertibility (the “0=” part follows again from causality),

$$0 = E[Z_{n+m}|X_n, X_{n-1}, \dots] \quad (3.4.7)$$

$$= E\left[\sum_{j=0}^{\infty} \pi_j X_{n+m-j} | X_n, X_{n-1}, \dots\right] \quad (3.4.8)$$

$$= \sum_{j=0}^{\infty} \pi_j E[X_{n+m-j}|X_n, X_{n-1}, \dots]. \quad (3.4.9)$$

Combining the previous two statements, yields

$$\tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}. \quad (3.4.10)$$

The equations can now be solved recursively for $m = 1, 2, \dots$. Note, however, that for any $m \geq 1$ the sequence $(X_{n+m+t} - \tilde{X}_{n+m+t} : t \in \mathbb{Z})$ does not consist of uncorrelated random variables. In fact, if $h \in \mathbb{N}_0$, it holds that

$$E[(X_{n+m} - \tilde{X}_{n+m})(X_{n+m+h} - \tilde{X}_{n+m+h})] \quad (3.4.11)$$

$$= E\left[\sum_{j=0}^{m-1} \psi_j Z_{n+m-j} \sum_{i=0}^{m+h-1} \psi_i Z_{n+m+h-i}\right] \quad (3.4.12)$$

$$= \sigma^2 \sum_{j=0}^{m-1} \psi_j \psi_{j+h}. \quad (3.4.13)$$

Finally, for practical purposes the given forecast needs to be truncated. This is accomplished by setting

$$\sum_{j=n+m}^{\infty} \pi_j X_{n+m-j} = 0.$$

The resulting equations (see Equation 3.4.10 for comparison) yield recursively the truncated m -step predictors X_{n+m}^* :

$$X_{n+m}^* = -\sum_{j=1}^{m-1} \pi_j X_{n+m-j}^* - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j}. \quad (3.4.14)$$

Contributors

- Alexander Aue (Department of Statistics, University of California, Davis)
- Integrated by Brett Nakano (statistics, UC Davis)

This page titled 3.4: Forecasting is shared under a not declared license and was authored, remixed, and/or curated by Alexander Aue.