Applied Mathematics and Computational Intelligence
Volume 11, No.1, Dec 2022 [123-132]

Robust Ridge Regression Approach for Combined Multicollinearity-Outlier Problem

Aliah Natasha Affindi[1], Sanizah Ahmad[2*]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam

[*] Corresponding author: saniz924@uitm.edu.my

Received: 25 October 2021; Accepted: 8 November 2021; Available online (In press); 29 April 2022

ABSTRACT

Ordinary least squares (OLS) offers good parameter estimates in regression if all assumptions are met. However, if the assumptions are not adhered to due to the presence of combined multicollinearity and outliers, parameter estimates may be severely distorted. Hence, robust parameter estimates were injected into the ridge regression method to produce robust ridge regression models. Therefore, the aim of this study is to investigate the performance of selected robust ridge estimators which include S, M, MM and Least Trimmed Squares (LTS) estimators via a simulation study. Laplace and Cauchy error distributions were introduced as outliers in the simulated data of various sample sizes and levels of multicollinearity. The performance of the estimation methods is investigated using criteria bias and root mean square error (RMSE). The finding indicates that Ridge LTS is the best robust ridge estimator in handling data containing both multicollinearity and outliers due to its smallest value in the RMSE. Applications of the estimators to two benchmark real-life datasets provide similar results.

Keywords: Multicollinearity, Outliers, Ridge regression, Laplace, Cauchy distribution

1 INTRODUCTION

Regression analysis is often used for parameter estimation. One problem usually occurs in estimating regression is multicollinearity, the condition when the independent or predictor variables are highly correlated to each other [1]. The impact of using regression with the occurrence of multicollinearity is it will cause the estimated parameter variance to be greater than the actual value resulting in low precision of estimation [2]. Since the problem of multicollinearity may affect the result in the estimation, ridge regression was introduced [3]. Apart from the multicollinearity issue, another problem that should be considered is the existence of outliers. Outliers are known as individual points that are distinct from other data in a dataset which are also called abnormalities or anomalies [4, 5]. To overcome this problem, some remedial measures such as a robust estimator approach could be considered. Ridge regression is used when multicollinearity exists while a robust estimator is suggested to be used in the occurrence of outliers. However, in the presence of both multicollinearity and outliers in the dataset, ridge regression and robust estimator cannot be used separately as ridge regression is unimmunized to the existence of outliers. This study combined two methods in handling the problem, later known as robust ridge regression. The robust ridge estimators that are recommended to be used include S-estimator, M-estimator, MM-estimator and Least Trimmed Squares (LTS). For that reason, this study will focus on discussing the performance of robust ridge estimators (S, M, MM, LTS) on data containing various level of multicollinearity and outliers.

1.1 Ridge Regression

Ordinary least squares (OLS) is commonly used to estimate regression due to its optimal properties and ability to ease certain computation [6]. The OLS estimator could be defined as in (1).

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

OLS method will only be efficient when all the regression assumptions are fulfilled. However, the existence of multicollinearity might affect the effectiveness of OLS. Hence, Ridge regression is suggested to be used if the dataset displays a multicollinearity problem. The concept of ridge regression method was introduced by Hoerl and Kennard [3] which suggests the introduction of biased into the regression model by adding a ridge parameter known as *k* [7]. Therefore, the ridge regression estimator can be defined as in (2)

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y \tag{2}$$

where *I* is the (*p*× *p*) identity matrix and *k* is the biasing constant.

The effectiveness of the ridge regression method is proven in the study by Kibria and Banik [8] where the value of the mean square error of ordinary least squares is much higher than ridge regression. There are several studies that prove ridge regression outperformed ordinary least squares in handling multicollinearity [9-11]. To determine the value of the scalar ridge parameter, there are several formulae can be used. In this study, the *k* value is determined using the formula suggested by Hoerl and Kennard [3] or also known as k_{HK}. The formula is presented in (3) and (4).

$$k_{HK} = \frac{ps_{LS}^2}{\beta_{LS}'\hat{\beta}_{LS}} \tag{3}$$

where

$$s_{LS}^2 = \frac{\left(Y - X\hat{\beta}_{LS}\right)'\left(Y - X\hat{\beta}_{LS}\right)}{n-p} \tag{4}$$

In using ridge regression method, the ridge regression estimator is said to perform well when the *k* value is greater than 0 and in a positive value [12]. If the value of biasing constant, *k* is equal to 0, it is said to have the same effectiveness as OLS. In mathematical terms, the properties of ridge regression can be described as when *k*> 0, the MSE_{RIDGE}< MSE_{OLS} and when *k* = 0, MSE_{OLS} = MSE_{RIDGE}.

1.2 Robust Estimators

Robust estimator is a method that could produce a reliable result if outliers exist in the dataset. This method functions by reducing and limiting the influence of outlying cases by applying weight to the observations [1]. The importance of applying this method is due to the existence of outliers that may affect the magnitude of the regression coefficient and the coefficient sign [12]. There are several robust estimators commonly used to rectify the outlier problem. In this study, four robust estimators which are S, M, MM, and LTS are investigated to know which estimator performs best.

The S estimator was introduced by [13], also known as an alternative method of M-estimation. While M-estimation uses the median as the estimator of σ, S-estimation defines the estimator, σ as dispersion of residual. S-estimates are the solution that finds the smallest possible dispersion of the residuals $min\ \hat{\sigma}(e_i(\hat{\beta}),...,e_n(\hat{\beta}))$. Rather than minimizing the variance of the residuals, robust S-estimation minimizes a robust M-estimate of the residual scale as in (5)

$$\frac{1}{n}\sum_{i=1}^{n} \quad \rho\left(\frac{e_i}{s}\right) = K \tag{5}$$

where K is a constant and $\rho\left(\frac{e_i}{s}\right)$ is the residual function. Peter Rouseeuw [13] suggested the Tukey's biweight function with the formula given in (6)

$$\rho(x) = \{\frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}\ for\ |x| \leq c\ \frac{c^2}{6}\ for\ |x| > c \tag{6}$$

by setting c=1.5467 and K=0.1995 which gives 50% breakdown point.

One of the most popular robust estimators is the M estimator which is introduced by [14]. The M-estimator technique is based on replacing the sum of squared errors in the least square method by another robust function to cope with the problem of outliers. Rather than minimizing the sum of squared errors, the M-estimate minimizes a function ρ of the errors. The objective function of the M-estimator is as shown in (7)

$$\min\sum_{i=1}^{n} \quad \rho\left(\frac{e_i}{s}\right) = \sum_{i=1}^{n} \quad \rho\ \left(\frac{y_i - X'\hat{\beta}_i}{s}\right) \tag{7}$$

where s is an estimate of scale often formed from a linear combination if the residuals and ρ is a function that assigns the contribution of the individual residual in the objective function. Differentiating the objective function with respect to the coefficients β, defining ψ = $\rho'$and setting the partial derivates to 0, the system of equations can be written as in (8)

$$\sum_{i=1}^{n} \quad \psi\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) x_i = 0 \tag{8}$$

where s is a robust estimate of scale.

The MM estimator is a special type of M estimator introduced by Yohai [15]. This estimator is a high breakdown and high-efficiency estimator described in three stages as follows [7]

Stage 1: Use S-estimate for high breakdown estimator

- Find the initial estimate, $\tilde{\beta}$.
- Then compute the residuals, $r_i(\beta) = y_i - x_i^T\tilde{\beta}$.

Stage 2: Calculate the M estimate for the scale of error.

By using the residuals in Stage 1 and the constant $k = \frac{1}{n}\sum_{i=1}^{n} \quad \rho\left(\frac{r_i}{s}\right)$ where ρ is the objective function an M estimate of scale with 50% breakdown point is computed. The s $\left(r_i(\tilde{\beta}), \dots, r_n(\tilde{\beta})\right)$ is denoted as s_n. The objective function used in this stage is labeled ρ_0.

Stage 3: M estimate of the regression parameters using a descending ψ function that assigns a weight of 0.0 to abnormally large.

The MM estimator is now defined as an M estimator of β using a redescending score function, $\varphi_1(u) = \frac{\partial\rho_1(u)}{\partial u}$ and the scale estimate s_n obtained from Stage 2. So, the MM estimator $\tilde{\beta}$ defined as a solution to (9)

$$\sum_{i=1}^{n} \quad x_{ij}\varphi_1\left(\frac{r_i(\beta)}{s_n}\right) = 0 \quad \text{j=1,…,p} \tag{9}$$

The LTS estimator, introduced by Rousseeuw and Van Driessen [16] is an estimator that is very robust to a small percentage of outliers. This estimator minimizes the sum of trimmed squared residuals and is written in (10) as

$$\hat{\beta}_{LTS} = min\sum_{i=1}^{h} \quad e_i^2(\beta) \tag{10}$$

such that $h = \frac{n}{2} + (\frac{p+1}{2})$ with n and p being sample size and number of parameters respectively, and $e_{(1)}^2 \leq e_{(2)}^2 \leq e_{(3)}^2 \leq \cdots \leq e_{(n)}^2$, the ordered squared residuals. LTS estimator may be very efficient based on the value of h and the outlier. The largest squared residuals are being excluded from the summation in this method. Therefore, it allows those outlier data points to be excluded completely. Contradictory, LTS estimator may not be efficient if the number of trimmed data points is more than the actual outlier as some good data will be excluded. Furthermore, if the exact numbers of outliers in the data set are trimmed, this method calculation is similar to OLS.

1.3 Robust Ridge Estimators

Since ridge regression can only be used in solving multicollinearity problems and is not immune to the existence of outliers, a new suggested method named robust ridge estimator was introduced. Robust ridge is the combination between ridge regression and robust estimator which can be used to solve two problems that usually occur in regression analysis which are multicollinearity and outliers [7]. This will dampen the effects of both problems in a classical linear regression model. To compute the robust ridge estimator, the formula is shown in (11)

$$\hat{\beta}_{RobustRidge} = (X'X + k_R I)^{-1}X'Y \tag{11}$$

where the $k_R$is the robust ridge parameter. The value is obtained from the robust regression methods explained before. In this study, the formula used to obtain the k_R value is shown in (12)

$$k_R = \frac{pS_{Robust}^2}{\beta'_{Robust}\hat{\beta}_{Robust}} \tag{12}$$

where p is the number of regressor and $S_{Robust}^2$is the robust scale estimator.

2 THE SIMULATION STUDY

We make Monte Carlo simulation comparing Least Squares Estimators (OLS), Ridge S-estimator, Ridge M-estimator, Ridge MM-estimator, and Ridge Least Trimmed Squares (LTS). We use R language to create our program to set up Monte Carlo simulation and this program is available if requested.

2.1 The Simulation Design

A Monte Carlo simulation study has been used to determine which robust ridge estimator performed best in the condition where both multicollinearity and outliers exist in the dataset. For this study, simulated data are generated for moderate and high values of collinearity (ρ= 0.50, 0.90, 0.95) with three different sample sizes (n= 25, 50, 100). The multivariate linear regression used for the simulation study is in (13)

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i \qquad (13)$$

where $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$.

The explanatory variables were generated using (14):

$$x_{ij} = (1 - \rho^2) z_{ij} + \rho z_{ij} \qquad i = 1,2,\dots,n \qquad j = 1,2,3 \qquad (14)$$

where z_{ij} are the independent standard normal random numbers that are held fixed for a given sample size n. The final factor that needs to be added to the model is the outliers, generated from two heavy-tailed distributions, which are the Laplace distribution with mean zero and variance two, and the Cauchy distribution with median zero and scale parameter one.

To access the performance of each robust ridge estimator, 1000 Monte Carlo trials were used by using bias and root mean square error (RMSE) with the formulas given in (15) and (16)

$$\text{Bias} = \underline{\beta_i} - \beta_i \quad \text{where} \quad \underline{\beta_i} = \frac{\sum_{i=1}^{m} \beta_i}{m}, \quad m{=}1000 \qquad (15)$$

and

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{n} \left(\hat{\beta}_i - \beta_i\right)^2} \qquad (16)$$

3 RESULTS AND DISCUSSION

The simulation study with 1000 trials were carried out for three sample sizes with a combination of various level of multicollinearity and outliers are computed. The performance of the methods considered are investigated for $\hat{\beta}_1$ based on the values of bias and RMSE (see Table 1). The performances of the estimators when referring to the bias values are inconsistent, unable to indicate which estimator is best. When referring to the RMSE value, it can be seen that Ridge LTS is the best estimator as compared to other estimators when both multicollinearity and outliers exist in the dataset since the value of RMSE is the lowest. Ridge LTS performs best as the sample size gets larger. The simulation results are similar for $\hat{\beta}_2$ and $\hat{\beta}_3$.

Figure 1 provides the density plots for $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ for 1000 Simulations for Laplace distribution and Cauchy Distribution for multicollinearity level of ρ=0.95 with sample size n=50. It shows that Ridge LTS estimator outperforms other estimators. Similar results apply for other sample sizes and levels of multicollinearity.

Table 1 : Bias and RMSE (in parentheses) of $\hat{\beta}_1$with Laplace and Cauchy Error Distribution for sample size, n=25, 50 and 100

n	**Method**	**Laplace**			**Cauchy**		
		ρ=0.50	ρ=0.90	ρ=0.95	ρ=0.50	ρ=0.90	ρ=0.95
	OLS	**0.0006** (0.4105)	0.0642 (1.6637)	**0.0185** (3.2114)	3.9811 (76.3161)	15.7149 (301.2479)	10.3837 (347.9403)
	Ridge S	0.0898 (0.3526)	0.0391 (0.8296)	0.0428 (1.5100)	0.5308 (0.7435)	0.3195 (1.1835)	0.1970 (1.7309)
25	Ridge M	0.0775 (0.3604)	0.0258 (0.8194)	0.0449 (1.4647)	**0.4825** (0.7733)	**0.2526** (1.3751)	0.1996 (2.0430)
	Ridge MM	0.0789 (0.3589)	0.0278 (0.8033)	0.0381 (1.4202)	0.5159 (0.7371)	0.2918 (0.9876)	0.1920 (1.5087)
	Ridge LTS	0.0965 **(0.3493)**	**0.0241** **(0.7738)**	0.0472 **(1.3622)**	0.5195 **(0.7294)**	0.3143 **(0.9629)**	**0.1717** **(1.4167)**
	OLS	**0.0087** (0.2754)	**0.0815** (1.1507)	0.1099 (2.1716)	**0.4014** (13.9885)	4.6226 (321.6129)	14.6776 (482.7921)
	Ridge S	0.0537 (0.2564)	0.0858 (0.6269)	0.0592 (1.0931)	0.5619 (0.7336)	0.4015 (0.7978)	0.2746 (1.1347)
50	Ridge M	0.0498 (0.2587)	0.0833 (0.6463)	**0.0524** (1.0387)	0.5408 (0.7441)	0.3898 (0.8396)	**0.2706** (1.1593)
	Ridge MM	0.0504 (0.2583)	0.0839 (0.6327)	0.0572 (1.0086)	0.5568 (0.7343)	0.4100 (0.7091)	0.2711 (0.8914)
	Ridge LTS	0.0553 **(0.2559)**	0.0916 **(0.6023)**	0.0661 **(0.9818)**	0.5577 **(0.7305)**	**0.3744** **(0.6813)**	0.2934 **(0.8531)**
	OLS	**0.0003** (0.1875)	**0.0092** (0.7772)	**0.0229** (1.4185)	0.6616 (14.4042)	0.4881 (75.5959)	6.6782 (286.1028)
	Ridge S	0.0223 (0.1807)	0.0296 (0.5020)	0.0319 (0.7547)	0.5443 (0.7187)	0.4484 (0.6651)	0.3211 (0.7507)
100	Ridge M	0.0213 (0.1814)	0.0264 (0.5027)	0.0273 (0.6906)	**0.5372** (0.7205)	0.4353 (0.6683)	0.2771 (0.7363)
	Ridge MM	0.0215 (0.1812)	0.0280 (0.4978)	0.0288 (0.6785)	0.5451 (0.7176)	0.4500 (0.6446)	0.2995 (0.6857)
	Ridge LTS	0.0229 **(0.1805)**	0.0248 **(0.4842)**	0.0429 **(0.6642)**	0.5401 **(0.7152)**	**0.4227** **(0.6320)**	**0.2848** **(0.6650)**

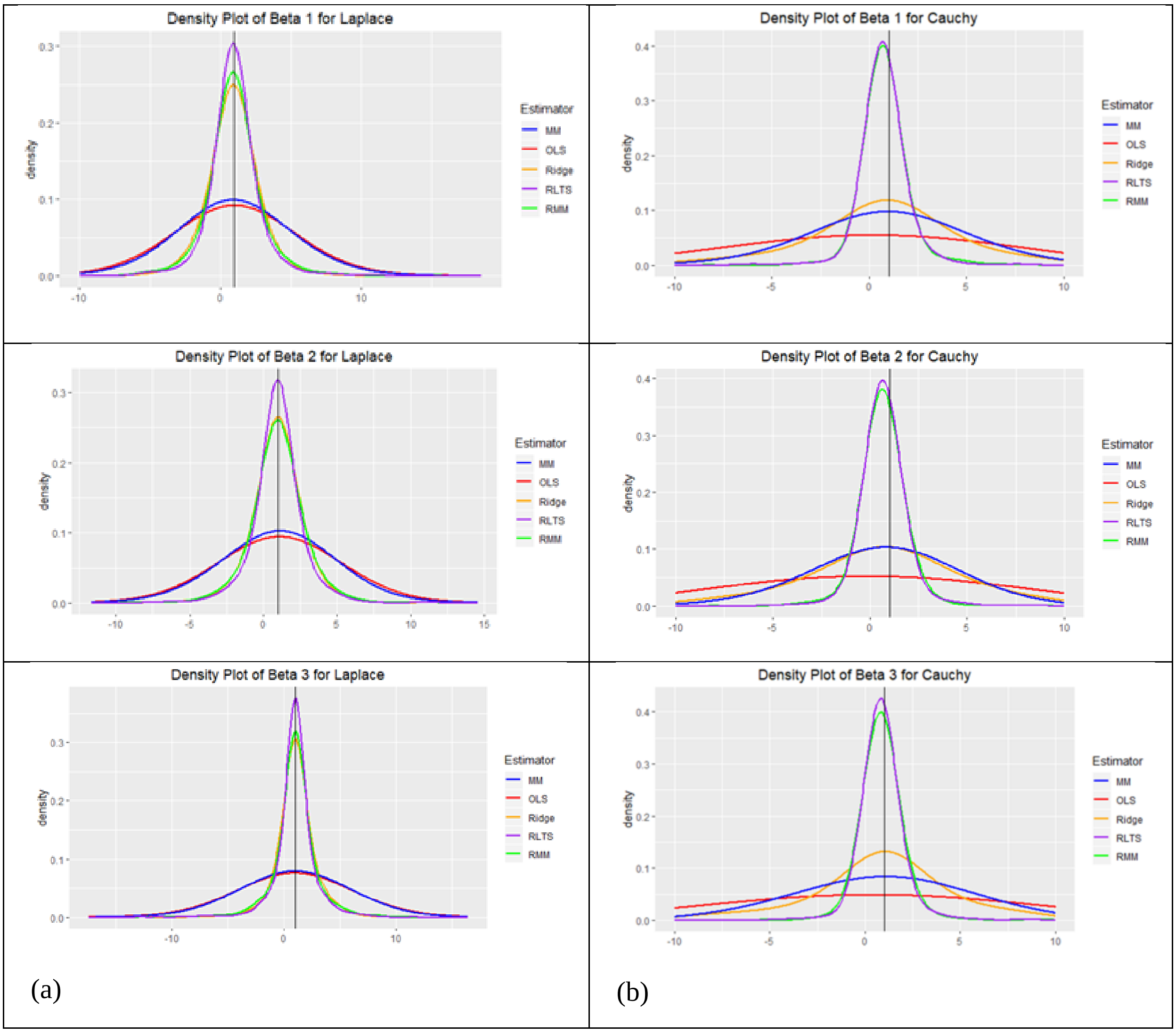

Figure 1 : Density Plots of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ for 1000 Simulations for (a) Laplace Distribution and (b) Cauchy Distribution for ρ=0.95 with n=50

3.1 Application of Real Data

For further investigation, the performance of the estimators are applied to Longley data [17]. This dataset is chosen since the data properties exhibit the interest of study where both multicollinearity and outliers exist in the dataset. All computation was made using R software. Longley data consists of six variables known as Employment, Prices, Unemployed, Military, GNP, and Population Size. GNP is the Gross National Product, employment is the number of people employed, price is the GNP implicit price deflator, unemployed is the number of unemployed, military is the number of people in the armed forces and population size is the non-institutionalized population of persons at age ≥14 years. Table 2 shows the result for the Longley dataset. Based on the result, the lowest standard error (SE) for all variables in Longley data were recorded by Ridge LTS. This proves that the result is parallel with the simulation study where Ridge LTS was found to be the best estimator in the presence of multicollinearity and outliers.

Table 2 : Estimated Parameter Coefficient and Standard Error (SE) of Longley Data for Different Estimators

Estimate	OLS	Ridge S	Ridge M	Ridge MM	Ridge LTS
$\hat{\beta}_1$	0.0151	-0.0040	-0.0060	-0.0049	-0.0068
SE	0.0849	0.0841	0.0840	0.0840	**0.0840**
$\hat{\beta}_2$	-0.0358	-0.0059	-0.0027	-0.0045	-0.0015
SE	0.0334	0.0276	0.0270	0.0274	**0.0268**
$\hat{\beta}_3$	-0.0202	-0.0157	-0.0153	-0.0155	-0.0151
SE	0.0048	0.0040	0.0039	0.0039	**0.0039**
$\hat{\beta}_4$	-0.0103	-0.0090	-0.0089	-0.0090	-0.0089
SE	0.0021	0.0020	0.0020	0.0020	**0.0020**
$\hat{\beta}_5$	-0.0511	-0.1529	-0.1636	-0.1575	-0.1678
SE	0.2260	0.2167	0.2159	0.2164	**0.2156**
$\hat{\beta}_6$	1.8292	1.3300	1.2776	1.3075	1.2566
SE	0.4554	0.3280	0.3146	0.3222	**0.3092**

4 CONCLUSION

Ordinary least squares (OLS) could not perform well in the condition where multicollinearity and outliers exist in the dataset. The suggested method to be used is robust ridge regression where various robust ridge estimators can be used to handle both multicollinearity and outlier problems. A simulation study to examine the performance of several robust ridge estimators proved that Ridge Least Trimmed Squares (LTS) was found to be the best estimator, followed by Ridge MM. Based on two real-life datasets, Longley and Portland Cement data, the results were found to be parallel with the simulation study where Ridge LTS estimator offers the most practical option over other estimators when both multicollinearity and outliers are present.

ACKNOWLEDGEMENT

The authors wish to thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia for their support and encouragement.

REFERENCES

[1] M. H. Kutner, C. J. Nachtsheim, and J. Neter, *Applied Linear Regression Models*, 4 ed. New York: McGraw-Hill Education, 2008.

[2] E. Setiawan, N. Herawati, K. Nisa, Nusyirwan, and S. Saidi, "Handling Full Multicollinearity and Various Numbers of Outliers Using Robust Ridge Regression," *Sci.Int.(Lahore),* vol. 31, no. 2, pp. 201-204, 2019.

[3] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics,* vol. 12, no. 1, 1970.

[4] J. Maervoet, C. Vens, G. V. Berghe, H. Blockeel, and P. D. Causmaecker, "Outlier detection in relational data: A case study in geographical information systems," *Expert Systems with Applications,* vol. 39, no. 5, pp. 4718-4728, 2012.

[5] C. C. Aggarwal, Outlier Analysis, 2 ed. Yorktown Heights, New York: Springer International Publishing, 2016. [Online]. Available: http://charuaggarwal.net/outlierbook.pdf

[6] H. Midi and M. Zahari, "A Simulation Study On Ridge Regression Estimators In The Presence Of Outliers And Multicollinearity," *Jurnal Teknologi,* vol. 47, no. 1, 2007.

[7] A. Lukman, O. Arowolo, and K. Ayinde, "Some Robust Ridge Regression for Handling Multicollinearity and Outlier," *International Journal of Sciences: Basic and Applied Research (IJSBAR),* vol. 16, no. 2, pp. 192-202, 2014.

[8] B. M. G. Kibria and S. Banik, "Some Ridge Regression Estimators and Their Performances," *Journal of Modern Applied Statistical Methods,* vol. 15, no. 1, pp. 206-238, 2016.

[9] L. Firinguetti, G. Kibria, and R. Araya, "Study of Partial Least Squares and Ridge Regression Methods," *Communications in Statistics -Simulation and Computation,* vol. 46, no. 8, pp. 6631-6644, 2017.

[10] G. A. Abdelgadir and H. Eledum, "A Comparison Study of Ridge Regression and Principle Component Regression with Application," *International Journalof Research,* vol. 3, no. 8, 2016.

[11] A. Bager, M. Roman, M. Algedih, and B. Mohammed, "Addressing multicollinearity in regression models: a ridge regression application," presented at the *International Conference of Applied Statistics*, Brasov, 2017.

[12] K. D. Pati, R. Adnan, and B. A. Rasheed, "Ridge Least Trimmed Squares Estimators in Presence of Multicollinearity and Outliers," *Nature and Science,* vol. 12, no. 12, 2014.

[13] P. Rousseeuw, and V. Yohai, “Robust Regression by Means of S-Estimators” in *Robust and Nonlinear Time Series Analysis,* vol. 26, J. Franke, W. Härdle and D. Martin, Eds. New York, NY: Springer, 1984, pp. 256-272.

[14] P. J. Huber, "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics,* vol. 36, pp. 73-101, 1964.

[15] V. J. Yohai, "High Breakdown Point and High Breakdown-Point and High Efficiency Robust Estimates for Regression," *The Annals of Statistics,* vol. 15, pp. 642-656, 1987.

[16] P. J. Rousseeuw and K. Van Driessan, "Computing LTS Regression for Large Data Sets," *Data Mining and Knowledge Discovery,* vol. 12, pp. 29-45, 1998.

[17] S. A. Ajiboye, A. Emmanuel, K. Ayinde and F. L. Adewale, “A Comparative Study of Some Robust Ridge And Liu Estimators,” *Science World Journal,* vol. 11, pp. 16-20, 2016.