Time Series

Loukia Meligkotsidou,
National and Kapodistrian University of Athens

MSc in Statistics and Operational Research,
Department of Mathematics

Partial Correlation in Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \epsilon_i, \ \ \epsilon_i \sim N(0, \sigma^2), \ \ i = 1, \ldots, n.$$

The coefficient β_j, $j = 1, \ldots, k$, is interpreted as the change in the expected value of Y corresponding to an increase in X_j by one unit with the values of all other covariates remaining fixed.

Partial Correlation in Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2), \;\; i = 1, \ldots, n.$$

The coefficient β_j, $j = 1, \ldots, k$, is interpreted as the change in the expected value of Y corresponding to an increase in X_j by one unit with the values of all other covariates remaining fixed.

- The coefficient β_j describes the effect of X_j on Y, having taken into account the effects of the remaining covariates (some of which may be correlated with X_j).

Partial Correlation in Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \epsilon_i, \ \ \epsilon_i \sim N(0, \sigma^2), \ \ i = 1, \ldots, n.$$

The coefficient β_j, $j = 1, \ldots, k$, is interpreted as the change in the expected value of Y corresponding to an increase in X_j by one unit with the values of all other covariates remaining fixed.

- The coefficient β_j describes the effect of X_j on Y, having taken into account the effects of the remaining covariates (some of which may be correlated with X_j).
- The coefficient β_j, though, is measured in some units, therefore it cannot directly quantify the extent of this effect or to be used for comparisons.

The Correlation Coefficient

Covariance: $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

Correlation: $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$

Sample Correlation: $r_{XY} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$

The Correlation Coefficient

Covariance: $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

Correlation: $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$

Sample Correlation: $r_{XY} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$

Partial Correlation: $\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$
(of X, Y given Z)

The Correlation Coefficient

Covariance: $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

Correlation: $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$

Sample Correlation: $r_{XY} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$

Partial Correlation: $\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$

(of X, Y given Z)

The coefficient of Partial Correlation, $\rho_{XY|Z}$, measures the correlation between Y and X after taking into account the information in Z, a variable correlated with both Y and X.

Partial Correlation in the Context of Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n.$$

Partial Correlation: $\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$

Partial Correlation in the Context of Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n.$$

Partial Correlation: $\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$

The partial correlation $\rho_{XY|Z}$ can be thought of as the correlation between the random errors, ϵ_X and ϵ_Y of the linear regression of X on Z and of the linear regression of Y on Z, respectively.

Partial Correlation in the Context of Linear Regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n.$$

Partial Correlation: $\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$

The partial correlation $\rho_{XY|Z}$ can be thought of as the correlation between the random errors, ϵ_X and ϵ_Y of the linear regression of X on Z and of the linear regression of Y on Z, respectively.

Estimation:

1. Estimation of the linear regression of Y on Z and calculation of the residuals $\hat{\epsilon}_Y$.
2. Estimation of the linear regression of X on Z and calculation of the residuals $\hat{\epsilon}_X$.
3. Estimate the partial correlation as: $r_{XY|Z} = r_{\hat{\epsilon}_X \hat{\epsilon}_Y}$.

Generalisation

The coefficient of Partial Correlation, $\rho_{XY|Z_1,\ldots,Z_p}$, measures the correlation between Y and X after taking into account the information in $Z_1, \ldots, Z_p$.

Generalisation

The coefficient of Partial Correlation, $\rho_{XY|Z_1,\ldots,Z_p}$, measures the correlation between Y and X after taking into account the information in $Z_1, \ldots, Z_p$.

In multiple regression, several methods of variable selection, such as, Forward, Backward and Stepwise Regression, are based on partial correlations. In order to decide which is the next variable to be included or excluded from the model, the partial correlations of Y with the available covariates, given those already included in the model, are examined.

Partial Autocorrelation

Recall:

Partial Correlation: $\rho_{YX|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$

(of X, Y given Z)

Partial Autocorrelation

Recall:
Partial Correlation: $\rho_{YX|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$
(of X, Y given Z)

In the context of Time Series: $Y \rightarrow Y_t$, $Z \rightarrow Y_{t-1}$, $X \rightarrow Y_{t-2}$

Partial Autocorrelation

Recall:
Partial Correlation: $\rho_{YX|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$
(of X, Y given Z)

In the context of Time Series: $Y \rightarrow Y_t$, $Z \rightarrow Y_{t-1}$, $X \rightarrow Y_{t-2}$

$\rho_{Y_t Y_{t-2}|Y_{t-1}} = \frac{\rho_{YY_{t-2}} - \rho_{Y_{t-1}Y_{t-1}}\rho_{Y_{t-1}Y_t}}{\sqrt{1-\rho_{Y_{t-2}Y_{t-1}}^2}\sqrt{1-\rho_{Y_{t-1}Y_t}^2}}$

Partial Autocorrelation

Recall:

Partial Correlation: $\rho_{YX|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$

(of X, Y given Z)

In the context of Time Series: $Y \to Y_t$, $Z \to Y_{t-1}$, $X \to Y_{t-2}$

$$\rho_{Y_t Y_{t-2}|Y_{t-1}} = \frac{\rho_{YY_{t-2}} - \rho_{Y_{t-1}Y_{t-1}}\rho_{Y_{t-1}Y_t}}{\sqrt{1-\rho_{Y_{t-2}Y_{t-1}}^2}\sqrt{1-\rho_{Y_{t-1}Y_t}^2}}$$

$$\alpha_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$$

Note: Here the partial autocorrelation is defined as a function of the correlations (in the same manner as the partial correlation above is defined as a function of the correlations).

Partial Autocorrelation in the Context of Autoregressive Models

The coefficient of partial autocorrelation at lag k is defined as the kth autoregressive coefficient in an AR(k) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_{\mathbf{k}} y_{t-k} + \epsilon_t, \;\; \epsilon_t \sim N(0, \sigma^2), \;\; t = 1, \ldots, T.$$

Partial Autocorrelation in the Context of Autoregressive Models

The coefficient of partial autocorrelation at lag k is defined as the kth autoregressive coefficient in an AR(k) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_{\mathbf{k}} y_{t-k} + \epsilon_t, \;\; \epsilon_t \sim N(0, \sigma^2), \;\; t = 1, \ldots, T.$$

- The coefficient ϕ_k describes the effect of Y_{t-k} on Y_t, having taken into account the effects of $Y_{t-1}, \ldots, Y_{t-k-1}$ (which are correlated with Y_t).

Partial Autocorrelation in the Context of Autoregressive Models

The coefficient of partial autocorrelation at lag k is defined as the kth autoregressive coefficient in an AR(k) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_{\mathbf{k}} y_{t-k} + \epsilon_t, \;\; \epsilon_t \sim N(0, \sigma^2), \;\; t = 1, \ldots, T.$$

- The coefficient ϕ_k describes the effect of Y_{t-k} on Y_t, having taken into account the effects of $Y_{t-1}, \ldots, Y_{t-k-1}$ (which are correlated with Y_t).
- The coefficient ϕ_{t-k}, does not have units, therefore it can directly quantify the extent of this effect.

Estimation of Partial Autocorrelations

1st way: Estimation through the autoregressive models

AR(1): $y_t = \phi_1 y_{t-1} + \epsilon_t$, then $\hat{\alpha}_1 = \hat{\phi}_1$
AR(2): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$, then $\hat{\alpha}_2 = \hat{\phi}_2$

.

.

.

AR(p): $\hat{\alpha}_p = \hat{\phi}_p$

Estimation of Partial Autocorrelations

1st way: Estimation through the autoregressive models

AR(1): $y_t = \phi_1 y_{t-1} + \epsilon_t$, then $\hat{\alpha}_1 = \hat{\phi}_1$
AR(2): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$, then $\hat{\alpha}_2 = \hat{\phi}_2$

.

.

.

AR(p): $\hat{\alpha}_p = \hat{\phi}_p$

2nd way: Estimation through the coefficients of autocorrelation

Solution of the Yule-Walker equations: for example $\hat{\alpha}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}$