Time Series Exercise Sheet 10

Sofia Olhede

May 1, 2025

Exercise 10.1

Let ε_t be a white noise process. Show that the best one-step-ahead predictors for the causal $AR(2)$ process

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

based on one observation X_1 and on two observations X_1, X_2 are

$$X_2^1 = \frac{\gamma_1}{\gamma_0} X_1 = \rho_1 X_1, \quad X_3^2 = \phi_1 X_2 + \phi_2 X_1.$$

Solution 10.1

For the one-step-ahead best linear predictor based on X_1, we use the theorem on prediction equations seen in the course, we get $\gamma_0 \beta_1 = \gamma_1$, which leads to

$$X_2^1 = \frac{\gamma_1}{\gamma_0} X_1 = \rho_1 X_1.$$

For the one-step-ahead best linear predictor based on X_1, X_2, we could try to reuse the theorem and build the matrix $\mathbf{\Gamma}_n$. Instead, we notice that $X_3 = \phi_1 X_2 + \phi_2 X_1 + \varepsilon_3$ implies that $X_3 - \phi_1 X_2 + \phi_2 X_1 = \varepsilon_3$. Because ε_t is a white noise process, we get

$$\mathbb{E}\left[\left(X_3 - \phi_1 X_2 + \phi_2 X_1\right) X_1\right] = \mathbb{E}\left[\varepsilon_3 X_1\right] = 0$$
$$\mathbb{E}\left[\left(X_3 - \phi_1 X_2 + \phi_2 X_1\right) X_2\right] = \mathbb{E}\left[\varepsilon_3 X_2\right] = 0,$$

which implies that $X_3^2 = \phi_1 X_2 + \phi_2 X_1$.

Exercise 10.2

Theorem. *For a stationary process $\{X_t\}$, X_{n+h}^n is found by solving for $\beta_0, \ldots, \beta_n$ the prediction equations*

$$\mathbb{E}\left[\left(X_{n+h} - X_{n+h}^n\right) X_k\right] = 0, \quad k = 0, \ldots, n,$$

where $X_0 = 1$, and $X_{n+h}^n = \beta_0 + \sum_{j=1}^n \beta_j X_j$. Let $\mu = \mathbb{E}\left[X_t\right]$.

The β_j are then given by the solution of the system of equations

$$\mathbf{\Gamma}_n \boldsymbol{\beta} = \boldsymbol{\gamma}_{[h]}, \quad \beta_0 = \mu\left(1 - \boldsymbol{\beta}^{\mathrm{T}} \mathbf{1}_n\right)$$

with

$$\mathbf{\Gamma}_n = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \text{ and } \boldsymbol{\gamma}_{[h]} = \begin{pmatrix} \gamma_{n+h-1} \\ \gamma_{n+h-2} \\ \vdots \\ \gamma_h \end{pmatrix}. \tag{1}$$

1. Prove the above theorem
   (a) Express the mean prediction error of X_{n+h}^n in matrix form in terms of $\mathrm{Var}\left(\boldsymbol{X}\right)$, $\boldsymbol{\beta}$ and β_0.

(b) Find the expression of β_0 by using the properties of the best linear predictor.

(c) Explain why there is no loss of generality in considering $\mu = 0$. What is then the value of β_0?

(d) Show that if $\mu = 0$, then the best linear predictor satisfies

$$\operatorname{Var}(\boldsymbol{X})\,\boldsymbol{\beta} = \operatorname{Cov}(X_{n+h}, \boldsymbol{X}). \tag{2}$$

(e) Show that the prediction equations for $k = 1, \ldots, n$ are equivalent to (2). Explain how you found the prediction equation for $k = 0$ in the previous steps.

2. Assume $\boldsymbol{\Gamma}_n$ is invertible.

(a) Express X_{n+h}^n in matrix form. How does it differ from the result seen in the lecture (e.g. $\mu = 0$)?

(b) What is the prediction mean squared error? How does it differ from the result seen in the lecture?

Solution 10.2

1. (a) The mean prediction error

$$\begin{aligned}
S(\beta_0, \ldots, \beta_n) &= \mathbb{E}\left[\left(X_{n+h} - \beta_0 - \sum_{j=1}^{n} \beta_j X_j\right)^2\right] \\
&= \mathbb{E}\left[\left(X_{n+h} - \beta_0 - \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}\right)^2\right] \\
&= \operatorname{Var}\left(X_{n+h} - \beta_0 - \beta^{\mathrm{T}} X\right) + \mathbb{E}\left[X_{n+h} - \beta_0 - \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}\right]^2 \\
&= \gamma_0 - 2\boldsymbol{\beta}^{\mathrm{T}} \operatorname{Cov}(X_{n+h}, \boldsymbol{X}) + \boldsymbol{\beta}^{\mathrm{T}} \operatorname{Var}(\boldsymbol{X})\,\boldsymbol{\beta} + \left(\mu - \beta_0 - \mu \boldsymbol{\beta}^{\mathrm{T}} \mathbf{1}_n\right)^2
\end{aligned}$$

is a quadratic function bounded below by zero, so it certainly has a minimum that satisfies the equations $\partial S / \partial \beta_r = 0$ for $r = 0, \ldots, n$.

(b) The first equation $\partial S / \partial \beta_0 = 0$ gives

$$\frac{\partial S}{\partial \beta_0} = 2\beta_0 - 2\left(\mu - \mu \boldsymbol{\beta}^{\mathrm{T}} \mathbf{1}_n\right) = 0 \quad \Longrightarrow \quad \beta_0 = \mu\left(1 - \boldsymbol{\beta}^{\mathrm{T}} \mathbf{1}_n\right).$$

(c) Hence the form of the best linear predictor is $X_{n+h}^n = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X} = \mu + \boldsymbol{\beta}^T (\boldsymbol{X} - \mu \mathbf{1}_n)$. Thus there is no loss of generality in considering $\mu = 0$ in which case $\beta_0 = 0$ and $X_{n+h}^n = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}$.

(d) When $\mu = 0$,

$$S(\boldsymbol{\beta}) = \gamma_0 - 2\boldsymbol{\beta}^{\mathrm{T}} \operatorname{cov}(X_{n+h}, \boldsymbol{X}) + \boldsymbol{\beta}^{\mathrm{T}} \operatorname{var}(\boldsymbol{X}) \boldsymbol{\beta}.$$

We differentiate $S(\boldsymbol{\beta})$ to find the matrix equation $\operatorname{var}(\boldsymbol{X})\boldsymbol{\beta} = \operatorname{cov}(X_{n+h}, \boldsymbol{X})$ as stated in the theorem.

(e) This matrix equation may be rewritten as the prediction equations, the first of which, for β_0, gives $\mu - \beta_0 - \sum_{j=1}^n \beta_j \mu = 0$, and hence the required form for β_0. The other equations for $k = 1, \ldots, n$ are obtained by noting that

$$\mathbb{E}\left[\left(X_{n+h} - \sum_{j=1}^{n} \beta_j X_j\right) X_k\right] = 0 \Longleftrightarrow \gamma_{n+h-k} = \sum_{j=1}^{n} \beta_j \gamma_{jk}, \quad k = 1, \ldots, n,$$

which is equivalent to the matrix equation (write $\operatorname{Var}(\boldsymbol{X})\,\boldsymbol{\beta} = \operatorname{Cov}(X_{n+h}, \boldsymbol{X})$ and identify each row equation).

2. (a) If $\boldsymbol{\Gamma}_n$ is invertible, then we have $\boldsymbol{\beta} = \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\gamma}_{[h]}$, which leads to

$$X_{n+h}^n = \mu + \boldsymbol{\gamma}_{[h]}^{\mathrm{T}} \boldsymbol{\Gamma}_n^{-1} (\boldsymbol{X} - \mu \mathbf{1}_n).$$

(b) Since

$$\mathbb{E}\left[X_{n+h} - X_{n+h}^n\right] = 0,$$

the prediction mean square error is

$$\operatorname{Var}\left(X_{n+h} - X_{n+h}^n\right) = \operatorname{Var}\left(X_{n+h} - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}\right) = \gamma_0 - 2\boldsymbol{\beta}^T\boldsymbol{\gamma}_{[h]} + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Gamma}_n\boldsymbol{\beta} = \gamma_0 - \boldsymbol{\gamma}^{h\ \mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_{[h]},$$

which is the same as in the zero-mean case.

Exercise 10.3

Prove the following theorem

Theorem. *The best linear predictor X_{n+h}^n for X_{n+h} in a causal ARMA process with general linear representation $\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ is*

$$X_{n+h}^n = \sum_{j=h}^{\infty}\psi_j\varepsilon_{n+h-j} = \psi_h\varepsilon_n + \psi_{h+1}\varepsilon_{n-1} + \cdots.$$

The corresponding prediction mean square error is $\sigma^2\sum_{j=0}^{h-1}\psi_j^2$.

Solution 10.3

We consider the linear predictor $X_{n+h}^n = \sum_{i=1}^{n}\beta_i X_i$, and on noting the linear representation $X_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, we see that we can write

$$X_{n+h}^n = \sum_{i=1}^{n}\beta_i\sum_{j=0}^{\infty}\psi_j\varepsilon_{i-j} = \sum_{j=0}^{\infty}c_j\varepsilon_{n-j},$$

say, i.e., X_{n+h}^n has a linear representation starting at ε_n. The prediction mean square error,

$$\mathbb{E}\left[\left(X_{n+h} - X_{n+h}^n\right)^2\right] = \mathbb{E}\left[\left(\sum_{j=0}^{\infty}\psi_j\varepsilon_{n+h-j} - \sum_{j=0}^{\infty}c_j\varepsilon_{n-j}\right)^2\right] = \sigma^2\left\{\sum_{j=0}^{h-1}\psi_j^2 + \sum_{j=h}^{\infty}\left(\psi_j - c_{j-h}\right)^2\right\}$$

is minimised by taking $c_j = \psi_{j+h}$, for $j = 0, 1, \ldots$, and the prediction is then

$$X_{n+h}^n = \sum_{j=0}^{\infty}c_j\varepsilon_{n-j} = \sum_{j=0}^{\infty}\psi_{j+h}\varepsilon_{n-j} = \sum_{j=h}^{\infty}\psi_j\varepsilon_{n+h-j}.$$

Since $X_{n+h} = \sum_{j=0}^{\infty}\psi_j\varepsilon_{n+h-j}$, we see that $X_{n+h} - X_{n+h}^n = \sum_{j=0}^{h-1}\psi_j\varepsilon_{n+h-j}$, and the mean square error is

$$P_{n+h}^n = \mathrm{E}\left\{\left(X_{n+h} - X_{n+h}^n\right)^2\right\} = \sigma^2\sum_{j=0}^{h-1}\psi_j^2.$$