Advances and Applications in Statistics

http://www.pphmj.com
http://dx.doi.org/10.17654/0972361723060
Volume 89, Number 2, 2023, Pages 243-261 **P-ISSN: 0972-3617**

CALCULATING THE SAMPLE SIZE FOR ORDINARY LEAST SQUARE ESTIMATION IN PRESENCE OF MULTICOLLINEARITY

O. Oyewole[1], Adedayo F. Adedotun[2,3,*], J. A. Adeyiga[4], A. A. Oyewole[2,3], Toluwalase J. Akingbade[2,3], Oluwatosin O. Onayemi[5] and Onuche G. Odekina[2,3]

[1]Department of Physical Sciences
Bells University of Technology
Ota 11001, Ogun State, Nigeria

[2]Department of Mathematics
Covenant University
Ota 11001, Ogun States, Nigeria
e-mail: adedayo.adedotun@covenantuniversity.edu.ng

[3]Department of Statistics
Allover Central Polytechnic
Ota 11001, Ogun States, Nigeria

Received: February 24, 2023; Revised: May 10, 2023; Accepted: June 7, 2023
2020 Mathematics Subject Classification: 62P20, 62R07, 62J05.
Keywords and phrases: multicollinearity, L2 ridge shrinkage method, and ordinary least squares (OLS) method.
*Corresponding author

How to cite this article: O. Oyewole, Adedayo F. Adedotun, J. A. Adeyiga, A. A. Oyewole, Toluwalase J. Akingbade, Oluwatosin O. Onayemi and Onuche G. Odekina, Calculating the sample size for ordinary least square estimation in presence of multicollinearity, Advances and Applications in Statistics 89(2) (2023), 243-261.
http://dx.doi.org/10.17654/0972361723060

Published Online: October 13, 2023

[4]Department of Computer Science and Information Technology
Ota 11001, Ogun States, Nigeria

[5]Department of Economics
Covenant University
Ota 11001, Ogun States, Nigeria
e-mail: oluwatosin.onayemi@covenantuniversity.edu.ng

Abstract

The relationship between a variable (the response variable) and the scores of several other variables (the independent variables) may be described using multiple linear regression analysis. This study compares the L2 (ridge) shrinkage method and least squares shrinkage method when multicollinearity is present in a dataset across various sample sizes. For different sample sizes ($n = 25$, $n = 50$, $n = 200$, and $n = 1000$), this process was repeated. The relationship between larger sample sizes and covariance was not linear in the simulated data. The results demonstrated that L2 regression is best and generates parsimonious models in the presence of multicollinearity; the higher the degree of multicollinearity, the smaller the shrinkage parameter. The L2 regularization technique also helps to reduce standard errors of regression coefficients and the prediction error of the generated model. This implies that for every change in the dataset values, there is always an optimal value of the shrinkage parameter (λ) that minimizes multicollinearity and produces more stable and reliable regression models. In moderation studies where we would like to keep all of the predictor variables, L2 regularization would be the best alternative. Increasing sample size gives stable results after estimation as it helps to reduce the standard errors of the regression coefficients of the predictor variables. It is also the best method to use for greatly inflated standard errors of OLS regression coefficients. OLS works best for independent samples, but correlated covariates should be handled with modern regression methods (L2).

The findings showed that the L2 approach is the best when there is significant multicollinearity in the dataset because it produces smaller

standard errors. However, the L2 regularization produced the lowest MSE across all sample sizes; as a result, combining the two techniques on a dataset at the same time was also examined. The level of the data's multicollinearity should be the main focus of future research, which should also determine the best robust approach for each level.

1. Introduction

According to regression theory, a variable and a group of other variables have a stochastic relationship [1]. In other words, the observed variable Y (also known as the dependent, endogenous, or explained variable) depends on other observed variables (also known as the independent, exogenous, or explanatory variables) [2]. However, this model's independent explanatory variables are one of its underlying premises. In terms of economic variables, this is not frequently the case. Age and the number of years of experience are two variables that do show some sort of linear relationship. When this presumption is incorrect, a multicollinearity issue arises [2].

When underfitting and reducing overfitting are controlled, a statistical model performs best. Stability, model performance, interpretability, and bias reduction properties should all be present in a good statistical model. The least-squares method is one of the most widely used statistical modeling techniques for fitting linear models [3]. Under multicollinearity conditions, this method, which is known as the best linear unbiased estimator (BLUE), falls short of the model fitting objectives.

Some suggested methods to reduce the prediction error include regularization methods [3]. In order to achieve a model with a lower Mean Square Error (MSE), which is more stable and precise in prediction scores, very little bias is introduced into the system. This study examines the idea of multicollinearity as well as methods that have been suggested for identifying and resolving the problems looked into. Using the MSE criterion, the study compares the L2 (ridge) regularization method to conventional least squares. Hoerl and Kennard [4] and Obadina et al. [5] first proposed ridge regression, and it has since gained popularity as a method for addressing the multicollinearity effect in multiple linear regression models. Several studies

have lately examined several techniques for dealing with the multicollinearity effect in multiple linear regression models [6-12]. In the study, methods for choosing the regularization constant are presented along with the ridge estimator and its properties.

2. Method and Materials

2.1. Ordinary least square

Consider the linear model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \xi_i$$

$$= \beta_0 + \sum_{p=1}^{k} \beta_p X_{ip} + \xi_i. \quad (1)$$

The above in matrix form is as follows:

$$Y = X\beta + \xi. \quad (2)$$

The expectation of Y is given as

$$E(Y_i) = E(X\beta + \xi) = X\beta. \quad (3)$$

The variance of Y is given as

$$Var(Y_i) = Var(X\beta + \xi) = \sigma_i^2. \quad (4)$$

The goal of ordinary least squares is to minimize the sum of squared differences between the observed and the predicted values of the L2 norm of the Beta vector [13].

From $Y = X\beta + \xi$, we have

$$\xi = Y - X\beta, \quad (5)$$

$$\xi^T \xi = (Y - X\beta)^T (Y - X\beta) \quad (6)$$

$$= (Y^T - \beta^T X^T)(Y - X\beta). \quad (7)$$

Therefore,

$$\xi^T\xi = (Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta). \tag{8}$$

Differentiating with respect to β and setting the derivative to zero helps us to find the optimal value for β:

$$\frac{\partial\xi}{\partial\beta} = -2X^TY + 2X^TX\beta, \tag{9}$$

$$\frac{\partial\xi}{\partial\beta} = -2X^TY + 2X^TX\beta \text{ but } \frac{\partial\xi}{\partial\beta} = 0. \tag{10}$$

These imply that

$$-2X^TY + 2X^TX\beta = 0. \tag{11}$$

Therefore,

$$X^TX\beta = X^TY. \tag{12}$$

Rearranging the formula gives

$$\hat{\beta} = (X^TX)^{-1}X^TY. \tag{13}$$

This equation is used to compute the β parameter estimates.

2.2. Model performance and accuracy of the OLS estimator

Expectation, variance and MSE of the estimator are, respectively, given as follows:

$$E(\hat{\beta}) = E[(X^TX)^{-1}X^TY] \tag{14}$$

$$= [(X^TX)^{-1}X^TE(Y)] \tag{15}$$

$$= (X^TX)^{-1}X^TX\beta. \tag{16}$$

Therefore,

$$E(\hat{\beta}) = \beta. \tag{17}$$

Hence, $\hat{\beta}$ is unbiased in estimating β, and we have

$$Var(\hat{\beta}) = Var[(X^T X)^{-1} X^T Y] \tag{18}$$

$$= (X^T X)^{-1} X^T Var(Y) X (X^T X)^{-1} \tag{19}$$

$$= (X^T X)^{-1} \sigma_i^2 X (X^T X)^{-1}, \tag{20}$$

$$Var(\hat{\beta}) = \sigma_i^2 (X^T X)^{-1}. \tag{21}$$

Again,

$$MSE(\hat{\beta}) = \hat{\sigma}^2 trace(X^T X)^{-1}. \tag{22}$$

Hence,

$$MSE(\hat{\beta}) = \hat{\sigma}^2 \sum_{i=1}^{p} \frac{1}{K_i}. \tag{23}$$

2.3. Ridge regression

Ridge regression helps us to reduce the impact of correlated inputs by regularizing the norm of the Beta vector [5, 13, 14]:

$$\hat{\beta} = J(\beta) = \lambda \| \beta_2^2 \|, \tag{24}$$

where λ is the ridge constant for any vector. The l_p norm is defined as

$$(\beta_1^p + \beta_2^p + \cdots + \beta_k^p)^{\frac{1}{p}} \equiv \| \beta \|_k. \tag{25}$$

Regularizing the l_2 norm for a linear model, we have

$$J(\beta) + \lambda \| \beta_2^2 \|. \tag{26}$$

In case of a linear regression, the loss function takes the form

$$J(\beta) = \sum_{i=1}^{N} (Y_i - \beta^T X_i)^2 + \lambda \| \beta_2^2 \|. \tag{27}$$

We want to find the value of β that minimizes the above function. To optimize the function, we have to differentiate the function with respect to β and set the derivative to zero. In particular, we know that

$$J(\beta) = (Y - X\beta)^T (Y - X\beta) + \lambda \beta^T \beta. \tag{28}$$

It implies that

$$\frac{\partial J(\beta)}{\partial \beta} = -2X^T (Y - X\beta) + 2\lambda\beta. \tag{29}$$

Now by setting $\frac{\partial J(\beta)}{\partial \beta} = 0,$ we have

$$-2X^T (Y - X\beta) + 2\lambda\beta = 0. \tag{30}$$

Solving for $\beta,$ we generate the solution

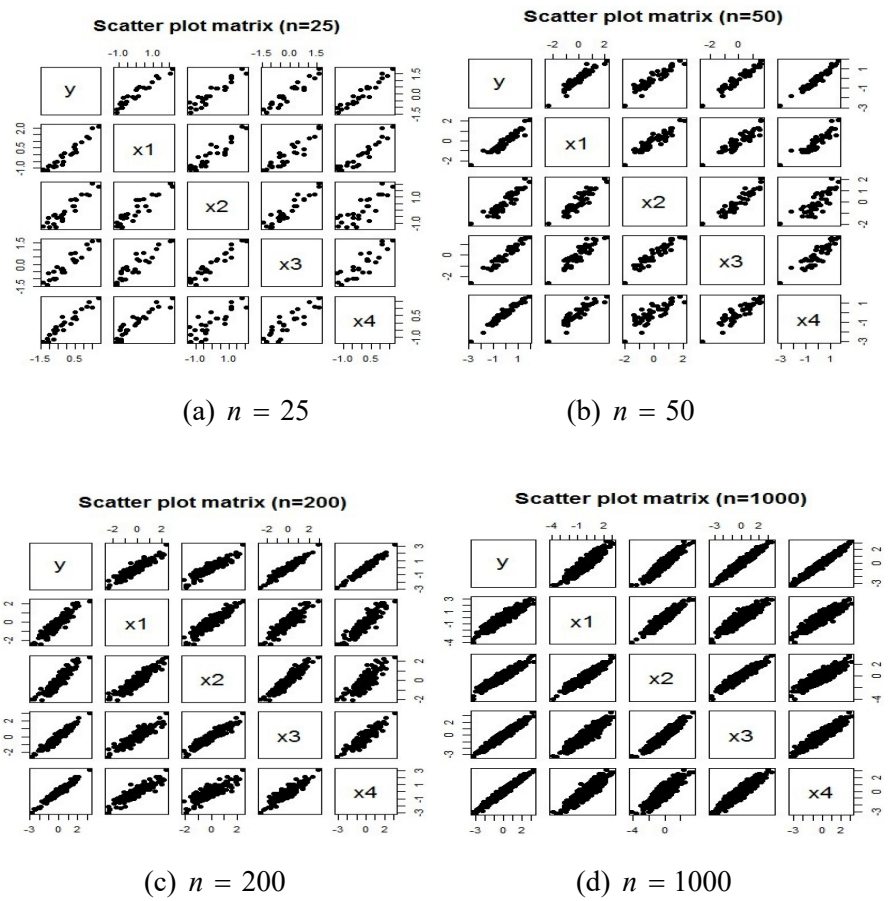
$$\hat{\beta}_{RR} = (X^T X + \lambda I)^{-1} X^T Y, \tag{31}$$

where I is an identity matrix.

3. Results and Discussion

3.1. Performance of estimators with increasing sample size

The simulated dataset's matrix plot for various sample sizes is shown in Figure 1. It displays the scatter plots of the simulated dataset's response and predictor variables.



(a) $n = 25$ (b) $n = 50$

(c) $n = 200$ (d) $n = 1000$

Figure 1. Scatter plot of simulated data for different sample sizes.

Figure 1 shows that all of the independent variables are highly positively correlated with one another and that there is a linear relationship between them. As the sample size rises, the data points in the matrix plot get smaller. Pragmatically, this might be the case if more samples are taken from a population that is very homogeneous because they might all show very similar traits.

Following are the correlation matrix tables for the simulated dataset for various sample sizes.

Table 1. Simulated dataset correlation matrix for $n = 25$

$n = 25$	Y	X_1	X_2	X_3	X_4
Y	1	0.88322	0.92835	0.92660	0.97103
X_1		1	0.91273	0.83321	0.91033
X_2			1	0.90425	0.87658
X_3				1	0.85925
X_4					1

Table 2. Simulated dataset correlation matrix for $n = 50$

$n = 50$	Y	X_1	X_2	X_3	X_4
Y	1	0.87771	0.92616	0.94428	0.97250
X_1		1	0.90807	0.84841	0.90175
X_2			1	0.91918	0.87299
X_3				1	0.88772
X_4					1

Table 3. Simulated dataset correlation matrix for $n = 200$

$n = 200$	Y	X_1	X_2	X_3	X_4
Y	1	0.90780	0.92770	0.95635	0.97259
X_1		1	0.91883	0.88993	0.90216
X_2			1	0.92713	0.85399
X_3				1	0.90826
X_4					1

Table 4. Simulated dataset correlation matrix for $n = 1000$

$n = 1000$	Y	X_1	X_2	X_3	X_4
Y	1	0.92105	0.94131	0.96338	0.97527
X_1		1	0.90807	0.84841	0.90175
X_2			1	0.91918	0.87299
X_3				1	0.85925
X_4					1

p-value $= 0.000$, $RSE = 0.09347$, $R^2 = 0.9908$, $Adj.\ R^2 = 0.989$.

Table 5. Simulated dataset OLS regression output for $n = 25$

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	0.00084	0.01929	0.044	0.96570
X_1	-0.01339	0.08425	-0.159	0.87535
X_2	0.30626	0.07540	4.282	0.0036
X_3	0.17686	0.07540	2.346	0.02942
X_4	0.57121	0.06732	8.485	0.00000

Even though all of the predictors had positive correlations with the response variable, Table 5 shows that X_1 has a negative coefficient (Y). We anticipate very little variability if we keep resampling because the standard error values are smaller. At a level of 0.05, some of the variables, like X_1, are not statistically significant.

The OLS output of the simulated dataset for a sample size of 50 is displayed in the table below.

Table 6. The results of the simulated data set's OLS regression for $n = 50$

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	0.00064	0.01219	0.052	0.95900
X_1	-0.06785	0.04776	-0.421	0.16200
X_2	0.29021	0.04765	6.068	0.00000
X_3	0.21466	0.04765	4.505	0.00000
X_4	0.60932	0.03664	16.631	0.00000

$$F4,\ 45 = 1469,\ p\text{-value} = 0.000,\ RSE = 0.08601,$$

$$R^2 = 0.9924,\ Adj.\ R^2 = 0.9917.$$

The standard error of the predator variables has decreased, and Table 6 shows that the number of significant variables has increased at $= 0.05,$ to use as an example.

The OLS of the simulated dataset for a sample size of 200 is displayed in the table below.

Table 7. The results of the simulated data set's OLS regression for $n = 200$

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	0.00184	0.00657	0.281	0.77900
X_1	-0.010502	0.02319	-0.4528	0.00001
X_2	0.34074	0.02581	13.201	0.00000
X_3	0.18364	0.02637	6.963	0.00000
X_4	0.61130	0.02003	30.516	0.00000

$$F4, 195 = 4986,\ p\text{-value} = 0.000,\ RSE = 0.09226,$$

$$R^2 = 0.9903,\ Adj.\ R^2 = 0.9901.$$

According to Table 7, all the variables would have been deemed significant at, let us say, a *p*-value of 0.05. Their standard errors have since risen to 9.2%, but the significance of the predictors has grown as *p*-values have dropped. The percentage of variation in the response variable that is explained by the predictor variables is indicated by the *R*-squared. It evaluates how well the model fits the data.

The OLS output of the simulated dataset for a sample size of 1000 is displayed in the table below.

Table 8. Simulated dataset's OLS regression output for $n = 1000$

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	-0.00425	0.00295	-1.438	0.15100
X_1	-0.13916	0.00988	-14.091	0.00000
X_2	0.33910	0.01110	30.564	0.00000
X_3	0.20930	0.01086	19.270	0.00000
X_4	0.61502	0.00887	69.360	0.00000

$$F4,\ 995 = 2.813 \times 104,\ p\text{-value} = 0.000,\ RSE = 0.09327,$$

$$R^2 = 0.9912,\ Adj.\ R^2 = 0.9912.$$

From Table 8, it is clear that a larger sample size resulted in significantly lower standard errors. This means that the sample size of the dataset has a significant impact on the stability of a model. Comparing the cases for $n = 25,\ n = 50,$ and $n = 200,$ the predictor variables' level of significance has increased.

Despite the fact that X_1 was positively correlated with Y across all sample sizes, it still had a negative sign. The data may have multicollinearity, which will be further examined using the VIFs.

The variance inflation factors for the four predictor variables of the simulated data for various sample sizes are displayed in Table 9. When compared to the situation where the independent variables are strictly uncorrelated, the variance inflation measures how much the variance of the regression coefficients have been inflated. Serious multicollinearity is indicated by VIF values greater than 5.

Table 9. VIFS' of the simulated dataset for various sample sizes are shown in Table 9

VIF	$n = 25$	$n = 50$	$n = 200$	$n = 1000$
X_1	17.9780	12.4855	10.5131	11.0825
X_2	13.5777	12.7499	12.5500	13.8634
X_3	12.2999	13.0577	13.9575	13.2335
X_4	8.9445	7.9278	8.3723	8.9871

Table 9 shows that across all sample sizes, the majority of the predictor variables have $VIFs > 10.$ This indicates that the simulated dataset exhibits severe multicollinearity.

3.2. OLS and ridge coefficients

Table 10. Coefficients of regression for 25

	Estimates (OLS)	**Estimates (RR)**
Intercept	0.00084	0.00407
X_1	-0.01339	0.18021
X_2	0.30626	0.20519
X_3	0.17686	0.22241
X_4	0.57121	0.38143

$$\hat{Y}_{OLS} = 0.00084 - 0.01339X_1 + 0.30626X_2 + 0.17686X_3 + 0.57121X_4. \tag{32}$$

The equation of the fitted RR model is

$$\hat{Y}_{RR} = 0.00407 + 0.18021X_1 + 0.20519X_2 + 0.22241X_3 + 0.38143X_4. \tag{33}$$

Table 11. 50-person regression coefficients

	Estimate (OLS)	**Estimate (RR)**
Intercept	0.00064	0.00176
X_1	-0.06785	0.14878
X_2	0.29021	0.19981
X_3	0.21466	0.25636
X_4	0.60932	0.41149

The fitted OLS model's equation is

$$\hat{Y}_{OLS} = 0.00064 - 0.06785X_1 + 0.29021X_2 + 0.21466X_3 + 0.60932X_4. \tag{34}$$

The fitted RR model's equation is

$$\hat{Y}_{RR} = 0.00176 + 0.14878X_1 + 0.19981X_2 + 0.25636X_3 + 0.41149X_4. \tag{35}$$

Table 12. 200-person regression coefficients

	Estimate (OLS)	**Estimate (RR)**
Intercept	-0.00426	-0.00604
X_1	-0.13916	0.09565
X_2	0.33910	0.21952
X_3	0.20930	0.27964
X_4	0.61502	0.39904

The fitted OLS model's equation is

$$\hat{Y}_{OLS} = -0.00426 - 0.13916X_1 + 0.33910X_2 + 0.20930X_3 + 0.61502X_4. \tag{36}$$

The fitted RR model's equation is

$$\hat{Y}_{RR} = -0.00604 + 0.09565X_1 + 0.21952X_2 + 0.27964X_3 + 0.39904X_4. \tag{37}$$

Table 13. Coefficients of regression for 1000

	Estimate (OLS)	**Estimate (RR)**
Intercept	-0.00951	0.00234
X_1	-0.19151	0.11208
X_2	0.38046	0.21660
X_3	0.17039	0.26522
X_4	0.66239	0.40529

The equation of the fitted OLS model is

$$\hat{Y}_{OLS} = -0.00951 - 0.19151X_1 + 0.38046X_2 + 0.17039X_3 + 0.66239X_4. \tag{38}$$

The fitted RR model's equation is

$$\hat{Y}_{RR} = 0.00234 + 0.11208X_1 + 0.21660X_2 + 0.26522X_3 + 0.40529X_4. \tag{39}$$

Table 14. Independent variable eigenvalues

Variable	X_1	X_2	X_3	X_4
Eigenvalues	0.1376	0.1105	0.0370	0.0062

Table 14 displays the eigenvalues of the predictor variables in the design matrix of the simulated dataset based on the information presented above. It can be seen that there is a significant difference between the maximum and minimum eigenvalues of the independent variables. This suggests that the explanatory variables exhibit severe multicollinearity.

The shrinkage parameters for RR after cross validation are shown in Table 15.

Table 15. Ridge shrinkage parameters

Sample size(n)	**RR** λ_{LSE}	**RR** $\lambda_{\min}$
25	0.1774	0.0925
50	0.1446	0.0971
200	0.1305	0.0987
1000	0.1063	0.1063

In particular, $\lambda(LSE)$ and $\lambda(\min)$ represent the range of λ values for which the regularized solutions have a smaller MSE than the OLS solution, where $\lambda(LSE)$ is the maximum and $\lambda(\min)$ is the minimum. $\lambda(LSE)$ is inversely proportional to the sample size increase whereas the $\lambda(\min)$ is directly proportional to the sample size. The minimum value is always chosen as our shrinkage parameter.

Table 16. MSEs and MAEs for OLS and RR

N	**MSE(OLS)**	**MAE(OLS)**	**MSE(RR)**	**MAE(RR)**
25	2.1050	1.2309	0.0102	0.0845
50	2.3977	1.2497	0.0121	0.0876
200	2.3237	1.2482	0.0144	0.0955
1000	2.3309	1.2339	0.0162	0.1016

According to Table 16, RR values have mean absolute errors (MAE) that are less than those of OLS. Additionally, RR values have lower MSEs for the regression coefficient than OLS. Ridge regression method, in this case for the four-predictor variable model, is therefore superior to OLS when the multicollinearity problem in a data exists.

3.3. The regression coefficients' standard errors

The standard error of $\hat{\beta}$ estimate is a measure of how consistent $\hat{\beta}$ will be if re-sampled repeatedly. It measures the sampling variation in estimating β according to [15]. Tables 17, 18, 19 and 20 show the standard errors of the estimates of the simulated dataset across different sample sizes.

Table 17. Standard errors for $n = 25$

$n = 25$	**OLS**	**RR**
X_1	0.08425	0.00761
X_2	0.07152	0.00489
X_3	0.07540	0.00213
X_4	0.06732	0.00546

Table 18. Standard errors for $n = 50$

$n = 50$	**OLS**	**RR**
X_1	0.04776	0.00337
X_2	0.04782	0.00239
X_3	0.04765	0.00148
X_4	0.03664	0.00276

Table 19. Standard errors for $n = 200$

$n = 200$	**OLS**	**RR**
X_1	0.04776	0.00337
X_2	0.04782	0.00239
X_3	0.04765	0.00148
X_4	0.03664	0.00276

Table 20. Standard errors for $n = 1000$

$n = 1000$	**OLS**	**RR**
X_1	0.00988	0.00021
X_2	0.01110	0.00017
X_3	0.01086	0.00013
X_4	0.00887	0.00018

4. Conclusion

We discussed the multicollinearity problem, techniques for identifying it, and how it affects the output of multiple regression models. According to the simulation study, multicollinearity is caused by the covariates' high covariances. We were able to use OLS to fit a multiple linear regression to the data because a graph of the matrix plot revealed a linear relationship between the response and each of the predictor variables. Despite having a positive correlation with Y as shown in Table 6, the coefficient of X_1 was negative, and the intercept had a very low value. Even though the standard errors were not very high, this was the first sign that multicollinearity might exist in the dataset. For different sample sizes $(n = 25,\ n = 50,\ n = 200$ and $n = 1000)$, this process was repeated. The level of significance of the predictor variables increased along with a reduction in the standard errors of the predictors for larger sample sizes. The overall model was always significant, even though some predictors might not have been at some given alpha 0.05, for example. Even without consulting the correlation matrix, this was the second sign that the predictor variables were correlated. To confirm the suspicion of the collinear variables, a more thorough multicollinearity check (VIF) was adopted. According to the general rule, Table 9's VIFs showed that the majority of the predictor variables had VIFs greater than ten, which denotes multicollinearity.

In order to address the issue of collinearity in the simulated dataset, L2 regularization technique was used. The study compared the MSEs and standard errors of OLS and L2 regularization to determine which of these methods produced a more accurate and reliable model as well as lowering the standard errors of the regression coefficients.

The prediction error decreases as an estimator's MSE decreases. This means that, despite its importance, unbiasedness should not be the sole factor to be considered while choosing between competing estimators. In the collinear simulated dataset, L2 also had the lowest MAE across all sample sizes. This indicates that regularized regression methods are more effective when multicollinearity is present. L2 regression should be used in

moderation studies because ridge regression does not remove parameters. Across all sample sizes, RR performed best in terms of their standard errors. As the sample size was increased, the OLS's standard errors shrank. Our simulated data showed that there was a skewed relationship between sample size and covariance.

Acknowledgement

The authors thank the anonymous referees for their valuable suggestions and comments which led to the improvement of the paper.

References

[1] T. O. Olatayo and A. F. Adedotun, On the test and estimate of fractional parameter in AFRIMA model, Applied Mathematical Science 8(96) (2014), 4783-4796. http://dx.doi.org/10.12988/ams.2014.46498

[2] S. Chatterjee, A. S. Hadi and B. Price, Regression Analysis by Examples, 3rd ed., Wiley VCH, New York, 2000. http://dx.doi.org/10.1002/0470055464

[3] D. A. Agunbiade and O. Oyewole, Using the Monte-Carlo method, regression techniques are compared in the presence of multicollinearity and autocorrelation phenomena, Computer Science Journal from the Annals Series XVIII (2020), 70-77.

[4] A. E. Hoerl and R. W. Kennard, Regression on the ridge: biased estimation for nonorthogonal, Technometrics 12(1) (1970), 55-67. http://dx.doi.org/10.1080/00401706.1970.10488634

[5] O. G. Obadina, O. A. Odusanya and A. F. Adedotun, Ridge estimation's effectiveness for multiple linear regression with multicollinearity: an investigation using Monte-Carlo simulations, Journal of the Society of Physical Sciences 3 (2021), 278-281. https://doi.org/10.46481/jnsps.2021.304

[6] C. C. Emioma and S. O. Edeki, Stock price prediction using machine learning on least-squares linear regression basis, J. Phys.: Conf. Ser. 1734 (2021), 012058. DOI 10.1088/1742-6596/1734/1/012058

[7] H. Duzan and N. S. B. M. Shariff, Review of techniques and models for ridge regression in the multi-collinearity problem, Applied Sciences Journal 15(3) (2015), 392.

[8] D. Hauser, Using stepwise regression techniques in geographical research has some issues, Le Géographe Canadien/Canadian Geographer 18(2) (1974), 148-158.

[9] Remi J. Dare, Olumide S. Adesina, Pelumi E. Oguntunde and Olasunmbo O. Agboola, Adaptive regression model for highly skewed count data, International Journal of Mechanical Engineering and Technology 10(1) (2019), 1964-1972.

[10] I. H. Okagbue, E. P. Oguntunde, C. M. Emmanuela and M. A. Elvir, Trends and usage pattern of SPSS and Minitab software in scientific research, Journal of Physics: Conference Series 1734 (2021), 012017. doi:10.1088/1742-6596/1734/1/012017

[11] P. Filzmoser and C. Croux, Explanatory variable dimension reduction in multiple linear regression, Pliska Studia Mathematica Bulgarica 14(1) (2003), 59-70.

[12] R. Grewal, H. Baumgartner and J. A. Cote, Implications for theory testing from multicollinearity and measurement error in structural equation models, Science of Marketing 23(4) (2004), 519-529.

[13] S.-H. Lee, H.-S. Park and J.-H. Lee and C.-H. Jun, The Selection of Variables and Quality Prediction using Partial Least Squares Regression, Published in Computers and Industrial Engineering, CIE, International Conference, IEEE, 2009, pp. 1302-1307.

[14] I.-G. Chong and C.-H. Jun, Performance of some variable selection techniques in the presence of multicollinearity, Chemometrics and Intelligent Laboratory Systems 78(1-2) (2005), 103-112.

[15] S. C. Ludvigson and S. Ng, A factor analysis of bond risk premia, Technical report, National Bureau of Economic Research, 2009. http://www.nber.org/papers/w15188