HANDLING MULTICOLLINEARITY; A COMPARATIVE STUDY OF THE PREDICTION PERFORMANCE OF SOME METHODS BASED ON SOME PROBABILITY DISTRIBUTIONS

Zakari Y., Yau S. A., Usman U.

Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria

Corresponding Author: Zakari Y., zakariyahaya007@gmail.com

ABSTRACT: This study used some probability distribution (Gamma, Beta and Chi-square distributions) to assess the performance of partial least square regression (PLSR), ridge regression (RR) and LASSO regression (LR) methods. Ordinary Least Squares may fail if the variables are almost collinear or related. As such, this methods (PLSR, RR, AND LR) were compared using simulated data that follows gamma, beta and chi-square distributions with number variables (*P*=4 and 10) and sample sizes (*n*=60 and 90). The comparison was carried out using Mean Square Log Error (MSLE), Mean Absolute Error (MAE) and R-Square (R^2) which shows that the results of RR is better when *P*=4 and n=60 using gamma distribution, but using chi square distribution PLRS is better methods. Also, when *P*=4 and n=90, RR shows better results with both gamma and beta distributions but with chi square distribution all methods have equal predictive ability. However, at *P*=10 and n=60 RR performed better with both gamma and chi square distributions while when data follows beta distribution all distributions have equal predictive ability. RR shows better results at both gamma and chi square distributions when *P*=10 and n=90 while PLSR performed better with beta distribution.

KEYWORDS**:** Regression; multicollinearity; ridge (RR); partial least square (PLSR); lasso regressions (LR).

1. INTRODUCTION

Econometrics were the first people who paid attention to the multicollinearity, after the international economic crisis in 1928, some of them discussed it as a problem, others took it as a study case especially in preparing budgets because the relation between the variables were very complicated. No precise definition of multicollinearity has been firmly established in the literature ([Bal70]).

Multicollinearity is a common problem in regression analysis. Handling multicollinearity in regression analysis is important because least squares estimations assume that predictor variables are not correlated with each other ([AAA06]). Also, this problem makes the estimated regression coefficients by least squares method to be conditional upon the correlated predictor variables in the model. It is also a condition in a set of regression data that have two or more regressors which are redundant and have the same information. Redundant information means what one variable explains about *Y* (dependent variable) is exactly what the other variable explains.

El-Fallah and El-Salam ([FS14]) compares the PLSR, RR and PCR as an alternative procedure for handling multicollinearity problem. A Monte Carlo simulation study was used to evaluate the effectiveness of these three procedures. Mean Squared Errors (MSE) was calculated for comparison purposes. From the results of their work, it shows that the RR is more efficient when the number of regressors is small, while the PLSR is more efficient than the others when the number of regressors is moderate or high.

Esra and Suleyman ([ES15]) compared partial least squares regression, principal component regression, ridge regression and multiple linear regression methods in modeling and predicting daily mean PM10 concentrations on the base of various meteorological parameters obtained for the city of Ankara, in Turkey. The analysed period is February 2007. Their results show that while multiple linear regression and ridge regression yield somewhat better results for fitting to this dataset, principal component regression and partial least squares regression are better than both of them in terms of prediction of PM10 values for future datasets. In addition, partial least squares regression was the remarkable method in terms of predictive ability as it had a close performance with principal component regression even with less number of factors.

Esra and Semra ([ES16]) compared the performance of robust biased Robust Ridge Regression (RRR), Robust Principal Component Regression (RPCR) and RSIMPLS methods with each other and their classical versions known as Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) in terms of predictive ability by using trimmed Root Mean Squared Error (TRMSE) statistic in case of both of multicollinearity and outliers existence in an unemployment data set of Turkey. Their analysis results show that RRR model is chosen as the best model for determining unemployment rate in Turkey for the period of 1985-2012. Robust biased RRR method showed that the most important independent variable affecting the unemployment rate is Purchasing Power Parities (PPP). The least important variables affecting the unemployment rate are Import Growth Rate (IMP) and Export Growth Rate (EXP). Hence, any increment in PPP cause an important increment in unemployment rate, however, any increment in IMP causes an unimportant increase in unemployment rate. Also, any increment in EXP causes an unimportant decrease in unemployment rate.

Toka ([Tok16]) compared some regression methods that handle multicollinearity such as partial least square regression (PLSR), ridge regression (RR) and principal component regression (PCR). The methods were compared by the simulation study. All results were compared with each other through MSE of their estimated beta values for different methods. The author showed that PLSR was better methods with large numbers of independent variable. Further, RR is better method when observation number and number of multicollinearity are large enough. He also, stated that PCR cannot be better method in simulation study scenarios.

Usman *et al.*, ([U+17]) used some regression techniques for prediction in the presence of multicollinearity which include: Ridge Regression (RR), Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR). They investigated the performance of these methods with the simulated data that follows lognormal and exponential distributions. Hence, Mean square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were obtained. And their result shows that PLSR and RR methods are generally effective in handling multicollinearity problems at both lognormal and exponential distributions.

Usman *et al.*, ([UZM17]) resort to biased regression methods which stabilize the variance of the parameter estimate and compared three biased regression methods for overcoming multicollinearity, Principal Components Regression (PCR), Partial Least Square Regression (PLSR) and Ridge Regression (RR) and used simulated data that follows normal and uniform distributions to estimate the regression coefficients by PCR, PLSR and RR methods. A comparison between the three methods were done by using symmetric loss functions like root mean square errors, mean absolute errors and mean absolute percentage errors. Based on their study, it is observed that PLSR has a lower measure of accuracy in normal distribution while RR shows better results in uniform distribution.

In this paper, the aim is to assess the performance of Ridge Regression, Partial Least Squares Regression and Lasso Regression using simulated data that follows Gamma, Beta and Chi Square Distributions.

2.1 METHODOLOGY

Several methods for handling multicollinearity problem have been developed Adnan *et al.*, ([AAA06]). Some of these methods include Principal Component Regression (PCR), Partial Least Square Regression (PLSR), Ridge Regression (RR) and Lasso regression etc.

2.2 SIMULATION STUDY

In this research work, Monte Carlos was performed with different levels of multicollinearity using R ([RCT16]). By following Mc Donald and Galarneau, ([DG75]), Wichern and Churchill, ([WC78]), Gibbons ([Gib81]), Kibria ([Kib03]), Arumairajan and Wijekoon ([AW14, AW15]) and Usman *et al.*, ([U+17]).

Also, four sets of correlations were considered corresponding to ρ = 0.7, 0.8, 0.9 and 0.99 as used by Khalaf ([Kha12]). Using the condition number, $CN = \frac{\lambda_{max}}{\lambda_{min}}$, it can be shown that these values of ρ will include a wide range of low, moderate and high correlations between variables. The *n* observations for the dependent variable *Y* are determined by:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \varepsilon_i, n = 1,2,\ldots,p \quad (1)$$

We simulated data that follows gamma and chi square distributions with P=4, 10 predictors variables Toka ([Tok16]) for each observations (*n* = 60 and 90) respectively. The goal is to develop a linear equation that relates all the predictor variables to a response variable. For the purpose of comparing the three methods under multicollinearity, the analysis was done using R software. Also, simulation was replicated 100 times in order to obtain the approximate distribution considered in the study in real life situation.

2.3 LASSO REGRESSION

The LASSO (Least Absolute Shrinkage and Selection Operator), Tibshirani ([Tib96]) is another regularization method, but here the penalty is applied to the sum of the absolute values of the regression coefficients, the L_1 norm. The penalty function is given by Pen(β) = $\lambda_i \sum_{j=1}^{p} |\beta_j|$

The objective is to minimize

$$\hat{\beta}_{LASSO} = argmin_{BERP}\left(Y - X\hat{\beta}\right)^T\left(Y - X\hat{\beta}\right) + \lambda_i \sum_{j=1}^{p} |\beta_j| \quad (2)$$

Where λ is a non-negative regularization parameter. Since the penalty term is no longer quadratic, there is no explicit formula for the mean squared error of the Lasso estimator.
Generally, the $Bias(\hat{\beta}_{LASSO})$ also increases as the tuning parameter λ increases. While the variance, $Var(\hat{\beta}_{LASSO})$ decreases.

2.4 PARTIAL LEAST SQUARES REGRESSION

The PLSR searches for a set of components (called latent vectors) that performs a simultaneous decomposition of X and Y with the constraints that this components explain as much as possible the covariance between $X$and Y. In this method, the component is extracted from which the rest of the components are extracted in such a way that they are uncorrelated (orthogonal). How this algorithm functions will now be described to show. Component is defined as:

$$t_i = W_{11}X_1 + W_{12}X_2 + \cdots + W_{ij}X_j \quad (3)$$

Where, X_j are the explanatory variables, Y is the dependent variables.
The W_{ij} is the coefficient:

$$W_{ij} = \frac{cov(X_j,Y)}{\sqrt{\sum_j^p cov(X_j,Y)^2}}, \text{ j=1, 2, 3…}p \quad (4)$$

From which it can be deduced that in order to obtain W_{ij} the scalar product (X_j, Y) must be calculated for each$J = 1, 2 \ldots P$.
Calculating the second component is justified when the single component model is inadequate i.e. when the explanatory power of regression is small and another component is necessary. The second component is denoted by t_2 and it will be a linear combination of the regression residues of X_j variables on components t_1 instead of the original variables. In this way, component orthogonality is assured. To do this, the residual for the single component regression must be calculated which will be,

$$e_1 = Y - \hat{Y} = Y - \beta_1 t_1$$

with

$$\beta_1 = \frac{cov(y_i,t_i)}{\|t_1\|^2} \quad (5)$$

The second component is obtained as:

$$t_2 = W_{21}e_{11} + W_{22}e_{12} + \cdots + W_{2p}e_{1p} \quad (6)$$

With

$$W_{2j} = \frac{cov(e_{ij},e_1)}{\sqrt{\sum_j^p cov^2(e_{ij},e_1)}}, \text{ j=1, 2, 3…}p \quad (7)$$

The residuals e_{ij} are calculated by computing the simple regression of $x_j$ont_1,
$X^*_j = \alpha_j t_j,\ j = 1, 2, \ldots, p$ therefore,

$$e_{ij} = X_j - X^*_j = X_j - \alpha_j t_j \quad (8)$$

Where, the estimators of the regression coefficients have been calculated thus:

$$\alpha_j = \frac{cov(x_j,t_1)}{\|t_1\|^2} \quad (9)$$

Now with e_i and e_{ij}, only the scalar products have to be computed $cov(e_i, e_{ij})$, for j = 1 . . .P, to be able to compute t_2.
To construct subsequent components, the same steps are performed as for the two previous components. This iterative procedure is continued until the number of components to be retained is significant.

2.5 RIDGE REGRESSION

When multicollinearity exists, the matrix X′ X, where X consists of the original regressors, becomes nearly singular. Since $Var(\beta) = \delta^2(X'X),^{-1}$ and the diagonal elements of $(X'X)^{-1}$ become quite large, this makes the variance of β to be large. This leads to an unstable estimate of β when OLS is used.
Consider the following regression model:

$$Y_i = \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon \quad (10)$$

Where, $\beta_1, \beta_2, \beta_3$ etc. are the parameters of the model and ε are random terms.
Also, standardize data by subtracting each x observation from its corresponding mean and dividing by its standard deviation i.e. $\frac{x_i - \mu_i}{\sqrt{\delta_i}}$
Arrange the predictors into convenient matrix. Suppose we have n observations of k predictors, this will be a $n \times k$ matrix x. And arrange the key parameters into a β. So that viewing the response variable as an n-vector, our model becomes:

$$y = x\beta + \varepsilon \quad (11)$$

Where, ε is now a vector of the random noise in the observed data vector Y.
Note: the least square parameter β_{LS} can be estimated by finding the parameter values which minimized the sum square residuals i.e.

$$SSR = \sum(Y - X\beta)'(Y - X\beta).$$

The solution turns out to be a matrix equation,

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (12)$$

Where, X' is the transpose of the matrix X.
According to Hoerl and Kennard ([HK70]), the potential instability in using the least squares estimator could be improved by adding a small constant λ to the diagonal entries of the $X'X$ marix before taking its inverse. The result is the Ridge regression estimator

$$\hat{\beta} = (X'X + \lambda I)^{-1}X'Y \qquad (13)$$

Where I is the $p \times p$ identity matrix and $X'X$ is the correlation matrix of independent variables values of lambda lie in the range (0 and 1). When $\lambda = 0$, $\hat{\beta}_{RIDGE}$ becomes $\hat{\beta}_{OLS}$. Obviously, a key aspect of ridge regression is determining what the best value of the constant that is added to the main diagonal of the matrix $X'X$should be to maximize prediction. There are many procedures for determining the best value. The simplest way is to plot the values of each $\hat{\beta}_{RIDGE}$versus λ. The smallest value for which each ridge trace plot shows stability in the coefficient is adopted ([Mye90]).

3. EVALUATION OF PERFORMANCE MEASURES

The validity test of the considered methods was evaluated by means of the estimated regression parameters $\hat{\beta}$. These values indicate to what extent the slope and intercept are correctly estimated. According to the value of Mean Square Log Error (MSLE), Mean Absolute Error (MAE), R-Square (R^2). that is close to zero, the slope and intercept are correctly estimated. The results of the performance measures are listed in the tables below:

Table 1: MSLE of LR, RR and PLSR when P=4 and n=60

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	0.54113	8.34561	2.167892
RR	0.53982*	0.09312*	1.990815
PLSR	0.59923	0.59309	1.98776*

Table 1 shows the results of Mean Square Log Error (MSLE) of the biased regression methods. It revealed that RR has minimum error under gamma and beta distributions. That is RR has higher predictive power compared to LR and PLSR in these distributions while PLSR has higher predictive power over both LR and RR when data follows chi square distribution. In summary, RR performed better under gamma and beta distributions. And PLSR is better in chi square distribution.

Table 2: MSLE of LR, RR and PLSR when P=4 and n=90

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	4.15214	0.14997*	2.479848
RR	1.13257*	0.56174	2.482757
PLSR	6.61317	0.98311	2.479466*

From Table 2, it shows that RR has the least values in gamma distribution than both LR and PLSR, when data follows beta distribution LR is the more efficient over both RR and PLSR and also, under chi square distribution PLSR has the minimum error then both LR and RR.

Table 3: MAE of LR, RR and PLSR when P=4 and n=60

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	0.91037	0.51774	13.99083
RR	0.50010*	0.220456	13.62415
PLSR	5.50016	0.22039*	13.62365*

The result obtained from Table 3 revealed that Ridge Regression (RR) has minimal error under gamma distribution. Also, in beta and chi square distributions PLSR is the more efficient. However, it shows that both only RR performed better in gamma distribution PLSR also performed better in both beta and chi square distributions.

Table 4: MAE of LR, RR and PLSR when P=4 and n=90

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	4.31415	10.13464	1.32001*
RR	0.53827*	0.15664*	9.93310
PLSR	6.03829	2.15688	18.50231

From the results presented in Table 4 Ridge Regression (RR) performed better than both Lasso Regression (LR) and Partial Least Squares Regression (PLSR) under gamma and beta distributions, while LASSO Regression (LR) performed better in chi square distribution than Ridge Regression (RR) and Partial Least Squares

Regression (PLSR). Thus, RR is better under gamma and beta distributions, and also, LR is better when data follows chi square distribution.

Table 5: R^2 of LR, RR and PLSR when P=4 and n=60

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	0.02819*	0.05853*	0.02477*
RR	0.03052	0.06899	0.06667
PLSR	0.09180	0.20775	0.07746

Table 5 represents R^2 (R- Square) of LASSO Regression (LR), Partial Least Squares Regression (PLSR) and Ridge Regression (RR). The results revealed that Lasso Regression (PCR) has minimum error under gamma, beta and chi square distributions, This means that, there is higher predictive power observed compared to Partial Least Squares Regression (PLSR) and Ridge Regression (RR) under these distributions. However, LR performed better under gamma, beta and chi square distributions.

Table 6: R^2 of LR, RR and PLSR when P=4 and n=90

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	0.00201*	0.08859	0.13471
RR	0.01638	0.08811	0.08811*
PLSR	0.08807	0.26518	1.98122

Table 6 which shows the values of R^2 when P=4 and n=90. It revealed that Lasso Regression (LR) has minimum error under gamma distribution. This means that, there is higher predictive power observed compared to Partial Least Squares Regression (PLSR) and Ridge Regression (RR) under gamma distribution, while Ridge Regression (RR) has minimum error under beta and chi square distributions than LASSO Regression (LR) and Partial Least Squares Regression (PLSR). However, LR performed best under gamma distribution while RR is the best under beta and chi square distributions.

Table 7: MSLE of LR, RR and PLSR when P=10 and n=60

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	15.46718	10.74506	5.66001
RR	9.46836*	10.11358*	0.55281*
PLSR	19.46713	16.78518	11.87111

Table 7 represents Mean Square Log Error (MSLE) of Lasso Regression (LR), Partial Least Squares Regression (PLSR) and Ridge Regression (RR). The results revealed that Ridge Regression (RR) has minimum error under gamma, beta and chi square distributions, This means that, there is higher predictive power observed compared to Partial Least Squares Regression (PLSR) and LASSO Regression (LR) under those distributions. However, RR performed better under gamma, beta chi square distributions.

Table 8: MSLE of LR, RR and PLSR when P=10 and n=90

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	21.03623	6.03049	9.11401
RR	13.03634	2.89217*	9.08934*
PLSR	4.96546*	5.546193	1.91891

Table 8 shows that Partial Least squares Regression (PLSR) has minimum error under gamma distribution, This means that, there is higher predictive power observed compared to LASSO Regression (LR) and Ridge Regression (RR) gamma distribution, while Ridge Regression (RR) has minimum error under in both beta and chi square distributions.

Table 9: MAE 0f LR, RR and PLSR when P=10 and n=60

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	0.51204	0.24000	1.76542*
RR	0.51144	0.23652	2.89200
PLSR	0.51142*	0.23645*	1.90408

From Table 9, the results shows that Partial Least Squares Regression (PLSR) has minimum error under both gamma and beta distributions, this means that, there is higher predictive power observed compared to LASSO Regression (LR) and Ridge Regression (RR) under those distributions, while Lasso Regression (LR) has minimum error under chi square distribution.

Table 10: MAE of LR, RR and PLSR when P=10 and n=90

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	0.66478	0.77729	6.23015
RR	0.66392*	0.23725	0.31900*
PLSR	0.66409	0.23505*	1.97061

Table 10 shows that Ridge Regression (RR) has minimum error under both gamma, chi square distributions, this signifies that, there is higher predictive power observed compared to Partial Least Squares Regression (PLSR) and LASSO Regression (LR), while in beta distribution Partial Least Squares Regression (PLSR) has minimum error.

Table 11: R^2 of LR, RR and PLSR when P=10 and n=60

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	2.18105	0.08623*	0.98990
RR	0.18048*	0.09355	0.30562*
PLSR	1.62261	2.83804	1.08611

Table 11 shows that Ridge Regression (RR) has predictive ability in both gamma and chi square distributions than both PLSR and LR. The results also revealed that when data follows beta distribution LASSO Regression performed better than PLSR and LR.

Table 12: R^2 of LR, RR and PLSR when P=10 and n=90

Regression	Probability Distributions		
	Gamma	Beta	Chi Square
LR	1.00923	0.95861	1.09812*
RR	0.09321*	0.99089	1.10101
PLSR	0.83872	0.87895*	1.96711

From Table 12 which presents the results of R^2 of LASSO Regression (LR), Partial Least Squares Regression (PLSR) and Ridge Regression (RR). The results shows that Ridge Regression (RR) has minimum error under gamma distribution, This means that, there is higher predictive power observed compared to Partial Least Squares Regression (PLSR) and LASSO Regression (LR), while Partial Least Squares Regression (PLSR) has minimum error under beta distribution. Also, when data follows chi square distributions LASSO regression is the most efficient.

4. CONCLUSION

In this research work, we compared three biased regression methods namely Lasso, Ridge and Partial Least Squares Regressions when there exists multicollinearity problem in the response variables. Also, we used Mean Square Log Error, Mean Absolute Error and R-Square (R^2) as a predictive ability. From the results of the analysis, we concluded that at P=4 and n=60, Ridge Regression is the most efficient when data follows gamma distribution but when follows beta distribution none of the methods dominate each other (i.e. the methods have equal performance). Also, when data follows chi square distribution PLSR is the most efficient. At P=4 and n=90, RR shows better result in both gamma and beta distributions while in the chi square distribution all methods have predictive ability. Furthermore, at P=10 and n=60, RR performed better in both gamma and chi square distributions while none of the methods dominate each other when data follows beta distribution and lastly, at P=10 and n=90, RR shows result when data follows gamma and chi square distributions while at the beta distribution PLSR performed better.

REFERENCES

[AW14] **Arumairajan S., Wijekoon P.** - *Improvement of Ridge Estimator when Stochastic Restrictions are Available in the Linear Regression Model.* Journal of Statistical and Econometric Methods, 3. 35-48, 2014.

[AW15] **Arumairajan S., Wijekoon P.** - *Optimal Generalized Biased Estimator in Linear Regression Model.* Open Journal of Statistics, 5, 403-411, 2015.

[AAA06] **Adnan N., Ahmad M. H., Adnan R.** - *A Comparative Study on some Methods for Handling Multicollinearity Problems.* Matematik. 22(2): 109-119, 2006.

[Bal70] **Balsely H. L.** - *Quantitative Research Method Business and Economics.* Random House Publisher, New York, 1970.

[DG75] **Mc Donald G. C., Galarneau A.** - *A Monte Carlo Evaluation of some Ridge Type Estimators.* Journal of American Statistical Association, 70, 407-416. http;//.dx.doi.org/10.1080/01621459.1975.10479882, 1975.

[ES15] **Esra P., Suleyman G.** - *The Comparison of Partial Least Squares Regression, Principal Component Regression and Ridge Regression with Multiple Linear Regression for Predicting PM10 Concentration Level Based on Meteorological Parameters.* Journal of Data Science. 13(4): 663-691. 29, 2015.

[ES16] **Esra P., Semra T.** - *The Comparison of Classical and Robust Biased Regression Methods for Determining Unemployment Rate in Turkey: Period of 1985-2012.* Journal of Data Science. 14(4): 739-768, 2016.

[FS14] **El-Fallah M., El-Salam A.** - *A Note on Partial Least Squares Regression for Multicollinearity (A Comparative*

Study). International Journal of Applied Science and Technology. 4(1): 163-169, 2014.

[Gib81] **Gibbons D. G. -** *A Simulation Study of some Ridge Estimators*. Journal of the American Statistical Association. 76(373): 131-139, 1981.

[HK70] **Hoerl A. E., Kennard R. W.** - *Ridge Regression Applications to Non-orthogonal Problems*, Technometrics. 12(1)**:** 69-82, 1970.

[Kha12] **Khalaf G. -** *Improved Estimator in the Presence of Multicollinearity*. Journal of Modern Applied Statistical Methods. 11(1), 152-157, 2012.

[Kib03] **Kibria B. M. G.** - *Performance of some new ridge regression estimators.* Communications in Statistics – Theory and Methods, 32, 419-435, 2003.

[Mye90] **Myers R. H.** - *Classical and Modern Regression with Applications.* 2nd edition, Duxbury Press, 1990.

[RCT16] **R Core Team** - *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org, 2016.

[Tib96] **Tibshirani R.** - *Regression Shrinage and Selection via the Lasso.* Journal of the Royal Statistical Society, Series B58, 267-288, 1996.

[Tok16] **Toka O.** - *A Comparative Study on Regression Methods in the presence of Multicollinearity.* Journal of Statisticians: Statistics and Actuarial Sciences IDIA. 9(2): 47-53, 2016.

[UZM17] **Usman U., Zakari Y., Musa Y.** - *A Comparative Study on the Prediction Performance Methods of Handling Multicollineariy Based on Normal and Uniform Distributions.* Journal of Basic and Applied Research International. 22(3): 111-117, 2017.

[U+17] **Usman U., Zakari Y., Suleman S., Manu, F.** - *A Comparison Analysis of Shrinkage Regression Methods of Handling Multicollinearity Problems Based on Lognormal and Exponential Distributions.* MAYFEB Journal of Mathematics. 3(2017): 45-52, 2017.

[WC78] **Wichern D., Churchill G.** - *A Comparison of Ridge Estimators.* Technometrics, 20, 301-311, 1978.