A Comparative Study between Ridge MM and Ridge Least Trimmed Squares Estimators in Handling Multicollinearity and Outliers

A N Affindi¹, S Ahmad² and M Mohamad

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Email: ¹aliahnatashaaffindi@gmail.com, ²sanizah@tmsk.uitm.edu.my

Abstract. Common problems found in multiple linear regression models are the existence of multicollinearity and outliers. These obstacles usually produce undesirable effects on least squares estimators. Ridge regression estimator is suggested in handling severe multicollinearity while robust estimators such as MM estimator and Least Trimmed Squares (LTS) estimator are recommended in tackling the outlier issues. An even worse scenario is when these two problems occur simultaneously. Combination of both leads to robust ridge regression methods which can be used to handle both conditions simultaneously. In this study, a comparative investigation is carried out to compare the performance between ridge MM and ridge LTS estimators. The Root Mean Square Error (RMSE) and Bias are obtained for each estimator to compare their performances. By using simulation study, Laplace and Cauchy distributions are used in introducing outliers to the simulated data with high multicollinearity \( \rho = 0.90, 0.95 \) and 0.98 for sample sizes \( n = 25, 50 \) and 100. From the results, it is found that Ridge LTS is the best estimator for many combinations of error distributions and degrees of multicollinearity. Similar results were obtained when using two sets of real data.

1. Introduction

Ordinary Least Squares (OLS) is a method commonly used in estimating parameters for linear regression. This method is said to be the Best Linear Unbiased Estimator (BLUE) when all the assumptions are fulfilled which include independence of error, normal distribution of error and equal variances of error. However, the estimation can be seriously disturbed by the existence of multicollinearity, a condition where there exists high correlation between two or more independent variables and when the independent variables have the same information given to the model [1]. There are many causes of multicollinearity such as the choice of model which is when polynomial model is used, the possibilities of having the multicollinearity problem might increase. Other than that, a model which include lag and carry over effects may also affect the presence of multicollinearity [2]. Additionally, an over defined model which happen when the number of predictors is higher than the number of observations may also cause multicollinearity. Problem of multicollinearity may cause the estimation using OLS become imprecise especially regarding the regression coefficient. Thus, numerous remedial techniques had been introduced to remedy this problem. One of the techniques is ridge regression introduced by [3].

Ridge regression is a method proposed to be used when multicollinearity exist between the independent variables. However, this method is not immune to the existence of outliers. Robust approach is suggested to be used in the situation where outliers exist in the data. Therefore, a method that may solve these two problems concurrently for a model have been applied in order to gain more...
reliable result and one is called robust ridge regression. Nevertheless, there are many estimators found in the literature and among the highly recommended to be used are ridge MM and ridge Least Trimmed Squares (LTS) estimators. Therefore, this paper aims to investigate and compare on the performance between these two robust ridge estimators. This paper is organized as follows: methodology in Section 2 followed result of Monte Carlo simulation results in Section 3, real dataset examples in Section 4 and the last section for concluding remarks.

2. Methodology
This section first discusses on the classical method of ordinary least squares and ridge regression followed by the two robust ridge regression estimators.

2.1 Ordinary Least Squares
Ordinary least squares is a statistical technique used to estimate the parameters of a linear regression equation. The aim of this technique is to determine a line that best fit the data by minimizing the sum of squared errors. The standard regression model is:

\[ Y = X\beta + \epsilon \]  

where \( Y \) is \( n \times 1 \) vector of the dependent variable, \( X \) is \( n \times p \) matrix with full rank, \( \beta \) is \( p \times 1 \) vector of unknown parameter and \( \epsilon \) is the error term. For the ordinary least square estimator, it is defined as

\[ \hat{\beta} = (X'X)^{-1}X'Y \]  

2.2 Ridge Regression
Ridge regression estimator is a method that is more efficient than ordinary least square when the data exhibit problem of multicollinearity. The principal use by this method is by adding a scalar ridge parameter known as \( k \). This \( k \) parameter is added to the main diagonal of \( X'X \) is also called biasing constant. In the presence of multicollinearity, the \( X'X \) matrix is singular and this ill-conditioned matrix may result in poor estimates.

Here, the \( k \) parameter is then added to improve the matrix condition. Hence, the ridge regression estimator is defined as,

\[ \hat{\beta}_R = (X'X+kI)^{-1}X'Y \]  

where \( I \) is the \( (p \times p) \) identity matrix and \( k \) is a scalar ridge parameter.

Various methods had been introduced in order to obtain the value of \( k \). In this study, the method that will be used in finding \( k \) in this study is introduced by [3]. It is shown as below:

\[ k_{HK} = \frac{p\hat{s}^2_L}{\hat{r}^2_{1,LS}} \]  

where

\[ \hat{s}^2_L = \frac{(r-X\hat{\beta}_{LS})'(r-X\hat{\beta}_{LS})}{n-p} \]  

When the value of \( k \) is equal to 0, the result of the ridge estimator will be equivalent to ordinary least square. It is believed that when the value of \( k \) is greater than 0, the result will be more precise and stable for the ridge regression estimator [4].

2.3 Robust Regression Estimator
Robust estimator is known as a method that is reliable to be used in the presence of outliers. This estimator is much more efficient than ordinary least square in analysing a model contaminates with
A usual statistical procedure may be run although there exist outliers in the model. However, it may affect the estimated coefficients, standard errors and test statistics. The procedure used in this robust regression estimator is by using estimators that dampen the impact of outliers. Various estimators had been proposed such as MM estimator and Least Trimmed Square (LTS) estimator.

2.3.1 MM Estimator
The MM estimator is introduced by [5] is a special type of M estimator. This estimator is a high breakdown and high efficiency estimator described in three stages. The first stage is high breakdown by using S estimate while the second stage calculates the M estimate for scale of error from the S estimate residual. For the third stage, it is an M estimate of the regression parameters using a descending ψ function that assigns a weight of 0.0 to abnormally large residuals [6]. Based on the study by [7], the summary for each stage is as follows:

Stage 1: A high breakdown estimator is used to find an initial estimate denoted as $\hat{\beta}$. Using this estimate, the residuals, $r_i(\beta) = y_i - x_i^T \hat{\beta}$ are computed.

Stage 2: By using the residuals and $\sum_{i=1}^{n} \rho \left( \frac{r_i(\beta)}{\hat{s}_n} \right) = k$ where $k$ is a constant and the objective function $\rho$, an M estimate of scale with 50% breakdown point is computed. The $s\left(r_1(\beta), \ldots, r_n(\beta)\right)$ is denoted as $\hat{s}_n$. The objective function used in this stage is labeled $\rho_n$.

Stage 3: The MM estimator is now defined as an M estimator of $\beta$ using a redescending score function, $\varphi_1(u) = \frac{\partial \rho_n(u)}{\partial u}$ and the scale estimate $\hat{s}_n$ obtained from stage 2. So, the MM estimator defined as a solution to

$$\sum_{i=1}^{n} x_{ij} \varphi_1 \left( \frac{r_i(\beta)}{\hat{s}_n} \right) = 0 \quad j=1, \ldots, p$$

(6)

2.3.2 Least Trimmed Squares (LTS) Estimator
LTS estimator, introduced by [8], is an estimator that is very robust to small percentage of outliers. This estimator minimizes the sum of trimmed squared residuals and is written as

$$\hat{\beta}_{LTS} = min \sum_{i=1}^{h} \varepsilon_i^2 (\beta)$$

(7)

such that $h = \left[ \frac{n}{2} + \frac{p+1}{2} \right]$ with $n$ and $p$ being sample size and number of parameters respectively, and $\varepsilon_{(1)} \leq \varepsilon_{(2)} \leq \varepsilon_{(3)} \leq \ldots \leq \varepsilon_{(n-p)}$ 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 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.

2.4 Robust Ridge Regression Estimator
Robust ridge regression which is the combination between ridge regression and robust estimator can be used to handle multicollinearity and outliers [7]. [9] discussed augmented robust estimators as a way of combining biased and robust regression techniques. The combined procedure is based on the fact that robust estimates can be computed using weighted least squares procedure. When both outliers and multicollinearity occur in a data set, it would seem preferred to combine methods for dealing with these problems simultaneously. This statement is supported by [10] which stated that since robust and ridge regression methods are unable to deal with the outliers and multicollinearity problems simultaneously, it seems worthwhile to combine both methods. Various studies had been made in combining both methods. In this study, two chosen best robust ridge estimator from previous study will be examined to know which of these two estimators will perform better. The chosen robust ridge estimators are Ridge MM estimator and Ridge Least Trimmed Square. To illustrate the idea of how the combination of ridge regression and robust estimators work, the formula uses to find the robust ridge parameter for both estimators is shown below.
For ridge MM estimator, MM parameter can be used to determine the biasing parameter \( k \) as:

\[
k_{MM} = \frac{\rho_{\hat{y}\hat{y}}}{\rho_{\hat{x}\hat{x}}^{p}}
\]  

(8)

The MM parameter is used in computing the \( k \) and \( s^2 \) values in order to reduce the effect of outliers. The \( s^2 \) for MM estimator is calculated using Equation (9).

\[
s^2_{MM} = \frac{(y - X\hat{x})^T(y - X\hat{x})}{n-p}
\]  

(9)

where \( s^2 \) is the estimated variance, \( n \) is the sample size and \( p \) is number of estimated parameter. Therefore, the ridge MM parameter is given by:

\[
\hat{\beta}_{RMM} = (X'X + kI)^{-1}X'Y
\]  

(10)

The same procedure is then applied for ridge LTS estimator where the estimation of parameter can be written as:

\[
\hat{\beta}_{RLTS} = (X'X + kI)^{-1}X'Y
\]  

(11)

where the value of \( k \) is determined using:

\[
k_{LTS} = \frac{p\hat{s}_{LTS}^2}{\hat{s}_{LTS}^2}
\]  

(12)

and

\[
s^2_{LTS} = \frac{(y - X\hat{x}_{LTS})^T(y - X\hat{x}_{LTS})}{n-p}
\]  

(13)

3. Simulation Study

A Monte Carlo simulation study has been carried out to assess the performance of Ridge MM and Ridge LTS estimators. The simulation study is designed with three levels of high multicollinearity (\( \rho = 0.90, 0.95 \) and \( 0.98) \) and three different sample sizes (\( n = 25, 50 \) and \( 100) \). Outliers were generated in the simulated data by introducing the non-normal error distributions which are Laplace distribution with mean zero and variance two, and Cauchy distribution with median zero and scale parameter one as suggested by [11].

The model used in the simulation study is:

\[
y_t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon_t
\]  

(14)

where \( \beta_0 = \beta_1 = \beta_2 = \beta_3 = 1 \).

The explanatory variables were generated using the following equation:

\[
x_{ij} = (1 - \rho^2)z_{ij} + \rho z_{ij} \\
\]  

(15)

where \( z_{ij} \) are the independent standard normal random numbers that is held fixed for a given sample of size \( n \). Laplace and Cauchy distributions which are prone to produce outliers were generated and combined with multicollinearity to examine the performance of robust ridge regression estimators. The simulation is repeated for 1000 trials using R. The performance of each estimator is examined using Bias and Root Mean Square Error (RMSE) with the respective formulas:

\[
\text{Bias} = \hat{\beta}_i - \beta_i
\]  

(16)

\[
\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{n} (\hat{\beta}_i - \beta_i)^2}
\]  

(17)
Table 1 presents the simulation result of $\hat{\beta}_1$ for different sample sizes and levels of correlation with multicollinearity and outliers generated from Laplace Error Distribution. It is evident that OLS has the larger bias and RMSE values out of all the estimators. The performances of the estimators when referring to the bias values are inconsistent, unable to indicate which estimator is best. However, the RMSE values confirm that the Ridge LTS consistently performs slightly better than Ridge MM estimator. The simulation results for $\hat{\beta}_2$ and $\hat{\beta}_3$ are similar to $\hat{\beta}_1$ (table is not provided). Next, Table 2 displays the simulation result for $\hat{\beta}_1$ when the model has high multicollinearity and outliers generated from Cauchy Error Distribution. Likewise, Ridge LTS performs better than other estimators followed closely by Ridge MM. This is evident when referring to the RMSE values where Ridge LTS produce the smallest. The result also shows consistency between the sample size 25, 50 and 100. Additionally, the RMSE value for large sample size is much lower than smaller sample size which concludes that larger sample size produces better and reliable results.

Table 1: Bias and RMSE of $\hat{\beta}_1$ with Laplace Error Distribution.

| Method   | $n$ | $\rho = 0.90$ | $\rho = 0.95$ | $\rho = 0.98$ |
|----------|-----|--------------|--------------|--------------|
|          | Bias | RMSE         | Bias         | RMSE         |
| OLS      | 25   | 0.0844       | 1.6659       | 0.0248       | 3.3401       | 0.1480       | 7.9279       |
|          | 50   | 0.0788       | 1.1121       | 0.0252       | 2.2352       | 0.0796       | 5.2808       |
|          | 100  | 0.0364       | 0.7975       | 0.0322       | 1.5222       | 0.0351       | 3.6474       |
| Ridge    | 25   | 0.0810       | 0.9359       | $\textbf{0.0099}$ | 1.7698       | 0.1008       | 4.0179       |
|          | 50   | $\textbf{0.0002}$       | 0.6807       | $\textbf{0.0114}$ | 1.2565       | 0.0428       | 2.6188       |
|          | 100  | 0.0350       | 0.5365       | 0.0297       | 0.8360       | 0.0350       | 1.8518       |
| Ridge MM | 25   | $\textbf{0.0661}$       | 0.8426       | 0.0127       | 1.5750       | $\textbf{0.0448}$ | 3.6372       |
|          | 50   | 0.0161       | 0.6359       | 0.0125       | 1.1880       | 0.0139       | 2.5763       |
|          | 100  | 0.0350       | 0.5168       | $\textbf{0.0073}$ | 0.7739       | 0.0265       | 1.6741       |
| Ridge LTS | 25   | 0.0707       | $\textbf{0.8313}$ | 0.0446       | $\textbf{1.5744}$ | 0.0816       | $\textbf{3.4529}$ |
|          | 50   | 0.0178       | $\textbf{0.6179}$ | 0.0200       | $\textbf{1.0879}$ | 0.0102       | $\textbf{2.2189}$ |
|          | 100  | $\textbf{0.0075}$       | $\textbf{0.5025}$ | 0.0282       | $\textbf{0.7318}$ | $\textbf{0.0019}$ | $\textbf{1.4816}$ |

Figures 1 and 2 provide the density plots for $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ for 1000 Simulations with Laplace and Cauchy Error Distribution, respectively for $n=100$. It can be seen that both the robust ridge estimators outperform the other estimators when multicollinearity and outliers are both present in the data. However, Ridge LTS estimator performs best for both cases.

Table 2: Bias and RMSE of $\hat{\beta}_1$ with Cauchy Error Distribution.

| Method   | $n$ | $\rho = 0.90$ | $\rho = 0.95$ | $\rho = 0.98$ |
|----------|-----|--------------|--------------|--------------|
|          | Bias | RMSE         | Bias         | RMSE         |
| OLS      | 25   | 26.9118      | 921.4280     | 30.8311      | 993.4940     | 27.5670      | 637.3137     |
|          | 50   | 4.2110       | 91.0743      | 14.6435      | 338.3836     | 5.7672       | 732.6458     |
|          | 100  | 1.8849       | 42.3147      | 1.6245       | 89.0728      | 0.7975       | 176.4795     |
| Ridge    | 25   | 0.0810       | 3.1186       | 0.1899       | 6.0036       | 0.1008       | 4.0179       |
|          | 50   | 0.0161       | 0.6359       | 0.0125       | 1.1880       | 0.0139       | 2.5763       |
|          | 100  | 0.0350       | 0.5365       | 0.0297       | 0.8360       | 0.0350       | 1.8518       |
| Ridge MM | 25   | 0.0707       | $\textbf{0.8313}$ | 0.0446       | $\textbf{1.5744}$ | 0.0816       | $\textbf{3.4529}$ |
|          | 50   | 0.0178       | $\textbf{0.6179}$ | 0.0200       | $\textbf{1.0879}$ | 0.0102       | $\textbf{2.2189}$ |
|          | 100  | $\textbf{0.0075}$       | $\textbf{0.5025}$ | 0.0282       | $\textbf{0.7318}$ | $\textbf{0.0019}$ | $\textbf{1.4816}$ |
| Ridge LTS | 25   | 0.0810       | 3.1186       | 0.1899       | 6.0036       | 0.1008       | 4.0179       |
|          | 50   | 0.0161       | 0.6359       | 0.0125       | 1.1880       | 0.0139       | 2.5763       |
|          | 100  | 0.0350       | 0.5365       | 0.0297       | 0.8360       | 0.0350       | 1.8518       |
Figure 1: Density Plots of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ from 1000 Simulations for Laplace Distribution for (a) $\rho=0.90$, (b) $\rho=0.95$ and (c) $\rho=0.98$, with $n=100$. 
Figure 2: Density Plots of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ from 1000 Simulations for Cauchy Distribution (a) $\rho=0.90$, (b) $\rho=0.95$ and (c) $\rho=0.98$ with $n=100$. 
4.0 Real Data Application

Two benchmark datasets are used to examine the performance of the estimators. The datasets are Car dataset [12] and Hawkins-Bradu-Kass dataset [13] which are both available in R.

4.1 Car Dataset

The Car dataset consists of a dependent variable and eleven independent variables which are displacement ($X_1$), horsepower ($X_2$), torque ($X_3$), compression ratio ($X_4$), real axle ratio ($X_5$), carburettor ($X_6$), transmission speeds ($X_7$), overall length ($X_8$), width ($X_9$), weight ($X_{10}$) and type of transmission ($X_{11}$). The purpose of the data is to examine the gasoline mileage performance of 30 different automobiles. To measure the amount of multicollinearity in the dataset, variance inflation factor (VIF) is examined. Six independent variables suffer from multicollinearity problem with variance inflation factor (VIF) value higher than 10 (refer Table 3). The boxplot in Figure 3 shows there are four outliers on the dependent variable. The results in Table 4 concludes that Ridge LTS as the best estimator in handling Car dataset with having lowest standard error for all the estimated parameters.

| Table 3: Variance Inflation Factor for of Car Dataset. |
|--------------------------------------------------------|
| Variable | VIF | Variable | VIF | Variable | VIF |
|----------|-----|----------|-----|----------|-----|
| $X_1$    | 119.4878 | $X_3$ | 7.7292 | $X_9$ | 9.3971 |
| $X_2$    | 42.8008 | $X_6$ | 5.3247 | $X_{10}$ | 85.7443 |
| $X_5$    | 149.2344 | $X_8$ | 11.7613 | $X_{11}$ | 5.1451 |
| $X_7$    | 2.0600 | $X_8$ | 20.9176 |       |       |

| Table 4: Estimated Parameter Coefficient and Standard Error (SE) of Car Dataset. |
|-----------------------------------------------------------------------------|
| Estimate | OLS | Ridge | Ridge MM | Ridge LTS |
|----------|-----|-------|----------|-----------|
| $\hat{\beta}_1$ | -0.0756 | -0.0564 | -0.0604 | -0.0474 |
| SE       | 0.0563 | -0.0564 | 0.0517 | 0.0439 |
| $\hat{\beta}_2$ | -0.0692 | -0.0510 | -0.0530 | -0.0448 |
| SE       | 0.0878 | 0.0799 | 0.0815 | 0.0684 |
| $\hat{\beta}_3$ | 0.1151 | 0.0855 | 0.0921 | 0.0699 |
| SE       | 0.0881 | 0.0805 | 0.0820 | 0.0686 |
| $\hat{\beta}_4$ | 1.4947 | 1.8864 | 2.0994 | 1.4076 |
| SE       | 3.1015 | 1.9056 | 2.1381 | 1.3358 |
| $\hat{\beta}_5$ | 5.8435 | 4.0683 | 4.7369 | 2.6360 |
| SE       | 3.1484 | 2.1797 | 2.5554 | 1.3719 |
| $\hat{\beta}_6$ | 0.3176 | 0.3057 | 0.2563 | 0.4023 |
| SE       | 1.2890 | 1.1682 | 1.2077 | 1.0040 |
| $\hat{\beta}_7$ | -3.2054 | -1.4585 | -2.0412 | -0.3595 |
| SE       | 3.1092 | 2.0143 | 2.4166 | 1.2381 |
| $\hat{\beta}_8$ | 0.1808 | 0.2010 | 0.2012 | 0.2020 |
| SE       | 0.1303 | 0.1135 | 0.1178 | 0.1015 |
| $\hat{\beta}_9$ | -0.3979 | -0.2024 | -0.2449 | -0.1119 |
| SE       | 0.3235 | 0.2394 | 0.2522 | 0.2155 |
| $\hat{\beta}_{10}$ | -0.0051 | -0.0069 | -0.0067 | -0.0074 |
| SE       | 0.0059 | 0.0047 | 0.0050 | 0.0041 |
| $\hat{\beta}_{11}$ | 0.6385 | 0.9821 | 1.0200 | 0.7092 |
| SE       | 3.0217 | 2.2264 | 2.5252 | 1.4771 |
Figure 3: Boxplot for the Dependent Variable (gasoline mileage).

4.2 Hawkins-Bradu-Kass Dataset

The Hawkins-Bradu-Kass dataset is an artificial dataset constructed by [12] consisting one dependent and three independent variables with 75 observations of which 14 are outliers. Checking for the existence of multicollinearity, Table 5 shows that all the independent variables have high multicollinearity problem since the VIF value is higher than 10. Hence, this dataset has both multicollinearity and outlier problems. Table 6 provides the estimated parameters and standard errors for Hawkins-Bradu-Kass dataset. Similarly, Ridge LTS is found to be the best estimator as the standard errors for all parameters is the lowest among the estimators.

Table 5: Variance Inflation Factor for Hawkins-Bradu-Kass Dataset.

| Variable | VIF   |
|---------|-------|
| $X_1$   | 13.4320 |
| $X_2$   | 28.8535 |
| $X_3$   | 33.4325 |

Table 6: Estimated Parameter Coefficient and Standard Error (SE) of Hawkins-Bradu-Kass Data for Different Estimators.

| Estimate | OLS    | Ridge  | Ridge MM | Ridge LTS |
|---------|--------|--------|----------|-----------|
| $\hat{\beta}_1$ | 0.2392 | 0.1174 | 0.0852   | 0.0432    |
| SE      | 0.2625 | 0.1441 | 0.0739   | 0.0174    |
| $\hat{\beta}_2$ | -0.3345 | -0.2493 | 0.0410   | 0.0379    |
| SE      | 0.1551 | 0.1183 | 0.0436   | 0.0183    |
| $\hat{\beta}_3$ | 0.3833 | 0.3546 | -0.0537  | 0.1569    |
| SE      | 0.1288 | 0.0881 | 0.0363   | 0.0167    |

5. Conclusion

A Monte Carlo simulation was implemented to examine the performance of two robust ridge regression estimators when high multicollinearity and outliers exist in the data. The simulation results proved that both Ridge MM and Ridge LTS perform better than other existing estimators. However, Ridge LTS estimator performs better than Ridge MM both multicollinearity and outliers exist in the data. Therefore, it can be concluded that Ridge LTS is suggested to be the best estimator for small sample size. Based on the result from the Car and Hawkins-Bradu-Kass dataset, Ridge LTS is also found to be the better estimator. Hence, it can be concluded that Ridge LTS should be the preferred estimator when a researcher discovers that the dataset contains multicollinearity and outliers. For future research, comparison between Ridge LTS can be made with other current robust methods such as generalized-M estimator (GM) [13].
References

[1] Adnan N, Ahmad M H and Adnan R 2006 A Comparative Study on Some Methods for Handling Multicollinearity Problems *MATHEMATIKA* **22**(2) 109-119
[2] Ofir C and Khuri A 1986 Multicollinearity in marketing models: Diagnostics and remedial measures *(Research In Marketing)* **3** 181-205
[3] Hoerl A E and Kennard R W 1970 Ridge Regression: Biased Estimation for Nonorthogonal Problems *Technometrics* **12**(1) 55-67
[4] Midi H and Zahari M 2007 A Simulation Study on Ridge Regression Estimators In The Presence of Outliers And Multicollinearity *Journal Teknologi* **47** 59-74
[5] Yohai V J 1987 High Breakdown-Point and High Efficiency Robust Estimates for Regression *The Annals of Statistics* **15** 642-656
[6] Wisnowski J W, Simpson J R and Montgomery D C 2002 An Improved Compound Estimator for Robust Regression *Communications in Statistical Simulations* **31**(4) 653-672
[7] Lukman A, Arowolo O and Ayinde K 2014 Some Robust Ridge Regression for Handling Multicollinearity and Outlier *International Journal of Sciences: Basic and Applied Research (IJSBAR)* **16**(2) 192-202
[8] Roussseeuw, P J and Van Driessan K 1998 Computing LTS Regression for Large Data Sets *Data Mining and Knowledge Discovery* **12** 29-45
[9] Arkin R G and Montogomery D C 1980 Augmented Robust Estimators *Technometrics* **22**(3) 333-341
[10] Gad A M and Qura M E 2016 Regression Estimation in the Presence of Outliers: A Comparative Study *International Journal of Probability and Statistics* **5**(3) 65-72
[11] Montgomery D C, Peck E A and Vining G C 2001 *Introduction to Linear Regression Analysis (3rd ed.)* (New York: John Wiley and Sons)
[12] Hawkins D M, Bradu D and Kass G V 1984 Location of Several Outliers in Multiple Regression Data Using Elemental Sets *Technometrics* **26** 197-208
[13] Md Shariff N S and Ferdous N A 2017 An Application of Robust Ridge Regression Model in the Presence of Outliers to Real Data Problem, *Journal of Physics: Conference Series* **890** 012150