STAT 153 - Introduction to Time Series Lecture Twenty

Fall 2022, UC Berkeley

Aditya Guntuboyina

November 4, 2022

An important diagnostic tool in time series analysis is the PACF (Partial Autocorrelation Function). In order to understand the PACF, we need to learn the notion of Partial Correlation. This is the topic for today's class. The Partial Correlation is defined as the correlation between Residuals, where Residuals are defined in terms of the Best Linear Predictors. We shall therefore first study the notion of Best Linear Prediictor, then Residuals, and then we define the Partial Correlation.

1 Best Linear Predictor

Consider random variables $Y, X_1, \dots, X_p$ that have finite variance. Suppose we want to predict Y based only on **linear** functions of $X_1, \dots, X_p$. Specifically, we consider predictions of the form $\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p = \beta_0 + \beta^T X$ (where $\beta := (\beta_1, \dots, \beta_p)^T$ and $X = (X_1, \dots, X_p)^T$). The Best Linear Predictor (BLP) of Y in terms of $X_1, \dots, X_p$ is the linear function

$$\beta_0^* + \beta_1^* X_1 + \dots + \beta_p^* X_p = \beta_0^* + (\beta^*)^T X \qquad \text{with } \beta^* := (\beta_1^*, \dots, \beta_p^*)^T$$

where $\beta_0^*, \dots, \beta_p^*$ minimize

$$L(\beta_0, \dots, \beta_p) = \mathbb{E}\left(Y - \beta_0 - \beta_1 X_1 - \dots - \beta_p X_p\right)^2$$

over $\beta_0, \beta_1, \dots, \beta_p$.

One can get an explicit formula for β_0^* and β^* by minimizing L directly via calculus. Taking partial derivatives with respect to $\beta_0, \beta_1, \dots, \beta_p$ and setting them equal to zero, we obtain the following equations:

$$\mathbb{E}(Y - \beta_0^* - \beta_1^* X_1 - \dots - \beta_p^* X_p) = 0 \tag{1}$$

and

$$\mathbb{E}(Y - \beta_0^* - \beta_1^* X_1 - \dots - \beta_p^* X_p) X_i = 0 \qquad \text{for } i = 1, \dots, p. \tag{2}$$

The first equation above implies that $Y - \beta_0^* - \beta_1^* X_1 - \dots - \beta_p^* X_p$ is a mean zero random variable. Using this, we can rewrite the second equation as

$$Cov(Y - \beta_0^* - \beta_1^* X_1 - \dots - \beta_p^* X_p, X_i) = 0 \qquad \text{for } i = 1, \dots, p$$

which is same as

$$Cov(Y - \beta_1^* X_1 - \cdots - \beta_p^* X_p, X_i) = 0 \qquad \text{for } i = 1, \ldots, p. \tag{3}$$

Rearranging the above, we obtain

$$\sum_{j=1}^{p} \beta_i^* Cov(X_i, X_j) = Cov(Y, X_i) \qquad \text{for } i = 1, \ldots, p.$$

In matrix notation, we can rewrite this as

$$Cov(X)\beta^* = Cov(X, Y) \qquad \text{with } \beta^* = (\beta_1^*, \ldots, \beta_p^*)^T.$$

Here $Cov(X, Y)$ is the $p \times 1$ vector with entries $Cov(X_1, Y), \ldots, Cov(X_p, Y)$. The above equation gives

$$\beta^* = (Cov(X))^{-1} Cov(X, Y)$$

assuming that $Cov(X)$ is invertible. This equation determines $\beta_1^*, \ldots, \beta_p^*$. We can then use (1) to write β_0^* as

$$\beta_0^* = \mathbb{E}(Y) - Cov(Y, X)(Cov(X))^{-1}\mathbb{E}(X).$$

Note that the term $Cov(Y, X)$ appearing above is the transpose of $Cov(X, Y)$. More generally, given two random vectors $W = (W_1, \ldots, W_p)$ and $Z = (Z_1, \ldots, Z_q)$, we define $Cov(W, Z)$ to be the $p \times q$ matrix whose (i, j)th entry is the covariance between W_i and Z_j.

The Best Linear Predictor (BLP) of Y in terms of $X_1, \ldots, X_p$ then equals

$$\begin{aligned}
\beta_0^* + \beta_1^* X_1 + \ldots \beta_p^* X_p &= \beta_0^* + (\beta^*)^T X \\
&= \mathbb{E}(Y) - Cov(Y, X)(Cov(X))^{-1}\mathbb{E}(X) + Cov(Y, X)(Cov(X))^{-1}X \\
&= \mathbb{E}(Y) + Cov(Y, X)(Cov(X))^{-1}(X - \mathbb{E}(X)). \qquad (4)
\end{aligned}$$

Here are some important properties of the BLP:

1. The BLP solves equations (1) and (3). These equations are called **normal equations**.
2. If $Cov(X)$ is invertible (equivalently, positive definite), then the BLP is uniquely given by (4).
3. $Y - BLP$ has mean zero (because of (1)) and $Y - BLP$ is uncorrelated with each $X_i, i = 1, \ldots, p$ (because of (3)). In fact, this property characterizes the BLP (see next).
4. If $Cov(X)$ is invertible, then it is clear from the form of the normal equations that the BLP is the unique linear combination of $X_1, \ldots, X_p$ such that $Y - BLP$ has mean zero and is uncorrelated with $X_1, \ldots, X_p$.

Example 1.1 (The case $p = 1$)**.** *When $p = 1$, the random vector X has only element X_1 so that $Cov(X)$ is just equal to the number $Var(X_1)$. In that case, the BLP of Y in terms of X_1 is given by*

$$BLP = \mathbb{E}(Y) + \frac{Cov(Y, X_1)}{Var(X_1)} (X_1 - \mathbb{E}(X_1)).$$

In other words, when $p = 1$,

$$\beta_1^* = \frac{Cov(Y, X_1)}{Var(X_1)} = Corr(Y, X_1)\sqrt{\frac{Var(Y)}{Var(X_1)}} = \rho_{Y, X_1}\sqrt{\frac{Var(Y)}{Var(X_1)}}.$$

In the further special case when $Var(Y) = Var(X_1)$ and $\mathbb{E}(Y) = \mathbb{E}(X_1) = 0$, we have

$$\beta_1^* = \rho_{Y,X_1}$$

so that the BLP is simply given by $\rho_{Y,X_1} X_1$.

Example 1.2. *Suppose $X_1, X_2, Z_3, \dots, Z_n, Z_{n+1}$ are uncorrelated random variables and mean zero random variables. Define random variables $X_3, \dots, X_{n+1}$ as*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t \qquad \text{for } t = 3, \dots, n+1.$$

What is the BLP of X_{n+1} in terms of $X_1, \dots, X_n$ for $n \geq 2$?

By definition,

$$X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + Z_{n+1}$$

which means that $X_{n+1} - \phi_1 X_n - \phi_2 X_{n-1} = Z_{n+1}$. It is now easy to see that each X_t depends only on $X_1, X_2, Z_3, \dots, Z_t$ for $t \geq 3$ which implies that Z_{n+1} is uncorrelated with all of $X_1, \dots, X_n$.

Therefore $\phi_1 X_n + \phi_2 X_{n-1}$ is a linear combination of $X_1, \dots, X_n$ such that $X_{n+1} - \phi_1 X_n - \phi_2 X_{n-1}$ is uncorrelated with each of $X_1, \dots, X_n$ (it also has mean zero). We deduce therefore that the BLP of X_{n+1} in terms of $X_1, \dots, X_n$ equals $\phi_1 X_n + \phi_2 X_{n-1}$.

2 Residual

The residual of a random variable Y in terms of $X_1, \dots, X_p$ will be denoted by $r_{Y|X_1,\dots,X_p}$ and defined as the difference between Y and the BLP of Y in terms of $X_1, \dots, X_p$. In other words,

$$r_{Y|X_1,\dots,X_p} = Y - BLP.$$

Using the formula for the BLP, we can write down the following formula for the residual:

$$r_{Y|X_1,\dots,X_p} = Y - \mathbb{E}(Y) - Cov(Y, X)(Cov X)^{-1}(X - \mathbb{E}(X)) \tag{5}$$

where X is the $p \times 1$ random vector with components $X_1, \dots, X_p$.

The residual has mean zero and is uncorrelated with each of $X_1, \dots, X_p$. This can be proved either directly from the formula (5) or from the properties of the BLP.

The variance of the residual can be calculated from the formula (5) as follows:

$$\begin{aligned}
Var(r_{Y|X_1,\dots,X_p}) &= Var\left(Y - \mathbb{E}(Y) - Cov(Y, X)(Cov X)^{-1}(X - \mathbb{E}(X))\right) \\
&= Var(Y) - 2Cov(Y, Cov(Y, X)(Cov X)^{-1}X) + Var(Cov(Y, X)(Cov X)^{-1}(X - \mathbb{E}(X))) \\
&= Var(Y) - 2Cov(Y, X)(Cov X)^{-1}Cov(X, Y) + Cov(Y, X)(Cov X)^{-1}Cov(X, Y) \\
&= Var(Y) - Cov(Y, X)(Cov X)^{-1}Cov(X, Y).
\end{aligned}$$

In other words, $Var(r_{Y|X_1,\dots,X_p})$ equals the **Schur complement** (recalled next) of $Var(Y)$ in the covariance matrix:

$$\begin{pmatrix} Cov(X) & Cov(X, Y) \\ Cov(Y, X) & Var(Y) \end{pmatrix}$$

of the $(n+1) \times 1$ random vector $(X_1, \dots, X_p, Y)^T$.

3 Detour: Schur Complements

Consider an $n \times n$ matrix A that is partitioned into four blocks as

$$A = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

where E is $p \times p$, F is $p \times q$, G is $q \times p$ and H is $q \times q$ (p and q are such that $p + q = n$).

We define

$$E^S := E - FH^{-1}G \quad \text{and} \quad H^S := H - GE^{-1}F$$

assuming that H^{-1} and E^{-1} exist. We shall refer to E^S and H^S as the *Schur complements* of E and H respectively (**Warning**: This is not standard terminology; it is more common to refer to E^S as the Schur complement of H and to H^S as the Schur complement of E. I find it more natural to think of E^S as the Schur complement of E and H^S as the Schur complement of H).

Note that both E and E^S are $p \times p$ while both H and H^S are $q \times q$.

Schur complements have many interesting properties such as:

1. $det(A) = det(E)det(H^S) = det(H)det(E^S)$.
2. If A is positive definite, then E, E^S, H, H^S are all positive definite.

and many others. Feel free to see the monograph titled *Schur Complements and Statistics* by Diane Ouellette for proofs and exposition of these facts (this is not really necessary for this course).

But one very important property of Schur Complements for our purpose is the fact that they arise naturally in inverses of partitioned matrices. A standard formula for the inverse of a partitioned matrix (see, for example, `https://en.wikipedia.org/wiki/Block_matrix#Block_matrix_inversion`) is

$$A^{-1} = \begin{pmatrix} (E^S)^{-1} & -E^{-1}F(H^S)^{-1} \\ -(H^S)^{-1}GE^{-1} & (H^S)^{-1} \end{pmatrix} \tag{6}$$

It must be noted from this formula that **the first (or $(1,1)^{th}$) block of A^{-1} equals the inverse of the Schur complement of the first block of A. Similarly, the last (or $(2,2)^{th}$) block of A^{-1} equals the inverse of the Schur complement of the last block of A.**

We shall use the expression (6) for the inverse of the partitioned matrix A but we will not see how to prove (6). You can find many proofs of this fact elsewhere (just google something like "inverse of partitioned matrices").

4 Partial Correlation

Given random variables Y_1, Y_2 and $X_1, \dots, X_p$, the partial correlation between Y_1 and Y_2 given $X_1, \dots, X_p$ is denoted by $\rho_{Y_1,Y_2|X_1,\dots,X_p}$ and defined as

$$\rho_{Y_1,Y_2|X_1,\dots,X_p} := Corr\left(r_{Y_1|X_1,\dots,X_p}, r_{Y_2|X_1,\dots,X_p}\right).$$

In other words, $\rho_{Y_1,Y_2|X_1,\dots,X_p}$ is defined as the correlation between the residual of Y_1 given $X_1, \dots, X_p$ and the residual of Y_2 given $X_1, \dots, X_p$.

$\rho_{Y_1,Y_2|X_1,\dots,X_p}$ is also termed the partial correlation of Y_1 and Y_2 after controlling for $X_1,\dots,X_p$. Since residuals are second order quantities, it follows that the partial correlation is a second order quantity as well. We shall now see how to explicitly write the partial correlation in terms of the covariances of Y_1, Y_2 and X.

As
$$r_{Y_1|X_1,\dots,X_p} = Y_1 - \mathbb{E}(Y_1) - Cov(Y_1, X)(Cov X)^{-1}(X - \mathbb{E}(X))$$
and
$$r_{Y_2|X_1,\dots,X_p} = Y_2 - \mathbb{E}(Y_2) - Cov(Y_2, X)(Cov X)^{-1}(X - \mathbb{E}(X)),$$
it can be checked (left as an exercise) that
$$Cov(r_{Y_1|X_1,\dots,X_p}, r_{Y_2|X_1,\dots,X_p}) = Cov(Y_1, Y_2) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_2).$$

This, along with the formula for the variance of the residuals from the previous subsections, gives the following formula for the partial correlation $\rho_{Y_1,Y_2|X_1,\dots,X_p}$:
$$\frac{Cov(Y_1, Y_2) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_2)}{\sqrt{Var(Y_1) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_1)}\sqrt{Var(Y_2) - Cov(Y_2, X)(Cov X)^{-1}Cov(X, Y_2)}}.$$

When $p = 1$ so that X equals the scalar random variable X_1, the above formula simplifies to (check this):
$$\rho_{Y_1,Y_2|X_1} = \frac{\rho_{Y_1,Y_2} - \rho_{Y_1,X_1}\rho_{Y_2,X_1}}{\sqrt{1-\rho^2_{Y_1,X_1}}\sqrt{1-\rho^2_{Y_2,X_1}}}.$$

It is instructive to put the variances of the residuals $r_{Y_1|X_1,\dots,X_p}$ and $r_{y_2|X_1,\dots,X_p}$ and their covariance in a matrix. Recall first that:
$$Var(r_{Y_1|X_1,\dots,X_p}) = Var(Y_1) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_1),$$
$$Var(r_{Y_2|X_1,\dots,X_p}) = Var(Y_2) - Cov(Y_2, X)(Cov X)^{-1}Cov(X, Y_2)$$
and
$$Cov(r_{Y_1|X_1,\dots,X_p}, r_{Y_2|X_1,\dots,X_p}) = Cov(Y_1, Y_2) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_2).$$

Let $R_{Y_1,Y_2|X_1,\dots,X_p}$ denote the 2×1 random vector consisting of the residuals $r_{Y_1|X_1,\dots,X_p}$ and $r_{Y_2|X_1,\dots,X_p}$. The formulae for the variances and covariances of the residuals allows us then to write the 2×2 covariance matrix of $R_{Y_1,Y_2|X_1,\dots,X_p}$ as
$$\begin{aligned} Cov(R_{Y_1,Y_2|X_1,\dots,X_p}) &= Cov\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} - \begin{pmatrix} Cov(Y_1, X) \\ Cov(Y_2, X) \end{pmatrix} (Cov X)^{-1} \begin{pmatrix} Cov(X, Y_1) & Cov(X, Y_2) \end{pmatrix} \\ &= Cov(Y) - Cov(Y, X)(Cov X)^{-1}Cov(X, Y) \end{aligned}$$
where
$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{pmatrix}.$$

The right hand side in the formula for $Cov(R_{Y_1,Y_2|X_1,\dots,X_p})$ equals precisely the Schur complement of $Cov(Y)$ in the matrix
$$\begin{pmatrix} Cov(X) & Cov(X, Y) \\ Cov(Y, X) & Cov(Y) \end{pmatrix} = Cov\begin{pmatrix} X \\ Y \end{pmatrix} =: \Sigma.$$

Thus if Σ denotes the covariance matrix of the $(p+2)\times 1$ random vector $(X_1, \dots, X_p, Y_1, Y_2)^T$, then $Cov(R_{Y_1,Y_2|X_1,\dots,X_p})$ equals precisely the Schur complement of $Cov(Y)$ in Σ.

There are two very nice connections between the partial correlation and the inverse covariance matrix, as well as the BLP. We shall see these in the next class.

5 Recommended Reading for Today

1. Read the wikipedia article `https://en.wikipedia.org/wiki/Partial_correlation` on partial correlation
2. Read the material on the PACF in Section 3.3 of the Shumway-Stoffer book.