Journal of Business Economics
ISSN 2682- 6933 | E-ISSN 2705-9944
2025 Volume 07 Issue 01: 13-22

MULTICOLLINEARITY ISSUES IN MODEL BUILDING PROCESS AND REMEDIAL MEASURES TO SOLVE

Aboobacker Jahufer
Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka
jahufer@seu.ac.lk

Abstract

One of the major assumptions of the multiple linear ordinary least squares (OLS) regression model is that there is no exact linear relationship between independent variables. If such a linear relationship does exist, it can be said that the independent variables are multicollinearity. When multicollinearity exists among the independent variables, a variety of interrelated problems are created. Specially, in the model building process multicollinearity causes high variance for parameters if ordinary least squares estimator is used. Depending on the situation, it may not be a problem for the model if only slight or moderate multicollinearity issue occurs. However, it is strongly advised to solve the issue if severe multicollinearity issue exists. Example if correlation >0.8 between two independent variables or Variance Inflation Factor (VIF >10) or Conditional Index Number > 1000 or Conditional Index > 30 or Variance Proportion > 0.5.

The main objective of this research paper is to analyze and detect the multicollinearity in the data set and recommend some important dealing methods and remedial methods for multicollinearity problems. A multicollinearity data set (Kim, 2019) consists of a dependent variable with six independent variables are used to illustrate the methodologies proposed in this paper. It is confirmed by the multicollinearity detecting methods the data set has severe multicollinearity issue. Two remedial measures (i) Principal Component Analysis (PCA) and (ii) Ridge Regression Analysis (RRA) also used for this data set. Comparison and Efficiency between OLS and remedial models (PAC and RRA) also estimated by studying the sum of square error. It is confirmed that the efficiency of PCA is 2.2 times than OLS whereas the efficiency of RRA is 70.8 times than OLS models for this data set.

Keywords: *Multicollinearity; Correlation Matrix; Eigen Analysis; Variance Inflation Factor; Conditional Indices; Variance Decomposition; Biased Estimation.*

1. Introduction

The use and interpretation of a multiple regression model often depends explicitly or implicitly on the estimates of the individual regression coefficients. If there is no linear relationship between the independent variables, they said to be orthogonal. When the independent variables are orthogonal, inferences such as: (i) identifying the relative effects of the independent variables, (ii) prediction and/or estimation, and (iii) selection of an appropriate set of variables for the model can be made relatively easily. Unfortunately, in most applications of regression, the independent variables are not orthogonal. Sometimes the lack of orthogonality is not serious. However, in some situations the independent variables are nearly perfectly linearly related, and in such cases the inferences based on the regression model can be misleading or erroneous. When there are near linear dependencies between the regressor between the independent variables, the problem of multicollinearity is said to be exist.

The presence of multicollinearity can make the usual least squares analysis of the regression model dramatically inadequate. The are four primary sources of multicollinearity: (i) The data collection methods employed, (ii)

Constraint on the model or in the population, (iii) Model specification and (iv) An over-defined model. It is important to understand the differences among these sources of multicollinearity, as the recommendations for analysis of the data and interpretation of the resulting model depend to some extent on the cause of the problems (see Mason, et. al., (1975), for further discussion of the source of multicollinearity).
This paper composed into five sections, section 2 illustrates review of literature, in section three methodology is clearly demonstrated. Results and discussion are given in section four. In section five conclusions are given.

2. Literature Review

Kim (2019) studied a data set to detect multicollinearity. He used multicollinearity detecting methods: Variance Inflation Factor, Condition Number and Condition Index and Variance Decomposition Proportion.
Jahufer (2011) analyzed and detected multicollinearity in the data sets by using the methods: Examination of the Correlation Matrix, Variance Inflation Factor (VIF) and Eigensystem Analysis of X'X. For dealing methods for multicollinearity problems he suggested, Collecting Additional data, Model Re-Specification and Biased Estimation methods. He used two multicollinearity data sets to illustrate the methodologies proposed in his paper.

Snee (1975) reviewed the theory of ridge regression and its relation to generalized inverse regression is presented along with the use of ridge regression in practice. Comments on variable selection procedures, model validation and ridge and generalized inverse regression computation procedures are included. The examples studied here show that when the independent variables are highly correlated, ridge regression procedures coefficients which predict and extrapolate better than least squares and is a safe procedure for selecting variables.

Hoerl and Kennard (1970) suggested that in multiple regression it is shown that parameter estimates based on minimum residual sum of squares have a high probability of being unsatisfactory, if not incorrect, if the prediction vectors are not orthogonal. Proposed is an estimation procedure based on adding small positive quantities to the diagonal of $\mathbf{X'X}$. Introduced is the ridge race, a method for showing in two dimension the effects of nonorthogonality. It is then shown how to augment $\mathbf{X'X}$ to obtain biased estimates with smaller mean square error.

El-Dereny and Rashwan (2011) introduced many different methods of ridge regression to solve multicollinearity problem. These methods include ordinary ridge regression (ORR), generalized ridge regression (GRR), and directed ridge regression (DRR). Properties of ridge regression estimators and methods of selecting biased ridge regression parameter are discussed. Authors used data simulation to make comparison between methods of ridge regression and ordinary least squares (OLS) method. According to a result of this study, Authors found that all methods of ridge regression are better than OLS method when the Multicollinearity exist.

Shariff and Duzan (2018) suggested that presence of multicollinearity will produce unreliable result in the parameter estimates if OLS is applied to estimate the model. Due to such reason, authors proposed ridge estimator as linear combinations of the coefficient of least squares regression of explanatory variables to the real application. The numerical example of stock market price and macroeconomic variables in Malaysia is employed using both methods with the aim of investigating the relationship of the variables in the presence of multicollinearity in the data set. The variables on interest are Consumer Price Index (CPI), Gross Domestic Product (GDP), Base Lending Rate (BLR) and Money Supply (M1). The obtained findings show that the proposed procedure is able to estimate the model and produce reliable result by reducing the effect of multicollinearity in the data set.

3. Methodology

3.1 Effects of Multicollinearity

The presence of multicollinearity has a number of potentially serious effects on the least squares estimates on the regression coefficients. Some of these effects may be easily demonstrated. If there is a strong multicollinearity between two variables say X_1 and X_2 results in large variances and covariances for the least squares estimators of the regression coefficients this means $\text{Var}(\hat{\beta}_i) \to \infty, i = 1, 2$ and $\text{Covar}(\hat{\beta}_1, \hat{\beta}_2) \to \mp\infty$. This implies that different samples taken at the same X levels could lead to widely different estimates of the model parameters.

Furthermore, the method of least squares will generally produce poor estimates of the individual model parameters when strong multicollinearity is present, this does not necessarily imply that the fitted model is a poor predictor.

3.2 Indication of Multicollinearity

When multicollinearity presents among the independent variables and fitting regression model using ordinary least squares method for those variables then the model reveals that higher coefficient of determination (R^2 ➔ 1), the ANOVA table F – statistics value is very high and probability value is near zero (p ➔ 0). These results indicate that the fitted model is significant. Whereas the statistical test results of the individual variables most of the variables are not significant. This is indicated that independent variables are correlated among them and created multicollinearity problems.

3.3 Multicollinearity Diagnostics

Several techniques have been proposed for detecting multicollinearity. In this paper most important three methods are discussed.

3.3.1 Examination of the Correlation Matrix

It is a very simple method to detect multicollinearity by inspection of the off-diagonal elements in $\mathbf{X'X}$ matrix, where the data matrix $\mathbf{X}$ is standardized form; that is each of the variable has been centered by subtracting the mean for that variable and dividing by the square root of the corrected sum of squares for that variable. If the variables X_i and X_j are highly correlated ($i \neq j$) and the correlation value is greater than 0.8, it is confirmed that multicollinearity presents among the independent variables.

3.3.2 Variance Inflation Factor

The diagonal elements of the $\mathbf{C} = (\mathbf{X'X})^{-1}$ matrix is very useful in detecting multicollinearity. Suppose C_{jj} is the j^{th} diagonal element of the $\mathbf{C}$, can be written as $C_{jj} = \left(1 - R_j^2\right)^{-1}$ where R_j^2 is the coefficient of determination obtained when X_j is regressed on the remaining (p -1) regressors. If X_j is nearly orthogonal to the remaining regressors, R_j^2 is small and C_{jj} is closed to unity, while if X_j is nearly linearly dependent on some subset on the remaining regressors, R_j^2 is near unity and C_{jj} is large. Since the variance of the j^{th} regression coefficients is $C_{jj}\sigma^2$ which the variance of $\hat{\beta}_j$ is increased due to near linear dependencies among the regressors.

The Variance Inflation Factor (VIF) was developed by Marquardt (1970) to measure the strength of multicollinearity and it was measured by

$$VIF_j = \left(1 - R_j^2\right)^{-1}. \quad (1)$$

The VIF for each term in the model measures the combined effect of the dependencies among the regressors on the variance of that term. One or more large VIFs indicate multicollinearity issues among the independent variables. Practical experience indicates that if any of the VIFs exceeds 5 or 10, it is an indication that the associated regression coefficients are poorly estimated because of multicollinearity.

3.3.3 Conditional Index Number

The eigenvalues of $\mathbf{X'X}$, say $\lambda_1, \lambda_2, \ldots, \lambda_p$, can be used to measure the extent of multicollinearity in the data. If there are one or more linear dependencies in the data, then one or more of the eigenvalues will be small. One or more small eigenvalues imply that there are near linear dependencies among the columns of $\mathbf{X}$. Some analysts prefer to examine the conditional index number of $\mathbf{X'X}$, defined as:

$$K_j = \frac{\lambda_{max}}{\lambda_j}, \quad j = 1, 2, \ldots, p. \tag{2}$$

This is just a measure of the spread in the eigenvalue spectrum of $\mathbf{X'X}$. Generally, if the conditional index number is less than 100, there is no serious problem with multicollinearity. Conditional index numbers between 100 and 1000 imply moderate to strong multicollinearity, and if conditional index number exceeds 1000, severe multicollinearity is indicated.

3.3.4 Variance Decomposition Proportion

Belsley, et. al., (1980) proposed variance decomposition proportion method to detect multicollinearity. A high proportion of the variance for two or more regression coefficients is associated with one singular value, multicollinearity is indicated. If a variance decomposition proportion is greater than 0.5 is recommended for multicollinearity.

3.4 Methods of Dealing with Multicollinearity

Several techniques have been proposed for dealing with multicollinearity. Generally, (i) Collecting additional data (ii) Model re-specification (iii) Use of estimation method other than least squares called biased estimators.

3.4.1 Collecting Additional Data

Collecting additional data has been suggested as the best method of combating multicollinearity (see Farrar and Glauber [1967] and Silvey [1969]). The additional data should be collected in a manner designed to break up the multicollinearity in the existing data. Unfortunately, collecting additional data is not always possible because of economic constrains or because the process being studied is no longer available for sampling. Even when additional data are available, it may be inappropriate to use if the new data extend the range of the regressor variables far beyond the analyst's region of interest. Furthermore, if the new data points are usually or atypical of the process being studied, their presence in the sample could be highly influential on the fitted model. Finally, note that collecting additional data is not a viable solution to the multicollinearity problems when the multicollinearity is due to constraints on the model or in the population.

3.4.2 Model Re-specification

Multicollinearity is often caused by the choice of model, such as when two highly correlated regressors are used in the regression equation. In these situations, some re-specification of the regression equation may lessen the impact of multicollinearity. One approach to model re-specification is to redefine the regressors. For example, if X_1, X_2 and X_3 are near linearly dependent, it may be possible to find some function such as $X = f(X_1, X_2, X_3)$ that preserves the information content in the original regressors but reduces the multicollinearity.

Another widely used approach to model re-specification is variable elimination. That is, if X_1, X_2 and X_3 are near linearly dependent, eliminating one variable may be helpful in combating multicollinearity. Variable elimination is often a highly effective technique. However, it may not provide a satisfactory solution if the variables dropped from the model have significant explanatory power relative to the response variable Y. That is, eliminating variables to reduce multicollinearity may damage the predictive power of the model. Care must be exercised in variable selection because many of the selection procedures are seriously destroyed by multicollinearity, and here is no assurance that the final model will exhibit any lesser degree of multicollinearity than was present in the original data.

3.4.3 Biased Estimators

In this paper two types of based estimators are used (i) Principal Component Estimator and (ii) Ridge Regression Estimator.

3.4.3.1 Principal Component Estimator

Biased estimators of regression coefficients can also be obtained by using a procedure known as principal component regression. Consider the canonical form of the model,

$$\mathbf{y} = \mathbf{Z\alpha} + \in \tag{3}$$

where, $\mathbf{Z} = \mathbf{XT}$, $\boldsymbol{\alpha} = \mathbf{T}'\boldsymbol{\beta}$, $\mathbf{T}'\mathbf{X}'\mathbf{XT} = \mathbf{Z}'\mathbf{Z} = \wedge$, $\mathbf{X}$ is a data matrix.

Recall that $\wedge = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_p)$ is a p x p diagonal of the eigenvalues of $\mathbf{X}'\mathbf{X}$ and $\mathbf{T}$ is a p x p orthogonal matrix whose columns are the eigenvectors associated with $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_p$. The columns of $\mathbf{Z}$, which define a new set of orthogonal regressors, such as

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \ldots \mathbf{Z}_p]$$

are referred to as principal components. The least squares estimator of $\boldsymbol{\alpha}$ **is** $\hat{\boldsymbol{\alpha}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$.

3.4.3.2 Ridge Type Estimators

When the method of least squares is applied to multicollinearity data, very poor estimates of the regression coefficients are usually obtained. The variance of the least squares estimates of the regression coefficients may be considerably inflated and the length of the vector of least squares parameter estimates is too long on the average. This implies that the absolute value of the least squares estimates is too large and that they are very unstable; that is, their magnitudes and signs may change considerably given a different sample.

The least squares estimator has minimum variance in the class of unbiased linear estimators, but there is no guarantee that this variance will be small when multicollinearity present in the independent variables. If multicollinearity presents among the independent variables the ideal method to fit model is biased estimators.

A number of procedures have been developed for obtaining biased estimators of regression coefficients. One of these procedures is ridge regression, originally proposed by Hoerl and Kennard (1970a, b). The ridge estimator is found by solving a slightly modified version of the normal equation. Specifically, it is defined as $\hat{\beta}_R$ and given in equation (4).

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y \tag{4}$$

where $k \geq 0$ is a constant selected by the analyst. The constant k is referred as *biasing parameter.*

Hoerl and Kennard (1970a, b) have suggested that an appropriate value of k may be determined by inspection of the *ridge race.* The ridge race is a plot of the elements of $\hat{\beta}_R$ versus k for the values of k usually in the interval 0-1. Marquardt and Snee (1975) suggested using up to about 25 values of k, spaced approximately logarithmically over the interval [0,1]. If multicollinearity is severe, the instability in the regression coefficients will be obvious from the ridge race. As k is increased, some of the ridge estimates will vary dramatically. At some value of k, the ridge parameters $\hat{\beta}_R$ will stabilize. The objective is to select a reasonably small value of kat which the ridge estimates $\hat{\beta}_R$ are stable. Hopefully, this will produce a set of estimates with smaller mean squared error (MSE) than the least squares. In this research paper the ridge race procedure is used to estimate basing parameter k.

4. Results and Discussion

In this research, to study the multicollinearity issues a real data set is used. This data set was used by Kim (2019) for his research and the same data is used for this study and it consists one dependent variable with six dependent variables and 36 observations (see Appendix-A).

4.1 Ordinary Least Squares Regression Model

Ordinary least square estimation technique is used to fit the regression model for this data and the results are given in Appendix – B. The statistical results in Appendix – B tables B1 and B2, it says that coefficient of determination $R^2 = 0.683$, ANOVA table F = 10.39 and p = 0.000. Hence, it is confirmed that the regression model is significance. Whereas the individual variable test results in table B3 confirmed that out of six independent variables only one variable is significant to the model. This is the indication of multicollinearity among the independent variables.

4.2 Multicollinearity Diagnostics

It is confirmed from the section 4.1 that the data set has multicollinearity issues. So multicollinearity diagnostics methods (i) Analyzing Correlation Matrix, (ii) Variance Inflation Factor (iii) Examination of Tolerance (iv) Conditional Index Number and Conditional Index and (v) Variance Proportion are used to study the multicollinearity issues for this data.

4.2.1 Correlation Matrix of Independent Variables

The correlation table for the independent variables is shown in Appendix – C, from this table it is confirmed that all independent variables are significantly correlated with other independent variables. Furthermore, two correlation values are greater than 0.8 this is an indication that multicollinearity problems present among the independent variables.

4.2.2 Results of Variance Inflation Factor

Variance inflation factor (VIF) is given in Appendix – D. From this table, two VIF values are greater than 5 and one VIF value is closed to 5. If any one of the VIF value is greater than 5 is indicating that independent variables are creating multicollinearity problems. Hence, it is confirmed that the data set used for this research study has multicollinearity problems.

4.2.3 Examination of Conditional Index Number and Conditional Index

The conditional index number (K) and conditional index (CI) values are given in Appendix – E. The highest conditional index number K=2239.466 which greater than 1000 and the highest conditional index CI=47.323 which is greater than 30. These two results reveal that sever multicollinearity exists among the independent variables.

4.2.4 Test Results of Variance Decomposition Proportion

The test results of variance decomposition proportion are shown in Appendix – F. If any one of the variance decomposition proportions is greater than 0.5 is an indication of multicollinearity of problems in the regressors. Accordingly, from the table in Appendix – F, five variance decomposition proportion values are greater than 0.5 and it is confirmed that multicollinearity exists among the independent variables.

4.3 Comparison of Biased Estimators with OLSE

In this research two biased estimators are used to study the multicollinearity problems. Accordingly, Principal Component Analysis and Ridge Regression techniques are used to fit model for the research data.

Sum of squared error is used to study the efficiency of the biased estimators: (i) principal component analysis and (ii) ridge regression model with unbiased estimator ordinary least square estimation (OLSE) model. The comparison results are given in Appendix – G. According to the results given in Appendix – G table, the efficiency of principal component model is two time than OLSE, whereas the efficiency of using ridge regression model is seventy times than OLSE. Hence, it is confirmed that when multicollinearity present among the independent variables using biased estimators are more efficient than OLSE model.

5. Conclusion

In this research study two biased estimators principal component analysis and ridge regression model are used to fit regression models for the multicollinearity data. The efficiency of the unbiased estimator called Ordinary Least Squared (OLS) regression Model was compered with biased estimators namely Principal Component Analysis and Ridge Regression model used in this research were studied by comparing Sum of Squared Error (SSE). It is confirmed that biased estimators are more efficient than OLS estimator when multicollinearity present among the independent data. Furthermore, comparing principal component analysis with ridge regression model for this data ridge regression is more efficient than principal component analysis model.

6. References

Belsley, D, A., Kuh, E., and Welsch, R.E., (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. *Wiley,* New York.

El-Dereny, M., and Rashwan, N.I., (2011). Solving Multicollinearity Problems Using Ridge Regression Models. *International Journal of Contemporary Mathematical Sciences.* 6(12), pp. 585-600.

Farrar, D. E. and Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The problem revisited. *Review Economic Statistics.* 49, pp. 92-107.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics,* 12(1), 55-67.

Hoerl, A. E. and Kennard, R. W. (1970a). Ridge Regression: Biased Estimation for Non-orthogonal Problems. *Technometrics,* 12, 55-67.

Hoerl, A. E. and Kennard, R. W. (1970b). Ridge Regression: Applications to Non-orthogonal Problems. *Technometrics,* 12, 69-82.

Jahufer, A., (2011). Collinearity Affects and It's Analysis in Data. *Journal of Management*, Vol. VII, No.1, pp. 101-113.

Kim, J. K. (2019). Multicollinearity and Misleading Statistical Results. *Korean Journal of Anesthesiology,* pp. 558-569, https://doi.org/10.4097/kja.19087.

Marquardt, D.W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation. *Technometrics*, 12, 591-612.

Marquardt, D.W. and Snee, R. D. (1975). Ridge Regression in practice. *American Statistics*, 29(1), pp. 3-20.

Shariff, N.S.M., and Duzan, H.M.B. (2018). An Application of Proposed Ridge Regression Methods to Real Data Problem. *International Journal of Engineering & Technology.* 7, pp. 106-108.

Silvey, S. D. (1969). Multicollinearity and Imprecise Estimation. J. R. Statist. Soc. Ser. B., 31, pp. 539-552.

Snee, R. (1975). Ridge Regression in Practice. *The American Statistician,* 29(1), pp. 1-20.

Appendix - A: Raw Data from Reference Kim (2019)

Serial number	PVV/GW (cm/s/100 g)	PSV/GW (cm/s/100 g)	EDV/GW (cm/s/100 g)	HVV/GW (cm/s/100 g)	GW/SLV (%)	GRWR (%)	Regeneration rate (%)
1	16.36	8.9	3.47	6.02	57.42	1.11	158.76
2	26.68	21.22	3.53	12.07	61.38	1.36	197.19
3	12.49	16.62	2	8.88	67.42	1.47	144.73

4	8.45	22.86	6.71	7.46	69.94	1.31	140.06
5	10.19	14.23	4.75	2.06	65.68	1.25	129.71
6	19.53	17.35	1.95	7.54	59.63	1.14	162.59
7	20.65	10.48	2.21	4.88	59.42	1.07	178.48
8	22.96	14.23	4.25	3.69	75.08	1.73	120.9
9	21.22	21.64	4.1	11.94	43.42	0.87	191.24
10	8.11	3.16	0.78	8.82	75.12	1.47	150.03
11	24.74	7.84	1.68	3.68	57.65	1.08	173.44
12	11.38	15.71	3.56	7.2	39.93	0.74	211.98
13	15.82	15.04	2.4	9.89	51.27	1.02	193.49
14	8.36	9.01	2.01	3.4	50.52	0.94	164.04
15	12.04	9.72	2.27	6.03	51.6	1.05	156.97
16	10.97	4.58	1.73	5.55	56.63	1.03	208.36
17	7.97	9.33	0.57	4.17	79.09	1.61	154.62
18	7.46	6.11	1.73	2.99	57.2	1.07	137.38
19	29.09	15.71	3.41	9.35	56.44	1.1	180.15
20	10.3	8.54	2.32	10.78	60.43	1.17	228.47
21	7.82	4.41	1.07	4.19	59.52	1	153.62
22	14.71	6.29	1.77	6.16	65.05	1.3	121.31
23	8.54	6.73	1.27	5.52	65.65	1.17	157.37
24	23.05	11.34	5.39	3	33.57	0.63	211.27
25	13.12	5.86	1.89	10.92	52.93	0.9	178.16
26	7.41	9.11	2.05	5.5	53.72	0.91	174.89
27	14.59	5.59	1.26	3.75	58.62	1.14	142.98
28	8.52	6.52	1	6.92	56.61	1.11	165.59
29	18.97	6.35	2.94	5.61	56.41	1.07	141.54
30	35.41	36.36	14.23	15	41.52	0.89	238.22
31	4.55	1.27	3.13	2.83	70.91	1.27	138.42
32	22.59	28.7	10.51	10.35	32.74	0.66	247.45
33	9.21	4.55	1.19	7.92	72.2	1.34	140.27
34	18.32	11.61	2.91	8.07	52.23	1.02	216.06
35	5.69	6.88	1.18	2.78	72.12	1.39	144.18
36	11.2	11.92	3.31	10.29	60.65	1.69	156.22

Appendix – B: Test Results of Ordinary Least Squares estimation

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.826[a]	0.683	0.617	20.36896

a. Predictors: (Constant), X6, X4, X1, X3, X2, X5

Table B2: ANOVA Table

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25865.388	6	4310.898	10.390	0.000[b]
	Residual	12031.944	29	414.895		
	Total	37897.332	35			

a. Dependent Variable: Y

b. Predictors: (Constant), X6, X4, X1, X3, X2, X5

Table B3: Individual Test Results

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	232.436	29.562		7.863	**0.000**
	X1	0.224	0.642	0.050	0.349	0.730
	X2	-0.041	1.026	-0.009	-0.040	0.968
	X3	0.670	2.505	0.055	0.267	0.791
	X4	4.091	1.411	0.397	2.900	**0.007**
	X5	-0.891	0.844	-0.300	-1.055	0.300
	X6	-38.112	32.581	-0.300	-1.170	0.252

a. Dependent Variable: Y

Appendix – C: Results of Correlation values between independent variables

Correlations

		X1	X2	X3	X4	X5	X6
X1	Pearson Correlation	1	.649**	.591**	.456**	-.459**	-.261
	Sig. (2-tailed)		.000	.000	.005	.005	.124
	N	36	36	36	36	36	36
X2	Pearson Correlation	.649**	1	**.841****	.610**	-.442**	-.216
	Sig. (2-tailed)	.000		.000	.000	.007	.205
	N	36	36	36	36	36	36
X3	Pearson Correlation	.591**	.841**	1	.450**	-.504**	-.331*
	Sig. (2-tailed)	.000	.000		.006	.002	.049
	N	36	36	36	36	36	36
X4	Pearson Correlation	.456**	.610**	.450**	1	-.310	-.107
	Sig. (2-tailed)	.005	.000	.006		.066	.534
	N	36	36	36	36	36	36
X5	Pearson Correlation	-.459**	-.442**	-.504**	-.310	1	**.886****
	Sig. (2-tailed)	.005	.007	.002	.066		.000
	N	36	36	36	36	36	36
X6	Pearson Correlation	-.261	-.216	-.331*	-.107	.886**	1
	Sig. (2-tailed)	.124	.205	.049	.534	.000	
	N	36	36	36	36	36	36

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix – D: Results of Variance Inflation Factor (VIF) and Tolerance Value

Model		t	Sig.	Collinearity Statistics	
				Tolerance	**VIF**
1	(Constant)	7.863	0.000		
	X1	0.349	0.730	**0.525**	1.906
	X2	-0.040	0.968	0.202	**4.955**
	X3	0.267	0.791	0.261	3.837
	X4	2.900	0.007	**0.585**	1.710
	X5	-1.055	0.300	0.135	**7.389**
	X6	-1.170	0.252	0.166	**6.018**

Appendix - E: Results of Conditional Index Number and Conditional Index

Eigen Value	6.164	0.555	0.119	0.099	0.043	0.017	0.003
K	1	11.102	51.984	62.142	142.420	358.799	**2239.466**
CI	1.000	3.332	7.210	7.883	11.934	18.943	**47.323**

Appendix – F: Results of Variance Decomposition Proportion

Variance Decomposition Proportion						
(Constant)	X1	X2	X3	X4	X5	X6
0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.02	0.07	0.00	0.00	0.00

0.00	0.09	0.00	0.25	0.45	0.00	0.00
0.00	**0.74**	0.02	0.02	0.24	0.00	0.00
0.01	0.01	**0.87**	**0.54**	0.22	0.00	0.00
0.47	0.09	0.08	0.11	0.04	0.00	0.15
0.52	0.06	0.02	0.00	0.04	**0.99**	**0.84**

Appendix – G: Comparison Results of OLSE, Ridge Regression and Principal Component Analysis

Estimator	**SSE**	**Efficiency comparing with OLSE**	**Model Type**
OLSE	7.8896		Unbiased
Ridge Regression	0.1114	**70.802**	Biased
Principal Component	3.6251	**2.176**	Biased