STAT 153 - Introduction to Time Series
Lecture Twenty One

Fall 2022, UC Berkeley

Aditya Guntuboyina

November 8, 2022

We shall continue our discussion of Partial Correlation, and define the diagnostic tool PACF.

1 Partial Correlation and Inverse Covariance

We defined partial correlation in the last lecture. 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 $X_1, \dots, X_p$.

Recall that the residuals $r_{Y_1|X_1,\dots,X_p}$ and $r_{Y_2|X_1,\dots,X_p}$ have the following expressions:

$$r_{Y_1|X_1,\dots,X_p} = Y_1 - \mathbb{E}(Y_1) - Cov(Y_1, X)(Cov X)^{-1}(X - \mathbb{E}(X))$$

and

$$r_{Y_2|X_1,\dots,X_p} = Y_2 - \mathbb{E}(Y_2) - Cov(Y_2, X)(Cov X)^{-1}(X - \mathbb{E}(X)),$$

In the last class, we computed the variances of $r_{Y_1|X_1,\dots,X_p}$ and $r_{Y_2|X_1,\dots,X_p}$ as well as the covariance between them. This gave us the formulae:

$$Var(r_{Y_1|X_1,\dots,X_p}) = Var(Y_1) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_1),$$

$$Var(r_{Y_2|X_1,\dots,X_p}) = Var(Y_2) - Cov(Y_2, X)(Cov X)^{-1}Cov(X, Y_2)$$

and

$$Cov(r_{Y_1|X_1,\dots,X_p}, r_{Y_2|X_1,\dots,X_p}) = Cov(Y_1, Y_2) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_2).$$

We can put these expressions together to get the following formula for the partial correlation between Y_1 and Y_2 given $X_1, \dots, X_p$:

$$\frac{Cov(Y_1, Y_2) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_2)}{\sqrt{Var(Y_1) - Cov(Y_1, X)(Cov X)^{-1}Cov(X, Y_1)}\sqrt{Var(Y_2) - Cov(Y_2, X)(Cov X)^{-1}Cov(X, Y_2)}}.$$

We shall now describe the connection between partial correlations and the inverse of the Covariance matrix. Let $R_{Y_1,Y_2|X_1,\dots,X_p}$ denote the 2×1 random vector consisting of the residuals $r_{Y_1|X_1,\ldots,X_p}$ and $r_{Y_2|X_1,\ldots,X_p}$. The formulae for the variances and covariances of the residuals allows us then to write the 2×2 covariance matrix of $R_{Y_1,Y_2|X_1,\ldots,X_p}$ as

$$\begin{aligned} Cov(R_{Y_1,Y_2|X_1,\ldots,X_p}) &= Cov\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} - \begin{pmatrix} Cov(Y_1,X) \\ Cov(Y_2,X) \end{pmatrix} (CovX)^{-1} \begin{pmatrix} Cov(X,Y_1) & Cov(X,Y_2) \end{pmatrix} \\ &= Cov(Y) - Cov(Y,X)(CovX)^{-1}Cov(X,Y) \end{aligned}$$

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 $Cov(R_{Y_1,Y_2|X_1,\ldots,X_p})$ 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 $(X_1,\ldots,X_p,Y_1,Y_2)^T$, then $Cov(R_{Y_1,Y_2|X_1,\ldots,X_p})$ equals precisely the Schur complement of $Cov(Y)$ in Σ.

But we know if we invert Σ, then the last diagonal block (or the $(2,2)^{th}$ block) of Σ^{-1} equals the inverse of the Schur complement of the $(2,2)^{th}$ block of Σ. This and the above connection between Schur complement and the covariance of $R_{Y_1,Y_2|X_1,\ldots,X_p}$ allows us to deduce that if

$$\Sigma^{-1} = \begin{pmatrix} (\Sigma^{-1})_{11} & (\Sigma^{-1})_{12} \\ (\Sigma^{-1})_{21} & (\Sigma^{-1})_{22}, \end{pmatrix}$$

then

$$(\Sigma^{-1})_{22} = \left(Cov(R_{Y_1,Y_2|X_1,\ldots,X_p})\right)^{-1} = \begin{pmatrix} Var(r_{Y_1|X_1,\ldots,X_p}) & Cov(r_{Y_1|X_1,\ldots,X_p}, r_{Y_2|X_1,\ldots,X_p}) \\ Cov(r_{Y_1|X_1,\ldots,X_p}, r_{Y_2|X_1,\ldots,X_p}) & Var(r_{Y_2|X_1,\ldots,X_p}) \end{pmatrix}^{-1}$$

The usual formula for the inverse of a 2×2 matrix then gives

$$(\Sigma^{-1})_{22} = \frac{1}{D}\begin{pmatrix} Var(r_{Y_2|X_1,\ldots,X_p}) & -Cov(r_{Y_1|X_1,\ldots,X_p}, r_{Y_2|X_1,\ldots,X_p}) \\ -Cov(r_{Y_1|X_1,\ldots,X_p}, r_{Y_2|X_1,\ldots,X_p}) & Var(r_{Y_1|X_1,\ldots,X_p}) \end{pmatrix}$$

where D is the determinant of $Cov(R_{Y_1,Y_2|X_1,\ldots,X_p})$.

From here it follows that the partial correlation $\rho_{Y_1,Y_2|X_1,\ldots,X_p}$ has the alternative expression:

$$\rho_{Y_1,Y_2|X_1,\ldots,X_p} = \frac{Cov(r_{Y_1|X_1,\ldots,X_p}, r_{Y_2|X_1,\ldots,X_p})}{\sqrt{Var(r_{Y_1|X_1,\ldots,X_p})Var(r_{Y_2|X_1,\ldots,X_p})}} = \frac{-(\Sigma^{-1})(n-1,n)}{\sqrt{(\Sigma^{-1})(n-1,n-1)\Sigma^{-1}(n,n)}}.$$

This shows the connection between partial correlation and inverse covariance matrices.

More generally, if $Y_1,\ldots,Y_n$ are random variables (no distributional assumptions are needed here) with covariance matrix given by Σ. Then the partial correlation between Y_i and Y_j given $Y_k, k \neq i, k \neq j$ can be written in terms of Σ^{-1} as

$$\rho_{Y_i,Y_j|Y_k,k\neq i,k\neq j} = \frac{-(\Sigma^{-1})(i,j)}{\sqrt{(\Sigma^{-1})(i,i)(\Sigma^{-1})(j,j)}}.$$

This implies, in particular, that

$$(\Sigma^{-1})(i,j) = 0 \iff \rho_{Y_i,Y_j|Y_k,k\neq i,k\neq j} = 0$$

Therefore $(\Sigma^{-1})(i,j) = 0$ is equivalent to the partial correlation between Y_i and Y_j given $Y_k, k \neq i, k \neq j$ being zero.

Also

$$(\Sigma^{-1})(i,j) \leq 0 \iff \rho_{Y_i,Y_j|Y_k,k\neq i,k\neq j} \geq 0 \quad \text{and} \quad (\Sigma^{-1})(i,j) \geq 0 \iff \rho_{Y_i,Y_j|Y_k,k\neq i,k\neq j} \leq 0$$

In other words, $\Sigma^{-1}(i,j)$ being nonpositive is equivalent to the partial correlation between Y_i and Y_j given $Y_k, k \neq i, k \neq j$ being nonnegative. Similarly, $\Sigma^{-1}(i,j)$ being nonnegative is equivalent to the partial correlation between Y_i and Y_j given $Y_k, k \neq i, k \neq j$ being nonpositive.

2 Partial Correlation and Best Linear Predictor

Consider random variables Y and $X_1, \dots, X_p$. Let $\beta_0^* + \beta_1^* X_1 + \cdots + \beta_p^* X_p$ denote the BLP of Y in terms of $X_1, \dots, X_p$.

We have seen before that If $p = 1$, then X is equal to the scalar random variable X_1 and the BLP then has the expression:

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

In other words, when $p = 1$, the slope coefficient of the BLP is given by

$$\beta_1^* = \frac{Cov(Y, X_1)}{Var(X_1)} = \rho_{Y,X_1}\sqrt{\frac{Var(Y)}{Var(X_1)}}. \tag{1}$$

When $p \geq 1$, we would have p "slope" coefficients $X_1, \dots, X_p$. In this case, one can write a formula analogous to (1) as follows:

$$\beta_i^* = \rho_{Y,X_i|X_k,k\neq i}\sqrt{\frac{Var(r_{Y|X_k,k\neq i})}{Var(r_{X_i|X_k,k\neq i})}} \tag{2}$$

In other words β_i^* equals the slope coefficient of BLP of $r_{Y|X_k,k\neq i}$ in terms of $r_{X_i|X_k,k\neq i}$.

We shall prove this fact now. We can assume without loss of generality $i = p$. The proof for other i can be completed by rearranging $X_1, \dots, X_p$ so that X_i appears at the last position. The formula for $\beta^* = (\beta_1, \dots, \beta_p)^*$ is

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

Let us write

$$X = \begin{pmatrix} X_{-p} \\ X_p \end{pmatrix}$$

where $X_{-p} := (X_1, \dots, X_{p-1})^T$ consists of all the X's except X_p. We can partition $Cov(X)$ as

$$Cov(X) = Cov\begin{pmatrix} X_{-p} \\ X_p \end{pmatrix} = \begin{pmatrix} Cov(X_{-p}) & Cov(X_{-p}, X_p) \\ Cov(X_p, X_{-p}) & Var(X_p) \end{pmatrix}.$$

The formula for β^* then becomes

$$\beta^* = (CovX)^{-1}Cov(X,Y) = \begin{pmatrix} Cov(X_{-p}) & Cov(X_{-p}, X_p) \\ Cov(X_p, X_{-p}) & Var(X_p) \end{pmatrix}^{-1} \begin{pmatrix} Cov(X_{-p}, Y) \\ Cov(X_p, Y) \end{pmatrix}$$

In order to derive an explicit formula for β_p^* from this expression, we need to figure out the last row of $(CovX)^{-1}$. A standard formula for the inverses of partitioned matrices states that

$$A = \begin{pmatrix} E & F \\ G & H \end{pmatrix} \implies A^{-1} = \begin{pmatrix} \text{something} & \text{something} \\ -(H^S)^{-1}GE^{-1} & (H^S)^{-1} \end{pmatrix}$$

where $H^S := H - GE^{-1}F$ is the Schur complement of H in A. We shall apply this formula to

$$E = Cov(X_{-p}), \quad F = Cov(X_{-p}, X_p), \quad G = Cov(X_p, X_{-p}), \quad \text{and } H = Var(X_p)$$

so that A equals $Cov(X)$. In this case,

$$H^S = Var(X_p) - Cov(X_p, X_{-p})(Cov(X_{-p}))^{-1}Cov(X_{-p}, X_p) = Var(r_{X_p|X_k, k\neq p})$$

so that

$$(H^S)^{-1} = \frac{1}{Var(r_{X_p|X_k,k\neq p})}.$$

We thus obtain

$$\begin{aligned} \beta_p^* &= -(H^S)^{-1}GE^{-1}Cov(X_{-p}, Y) + (H^S)^{-1}Cov(X_p, Y) \\ &= \frac{Cov(X_p, Y) - Cov(X_p, X_{-p})(CovX_{-p})^{-1}Cov(X_{-p}, Y)}{Var(r_{X_p|X_k,k\neq p})} \\ &= \frac{Cov(r_{Y|X_k,k\neq p}, r_{X_p|X_k,k\neq p})}{Var(r_{X_p|X_k,k\neq p})} = \rho_{Y,X_p|X_k,k\neq p}\sqrt{\frac{Var(r_{Y|X_k,k\neq p})}{Var(r_{X_p|X_k,k\neq p})}}. \end{aligned}$$

which proves the result for $i = p$. One can prove it for other i by simply rearranging $X_1, \dots, X_p$ so that X_i appears as the last variable.

An important consequence of (2) is:

$$\beta_i^* = 0 \iff \rho_{Y,X_i|X_k,k\neq i} = 0 \tag{3}$$

In other words, the coefficient of X_i in the BLP of Y based on $X_1, \dots, X_p$ equals zero if and only if the partial correlation between Y and X_i given $X_k, k \neq i$ equals 0.

3 The Partial Autocorrelation Function (pacf)

3.1 First Definition

Let $\{Y_t\}$ be a stationary process. The Partial Autocorrelation at lag h, denoted by $pacf(h)$, is defined as the partial corrlelation between Y_t and Y_{t-h} given the intervening variables $Y_{t-1}, \dots, Y_{t-h+1}$:

$$pacf(h) := \rho_{Y_t, Y_{t-h}|Y_{t-1},\dots,Y_{t-h+1}}.$$

Note that the right hand side involves both t and h. However, because we assumed that the process $\{Y_t\}$ is stationary, the right hand side actually does not depend on t.

3.2 Second Definition

Using the connection (2) between partial correlation and BLP coefficients, we can also define the $pacf(h)$ for a stationary process $\{X_t\}$ in the following alternative way: $pacf(h)$ is defined as the coefficient of Y_{t-h} in the BLP of Y_t in terms of $Y_{t-1}, \dots, Y_{t-h}$.

To see why these two definitions are equivalent, suppose the BLP of Y_t in terms of $Y_{t-1}, \dots, Y_{t-h}$ equals $\beta_0^* + \beta_1^* Y_{t-1} + \cdots + \beta_h^* Y_{t-h}$. Then by (2)

$$\beta_h^* = \rho_{Y_t, Y_{t-h} | Y_{t-1}, \dots, Y_{t-h+1}} \sqrt{\frac{Var(r_{Y_t | Y_{t-1}, \dots, Y_{t-h+1}})}{Var(r_{Y_{t-h} | Y_{t-1}, \dots, Y_{t-h+1}})}}.$$

By stationarity, the covariance matrix of $(Y_t, Y_{t-1}, \dots, Y_{t-h+1})^T$ and that of $(Y_{t-h}, Y_{t-h+1}, \dots, Y_{t-1})^T$ is exactly the same. As a result

$$Var(r_{Y_t | Y_{t-1}, \dots, Y_{t-h+1}}) = Var(r_{Y_{t-h} | Y_{t-1}, \dots, Y_{t-h+1}})$$

which gives

$$\beta_h^* = \rho_{Y_t, Y_{t-h} | Y_{t-1}, \dots, Y_{t-h+1}} = pacf(h).$$

4 $pacf(h)$ for $AR(p)$

Consider the causal stationary $AR(p)$ model (i.e., the roots of the AR polynomial have modulus strictly larger than 1): $Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$. Then

$$pacf(h) = \begin{cases} \phi_p & \text{when } h = p \\ 0 & \text{for } h > p \end{cases}$$

For $h < p$, the $pacf(h)$ will be a slightly complicated function of $\phi_1, \dots, \phi_p$.

This form for $pacf(h)$ can be proved in the following way. For $AR(p)$, when $h \geq p$, the BLP of Y_t in terms of $Y_{t-1}, \dots, Y_{t-h}$ is simply equal to $\mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}$. This is because

$$Y_t - (\mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}) = Z_t$$

which is uncorrelated with $Y_{t-1}, \dots, Y_{t-p}$. Thus, the second definition of $pacf(h)$ which says that $pacf(h)$ equals the coefficient of Y_{t-h} in the BLP of Y_t in terms of $Y_{t-1}, \dots, Y_{t-h}$ immediately yields the above formula for $pacf(h)$.

5 Estimating $pacf$ from Data

How does one estimate $pacf(h)$ from data for different lags h? In order to calcuate the BLP of X_t in terms of $X_{t-1}, \dots, X_{t-h}$, we need to know all the covariances $\mathrm{Cov}(X_{t-i}, X_{t-j})$ for $i, j \in \{0, 1, \dots, h\}$. These covariances are estimated by sample covariances which are then used to estimate the BLP. The coefficient of X_{t-h} in the estimated BLP gives the estimate of $pacf(h)$.

6 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.
3. The theory of graphical models crucially uses inverse covariance matrices. See, for example, the book on *Graphical Models* by Lauritzen.