MULTICOLLINEARITY

By:
Dr. Akash Asthana
Assistant Professor,
Dept. of Statistics,
University of Lucknow, Lucknow

MULTICOLLINEARITY

- In Linear models an assumption is taken that the regressors must be independent of each other. But if they are not independent of each other then multicollinearity is said to be present. The regression parameters are highly affected by the presence of multicollinearity.

DETECTION OF MULTICOLLINEARITY

- There are several methods used for the detection of multicollinearity among which main methods are:
- Use of R^2 and t-statistic
- Use of pair-wise correlations
- Use of Auxiliary regression
- Use of Eigen-Value and condition index
- Tolerance and Variance inflation factor

USE OF R^2 AND T-STATISTIC

- Generally it is said that a regression model is good if the value of coefficient of determination (R^2) is high. It is also a useful tool for detection of multicollinearity. If the value of R^2 is high (more than 0.8) and a very few regression parameters comes out to be significant (using t-test for individual regression parameters) then multicollinearity is said to be present.

USE OF PAIR-WISE CORRELATION

- It is another method used for the detection of multicollinearity.
- In this method the Karl Pearson’s product moment correlation coefficient is computed among each pair of regressors.
- As a rule of thumb if the pair-wise correlation exceeds the value of 0.8 then multicollinearity is said to be present.
- Some researchers prefer the use of partial correlation instead of Pearson’s correlation as it removes the effect of other regressors.

USE OF AUXILIARY REGRESSION

- Let's consider the regression model :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \epsilon_i$$

- Then the regression equations:

$$X_{1i} = \beta_{10} + \beta_{12} X_{2i} + \beta_{13} X_{3i} + \cdots + \beta_{1p} X_{pi} + \epsilon_{1i}$$

$$X_{2i} = \beta_{20} + \beta_{21} X_{1i} + \beta_{23} X_{3i} + \cdots + \beta_{2p} X_{pi} + \epsilon_{2i}$$

.

.

.

$$X_{pi} = \beta_{10} + \beta_{p1} X_{1i} + \beta_{p2} X_{2i} + \cdots + \beta_{p(p-1)} X_{(p-1)i} + \epsilon_{1i}$$

are called as Auxiliary regressions

USE OF AUXILIARY REGRESSION

- In this method each of the regressor is regressed over remaining regressors.
- For each Auxiliary regression the value of coefficient of determination (R_j^2)is computed.
- Also the value of R^2 is computed for the regression model of Y on all the regressors.
- As Klien's rule of thumb if the R_j^2 exceeds the value of R^2 then multicollinearity s said to be present.

EIGEN-VALUE AND CONDITION INDEX

- It is another method used for the detection of multicollinearity.
- In this method first the Eigen-Value for the matrix X'X is computed.
- Then the condition number (k) and condition Index (CI) are obtained by:

$$k = \frac{Maximum\ Eigen\ Value}{Minimum\ Eigen\ Value};\ CI = \sqrt{\frac{Maximum\ Eigen\ Value}{Minimum\ Eigen\ Value}}$$

- As a rule of thumb:

1. If k < 100 or CI < 10 then multicollinearity is absent.
2. If 100<k<1000 or 10<CI<30 then moderate to strong multicollinearity is present.
3. If k>1000 or CI>30 the sever multicollinearity is present.

VARIANCE INFLATION FACTOR AND TOLERANCE

- It is another method of detection of multicollinearity.
- The Variance Inflation Factor (VIF) is given by:

$$VIF = 1/(1-R_j^2)$$

  where R_j^2 is the coefficient of determination for the auxiliary regression of X_j on remaining regressors.
- As a rule of thumb if VIF exceeds 10 then multicollinearity is said to be present.
- VIF has a drawback that it is unbounded therefore use of tolerance (TOL) is preferred which is defined by:

$$TOL = 1/VIF$$

- The value of TOL lies between 0 and 1.
- As a rule of thumb if value of TOL is closer to zero then multicollinearity is said to be present.

REFRENCES

1. Gujrati DN, Basic Econometrics, 4th edition (2004) , The McGraw-Hill Companies.
2. Draper NR & Smith H, Applied Regression Analysis, 3rd edition (1998), John Wiley & Sons Inc.
3. Johnston J & Dinardo J, Econometric Methods, 4th edition (1997), McGraw-Hill Companies.