# Lecture 5: Multivariate linear prediction and partial autocorrelation

Josep Vives

May 11, 2023

## 1 Multivariate linear prediction

Assume now we have $n$ predictors $X_1, \ldots, X_n$ to predict $Y$. In this case, we have to impose

$$Y - b_0 - b_1 X_1 - b_2 X_2 - \cdots - b_n X_n \perp 1, X_1, X_2, \ldots, X_n,$$

that is,

$$b_0 + \mathbb{E}(X_1) b_1 + \cdots + \mathbb{E}(X_n) b_n = \mathbb{E}(Y)$$

and

$$\begin{aligned} \mathbb{E}[X_1] b_0 + \mathbb{E}[X_1^2] b_1 + \cdots + \mathbb{E}[X_1 X_n] b_n &= \mathbb{E}[X_1 Y] \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathbb{E}[X_n] b_0 + \mathbb{E}[X_n X_1] b_1 + \cdots + \mathbb{E}(X_n^2) b_n &= \mathbb{E}[X_n Y] \end{aligned} \tag{1.1}$$

Then, if we denote $b^t := (b_1, \ldots, b_n)$, $\mu_X^t := (\mathbb{E}X_1, \ldots, \mathbb{E}X_n)$ and $\mu_Y = \mathbb{E}Y$ we have, on one hand,

$$b_0 = \mu_Y - b^t \mu_X.$$

Multiplying this equation by any $\mathbb{E}(X_j)$ we obtain the family of equations

$$\mathbb{E}[X_j] b_0 = \mathbb{E}(X_j) \mu_Y - \mathbb{E}(X_j) b^t \mu_X, \quad \forall j = 1 \ldots, n$$

and substituting every $\mathbb{E}[X_j] b_0$ in every equation of the system (1.1) and reordering terms we obtain the system

$$\mathbb{C}(X_j, X_1) b_1 + \cdots + \mathbb{C}(X_j, X_n) b_n = \mathbb{C}(X_j, Y), \; j = 1, \ldots, n.$$

That is,

$$\Sigma\, b = \mathbb{C}(Y, X) \tag{1.2}$$

where $\Sigma : \mathbb{E}[(X - \mu_X)(X - \mu_X)^t]$ is the covariance matrix of vector $X$.

Therefore, the general formula is given by

$$P(Y \,|\, X_1, \ldots, X_n) = (\mu_Y - \widehat{b}^{\,t} \mu_X) + \widehat{b}^{\,t} X = \mu_Y + \widehat{b}^{\,t} (X - \mu_X)$$

with

$$\widehat{b} = \Sigma^{-1}\mathbb{C}(Y, X)$$

Recall that given two vectors $v$ and $w$ we have $v^t w = w^t v$. Using this property and Equation (1.2) it is immediate to see that the prediction error is given by

$$\begin{aligned}
& \mathbb{E}[(Y - P(Y \mid X_1, \cdots, X_n))^2] \\
= \; & \mathbb{E}[(Y - \mu_Y - \widehat{b}^{\,t}(X - \mu_X))^2] \\
= \; & \mathbb{V}(Y) + \mathbb{E}[(\widehat{b}^{\,t}(X - \mu_X))^2] - 2\mathbb{E}[(Y - \mu_Y)\widehat{b}^{\,t}(X - \mu_X)] \\
= \; & \mathbb{V}(Y) + \widehat{b}^t \Sigma \widehat{b} - 2\mathbb{C}(Y, X)^t \widehat{b} \\
= \; & \sigma_Y^2 - \mathbb{C}(Y, X)^t \widehat{b}.
\end{aligned}$$

**Remark 1.1** *The optimal predictor $E(Y \mid X_1, \ldots, X_n)$ can be much more difficult to compute.*

Recalling that $P(Y|X_1, \ldots, X_n)$ is the orthogonal projection over the linear subspace generated by $X_1, \ldots X_n$, the following properties are immediate

1. $P(X_j \mid X_1, \ldots, X_n) = X_j, \;\; j = 1, \ldots, n.$
2. $P(\alpha + \beta Y_1 + \gamma Y_2 \mid X_1, \ldots, X_n) = \alpha + \beta P(Y_1 \mid X_1, \ldots, X_n) + \gamma P(Y_2 \mid X_1, \ldots, X_n).$
3. If $\mathbb{C}(Y, X_j) = 0, \;\; \forall j = 1, \ldots, n$, then $P(Y \mid X_1, \ldots, X_n) = E(Y)$.
4. $P(P(Y \mid X_1, X_2) \mid X_1) = P(Y \mid X_1).$

**Remark 1.2** *If $Z_i := X_i - \mu_{X_i}$, we have*

$$P(Y \mid X_1, \ldots, X_n) = P(Y \mid Z_1, \ldots, Z_n).$$

*Indeed we have*

$$P(Y \mid X_1, \ldots, X_n) = \mu_Y + \widehat{b}^{\,t} \cdot Z$$

*and*

$$P(Y \mid Z_1 \ldots Z_n) = \mu_Y + \widehat{b}_*^{\,t} \cdot Z$$

*with $\widehat{b} = \Sigma_X^{-1}\mathbb{C}(Y, X)$ and $\widehat{b}_* = \Sigma_Z^{-1}\mathbb{C}(Y, Z)$.*

*But, on one hand,*

$$\mathbb{C}(Y, X) = \mathbb{C}(Y, \mu_X + Z) = \mathbb{C}(Y, \mu_X) + \mathbb{C}(Y, Z) = \mathbb{C}(Y, Z)$$

*and on the other hand,*

$$\begin{aligned}
\mathbb{C}(X_i, X_j) &= \mathbb{C}(\mu_X + Z_i, \mu_X + Z_j) \\
&= \mathbb{C}(\mu_X, \mu_X) + \mathbb{C}(\mu_X, Z_j) + \mathbb{C}(Z_i, \mu_X) + \mathbb{C}(Z_i, Z_j) \\
&= \mathbb{C}(Z_i, Z_j).
\end{aligned}$$

## 2 Prediction in Time Series

Let $X = \{X_j,\ j \in \mathbb{Z}\}$ a second order stationary time series with mean $\mu$ and covariance function $\gamma$. Consider the variables $X_1, \ldots, X_n$ and denote $X := (X_1, \ldots, X_n)^t$. According to the formulas obtained in the previous section, we have

$$P(X_{n+1} \,|\, X_1, \ldots, X_n) = \mu + \widehat{b}^{\,t}(X - \mu)$$

with

$$\widehat{b} = \Sigma^{-1}\mathbb{C}(X_{n+1}, X). \tag{2.3}$$

Note that

$$\mathbb{C}(X_{n+1}, X)^t = (\gamma(n), \gamma(n-1), \cdots, \gamma(1))$$

and on other hand,

$$\Sigma = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \gamma(n-1) & \cdots & \cdots & \cdots & \gamma(0) \end{bmatrix}.$$

From Equation (2.3), dividing by $\gamma(0)$, we have

$$R\,\widehat{b} = \begin{bmatrix} \rho_n \\ \vdots \\ \rho_1 \end{bmatrix}$$

with

$$R = \begin{bmatrix} \rho_0 & \cdots & \cdots & \rho_{n-1} \\ \rho_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho_{n-1} & \cdots & \cdots & \rho_0 \end{bmatrix}.$$

Therefore,

$$\widehat{b} = R^{-1} \begin{bmatrix} \rho_n \\ \vdots \\ \rho_1 \end{bmatrix}$$

and then, we can write

$$P(X_{n+1} \,|\, X_1, \ldots, X_n) = \mu + (\rho_n, \ldots, \rho_1) R^{-1} \begin{bmatrix} X_1 - \mu \\ \vdots \\ X_n - \mu \end{bmatrix}.$$

# 3 The partial autocorrelation function

Let $X = \{X_j,\ j \in \mathbb{Z}\}$ a centered second order stationary time series. The partial auto-correlation function is defined as an application

$$\alpha\colon \mathbb{N} \to [-1,1]$$

such that $\alpha(0) = 1$, $\alpha(1) = \rho_1$ and for $l \geq 2$,

$$\alpha(l) = \rho(X_1 - P(X_1 \,|\, X_2, \dots, X_l), \quad X_{l+1} - P(X_{l+1} \,|\, X_2, \dots, X_l)),$$

that is, $\alpha(l)$ is the correlation between $X_1$ and $X_{l+1}$ after erasing the influence of variables $X_2, \dots, X_l$.

To compute $\alpha(2)$ we have to compute $P(X_1 \,|\, X_2)$ and $P(X_3 \,|\, X_2)$.

We have

$$P(X_1 \,|\, X_2) = bX_2$$

such that

$$\mathbb{E}[(X_1 - bX_2)^2]$$

is minimal.

We know

$$\widehat{b} = \frac{\mathbb{C}(X_1, X_2)}{\mathbb{V}(X_2)} = \frac{\gamma(1)}{\gamma(0)} = \rho(1)$$

and so

$$P(X_1 \,|\, X_2) = \rho_1 \, X_2.$$

Obviously, the same happens with $P(X_3 \,|\, X_2)$. Then,

$$\begin{aligned}
\alpha(2) &= \rho(X_1 - \rho_1 X_2,\ X_3 - \rho_1 X_2) \\
&= \frac{\mathbb{C}(X_1 - \rho_1 X_2,\ X_3 - \rho_1 X_2)}{\{\mathbb{V}(X_1 - \rho_1 X_2) \cdot \mathbb{V}(X_3 - \rho_1 X_2)\}^{1/2}} \\
&= \frac{\gamma(2) - \rho_1 \gamma(1) \cdot 2 + \rho_1^2 \gamma(0)}{[\gamma(0) + \rho_1^2 \gamma(0) - 2\rho_1 \gamma(1)]} \\
&= \frac{\rho(2) - \rho(1)^2 \cdot 2 + \rho(1)^2}{1 + \rho(1)^2 - 2\rho(1)^2} \\
&= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.
\end{aligned}$$

Bigger partial auto-correlations can be computed similarly.

An alternative method to compute the partial auto-correlation function is the following. Let $(x_1, \dots, x_n)^t$ be the solution of the system

$$R_n \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_n \end{bmatrix}.$$

Then,

$$\alpha(n) = x_n.$$

We check this in cases $n = 1$ and $n = 2$. If $n = 1$, we have to see

$$R_1 \cdot \alpha(1) = \rho(1)$$

and this is obvious because $R_1 = (\rho_0) = (1)$.

If $n = 2$ we have

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

and so

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \frac{1}{1-\rho_1^2} \begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \\ &= \begin{bmatrix} \rho_1 - \rho_1\rho_2 \\ \rho_2 - \rho_1^2 \end{bmatrix} \frac{1}{1-\rho_1^2} \\ &= \begin{bmatrix} \rho_1(1-\rho_2)/(1-\rho_1)^2 \\ \rho_2 - \rho_1^2/1-\rho_1^2 \end{bmatrix}. \end{aligned}$$

Therefore, as expected,

$$\alpha(2) = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}.$$

To check the general case $\alpha(k)$ is analogous.

**Remark 3.1** *Observe that on one hand*

$$R_n \,\widehat{b} = \begin{bmatrix} \rho_n \\ \vdots \\ \rho_1 \end{bmatrix},$$

*or equivalently,*

$$R_n \begin{bmatrix} \widehat{b}_n \\ \vdots \\ \widehat{b}_1 \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_n \end{bmatrix}.$$

*On other hand*

$$R_n \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_n \end{bmatrix}.$$

*Therefore,*

$$\widehat{b}_1 = \alpha(n),$$

*that is, $\alpha(n)$ is the coefficient of $X_1$ of the optimal linear predictor of $X_{n+1}$ over the linear subspace generated by random variables $X_1, \ldots, X_n$. This gives an interesting interpretation of $\alpha(n)$ for a centered second order stationary time series model.*