# Analyzing the problems of multicollinearity in the data

Aboobacker Jahufer[1] and Chen Jianbao[1]

Regression Analysis is one of the most widely used statistical techniques for analyzing multifactor data. Its broad appeal results from the conceptually simple process of using an equation to express the relationship between a set of variables. Regression analysis is also interesting theoretically because of the elegant underlying mathematics. Successful use of regression analysis requires an appreciation of both the theory and the practical problems that often arise when the technique is employed with real world data.

In regression model fitting process the most frequently applied and most popular estimation procedure is the Ordinary Least Square Estimation (OLSE). The significant advantage of OLSE is that it provides minimum variance unbiased linear estimates for the parameters in the linear regression model.

In many situations both experimental and non-experimental, the independent variables tend to be correlated among themselves. Then inter-correlation or multicollinearity among the independent variables is said to be exist. A variety of interrelated problems are created when multicollinearity exists. Specially, in the regression model building process, multicollinearity among the independent variables causes high variance (if OLSE is used) even though the estimators are still the minimum variance unbiased estimators in the class of linear unbiased estimators.

The main objective of this paper is to analyze the multicollinearity among the independent variables. For this research object, five independent and one dependent standard normal pseudo-random (numbers) variables of multicollinearity data were generated by using Monte Carlo Simulation method with the correlation $r = 0.99$ between the independent variables. For data analysis purposes 100 observations for each variable are randomly selected. The diagnostics and dealing methods of multicollinearity were analyzed and proposed respectively.

**Key words:** Multicollinearity, Correlation matrix, Eigen value, VIF, Conditional indices.

[1] Department of Planning and Statistics, School of Economics, Xiamen University, Xiamen 361005, P. R. China