Science Journal of Applied Mathematics and Statistics
2021; 9(6): 141-153
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20210906.12
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

Effect of Multicollinearity on Variable Selection in Multiple Regression

Harrison Oghenekevwe Etaga, Roseline Chibotu Ndubisi, Ngonadi Lilian Oluebube

Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria

Email address:
ho.etaga@unizik.edu.ng (H. O. Etaga)

To cite this article:
Harrison Oghenekevwe Etaga, Roseline Chibotu Ndubisi, Ngonadi Lilian Oluebube. Effect of Multicollinearity on Variable Selection in Multiple Regression. *Science Journal of Applied Mathematics and Statistics*. Vol. 9, No. 6, 2021, pp. 141-153.
doi: 10.11648/j.sjams.20210906.12

Received: October 25, 2021; **Accepted**: November 17, 2021; **Published**: December 9, 2021

Abstract: When Multicollinearity exists in a data set, the data is considered deficient. Multicollinearity is frequently encountered in observational studies. It creates difficulties when building regression models. It is a phenomenon whereby two or more explanatory variable in a multiple regression model are highly correlated. Variable selection is an important aspect of model building as such the choice of the best subset among many variables to be included in a model is the most difficult part of model building in regression analysis. Data was obtained from Nigerian Stock Exchange Fact Book, Nigerian Stock Exchange Annual Report and Account, CBN Statistical Bulletin and FOS Statistical bulletin from 1987 to 2018. Variance Inflation Factor (VIF) and correlation matrices were used to detect the presence of multicollinearity. Ridge regression and Least Square Regression were applied using R-package, Minitab and SPSS Packages. Ridge Models with constant range of $0.01 \leq K \leq 1.5$ and Least Square Regression models were considered for each value of $P = 2, 3, \ldots, 7$. The optimal Ridge and Least Square model from the Ridge and Least Square Regression models were obtained by taking the average rank of the Coefficient of Determination and Mean Square Error. The result showed that the choices of variable selection were affected by the presence of multicollinearity as different variables were selected under Ridge and Least Square Regression for same level of P.

Keywords: Regression, Multicollinearity, Ridge Regression, Partial Least Square, Extra Sum of Squares

1. Introduction

In Regression, Modeling of the correlated binary responses may suffer from some problems in modeling like collinearity Dormann et al [6]. The term Collinearity is often referred to a situation where there is high linear relationship between predictor variables. Gujarati and Porter [10] observed that the inherent relationship between the predictor variables in the real world, as well as small sample size, design of model, and the trend of predictor variables can cause collinearity. When Multicollinearity exist in a data set, the data is considered deficient. Multicollinearity is frequently encountered in observational studies. It creates difficulties when building regression models. It is a phenomenon whereby two or more explanatory variable in a multiple regression model are highly correlated. Variable selection is an important aspect of model building as such the choice of the best subset among many variables to be included in a model is the most difficult part of model building in regression analysis. Multicollinearity, which exists when two or more explanatory variables in a regression model are highly correlated, is a frequently encountered problem in multiple regression analysis [Farrar and Glauber [7], Gunst and Webster [11], Mansfield and Helms [15]. The interrelationship among explanatory variables obscures their relationship with the explained variable, which in turns leads to computational instability in estimation of the parameters of the model. The reliability of the regression analysis is decreased in the presence of multicollinearity by the low quality of the resultant estimates. Several approaches can be used to avoid the deleterious effects of multicollinearity Chatterjee and Hadi [5]. Ridge Regression method was proposed by Hoerl and Kennard [13] to solve the problem Least Methods in estimating regression coefficient when multicollinearity exist. The method involves choosing a constant to achieve a satisfactory balance between bias and variance. Ridge estimators are biased but with smaller Mean Square Error.

Multicollinearity can also to viewed a condition of close to linear relationship among two or more explanatory variables. Multicollinearity must not be exact before it causes a problem in regression analysis and there are other approaches to solving the problem of multicollinearity. One of such approach is orthogonal transformation through procedures such as principal component regression Jolliffe [14], Massy [16], and Partial Least Squares Regression Vandenberghe and Boyd [20], Wold [21]. In the approach, a set of correlated variables is transformed into a set of linearly uncorrelated variables (i.e., principal components) for use in a regression model. Orthogonal transformation can enhance the computational stability of model estimation but often leads to worse predictive performance and results that are strongly influenced by the presence of outliers Frank and Friedman [8], Hadi and Ling [12]. Another approach is penalized regression, such as ridge regression, Hoerl and Kennard [13], lasso, Tibshirani [19], and elastic net Zou [22]. This approach introduces a penalty function to shrink regression coe cient estimates toward zero. Penalized regression helps prevent regression models from overfitting noisy datasets and, accordingly, is effective for achieving high predictive performance. However, the penalty functions produce biased estimates, which are undesirable from the standpoint of model interpretation Bertsimas and King [3], Bartsimas et al [4].

A promising technique called the lasso was proposed by Tibshirani [19]. The lasso is a penalized least squares method imposing an L1-penalty on the regression coefficients. Owing to the nature of the L1-penalty, the lasso does both continuous shrinkage and automatic variable selection simultaneously. Tibshirani [19] and Fu [9] compared the prediction performance of the lasso, ridge and bridge regression, Frank and Friedman, [8] and found that none of them uniformly dominates the other two. However, as variable selection becomes increasingly important in modern data analysis, the lasso is much more appealing owing to its sparse representation. Although the lasso has shown success in many situations, it has some limitations. Consider the following three scenarios. (a) In the p>n case, the lasso selects at most n variables before it saturates, because of the nature of the convex optimization problem. This seems to be a limiting feature for a variable selection method. Moreover, the lasso is not well defined unless the bound on the L1-norm of the coefficients is smaller than a certain value. (b) If there is a group of variables among which the pairwise correlations are very high, then the lasso tends to select only one variable from the group and does not care which one is selected.

According to Alin [1], multicollinearity is a situation where two or more explanatory variables are highly related. Meloun, et al [17], Murray et al [18] and Brue [2] have worked on multicollinearity.

Table 1. *Complied annual reports from various source.*

YEAR	REAL GDP	GMC	ASI	VALT	TLNSE	OOTE	TMT	TNI
1987	315458.100	4464.200	88.000	388.700	157.000	0.047	0.230	423.500
1988	205222.100	4979.800	87.000	304.800	194.000	0.062	0.190	455.200
1989	199688.200	4025.700	94.000	214.800	205.000	0.077	0.210	533.400
1990	185598.100	5768.000	111.000	397.900	212.000	0.057	0.260	448.500
1991	183563.000	5514.900	100.000	418.200	213.000	0.099	0.250	159.800
1992	201036.300	6670.700	127.300	319.600	220.000	0.093	0.310	817.200
1993	205971.400	6794.800	163.800	494.400	240.000	0.072	0.490	833.000
1994	204806.500	8297.600	190.900	348.000	244.000	0.235	0.290	450.700
1995	219876.800	10020.800	233.600	137.600	253.000	0.239	0.250	400.000
1996	263729.600	12848.600	325.300	521.600	267.000	0.375	0.650	1629.900
1997	267660.000	16358.400	513.800	265.500	295.000	0.582	0.310	9964.500
1998	265379.100	23125.000	783.000	136.000	239.000	0.795	0.230	1870.000
1999	274833.300	31272.600	1107.600	313.500	251.000	1.285	0.490	3306.300
2000	275450.600	47436.100	1548.800	402.300	272.000	1.399	0.660	2636.900
2001	281407.400	663680.000	2205.000	569.700	276.000	1.339	0.990	2161.700
2002	293745.400	180305.100	5092.200	1838.800	276.000	6.373	1.840	4425.600
2003	302022.500	281815.800	6992.100	7062.700	276.000	6.373	7.060	5858.200
2004	310890.100	281887.200	6440.500	11072.700	264.000	6.911	11.070	10875.700
2005	312183.500	262517.300	5716.000	13572.300	264.000	5.112	13.500	15018.100
2006	329978.700	300041.100	5266.400	14027.400	268.000	6.571	14.100	12038.500
2007	356994.300	427290.000	8111.000	28154.600	260.000	8.903	28.150	17207.800
2008	433203.500	662561.300	10965.000	57637.200	261.000	9.037	57.680	37198.800
2009	477833.000	764975.800	12137.700	60088.600	258.000	7.518	59.410	61284.000
2010	527576.000	1359274.200	21222.600	120703.000	277.000	10.823	120.400	180079.900
2011	561931.400	2112549.600	23844.500	225820.600	288.000	12.491	225.800	195418.400
2012	595821.600	2900062.100	24085.800	470257.000	294.000	17.880	262.940	552782.000
2013	634251.000	5120000.000	33189.300	1076020.400	310.000	18.020	470.250	707400.000
2014	674889.000	13294059.000	57990.200	1679143.700	301.000	19.721	2086.290	1935080.000
2015	716949.700	9562970.000	31450.800	68572000.000	266.000	23.257	2379.140	1509230.000
2016	801700.000	9920000.000	46437.640	79755000.000	264.000	23.734	2388.340	1894374.500
2017	901300.000	10280000.000	59365.750	63492000.000	250.000	25.224	2511.670	1735623.340
2018	1067650.000	89000000.000	64768.550	62758000.000	198.000	27.555	2676.240	1843274.870

Source: The Nigerian stock exchange annual reports and account, various years; sec annual reports and accounts; CBN statistical bulletin and FOS Statistical bulletin from 1987 to 2018.

2. Methods

2.1. Methodology

The data was obtained from the annual reports in Nigerian Economy and Nigerian Stock Market through the Nigeria Stock Exchange Fact books, Nigerian Stock Exchange Annual Reports and Accounts, CBN Statistical bulletins, FOS Statistical bulletin from 1987 to 2018.

2.2. Model Specification

The research model is specified thus:

GDP= f (GMC, ASI, TLNSE, TNI, OOTE, VALTRANS, TMT),

where; GDP = Gross Domestic Product. GMC = Growth of Market Capitalization. ASI = All-Share Index,

TLNSE = Total Listing on the Nigerian Stock Exchange. TNI = Total New Issues,

OOTE = Openness of Nigeria Trade Economy, VALTRANS = Value of Transactions,

TMT = Total Market Turnover.

The Coefficient of Multiple Determinations is the proportion of variation in the dependent variable (response variable) that can be explained by the model. It can be expressed mathematically as follows.

$$R^2 = 1- \frac{SSE}{SST} \; OR \; \frac{SSR}{SST} \text{ since } SST = SSR + SSE$$

where SSE is the unexplained variation, SSR is the explained variation and SST is the total variation in the response variable.

***Table 2.** Analysis of Variance for Partial Least Square Table Showing Test of Model Adequacy.*

SOURCE OF VARIANCE	DEGREE OF FREEDOM	SUM OF SQUARES	MEAN SUM OF SQUARE	F-RATIO
REGRESSION	K-1	SSR	MSR	$\frac{MSR}{MSE}$
RESIDUAL	n-(K + 1)	SSE	MSE	
TOTAL	n -1	SST	MST	

Test of hypothesis

H_0: The model is not adequate

H_1: The model is adequate using a 5% level of significance

Test statistic

$$F_{cal} = \frac{MSR}{MSE} \sim F^{(a)}_{k,n-(k+1)}$$

Decision rule

Reject $H_o: If\ F_{cal} > F_{tab,}$ accept if otherwise

2.3. Criteria for Evaluating the Optimal Regression Model

Firstly, in obtaining the optimal ridge regression models for each value of p (where p is the number of parameters p = 2,3,...,8. The selection will be based on Mean Square Error and explained variation (R^2).

Secondly, obtaining the best optimal ridge regression model from the optimal ridge regression models gotten from one above, we use only the coefficient of determination. It is unfair to judge them by their Mean Square Error since the mean square error is observed at different values of p.

Lastly, the best model for the estimation of the data will be obtained by comparing the optimal ridge regression model and the least square regression model obtained by;

1) Explained variation of response variable (Coefficient of multiple determination) (R^2).
2) Precision of the regression coefficients (average standardized mean square error of the coefficient).

2.4. Multicollinearity Diagnostics

Several techniques have been proposed for detecting multicollinearity, but three techniques will be considered. We have

1) Principal Component Analysis,
2) The evaluation of the correlation matrix and,
3) The variance inflation factor was used to account for the effect of multicollinearity on the various subset regression models.

3. Results

Minitab Package, R -Package, and SPSS Package were used to compute for the Least Squares Method in the Parameters, Ridge Regression Methods in selecting variables and Concept of Extra Sum of Square in Partial F-ratio and the results of all the computations were tabulated below.

***Table 3.** The Evaluation of correlation results.*

CORRELATION	GMC	ASI	VALT	TLNSE	OOTE	TMT	TNI
GMC	1.000	0.660	0.565	-0.211	0.586	0.650	0.614
ASI	0.660	1.000	0.726	0.228	0.949	0.911	0.946
VALT	0.567	0.726	1.000	-0.075	0.757	0.902	0.833
TLNSE	-0.211	0.228	-0.075	1.000	0.319	0.049	0.139
OOTE	0.586	0.949	0.757	0.319	1.000	0.868	0.895
TMT	0.650	0.911	0.902	0.049	0.868	1.000	0.981
TNI	0.614	0.946	0.833	0.139	0.895	0.981	1.000

The Evaluation of correlation matrix results among the variable of Annual report of the Nigerian Stock Exchange; The variables studied included: - Growth of Market Capitalization (GMC), All Share Index (ASI), Value of Transactions (VALT), Total Listing on the Nigeria Stock Exchange (TLNSE), Openness of Nigeria Trade Economy (OOTE), Total Market Turnover (TMT) and Total New Issues.

The result of Correlation table above shows high Correlation between most of the Independent variables.

3.1. Testing the Hypothesis for Regression Results

Table 4. *Analysis of Variance results for Real GDP in Regression (with α =0.05).*

SOURCES OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F –VALUES	P-VALUES
REGRESSION	7	1.61049E+12	2.30071E +11		
RESIDUAL ERROR	24	3.52988E +10	1.47078E + 09	156.43	0.000
TOTAL	31	1.64579E +12			

REAL GDP = 232231 + 0.001207GMC + 9.07ASI + 0.00269VALT – 42TLNSE + 112370OOTE -53.6TMT – 0. 1004TNI.

To test H_0: $\beta_1 = \beta_2 = 0$, we calculate the statistic

$$F^* = \frac{MSR}{MSE} = \frac{2.30071E+11}{1.47078E+09} = 156.43$$

Since $F_{cal} = 156.43 > F_{tab=0.05,7,24} = 2.42$ (or since the P– Value is considerably smaller than α= 0.05), the null hypothesis was rejected and conclude that Real GDP is linearly related to either GMC, ASI, VALT, TLNSE, OOTE, TMT, AND TNI.

However, note that this does not necessarily imply that the relationship found is an appropriate model for predicting Real GDP as a function of other variables. (GMC, ASI, VALT, TLNSE, OOTE, TMT, and TNI).

Table 5. *Coefficients of Multicollinearity.*

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	232231	66928	3.47	0.002	
GMC	0.001207	0.000712	1.70	0.103	2.67
ASI	9.07	2.10	4.32	0.000	34.88
VALT	0.00269	0.00109	2.47	0.021	13.34
TLNSE	-42	282	-0.15	0.882	1.99
OOTE	11237	3353	3.35	0.003	18.30
TMT	-53.6	73.1	-0.73	0.470	87.23
TNI	-0.1004	0.0864	-1.16	0.257	67.56

From Table 5, it can be observed that only the constant, ASI, VALT and Openness of Nigeria trade economy OOTE, TMT and TNI being the only significant variables in the model. Even though the model is adequate. As predicted from the collinearity column, it can be observed that in Variance Inflation Factors Results (VIFs), 5 of the independent variable's VIF's exceeded 10, which indicate very strong presence of multicollinearity or a good indication that Multicollinearity is present.

The results the Least Square Method are shown in Table 6.

Table 6. *Results of Least Square.*

Model	Unstandardized coefficient		Standardized coefficients Beta	T	P-values or significance	95% Confidence Interval for B		Collinearity statistics
	B	Std. Error				lower Boundary	Upper boundary	VIF
CONSTANT	232231	66928	-	3.47	0.002	94099	370363	-
GMC	0.001	0.001	0.083	1.70	0.103	0	0.003	2.67
ASI	9.07	2.10	0.763	4.32	0.000	4.735	13.398	34.88
VALT	0.003	0.001	0.270	2.47	0.021	0	0.005	13.34
TLNSE	-42	282	-0.006	-0.15	0.882	-624.872	540.14	1.99
OOTE	11237	3353	0.429	3.35	0.003	4317.6	18156.84	18.30
TMT	-54.6	73.1	-0.205	-0.73	0.470	-204.506	97.241	87.23
TNI	-0.1	0.086	-0.285	-1.16	0.257	-0.279	0.078	67.56

3.2. Selection of Optimal Models for Ridge Regression for Each Value of p=2…8

Having establish the presence of Multicollinearity in the data set, the Ridge and Partial Least Square Regression was applied at various level of p (number of independent variables considered). The result was then ranked using R-square and MSE, the average of the ranks was then taken to guide the selection of the best model.

Table 7. *Ridge Regression Model for P = 2 and ranks.*

C_2^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS		R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
				1	2					
1	21995	25094150	ASI,OOTE	48695880	11231250	0.956	455000000	1	1	1
2	21995	258494.4463	OOTE,TMT	14292.6916	83.654	0.9383213	4035000000	3	2	2.5
3	21995	296054.9884	TLNSE,OOTEE	-210.7256	20987.6936	0.9412349	4428000000	2	5	3.5
4	21995	25929400	OOTE,TNI	13628910	11630220	0.9354083	4369000000	4	3	3.5
5	21995	25500490	VALT,OOTE	23322600	16673320	0.9353928	4423000000	5	4	4.5
6	21820	28338970	ASI,VALT	70797750	26243500	0.9308217	4647000000	6	6	6
7	21820	22109940	ASI,TLNSE	91194260	2274350	0.9183316	6282000000	8	8	8
8	21820	28739890	ASI,TMT	59431970	81966860	0.913353	5878000000	9	7	8
9	21820	28395910	ASI,TNI	59381800	11207080	0.9085469	6886000000	10	9	9.5
10	21820	37424060	GMC,ASI	12800250	15879550	0.9218364	18020000000	7	15	11
11	19939	119147.0848	TLNSE,TMT	858.3347	155.576	0.798443	13930000000	12	11	11.5
12	20272	18073550	TLNSE,TNI	60130820	20353700	0.8026522	14120000000	11	13	12
13	38880	33171580	VALT,TNI	25907710	14079150	0.7852588	13660000000	14	10	12
14	26799	32261610	TMT,TNI	87612250	12639700	0.792549	13940000000	13	12	12.5
15	19939	32855480	VALT,TMT	25140310	12275610	0.7609195	15010000000	15	14	14.5
16	27937	12685070	VALT,TLNSE	43549740	93395570	0.6657331	23420000000	16	17	16.5
17	520078	36369280	GMC,OOTE	14781520	40960780	0.5883149	21930000000	17	16	16.5
18	577360	37801540	GMC,TNI	14757680	49737190	0.4958349	24510000000	18	18	18
19	623243	37940140	GMC,TMT	14854310	36679730	0.4623827	26220000000	19	19	19
20	724965	38683970	GMC,VALT	12773510	10104010	0.3742466	28110000000	20	20	20
21	531214	33074160	GMC,TLNSE	15694370	25061990	0.276719	34450000000	21	21	21

Table 8. *Least Square Regression Model for P = 2 and ranks.*

C_2^7	INTERCEPT	VARIABLES	PARAMETERS		R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
			1	2					
1	217592.332	GMC,OOTE	0.003	22746.16	0.961	2231501763	1	1	1
2	220913.494	ASI,OOTE	4.998	14968.783	0.958	2361918160	2	2	2
3	377544.745	TLNSE,OOTE	-690.464	26289.632	0.95	2825482498	3	3	3
4	218537.173	OOTE,TMT	22009.478	39.296	0.946	3052942849	4	4	4
5	214859.086	VALT,OOTE	0.001	23609.309	0.944	3164414215	5.5	5	5.25
6	216708.972	OOTE,TNI	22398.599	0.045	0.944	3179509266	5.5	6	5.75
7	252276.729	ASI,VALT	10.066	0.002	0.938	3538587608	7	7	7
8	250852.244	GMC,ASI	0.001	10.767	0.93	3981128641	8	8	8
9	242421.106	ASI,TNI	13.18	-0.55	0.928	4071453793	9	9	9
10	248223.63	ASI,TMT	11.246	4.606	0.926	4211918172	10.5	10	10.25
11	251400.861	ASI,TLNSE	11.442	-15.411	0.926	4214629927	10.5	11	10.75
12	297093.724	GMC,TNI	0.003	0.268	0.828	9775017887	12	12	12
13	136477.057	TLNSE,TNI	635.091	0.31	0.808	10903394732	13	13	13
14	9684.429	TLNSE,TMT	1171.215	227.981	0.804	11150343492	14	14	14
15	297065.315	VALT,TNI	0.001	0.286	0.802	11243874615	15	15	15
16	296387.859	TMT,TNI	17.192	0.292	0.799	11394451767	16	16	16
17	306414.153	GMC,TMT	0.003	197.459	0.794	11671410411	17	17	17
18	304514.73	VALT,TMT	-0.001	254.142	0.775	12772154418	18	18	18
19	328017.143	GMC,VALT	0.005	0.006	0.687	17787436868	19	19	19
20	-136875.627	VALT,TLNSE	0.008	1852.35	0.675	18467980497	20	20	20
21	-292368.37	GMC,TLNSE	0.011	2534.772	0.603	22527885642	21	21	21

From the results in the Table 7, and Table 8 shows that for combination of two regressors out of seven regressors (p= 2), the optima model for value p=2 was selected by ranking the Coefficient of Determination in descending order (highest coefficient of determination has the lowest average ranking and also ranking the Mean Square Error of the standardised value of regression coefficient in ascending order (lowest Mean Square Error has the lowest rank), then the model with the lowest average rank was selected. The optima model with the best subset of Ridge regressors result for p = 2 were ASI, OOTE with the lowest average rank of 1, also the optima model with the best subset of Least Square regressors result for p = 2 were GMC, OOTE with the lowest average rank of 1.

Table 9. *Least Square Regression Model for P = 3 and ranks.*

C_3^7	INTERCEPT	VARIABLES	PARAMETERS			R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
			1	2	3					
1	223470.496	GMC,ASI,OOTE	0.002	3.406	16315.848	0.968	1907056726	1	3	2
2	236465.939	ASI,VALT,TNI	15.692	0.004	-0.231	0.963	2145840442	3.5	4	3.75
3	236568.756	ASI,VALT,TMT	14.73	0.005	-189.697	0.963	2168072114	3.5	5	4.25
4	273881.512	GMC,TLNSE,OOTE	0.002	-234.908	23366.547	0.961	2263755286	8.5	6	7.25
5	219971.168	GMC,VALT,OOTE	0.002	0	22005.896	0.961	2266171353	8.5	7	7.75
6	226102.292	ASI,VALT,OOTE	4.923	0.001	13454.803	0.961	2268917290	8.5	8	8.25
7	220573.317	GMC,OOTE,TMT	0.002	21647.716	14.783	0.961	2270886850	8.5	9	8.75
8	220410.196	GMC,OOTE,TNI	0.002	21606.872	0.019	0.961	2280082609	8.5	10	9.25
9	216283.053	ASI,OOTE,TNI	6.647	14888.613	-0.051	0.961	2319011142	8.5	11	9.75
10	265323.24	ASI,TLNSE,OOTE	4.807	1-49.524	11101.664	0.932	1017225555	22.5	1	11.75
11	220996.812	ASI,OOTE,TMT	4.976	14966.193	0.561	0.958	2446226367	12	12	12
12	352500.274	TLNSE,OOTE,TMT	-569.43	24523.424	18.571	0.951	2870968702	13.5	13	13.25
13	268281.226	ASI,TLNSE,TNI	13.353	-105.784	0.059	0.964	4203785643	2	25	13.5
14	365071.459	TLNSE,OOTE,TNI	-624.724	24800.24	2748.312	0.951	2889526603	14	14	14
15	367164.379	VALT,TLNSE,OOTE	0	-642.644	25827.85	0.95	2918813314	15	15	15
16	217389.162	OOTE,TMT,TNI	22755.509	74.149	-0.057	0.947	3115710281	16	16	16
17	85191.926	TLNSE,TMT,TNI	849.113	93.598	0.185	0.812	1106561425	31	2	16.5
18	218549.013	VALT,OOTE,TMT	0	22045.477	35.864	0.946	3160081008	17	17	17
19	217794.51	VALT,OOTE,TNI	0.001	22299.17	0.028	0.945	3223126440	18	18	18
20	241805.588	ASI,TMT,TNI	14.361	159.531	-0.298	0.941	3446858873	19	19	19
21	254332.161	GMC,ASI,VALT	0.001	9.676	0.001	0.94	3538772069	20	20	20
22	246969.759	ASI,TLNSE,TMT	11.236	5.16	4.794	0.962	43623155915	5	35	20
23	180559.974	ASI,VALT,TLNSE	9.757	0.002	291.72	0.939	3572881269	21	21	21
24	245851.354	GMC,ASI,TNI	0.001	12.445	-0.052	0.932	3989093752	22.5	22	22.25
25	185237.353	GMC,ASI,TLNSE	0.002	10.457	268.45	0.931	4055673694	24	23	23.5
26	250600.649	GMC,ASI,TMT	0.001	10.843	-1.976	0.93	4122755329	25	24	24.5
27	-16451.051	GMC,TLNSE,TNI	0.004	1250.679	0.24	0.857	8404221437	26	26	26
28	-111427.676	GMC,TLNSE,TMT	0.004	1654.799	176.594	0.849	8876538238	27	27	27
29	295632.201	GMC,TMT,TNI	0.003	-59.376	0.343	0.83	10020185136	28	28	28
30	297751.131	GMC,VALT,TNI	0.003	0.001	0.252	0.829	10063987561	29	29	29
31	81109.135	VALT,TLNSE,TNI	0.002	863.852	0.257	0.816	10802178551	30	30	30
32	-274826.433	GMC,VALT,TLNSE	0.007	0.005	2362.495	0.805	11459147086	32	31	31.5
33	295670.212	VALT,TMT,TNI	0.002	-101.445	0.389	0.804	11516987976	33.5	32	32.75
34	7776.653	VALT,TLNSE,TMT	0	1179.229	225.589	0.804	11547538828	33.5	33	33.25
35	305453.816	GMC,VALT,TMT	0.003	-0.001	217.13	0.796	12015737689	35	34	34.5

Table 10. *Ridge Regression Model for P = 3 and ranks.*

C_3^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS			R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
				1	2	3					
1	21995	23754180	GMC,ASI,OOTE	20475960	44207060	12552160	0.9643892	2,387,000,000	1	1	1
2	21995	25426840	ASI,VALT,OOTE	43519030	16956770	97640050	0.9587	2,837,000,000	3	2	2.5
3	21995	273069.6652	ASI,TLNSE,OOTE	4982272	-106571388	11676.9455	0.9593331	3,323,000,000	2	4	3
4	21995	25837340	ASI,OOTE,TMT	37689200	92781040	53262000	0.9533765	3,323,000,000	4	3	3.5
5	21995	25941260	ASI,OOTE,TNI	36612280	90636500	72096720	0.9519391	3,711,000,000	7	5	6
6	21995	26017610	VALT,OOTE,TNI	14213760	12947380	92216880	0.9352838	4,473,000,000	12	6	9
7	21995	26047380	VALT,OOTE,TMT	12040580	13638590	65633590	0.9370336	4,593,000,000	10	8	9
8	21995	26054290	TLNSE,OOTE,TNI	-30602810	14599040	11381640	0.9382839	4,659,000,000	9	9	9
9	21995	221404.5713	TLNSE,OOTE,TMT	142.7448	14361.1711	84.5697	0.9391039	4,849,000,000	8	10	9
10	21995	25607900	OOTE,TMT,TNI	13177760	51892280	71714860	0.9350494	4,551,000,000	13	7	10
11	21995	20677990	VALT,TLNSE,OOTE	24330050	19748990	16313890	0.9366238	5,428,000,000	11	12	11.5
12	21995	35170370	GMC,OOTE,TMT	13834940	39022730	33377050	0.9525565	11,560,000,000	5	21	13
13	21820	14235570	ASI,VALT,TLNSE	70747890	27813210	55115770	0.9327874	5,581,000,000	14	14	14
14	21820	28978200	ASI,VALT,TNI	50021240	18900550	8328866	0.9244826	5,525,000,000	16	13	14.5
15	21820	28559070	ASI,VALT,TMT	57844570	17281530	55330600	0.9259539	5,703,000,000	15	15	15
16	21995	20474570	TLNSE,TMT,TNI	52052480	73785620	10154450	0.8064782	5,138,000,000	24	11	17.5
17	21820	27162110	GMC,ASI,TMT	18079800	72894010	56306780	0.9174266	6,700,000,000	18	17	17.5
18	21820	17764210	ASI,TLNSE,TMT	54510480	45550140	84123920	0.9135299	6,688,000,000	20	16	18
19	21820	28781200	ASI,TMT,TNI	48084380	56290180	76102480	0.908765	6,719,000,000	21	18	19.5
20	21820	27251320	GMC,ASI,TNI	20784140	68222870	82167560	0.9142586	9,346,000,000	19	20	19.5

C_3^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS			R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
				1	2	3					
21	21820	20873110	ASI,TLNSE,TNI	53971270	32438050	11287680	0.908308	7,901,000,000	22	19	20.5
22	21995	25117640	GMC,OOTE,TNI	25811740	14351550	88732850	0.9522295	72,740,000,000	6	35	20.5
23	21820	32273250	GMC,ASI,TLNSE	12954610	15745610	20359310	0.9232097	18,420,000,000	17	27	22
24	32279	16334430	VALT,TLNSE,TNI	26105900	68182880	12689390	0.807059	14,060,000,000	23	23	23
25	61908	33038720	VALT,TMT,TNI	17210050	61209590	91395030	0.7807585	13,510,000,000	26	22	24
26	19939	18842970	VALT,TLNSE,TMT	22360240	63847300	76317500	0.7942943	17,750,000,000	25	24	24.5
27	473876	34875300	GMC,VALT,OOTE	15973720	12396990	45724930	0.6903968	17,940,000,000	29	25	27
28	362599	35931040	GMC,TMT,TNI	16349580	39217040	54643810	0.6906401	17,940,000,000	28	26	27
29	526069	37118540	GMC,VALT,TNI	13502410	10404470	4484838	0.7306	20,420,000,000	27	28	27.5
30	682031	36806090	GMC,ASI,VALT	11785180	14758880	93768320	0.5939074	20,550,000,000	31	29	30
31	473876	30229200	GMC,TLNSE,OOTE	16004540	23256260	43437360	0.6236711	20,970,000,000	30	30	30
32	623243	37433750	GMC,VALT,TMT	12797260	98158000	30611810	0.5457432	22,110,000,000	33	31	32
33	479335	30796070	GMC,TLNSE,TNI	16008070	26972770	52995110	0.5643404	22,120,000,000	32	32	32
34	517427	30409390	GMC,TLNSE,TMT	16111130	29058650	39214880	0.5354794	23,730,000,000	34	33	33.5
35	601879	32231800	GMC,VALT,TLNSE	13902010	10915390	24965580	0.4493422	26,790,000,000	35	34	34.5

From the results in the Table 9, and Table 10 shows that for combination of three regressors out of seven regressors (p = 3), the optima model for value p =3 was selected by ranking the Coefficient of Determination in descending order (highest coefficient of determination has the lowest average ranking and also ranking the Mean Square Error of the standardised value of regression coefficient in ascending order (lowest Mean Square Error has the lowest rank), then the model with the lowest average rank was selected. The optima model with the best subset of Ridge regressors result for p = 3 were GMC, OOTE, TMT with the lowest average rank of 1, also the optima model with the best subset of Least Square regressors result for p = 3 were GMC, ASI, OOTE with the lowest average rank of 2.

Table 11. *Ridge Regression Model for P = 4 and ranks.*

C_4^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS			
				1	2	3	4
1	21995	24298720	GMC,ASI,VALT,OOTE	17822700	41793110	11255850	11130450
2	24140	24485760	GMC,ASI,OOTE,TMT	1848660	37703010	11128670	31552470
3	21995	22237870	GMC,ASI,TLNSE,OOTE	20915090	44028260	58660900	12595610
4	26493	24729850	GMC,ASI,OOTE,TNI	19932870	36021900	10657720	47222570
5	21995	21745340	ASI,VALT,TLNSE,OOTE	43926610	17384010	14313280	97119820
6	21995	26616110	ASI,VALT,OOTE,TNI	33221240	13656670	78688140	57152210
7	21995	23180120	ASI,TLNSE,OOTE,TMT	38059970	10025630	93542300	53453530
8	21995	25799550	ASI,OOTE,TMT,TNI	33647060	87577540	38373170	47529100
9	21995	25777890	ASI,TLNSE,OOTE,TNI	38964990	-24458990	99236010	66608510
10	29077	26443670	ASI,VALT,OOTE,TMT	35046040	12407650	82487820	40038030
11	21995	19239640	VALT,TLNSE,OOTE,TMT	12686420	2699470	13474370	66399510
12	21995	12584280	TLNSE,OOTE,TMT,TNI	13440490	12584280	54139040	73017020
13	21995	17205960	GMC,TLNSE,OOTE,TNI	28713870	32544250	13668860	90034890
14	21995	26344070	VALT,OOTE,TMT,TNI	95617920	11744500	43089130	68247190
15	21995	25298760	GMC,OOTE,TMT,TNI	23243400	13609870	35239590	5990000
16	21995	14204480	GMC,TLNSE,OOTE,TNI	24180340	99156990	13151040	75930660
17	21995	20168260	VALT,TLNSE,OOTE,TNI	16779180	27409740	11188100	92480390
18	21820	65739870	GMC,ASI,VALT,TLNSE	25033860	691638700	2346338000	831154200
19	21820	14469280	ASI,VALT,TLNSE,TMT	49907880	19041210	58621260	58760610
20	21995	34745320	GMC,VALT,OOTE,TMT	12773730	95187860	13151040	29865930
21	21995	14204480	GMC,TLNSE,OOTE,TMT	28210860	46429380	13315680	67135580
22	21995	30654920	GMC,VALT,TLNSE,OOTE	12856060	98908290	21376020	34664280
23	21820	95756300	GMC,ASI,TLNSE,TMT	25408660	61959410	72302830	65662400
24	21820	17924610	ASI,VALT,TLNSE,TNI	41141560	19794850	48188670	81911780
25	21820	27469990	GMC,ASI,OOTE,TNI	19932870	62644800	38150950	53420140
26	21820	12140980	GMC,ASI,TLNSE,TNI	26904360	62432560	61247440	84533260
27	21820	30037890	ASI,VALT,TMT,TNI	36164390	14093680	43890280	65242190
28	21820	19979470	ASI,TLNSE,TMT,TNI	37215570	40413590	57020860	7575310
29	21820	27412070	GMC,ASI,VALT,TNI	17847310	65763690	16235780	49208140
30	118735	24151830	VALT,TLNSE,TMT,TNI	14697110	41583310	46950900	66264640
31	301036	25099610	GMC,TLNSE,TMT,TNI	19430550	40063330	45356570	6287967

C_4^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS			
				1	2	3	4
32	397952	28224720	GMC,VALT,TLNSE,TNI	16414370	12443450	32704450	53503520
33	526069	35866160	GMC,VALT,TMT,TNI	13569340	10009880	31323090	44247250
34	621441	35568510	GMC,ASI,VALT,TMT	11964660	15241190	92257580	28289390
35	471460	28642950	GMC,VALT,TLNSE,TMT	15617990	11772200	32477230	36775710

Table 11. *Continue.*

C_4^7	R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
1	0.9652346	2.31E+09	1	1	1
2	0.9611727	2.55E+09	3	2	2.5
3	0.9641735	3.34E+09	2	5	3.5
4	0.9595736	2.89E+09	4	3	3.5
5	0.9590909	3.54E+09	5	6	5.5
6	0.9577315	3.68E+09	6	8	7
7	0.9551898	3.66E+09	7	7	7
8	0.9524906	2.90E+09	10	4	7
9	0.9550075	3.91E+09	8	9	8.5
10	0.9542643	3.93E+09	9	10	9.5
11	0.9370903	5.32E+09	18	13	15.5
12	0.9360772	4.89E+09	20	11	15.5
13	0.9513864	5.96E+09	15	17	16
14	0.9346509	5.14E+09	21	12	16.5
15	0.9506013	6.04E+09	16	18	17
16	0.9515659	6.90E+09	13	22	17.5
17	0.9361718	5.63E+09	19	16	17.5
18	0.9391149	6.32E+09	17	20	18.5
19	0.9274244	5.54E+09	22	15	18.5
20	0.9517943	1.11E+10	11	27	19
21	0.9514518	7.85E+09	14	25	19.5
22	0.9515671	1.42E+10	12	28	20
23	0.921258	5.35E+09	26	14	20
24	0.9269285	7.08E+09	23	23	23
25	0.9135132	6.07E+09	28	19	23.5
26	0.9172913	6.47E+09	27	21	24
27	0.9231697	7.21E+09	25	24	24.5
28	0.9091006	7.92E+09	29	26	27.5
29	0.9268593	6.546+09	24	35	29.5
30	0.7867046	1.57E+10	30	30	30
31	0.7453154	1.44E+10	31	29	30
32	0.6854422	1.76E+10	32	32	32
33	0.665233	1.72E+10	34	31	32.5
34	0.6830887	1.77E+10	33	33	33
35	0.6357043	1.89E+10	35	34	34.5

Table 12. *Least Square Regression Model for P = 4 and ranks.*

C_4^7	INTERCEPT	VARIABLES	PARAMETERS			
			1	2	3	4
1	220550.46	ASI,VALT,OOTE,TNI	10.593	0.003	10016.591	-0.179
2	219197.513	GMC,ASI,OOTE,TNI	0.002	4.941	16216.745	-0.046
3	223779.007	ASI,VALT,OOTE,TMT	9.936	0.004	8950.86	-133.868
4	217402.795	ASI,OOTE,TMT,TNI	8.029	13961.311	131.717	-0.252
5	226473.214	GMC,ASI,VALT,OOTE	0.002	3.494	0.001	15262.594
6	271405.231	GMC,ASI,TLNSE,OOTE	0.002	3.355	-200.409	16941.135
7	222058.833	GMC,ASI,OOTE,TMT	0.002	3.755	16399.352	-9.985
8	360948.07	ASI,TLNSE,OOTE,TNI	6.58	-609.137	17306.805	-0.073
9	238691.483	GMC,ASI,VALT,TMT	0.001	14.345	0.005	-191.681
10	235115.733	ASI,VALT,TMT,TNI	15.688	0.005	-99.372	-0.13
11	168307.357	ASI,VALT,TLNSE,TNI	15.383	0.004	277.429	-0.23
12	360762.642	ASI,TLNSE,OOTE,TMT	5.065	-593.908	17462.073	-21.749
13	237397.809	GMC,ASI,VALT,TNI	0	15.468	0.004	-0.226
14	315064.818	ASI,VALT,TLNSE,OOTE	4.455	0	-379.878	15730.522
15	217696.034	ASI,VALT,TLNSE,TMT	14.592	0.005	77.544	-187.391
16	262382.952	GMC,TLNSE,OOTE,TMT	0.002	-178.376	22462.726	10.154
17	258323.09	GMC,VALT,TLNSE,OOTE	0.002	0	-163.388	22686.18

C_4^7	INTERCEPT	VARIABLES	PARAMETERS			
			1	2	3	4
18	267190.973	GMC,TLNSE,OOTE,TNI	0.002	-198.472	22445.446	0.014
19	220936.777	GMC,TLNSE,OOTE,TNI	0.002	0	21566.53	0.01
20	220610.331	GMC,VALT,OOTE,TMT	0.002	0	21724.379	7.15
21	220453.228	GMC,OOTE,TMT,TNI	0.002	21721.382	18.259	-0.005
22	358679.054	VALT,TLNSE,OOTE,TMT	0	-595.808	24558.042	25.413
23	347621.785	TLNSE,OOTE,TMT,TNI	-550.112	24654.932	29.402-0.016	-0.016
24	364741.341	VALT,TLNSE,OOTE,TNI	8.42E-06	-623.272	24793.311	0.021
25	216838.892	VALT,OOTE,TMT,TNI	0	22960.879	101.559	-0.0082
26	112942.551	GMC,ASI,VALT,TLNSE	0.002	8.746	0.002	581.838
27	176417.767	ASI,TLNSE,TMT,TNI	14.083	267.141	180..813	-0.32
28	242719.783	GMC,ASI,OOTE,TNI	0	14.102	149.976	-0.283
29	205076.633	GMC,ASI,TLNSE,TNI	0.001	12.022	169.597	-0.045
30	178364.927	GMC,ASI,TLNSE,TMT	0.002	10.149	300.213	6.997
31	173157.575	GMC,VALT,TLNSE,TNI	0.004	0.002	1484.925	0.185
32	-34485.328	GMC,TLNSE,TMT,TNI	0.004	1326.378	38.326	0.19
33	-133333.187	GMC,VALT,TLNSE,TMT	0.004	0.001	1746.011	152.204
34	294409.381	GMC,VALT,TMT,TNI	0.004	0.003	-255.389	0.501
35	86657.219	VALT,TLNSE,TMT,TNI	0.002	840.414	-22.624	0.28

Table 12. *Continue.*

C_4^7	R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
1	0.975	1503272992	1	2	1.5
2	0.969	868090178.1	4	1	2.5
3	0.971	1742103059	2	3	2.5
4	0.969	1868177615	4	4	4
5	0.969	1910902660	4	5	4.5
6	0.968	1941971766	6.5	6	6.25
7	0.968	1963009037	6.5	7	6.75
8	0.967	2007145474	8	8	8
9	0.966	2086519874	9.5	9	9.25
10	0.966	2097492040	9.5	10	9.75
11	0.965	2138961034	11	11	11
12	0.964	2208776052	12.5	12	12.25
13	0.964	2212989526	12.5	13	12.75
14	0.963	2232434678	14.5	14	14.25
15	0.963	2241852749	14.5	15	14.75
16	0.962	2330722796	18	16	17
17	0.962	2331575623	18	17	17.5
18	0.962	2331866337	18	18	18
19	0.962	2343345889	18	19	18.5
20	0.962	2345640059	18	20	19
21	0.961	2354565741	21	21	21
22	0.951	2967789678	22	22	22
23	0.951	2973596390	23.5	23	23.25
24	0.951	2996535023	23.5	24	23.75
25	0.947	3216681289	23.5	25	24.25
26	0.945	3374304870	26	26	26
27	0.943	3502243803	27	27	27
28	0.942	3561933444	28	28	28
29	0.933	4111402712	29	29	29
30	0.931	4199628316	30	30	30
31	0.866	8189908013	31	31	31
32	0.858	8677111513	32	32	32
33	0.851	9110478220	33	33	33
34	0.841	9693754796	34	34	34
35	0.816	11196277717	35	35	35

From the results in the Table 11, and Table 12 shows that for combination of four regressors out of seven regressors (p = 7), the optima model for value p = 7 was selected by ranking the Coefficient of Determination in descending order (highest coefficient of determination has the lowest average ranking and also ranking the Mean Square Error of the standardised value of regression coefficient in ascending order (lowest Mean Square Error has the lowest rank), then the model with the lowest average rank was selected. The optima model with the best subset of Ridge regressors result for p = 4 were GMC, ASI, VALT, OOTE with the lowest average rank of 1, also the optima model with the best subset of Least Square regressors result for p = 4 were ASI, VALT, OOTE, TNI with the lowest average rank of 1.5.

Table 13. *Ridge Regression Model for P = 5 and ranks.*

C_5^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS				
				1	2	3	4	5
1	21995	24492400	GMC,ASI,VALT,OOTE,TMT	17474340	39565320	95869370	10816510	12135040
2	29077	6166550	GMC,ASI,VALT,TLNSE,OOTE	21307320	39950150	13998740	35123930	99098870
3	29077	18257370	GMC,ASI,TLNSE,OOTE,TMT	20939760	35900320	26847530	10195990	37734900
4	26493	24941750	GMC,ASI,VALT,OOTE,TNI	18007700	36498470	10143930	10015990	29833560
5	24140	19960930	GMC,ASI,TLNSE,OOTE,TNI	21671520	35709540	19084600	10557640	46270090
6	31912	25190520	GMC,ASI,OOTE,TMT,TNI	18448930	33001970	97845590	26383920	34626130
7	21995	20172660	ASI,VALT,TLNSE,OOTE,TMT	35273210	13070680	2468270	81472280	40081390
8	21995	21277480	ASI,VALT,TLNSE,OOTE,TNI	33618470	14305840	20845700	78426130	56064790
9	21995	26518940	ASI,VALT,OOTE,TMT,TNI	31143110	10416690	76599880	30222460	44012060
10	21995	22619270	ASI,TLNSE,OOTE,TMT,TNI	30018930	16925050	74663880	41882320	53760750
11	21820	6359740	GMC,ASI,VALT,TLNSE,TMT	24677580	60808300	18829960	85284600	29033480
12	21995	19332790	GMC,VALT,TLNSE,OOTE,TNI	21951240	14825870	4459440	65199540	67768480
13	21995	26317420	GMC,VALT,OOTE,TMT,TNI	22246890	75858490	11547920	32835490	60749970
14	23947	76992570	GMC,ASI,VALT,TLNSE,TNI	25711670	51805940	19009420	81658570	57690340
15	21995	19897450	VALT,TLNSE,OOTE,TMT,TNI	10303720	25530010	11663920	43726760	66576390
16	21995	25646410	GMC,VALT,TLNSE,OOTE,TMT	15845980	11421500	31196860	44527410	36058590
17	21820	27572170	GMC,ASI,VALT,TMT,TNI	17575170	62159030	14175890	16653990	43955340
18	23947	11378220	GMC,ASI,TLNSE,TMT,TNI	25333940	42602850	69387880	49706780	65816170
19	21820	19727260	ASI,VALT,TLNSE,TMT,TNI	30071370	14372340	44425740	43341180	61334690
20	520078	28340460	GMC,TLNSE,OOTE,TMT,TNI	13381880	22726590	36560210	30341900	41749870
21	362599	26779990	GMC,VALT,TLNSE,TMT,TNI	15374690	11050530	34100910	34366690	48419550

Table 13. *Continue.*

C_5^7	R-SQUARES	OPT MSE	R-SQUARE RANK	MSE RANK	AVERAGE RANK
1	0.9649479	2.60E+09	1	1	1
2	0.9630511	2.70E+09	3	2	2.5
3	0.9641671	3.05E+09	2	5	3.5
4	0.9623415	3.01E+09	4	4	4
5	0.9599222	3.21E+09	5	6	5.5
6	0.9563697	2.99E+09	9	3	6
7	0.9576816	3.74E+09	7	7	7
8	0.9581357	4.07E+09	6	8	7
9	0.957253	4.08E+09	8	9	8.5
10	0.9543286	5.83E+09	10	12	11
11	0.9356682	5.65E+09	14	10	12
12	0.950409	6.78E+09	12	13	12.5
13	0.95045	6.90E+09	11	15	13
14	0.9329078	5.82E+09	16	11	13.5
15	0.9349993	6.86E+09	15	14	14.5
16	0.9502732	9.33E+09	13	19	16
17	0.9265996	7.65E+09	17	17	17
18	0.9164111	7.57E+09	19	16	17.5
19	0.9251568	8.96E+09	18	18	18
20	0.7622799	1.28E+10	20	20	20
21	0.7482209	1.41E+10	21	21	21

Table 14. *Least Square Regression Model for P = 5 and ranks.*

C_5^7	INTERCEPT	VARIABLES	PARAMETERS				
			1	2	3	4	5
1	219045.791	GMC,ASI,OOTE,TMT,TNI	0.001	6.425	15163.211	90.195	-0.185
2	221559.432	GMC,ASI,VALT,OOTE,TNI	0.001	8.888	0.002	11654.674	-0.154
3	224271.083	GMC,ASI,VALT,OOTE,TMT	0.002	8.306	0.003	10888.713	-124.83
4	268579.89	ASI,VALT,TLNSE,OOTE,TNI	10.134	0.002	-204.222	11365.85	-172
5	220703.63	ASI,VALT,OOTE,TMT,TNI	10.7	0.003	9806.184	-13.332	-0.166
6	300726.711	ASI,VALT,TLNSE,OOTE,TMT	9.413	0.003	-328.338	11024.556	-130.696
7	300628.436	ASI,TLNSE,OOTE,TMT,TNI	7.662	-351.553	15576.755	100.493	-0.217
8	299446.376	GMC,ASI,TLNSE,OOTE,TNI	0.001	5.313	-340.852	17250.59	-0.06
9	293027.288	GMC,ASI,TLNSE,OOTE,TMT	0.002	4.011	-302.346	17422.44	-19.522
10	247659.456	GMC,ASI,VALT,TLNSE,OOTE	0.002	3.457	0	-90.55	15710.639
11	157815.898	GMC,ASI,VALT,TLNSE,TMT	0.001	13.585	0.005	335.893	-182.511
12	237104.456	GMC,ASI,VALT,TMT,TNI	0.001	15.11	0.005	-129.05	0.09
13	134016.466	GMC,ASI,VALT,TLNSE,TNI	0.001	14.545	0.004	428.293	-0.217

C_5^7	INTERCEPT	VARIABLES	PARAMETERS				
			1	2	3	4	5
14	188870.502	ASI,VALT,TLNSE,TMT,TNI	15.478	0.005	189.214	-81.654	-0.148
15	258429.961	GMC,VALT,TLNSE,OOTE,TNI	0.002	0	-159.975	22258.804	0.009
16	257964.256	GMC,VALT,TLNSE,OOTE,TMT	0.002	0	-159.436	22419.202	6.363
17	263484.931	GMC,TLNSE,OOTE,TMT,TNI	0.002	-182.676	22424.586	7.316	0.004
18	221062.998	GMC,VALT,OOTE,TMT,TNI	0.002	0	21497.02	-9.275	0.02
19	350821.329	VALT,TLNSE,OOTE,TMT,TNI	-0.001	-566.882	25000.442	66.424	-0.05
20	150894.919	GMC,ASI,TLNSE,TMT,TNI	0.001	13.337	380.486	166.636	-0.292
21	-45376.328	GMC,VALT,TLNSE,TMT,TNI	0.005	0.003	1364.886	-168.316	0.355

Table 14. *Continue.*

C_5^7	R-SQUARES	OPT MSE	R-SQUARE RANK	MSE RANK	AVERAGE RANK
1	0.986	17339573.39	1	1	1
2	0.978	138856408	2	2	2
3	0.977	1437077143	3	3	3
4	0.976	1526102468	4	4	4
5	0.975	1559032708	5	5	5
6	0.973	1715967334	6	6	6
7	0.971	1833771152	7.5	7	7.25
8	0.971	1842825652	7.5	8	7.75
9	0.969	1967994606	9.5	9	9.25
10	0.969	1978539950	9.5	10	9.75
11	0.967	2068252258	11.5	11	11.25
12	0.967	2105665296	11.5	12	11.75
13	0.966	2134366888	13.5	13	13.25
14	0.966	2140670923	13.5	14	13.75
15	0.962	2415054376	17	15	16
16	0.962	2417592050	17	16	16.5
17	0.962	2420107122	17	17	17
18	0.962	2432544362	17	18	17.5
19	0.952	3058834151	17	19	18
20	0.944	3573353978	20	20	20
21	0.871	8185024485	21	21	21

From the results in the Table 13, and Table 14 shows that for combination of five regressors out of seven regressors (p = 5), the optima model for value p = 5 was selected by ranking the Coefficient of Determination in descending order (highest coefficient of determination has the lowest average ranking and also ranking the Mean Square Error of the standardised value of regression coefficient in ascending order (lowest Mean Square Error has the lowest rank), then the model with the lowest average rank was selected. The optima model with the best subset of Ridge regressors result for p = 5 were GMC, ASI, VALT, OOTE, TMT with the lowest average rank of 1, also the optima model with the best subset of Least Square regressors result for p = 5 were GMC, ASI, OOTE, TMT, TNI with the lowest average rank of 1.

Table 15. *Ridge Regression Models for P=6 and Ranks.*

C_6^7	LAMBDA	INTERCEPT	VARIABLES	PARAMETERS					
				1	2	3	4	5	6
1	2693	25462320	GMC,ASI,VALT,TNI,OOTE,TMT	17502110	33186470	86288610	30499640	91705880	16769570
2	26493	15463720	GMC,ASI,VALT,TLNSE,OOTE,TMT	20783190	35718390	11280880	39635170	91899060	22448070
3	29077	15791670	GMC,ASI,VALT,TLNSE,OOTE,TNI	21261690	32754370	12349350	40126170	84333550	3987440
4	38438	17548700	GMC,ASI,TLNSE,OOTE,TMT,TNI	20780530	29103490	34667190	80386800	33520300	43399810
5	21995	20458180	ASI,VALT,OOTE,TLNSE,TNI,TMT	29862990	11275180	71573780	2542660	44742660	31846610
6	21995	12668250	GMC,VALT,TLNSE,OOTE,TNI,TMT	26054060	97351450	55966160	10441430	57479360	33915600
7	21820	81754040	GMC,ASI,VALT,TLNSE,TNI,TMT	24509600	44323580	15229680	81317420	49667130	30538150

Table 15. *Continue.*

C_6^7	R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
1	0.9619524	2.97E+09	2	1	1.5
2	0.9621445	3.23E+09	1	2	1.5
3	0.9606406	3.38E+09	3	4	3.5
4	0.9527003	3.34E+09	5	3	4
5	0.9576264	4.51E+09	4	5	4.5
6	0.9492047	6.21E+09	6	6	6
7	0.9337035	6.49E+09	7	7	7

Table 16. *Least Square Regression Model for P = 6 and ranks.*

C_6^7	INTERCEPT	VARIABLES	PARAMETERS					
			1	2	3	4	5	6
1	222297.059	GMC,ASI,VALT,TNI,OOTE,TMT	0.001	9.098	0.003	11023.033	-53.122	-0.101
2	229228.851	GMC,ASI,VALT,TLNSE,OOTE,TNI	0.001	8.863	0.002	-32.731	11824.841	-0.153
3	239840.159	GMC,ASI,VALT,TLNSE,OOTE,TMT	0.001	8.269	0.003	-66.522	11227.016	-124.569
4	271836.232	ASI,VALT,OOTE,TLNSE,TNI,TMT	10.284	0.003	11097.93	-216.978	-0.151	-22.316
5	259239.106	GMC,VALT,TLNSE,OOTE,TNI,TMT	0.002	0	-162.741	22182.251	0.022	-11.812
6	266643.662	GMC,ASI,TLNSE,OOTE,TMT,TNI	0.001	6.451	-202.106	15914.171	78.385	-0.175
7	150190.493	GMC, ASI, VALT, TLNSE, TNI,TMT	0.001	14.381	0.005	360.23	-112.164	-0.1

Table 16. *Continue.*

C_6^7	R-SQUARES	OPT MSE	R-SQUARE RANK	OPT MSE RANK	AVERAGE RANK
1	0.979	1413275837	1	1	1
2	0.978	1443617706	2	2	2
3	0.977	1491274087	3	3	3
4	0.976	1581290271	4	4	4
5	0.962	1581290271	7	7	7
6	0.973	1771479255	5	5	5
7	0.969	2072853708	6	6	6

From the results in the Table 15, and Table 16 shows that for combination of six regressors out of seven regressors (p = 6), the optima model for value p = 6 were selected by ranking the Coefficient of Determination in descending order (highest coefficient of determination has the lowest average ranking and also ranking the Mean Square Error of the standardised value of regression coefficient in ascending order (lowest Mean Square Error has the lowest rank), then the model with the lowest average rank were selected. The optima model with the best subset of Ridge regressors result for p = 6 was GMC, ASI, VALT, TNI, OOTE, TMT with the lowest average rank of 1.5, also the optima model with the best subset of Partial Least Square regressors result for p= 6 was all the combination of six variables GMC, ASI, VALT, TNI, OOTE, TMT with the lowest average rank of 1.

Table 17. *Summary of all Ridge Regression and Least Regression Models Result for all values of $P_2,\ldots,P_n$.*

PARAMETER (p)	RIDGE REGRESSION VARIABLES	LOWEST AVERAGE RANK
2	ASI, OOTE	1
3	GMC, ASI, OOTE	1
4	GMI, ASI, VALT, OOTE	1
5	GMC,ASI,VALT,OOTE,TMT	1
6	GMC,ASI,VALT,TNI,OOTE,TMT	1.5

Table 17. *Continue.*

PARAMETER (p)	LEAST SQUARE REGRESSION VARIABLE	LOWEST AVERAGE RANK	REMARKS
2	GMC, OOTE	1	Selected Different variables
3	GMC, ASI,OOTE	2	Selected Same variables
4	ASI, VALT, OOTE, TNI	1.5	Selected Different variables
5	GMC,ASI,OOTE,TMT,TNI	1	Selected Different variables
6	GMC,ASI,VALT,TNI,OOTE,TMT	1	Selected Same variables

The Variable selection techniques in the presence of multicollinearity, was calculated by use of Least Square Methods in the parameter estimation, Ridge Regression Methods in Selecting variables.

From the Table 17, the Result of Ridge and Least Square Regression it shows that the Parameter P = 6, and P = 3 the same set of variables were selected as GMC, ASI, VALT, TNI, OOTE, TMT and GMC, ASI, OOTE respectively.

For Parameter P = 2, P = 4 and P = 5, different sets of variables were selected as ASI, OOTE was selected under Ridge while GMC, OOTE was selected under Least Squares Method, GMI, ASI, VALT, OOTE was selected under Ridge for P = 4 and ASI, VALT, OOTE, TNI under Least Squares Method. For P = 5, GMC, ASI, VALT, OOTE, TMT was selected under Ridge while GMC, ASI, OOTE, TMT, TNI were selected under Least Squares Method.

The computations were done using Minitab packages, R-packages and SPSS packages after computing for the Least Square Method in the parameters, Ridge Regression Methods in selecting variables.

4. Conclusion

Multicollinearity is a serious problem in Multiple regression as it has so many undesirable effects on the estimates of the Multiple Regression model especially when the parameters are estimated with the aid of Least Square Method. Multicollinearity is a condition of deficient data, which frequently encountered in observational studies in which the investigator does not interfere with the study. Multicollinearity creates difficulties in which one builds a regression model between response variables and explanatory variables. Multicollinearity is a phenomenon in which two or more predictors (explanatory) variable in a multiple regression model are highly correlated. Thus in multiple regression, identifying the best subset among many variables to include in a model is arguable the hardest part of model building in regression Analysis. There exist various variable selection techniques. This research work has investigated the effects of multicollinearity in variable selection and have observed that multicollinearity affects the choice of variables selected. Multicollinearity must therefore be treated or handled whenever it exists in data sets before proceeding with Variable selection.

References

[1] Alin, A. (2010). *Multicollinearity.* – WIREs Computational Statistics.

[2] Brue Ratner (2009). *Variable selection method in regression: Ignorable Problem, Outing notable solution*. 574 Flander drive north, woodmere NT 11581, USA.

[3] Bertsimas D. and King A.: (2016) OR forum—An algorithmic approach to linear regression. Operations Research, 64, 2–16.

[4] Bertsimas D., King A., and Mazumder R. (2016) Best subset selection via a modern optimization lens. The Annals of Statistics, 44, 813–852.

[5] Chatterjee S. and Hadi A. S (2012).: Regression Analysis by Example, Fifth Edition (Wiley, Hoboken).

[6] Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36 (1): 27–46.

[7] Farrar D. E. and Glauber R. R (1967).: Multicollinearity in regression analysis: The problem revisited. The Review of Economic and Statistics, 49, 92–107.

[8] Frank, I. and Friedman, J. (1993) A statistical view of some chemometrics regression tools. Technometrics, 35, 109–148.

[9] Fu, W. (1998) Penalized regression: the bridge versus the lasso. J. Computnl Graph. Statist., 7, 397–416.

[10] Gujarati DN, and Porter DCn (2009) Basic Econometrics. New York: McGraw Hill Inc.

[11] Gunst R. F. and Webster J. T. (1975),: Regression analysis and problems of multicollinearity. Communications in Statistics—Theory and Methods, 4 277–292.

[12] Hadi A. S. and Ling R. F. (1998): Some cautionary notes on the use of principal components regression. The American Statistician, 52, 15–19.

[13] Hoerl, A. E and Kennard, R. W (1970). *Ridge regression: Biased estimation of non-orthogonal problems.* Techno metrics, 12 (1): 55-67.

[14] Jolliffe I. T. (1982): A note on the use of principal components in regression. Applied Statistics, 31 300–303.

[15] Mansfield E. R. and Helms B. P.: Detecting multicollinearity. The American Statistician, 36 (1982), 158–160.

[16] Massy W. F (1965). Principal components regression in exploratory statistical research. Journal of the American Statistical Association, 60, 234–256.

[17] Meloun, M., M. Meloun, J. Militky, M. Hill, R. G. Brereton (2002). *Crucial problems in regression modelling and their solutions.* – Analyst 127: 433–450.

[18] Murray, C. J. L. (2006). Eight Americas: investigating mortality disparity across races, counties, and race-counties in the United States.

[19] Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B, 58, 267–288.

[20] Vandenberghe L. and Boyd S (1996), Semidefinite programming. SIAM Review, 38 49–95.

[21] Wold H. (1966): Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiaah (ed.): Multivariate Analysis (Academic Press, New York,), 391–420.

[22] Zou H. and Hastie T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 67 (2005), 301–320.