Ridge Least Trimmed Squares Estimators in Presence of Multicollinearity and Outliers

Kafi Dano Pati [1], Robiah Adnan [2], Bello Abdulkadir Rasheed [3]

[1,2,3]Department of Mathematical Sciences, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
kafi_dano@yahoo.com

Abstract: The multicollinearity and non-normal errors are common problems in multiple regression models, that produce inappropriate effects on the least squares estimators. So, it is important to use methods of estimation designed to tackle these problems. The proposed method in this paper is the Ridge Least Trimmed of Squares (RLTS). The performance of this method is compared with the Ordinary Least Squares (LS); Ridge Regression (RR) and Ridge Least Absolute Value (RLAV). Bias, Standard Error (SE) and Root Mean Square Error (RMSE) are obtained for all the methods. The efficiency of the proposed method is compared with the alternatives using Mean Squared Error (MSE) ratios. The experimental evidence displays that RLTS is the best among the three estimators for many combinations of errors distribution and degree of multicollinearity.
[Pati K.D., Adnan R., Rasheed B.A. **Ridge Least Trimmed Squares Estimators in Presence of Multicollinearity and Outliers.** *Nat Sci* 2014;12(12):1-8]. (ISSN: 1545-0740). http://www.sciencepub.net/nature. 1

Keywords: Multicollinearity; outliers; ridge regression; robust ridge regression estimation

1. Introduction

Multicollinearity and non-normal errors are important problems considered in a multiple linear regression analysis. The ordinary least squares (OLS) estimators of coefficients are known to have certain best properties when the explanatory variables are not correlated among themselves. Consider the linear regression model as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (1)$$

and the least squares estimates of the parameter β is given by

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2)$$

The least squares estimate of β is determined by minimizing the function, $\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 = \sum_{i=1}^{n}(r_i)^2$ (3)

The property of least squares (LS) is unbiased with minimum variance among all unbiased linear estimators if the errors are independent, identically and normally distributed. But, it is commonly impossible to interpret the estimates of the individual coefficients if the explanatory variables are highly correlated. Such problem is known as multicollinearity problem. Although LS estimates are unbiased in the presence of multicollinearity.

Multicollinearity increases the standard errors of the coefficients. Increased standard errors in turn means that coefficients for some independent variables may be found not to be significantly different from 0, whereas without multicollinearity and with lower standard errors, these same coefficients might have been found to be significant and the researcher may not have come to null findings in the first place.

In other words, multicollinearity misleadingly inflates the standard errors. Thus, it makes some variables statistically unimportant while they should be otherwise important. In the literature, there exist many procedures to tackle the problem of multicollinearity. One of them is the ridge regression which is alternative to the LS. It shows that the ridge regression estimator enhances the precision of the regression coefficient estimates, (Hatice and Ozlem Alpu, 2010). (Hoerl and Kennard, 1970 a, b) suggested to estimate the parameter K. (Hoerl and Baldwin, 1975) introduced that ridge regression shrinks the parameter β by imposing a penalty on the size of the coefficients. (Alkhamisi and Shukur, 2007) proposed a new method to estimate the ridge parameter by introducing some new estimators by adding $1/\lambda_{\max}$ to some well-known estimators, where $\lambda_{\max}$ is the largest eigenvalue of $X'X$. However, (Al-Hassan, 2010) applied the modification mentioned in (Alkhamisi and Shukur, 2007) to the estimator suggested by (Hocking et al., 1976) in order to define a new estimator. In these and other studies, the achievement of the ridge estimators was mainly compared using real data and simulation studies.

In addition, the least squares estimator can produce extremely poor estimates in the presence of non-normal error distribution. Nevertheless, (Midi et al., 2007) expressed that these procedures are not strong enough to deal with non-normal heavy-tailed distribution as a result of outliers. Outliers which arise from bad data points can have an influence on the ridge estimates and LS estimates. However, the problem is more complex when both multicollinearity and outliers are present in the data. In recent years, several authors have proposed reliable estimates, especially in the presence of multicollinearity and also non-normal error distribution. Although, we typically think these two problems occur separately, however in practical situations, these problems occur simultaneously.

(Montgomery and peck, 1982) has proposed that either robust or ridge estimation methods alone can be enough to tackle with the combined problem. To tackle these two problems together many robust ridge regression estimators have been used, so these estimates are much less influenced by non-normality and multicollinearity. (Askin and Montgomery, 1980) and (Pfaffenberger & Dielman, 1984, 1985) proposed combining the ridge and the least absolute deviation (LAD) robust regression techniques. (Maronna, 2011) suggested robust MM estimator in ridge regression for high dimensional data. (Siti Meriam et al., 2012) used Weighted Ridge MM-estimators (WRMM) to remedy the problem of multicollinearity only.

In this paper, an improved and more robust technique to overcome multicollinearity and outliers problems is proposed. This method is a combination of the ridge regression and a high breakdown point and efficient estimator, the least trimmed of squares (LTS). From here onwards the proposed method is referred to as the Ridge Least Trimmed Squares (RLTS) estimator that should be able to give a good estimate of the regression coefficients even in the presence of multicollinearity and outliers.

This paper consists of six sections. The background is discussed in Section 1. The ridge regression estimator is discussed in Section 2, and a search for the robust estimation techniques will be explained in Section 3. In Section 4, the proposed method RLTS is discussed. Section 5 presents the results of numerical example and a simulation study to investigate the performance of the proposed method. The concluding remarks are presented in Section 6.

2. Ridge Regression Estimators

The Least Squares (LS) estimator in equation (2) is very weak and inaccurately estimating the regression coefficients in the presence of multicollinearity in $\mathbf{X'X}$ matrix. The degree of multicollinearity is often denoted by the variance inflation factor (VIF) value of the $\mathbf{X'X}$ matrix. We call this is variance inflation factor or VIF, written as in equation (4), Marquardt (1970).

$$VIF = (1 - R^2)^{-1} \tag{4}$$

The VIF for each term in the model measures the combined effect of the dependencies among the regressors on the variance of that term. One or more large VIF indicate multicollinearity.Practical experience indicates that if any of the VIFs exceeds 5 or 10, it is an indication that the associated regression coefficients are poorly estimated because of multicollinearity. Complete removal of multicollinearity is not possible but we can reduce the degree of multicollinearity present in the data, (Marquardt, and Snee, 1975).

The technique of Ridge Regression (RR) is one of the most popular techniques and it was introduced by Hoerl and Kennard (1970 a, b). It is the best performing alternatives to the LS methods. LS has no bias, but it has a bigger variance than the ridge regression estimator in the presence of multicollinearity. The Ridge regression estimator can improve the estimation of β by adding a small constant to the diagonal of the matrix, which will reduce significantly the variance influential factor in the matrix. Ridge regression is proven as an effective and efficient remedial method to deal with the general problems caused by multicollinearity.

The ridge regression is defined as follows:

$$\hat{\boldsymbol{\beta}}_{Ridge} = (\mathbf{X'X} + K\mathbf{I})^{-1}\mathbf{X'Y} \tag{5}$$

where K is the unknown biasing constant and $\mathbf{I}$ is the ($p \times p$) identity matrix. There are several methods for determining K value and have been shown in the literature such as described by (Hoerl and Kennard, 1970) and (Gibbons, 1981). The estimator of the constant K is given by

$$K_{HK} = \frac{pS^2_{LS}}{\hat{\boldsymbol{\beta}}'_{LS}\boldsymbol{\beta}_{LS}} \tag{6}$$

here

$$S^2_{LS} = \frac{(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}_{LS})'(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}_{LS})}{n-p} \tag{7}$$

The main advantage of this method in multiple regression is to reduce MSE of the regression parameter by adding a positive value of ridge parameter, such that the increase of the bias is less than the reduced of the variance. (Hoerl and Kennard, 1970 a, b) have shown that there always exists a positive value $K>0$ such that $MSE(\hat{\boldsymbol{\beta}}_{\mathbf{R}}) < MSE(\boldsymbol{\beta}_{\mathbf{LS}})$. It is also true that $\hat{\boldsymbol{\beta}}_{Ridge} \to 0$ when the estimator $K \to \infty$. When $K=0$, $\hat{\boldsymbol{\beta}}_{Ridge} = \boldsymbol{\beta}_{LS}$ and when $K > 0$, $\hat{\boldsymbol{\beta}}_{Ridge}$ is biased, but more accurate and stable.

3. Robust Regression Estimators

Outliers are points that are lying far away from the pattern formed by the good points. OLS is affected by the extreme leverage points having randomly very large residuals, (Alamgir, 2013). The presence of outliers and influential cases can drastically change the magnitude of regression coefficients and even the direction of coefficient signs. Robust regression estimators are used to decrease the effect of outliers. It has been shown to be efficient, more reliable and provide fixed stable results than the LS estimator, especially so when error are non-normal that have heavier tails than the normal distribution, (Midi, 2007). Outliers greatly influence the estimation of coefficients, standard errors, test statistics and confidence interval.

There are more robust estimators that can reduce the sum of squares residuals such as least trimmed squares (LTS). (Betül Kan, 2013) applied robust ridge and robust Liu estimator for regression based on the LTS estimator. This technique is used to fit a regression by using estimators that dampen the impact of influential points and it is resistant to the presence of outliers in both dependent variable and the explanatory variables. In fact, the LTS estimates are very inefficient if the data are actually normally distributed. The robust regression estimators performed worse than the LS estimates when the data are clean (without outliers). Thus, it is better to determine whether the error are normally distributed or not and then use the appropriate technique.

3.1. Measurement of Robustness

There are two common measures of robustness, breakdown point and influence curve. In this paper we discussed a breakdown point, which measures how well an estimate can resist bad data before it fails, in other words, it is a measure of stability of the estimator when the sample contains a large fraction of outliers. (Hampel, 1975) explained that breakdown point (BDP) of a regression estimator is the smallest fraction of contamination that can cause the estimator to (break down) and no longer represent the trend in the bulk of the data. When an estimator breaks down, the estimate produced from the contaminated data can become arbitrarily far from the estimate it would give when the data was uncontaminated. For example the breakdown point of the sample mean, least squares estimates and Least absolute deviation is 1/n. This indicates that only one outlier can make the estimate useless. But the estimators of the LMS and LTS have breakdown points near 1/2. (Rousseeuw and Leroy, 1987) introduced, the most of robust estimators having the highest breakdown point, that are known as Least Median Squares (LMS) and Least Trimmed Squares (LTS).

The most commonly used robust estimator is the Least Trimmed Squares (LTS), which was proposed by (Rousseeuw, 1984). The estimated $\hat{\boldsymbol{\beta}}_{LTS}$ can be defined as the solution of the following

$$\min \sum_{i=1}^{h} \boldsymbol{\varepsilon}_i^2 \quad (8)$$

where $\varepsilon_{(1)}^2 \le \varepsilon_{(2)}^2 \le \varepsilon_{(3)}^2 \le ,\dots, \le \varepsilon_{(n)}^2$ denotes the order statistics of a set of residuals, from smallest to largest. LTS are calculated by minimizing the *h* ordered squares residuals, where $h = \left[\frac{n}{2}\right] + \left[\frac{(p+1)}{2}\right]$, with *n* and *p* being the sample size and the number of parameters, respectively. The objective function of LTS estimator is the sum of h smallest squared residuals and was indeed proposed as a remedy to the low asymptotic efficiency of LMS. The largest squared residuals are excluded from the summation in this method, that allows those outlier data points to be excluded completely. According to the value of *h* and the outlier data configuration, LTS can be very efficient.

If indeed the exact numbers of outlying data points are trimmed, this method is mathematically equivalent to OLS. On the contrary, LTS is not efficient if the number of trimmed data points is more than the actual outliers as some good data will be excluded. LTS is considered a high breakdown method with 50% breakdown value, (Rousseeuw and Leroy, 1987; Rousseeuw Van Driessen, 1998). Thus, the effect of outliers on the LTS estimates will be less than that of LS estimate.

4- Proposed Method (RLTS)

Despite $\hat{\boldsymbol{\beta}}_{Ridge}$ working better in the presence of multicollinearity, nevertheless, it is not robust when there are deviation from normality for the disturbances. Thus, we need to combine this procedure with some robust estimation procedures to produce robust ridge regression estimator. It is hoped that by combining the robust and ridge regression techniques, the problems of outliers and multicollinearity can be solved.

There have been some studies related to estimation using the robust ridge regression estimators in the literature such as (Hoerl, 1975; Askin and Montgomery, 1980; Moawad El-Fallah et al., 2013 and Betül Kan, 2013). Our proposed method to estimate the regression parameters is known as RLTS. We would expect the modified method to be more robust than the Ridge Least Absolute Value (RLAV) estimator.

The RLTS estimator of the parameter β can be calculated using the following formula.

$$\hat{\beta}_{LTS} = \left(X'X + K_{LTS} I\right)^{-1} X'Y \quad (9)$$

where the value of K is given by

$$K = \frac{pS^2_{LTS}}{\hat{\beta}'_{LTS}\hat{\beta}_{LTS}} \quad (10)$$

and

$$S^2_{LTS} = \frac{\left(y - x\hat{\beta}_{LTS}\right)'\left(y - x\hat{\beta}_{LTS}\right)}{n-p} \quad (11)$$

The S^2_{LTS} is the estimated variance, n is the sample size and p is the number of estimated parameters. The value of K is calculated using equation (10) with two changes. First, the LTS estimator of β is used rather than the LS estimator. Second, the estimator of σ^2 used in equation (11) is modified by using the LTS coefficient estimates rather than the least squares estimates. With these changes we are able to reduce the effect of extreme points on the value chosen for the biasing parameter.

5. Application

A numerical example is provided here to illustrate the application of the proposed method, the dataset used is the body fat data, (Penrose et al., 1985). This data contained 152 outliers, and VIF for the multicollinearity are shown in table 1. This data set consist of 14 variables and 252 observations. The variables in the data set are:

y= PCTBF, x1= Density, x2= Age, x3= Weight, x4= Height, x5= Neck, x6= Chest, x7= Abdomen, x8= Hip, x9= Thigh, x10= Knee, x11= Ankle, x12= Biceps, x13= Forearm, x14= Wrist

Table 1. The VIF for the boy fat data

Var.	VIF
X1	3.8183
X2	2.2747
X3	34.032
X4	1.6778
X5	4.3965
X6	9.4722
X7	18.120
X8	14.961
X9	7.8877
X10	4.6123
X11	1.9200
X12	3.6516
X13	2.2370
X14	3.5215

Here the maximum VIF is 34.031683. So it is clear that the multicollinearity problem exists. The summary of the parameter estimates and the standard error for the different methods can be seen in Table 2. From this table it can be seen that, the proposed method RLTS performed the best compared to the other methods.

Table 2 Estimated parameters and SE of $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_{14}$ for the different methods

	Estimate	LS	RIDGE	RLAV	RLTS
$\hat{\beta}_1$	parameter	-411.2	-1.0014	-0.9956	-0.9301
	s.e.	8.258	0.0196	0.0197	0.0185
$\hat{\beta}_2$	parameter	0.0126	0.0029	-0.0084	0.0193
	s.e.	0.0096	0.0152	0.0153	0.0143
$\hat{\beta}_3$	parameter	0.0101	0.0058	0.0094	0.0323
	s.e.	0.0160	0.0561	0.0563	0.0528
$\hat{\beta}_4$	parameter	-0.0080	-0.0030	-0.0038	-0.0036
	s.e.	0.0284	0.0130	0.0131	0.0123
$\hat{\beta}_5$	parameter	-0.0285	0.0006	0.0037	-0.0088
	s.e.	0.0694	0.0211	0.0212	0.0199
$\hat{\beta}_6$	parameter	0.0268	0.0036	-0.0015	0.0269
	s.e.	0.0294	0.0305	0.0307	0.0288
$\hat{\beta}_7$	parameter	0.0186	0.0002	0.0037	0.0305
	s.e.	0.0318	0.0419	0.0421	0.0395
$\hat{\beta}_8$	parameter	0.0192	0.0076	0.0000	0.0156
	s.e.	0.0434	0.0380	0.0382	0.0358
$\hat{\beta}_9$	parameter	-0.0168	-0.0046	-0.0112	-0.0095
	s.e.	0.0430	0.0282	0.0283	0.0266
$\hat{\beta}_{10}$	parameter	-0.0046	-0.0019	0.0037	-0.0012
	s.e.	0.0716	0.0216	0.0218	0.0204
$\hat{\beta}_{11}$	parameter	-0.0857	0.0006	-0.0027	-0.0170
	s.e.	0.0658	0.0140	0.0141	0.0132
$\hat{\beta}_{12}$	parameter	-0.0550	-0.0028	-0.0035	-0.0193
	s.e.	0.0509	0.0193	0.0194	0.0182
$\hat{\beta}_{13}$	parameter	0.0339	0.0008	-0.0048	0.0087
	s.e.	0.0595	0.0151	0.0153	0.0143
$\hat{\beta}_{14}$	parameter	0.0073	-0.0041	0.0027	-0.0001
	s.e.	0.1617	0.0190	0.0191	0.0179

5.1 Simulation Study

We carry out a simulation study to compare the performance of the different methods LS, RR and RLAV with the proposed estimator RLTS. The simulation is designed to allow both multicollinearity and non-normality to exist simultaneously. The non-normal distributions are used to generate outliers.

Suppose, we have the following linear regression model:

$$y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

where i=1, 2, 3 (12)

The parameter values $\beta_0, \beta_1, \beta_2$ and β_3 are set equal to one. The explanatory variables

x_{i1}, x_{i2} and x_{i3} are generated using the equation (13)

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

i=1, 2,…, n, j =1, 2, 3 (14)

where, z_{ij} are independent standard normal random numbers generated by the normal distribution. The explanatory variable values were generated for a given sample size *n*. The sample sizes used were 25 and 50. The value of ρ representing the correlation between any two explanatory variables, and its values were chosen as: 0.0, 0.5 and 0.99. One important factor in this study is the error distribution, we have taken 25% of ourtliers. The standard normal distribution is used with 500 trials for each sample size. The statistics computed are the bias,

RMSE, standard error (SE), and 6 pairwise MSE ratios of the estimators.

The bias and MSE are given by:

$\text{Bias} = \bar{\beta}_i - \beta_i$ where $\bar{\beta}_i = \frac{\sum_{i=1}^{k} \beta_i}{k}$, $k = 500$ and the mean squared error is $MSE = \frac{1}{500}\sum_{i=1}^{500}(\hat{\beta}_i - \beta_i)^2$, therefore, the RMSE is given by

$$[MSE(\hat{\beta}_j)]^{1/2}$$

where j= 0, 1, 2, 3

The VIF for the simulated data are shown in Table 3.

Here maximum VIF is 284.5107 when the correlation between independent variables is very high with different size number of observations. So it is clear that the multicollinearity problem exists. Thus, the VIF can help identify which regressors are involved in the multicollinearity. The measure of convergence was computed as the number of times estimator 1 was closer than estimator 2 or 3 to the true parameter β while the value in Table 4 and Table 5 showed the summary statistics such as bias, RMSE and SE of the estimators of the normal distributions for sample size 25 and 50 with 0% and 25% of outliers and different value of ρ .

Table 3 The VIF for the simulated data

$\rho = 0.99$			
Variable	**X1**	**X2**	**X3**
VIF (N=25)	125.9631	284.5107	108.2148
VIF (N=50)	91.68085	215.92021	106.24795

Table 6 shows the efficiency of the estimators by looking at the MSE ratios of the estimators. Values less than 1 denote that the estimator is more efficient, however, values more than 1 denote that the other estimator is more efficient.

From Table 4 and Table 5 we can see that the RMSE of the LS is relatively smaller than the other estimators when the errors are normally distributed without outliers and no multicollinearity. As expected, the LS give the best results in the normal situation. Also, the result in Table 6 is in favor of LS. However, we see in table 6 that MSE ratios of RLTS to OLS is greater than 1.00 denoting that the LS is more efficient than the RLTS when no outliers and no multicolliearity. On the other hand, in the same Table 4 and Table 5, we can see that the RMSE of the RIDGE is relatively smaller than the RLAV also, looking at the MSE ratios of the estimators show that the values of ridge more than 1 indicate that this estimator is more efficient than RLAV and RLTS when the errors are normally distributed without outliers and no multicollinearity. However, for non-normal error distribution and when correlation and outliers are present in the data, RLTS is better than LS and RIDGE, RLAV and its performance is almost as good as RIDGE and LS. Else, the LS is superior. The MSE in table 6 supported the result obtained from Table 4 and Table 5. These ratios indicate the efficiency of RLTS relative to other estimators. Values less than one indicate that RLTS is more efficient, however, values greater than one denote that the other estimators are more efficient than RLTS.

Consequently, we can see that the RMSE of the RLTS is relatively smaller than the other estimators,

when the errors are normally distributed in the presence of outliers and multicollinearity, it obviously shows that RLTS is almost as efficient as RLAV and RIDGE but certainly more efficient than with the presence of outliers and multicollinearity.

Table 4 Bias, RMSE and SE of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ with error normal (0,1) distribution of the sample size n=25 correlation 0.0, 0.5 and 0.99 outliers 0% and 25%.

0% outliers						
	Coef.	**Parameter**	**LS**	**RIDGE**	**RLAV**	**RLTS**
Value of ρ		**Bias**	-0.0735	-0.7065	-0.6593	-0.8253
	$\hat{\beta}_1$	**RMSE**	0.2160	0.9894	1.5704	1.3199
		S.e	0.2031	0.6926	1.4253	1.0300
		Bias	0.0078	-0.6791	-0.7273	-0.6498
0.0	$\hat{\beta}_2$	**RMSE**	0.1491	1.2014	2.0982	1.5016
		S.e	0.1489	0.9910	1.9681	1.3537
		Bias	0.0574	-0.6636	-0.6619	-0.5771
	$\hat{\beta}_3$	**RMSE**	0.1823	0.9574	1.4377	1.1064
		S.e	0.1730	0.6901	1.2762	0.9439
25%outiers						
	Coef.	**Parameter**	**LS**	**RIDGE**	**RLAV**	**RLTS**
		Bias	-6.5247	-1.1478	-0.9751	-0.9775
	$\hat{\beta}_1$	**RMSE**	78.537	2.3385	0.9802	0.9775
		S.e	78.265	2.0374	0.1005	0.0018
		Bias	7.7103	-0.7565	-0.9759	-0.9774
0.5	$\hat{\beta}_2$	**RMSE**	112.78	3.0444	0.9811	0.9774
		S.e	112.51	2.9489	0.1008	0.0016
		Bias	-1.4323	-1.0158	-0.975	-0.9773
	$\hat{\beta}_3$	**RMSE**	80.608	2.3722	0.9805	0.9773
		S.e	80.595	2.1437	0.1002	0.0018
25%outiers						
	Coef.	**Parameter**	**LS**	**RIDGE**	**RLAV**	**RLTS**
		Bias	-7.8369	-1.1523	-0.9792	-0.9814
	$\hat{\beta}_1$	**RMSE**	94.332	2.3421	0.9843	0.9814
		S.e	94.006	2.0391	0.1001	0.0012
	$\hat{\beta}_2$	**Bias**	9.2610	-0.7605	-0.9801	-0.9813
0.99		**RMSE**	135.46	3.0475	0.9852	0.9813
		S.e	135.14	2.9511	0.1006	0.0012
	$\hat{\beta}_3$	**Bias**	-1.7204	-1.0206	-0.9795	-0.9812
		RMSE	96.819	2.3761	0.9847	0.9812
		S.e	96.804	2.1457	0.1008	0.0012

From Table 6 values less than one indicate that the first estimator (first column) is more efficient than the other estimator (other column); values greater than one indicate that the second estimator (other column) is more efficient than the first estimator (first column).

The simulation results for larger samples, that is for n = 50 are consistent with the results of smaller samples. The results also verify that the estimator for larger samples are more efficient than those of smaller samples it is clear by the smaller values of RMS.

6. Conclusion

A simulation was designed to compare the performance of RLTS with some existing methods in dealing with multicollinearity and non-normal errors. The results of the comparisons show that ridge least trimmed squares (RLTS) estimator is better than other methods (ridge and robust regression estimators) for the different combinations of multicollinearity and outliers. From Table 4 and Table 5, RLTS is always better than LS. Only LS outperform RLTS in the cases when disturbances are normal with no multicollinearity and no outliers. In addition, The numerical analysis and simulation studies show clearly that RLTS estimator is the most suitable option over other estimators when both multicollinearity and outliers are present.

Table 5 Bias, RMSE and SE of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ with normal (0,1) error distribution of the sample size n=50 correlation 0.0, 0.5 and 0.99 outliers 0% and 25%

0% outliers						
Value of ρ	**Coef.**	**Parameter**	**LS**	**RIDGE**	**RLAV**	**RLTS**
		Bias	-0.0156	-0.6906	-0.7078	-0.6995
	$\hat{\beta}_1$	**RMSE**	0.4665	0.8367	1.3350	0.9733
		S.e	0.4663	0.4724	1.1320	0.6768
		Bias	-0.0703	-0.7040	-0.6882	-0.6875
0.0	$\hat{\beta}_2$	**RMSE**	0.1249	0.9713	1.7888	1.1494
		S.e	0.1032	0.6692	1.6511	0.9211
		Bias	0.0854	-0.6542	-0.6532	-0.6632
	$\hat{\beta}_3$	**RMSE**	0.4478	0.8001	1.3420	0.8950
		S.e	0.4396	0.4607	1.1723	0.6009
25%outiers						
	Coef.	**Parameter**	**LS**	**RIDGE**	**RLAV**	**RLTS**
		Bias	0.7658	-0.9364	-0.9731	-0.9742
	$\hat{\beta}_1$	**RMSE**	54.112	1.7432	0.9746	0.9742
		S.e	54.107	1.4703	0.0539	0.0016
		Bias	0.8508	-0.9620	-0.9724	-0.9741
0.5	$\hat{\beta}_2$	**RMSE**	71.982	2.1755	0.9744	0.9741
		S.e	71.977	1.9512	0.0629	0.0016
		Bias	-1.7518	-1.0224	-0.9736	-0.974
	$\hat{\beta}_3$	**RMSE**	51.756	1.7561	0.9742	0.9740
		S.e	51.726	1.4279	0.0337	0.0017
25%outiers						
	Coef.	**Parameter**	**LS**	**RIDGE**	**RLAV**	**RLTS**
		Bias	-1.2362	-0.9410	-0.9776	-0.9779
	$\hat{\beta}_1$	**RMSE**	67.589	1.7468	0.9782	0.9780
		S.e	67.578	1.4717	0.0337	0.0054
	$\hat{\beta}_2$	**Bias**	-4.2877	-0.9664	-0.9770	-0.9777
0.99		**RMSE**	91.327	2.1795	0.9784	0.9775
		S.e	91.226	1.9536	0.0527	0.0050
	$\hat{\beta}_3$	**Bias**	5.6092	-1.0271	-0.9782	-0.9775
		RMSE	62.164	1.7603	0.9796	0.9775
		S.e	61.910	1.4295	0.0525	0.0049

Table 6 MSE ratios of 6 pairwise estimators of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ with errors normal (0,1) distribution and 0% and 25% of outliers

		$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
Estimator 1 vs Estimator 2 vs Estimator 3		**Values of ρ**								
		0.0	**0.5**	**0.99**	**0.0**	**0.5**	**0.99**	**0.0**	**0.5**	**0.99**
	LS	37.34	0.00	0.00	101.36	0.00	0.00	36.85	0.00	0.00
		4.352	**0.00**	**0.00**	**84.70**	**0.00**	**0.00**	**3.994**	**0.00**	**0.00**
RLTS	**RIDGE**	1.780	0.175	0.176	1.562	0.103	0.104	1.335	0.170	0.171
		1.353	**0.312**	**0.313**	**1.400**	**0.201**	**0.201**	**1.251**	**0.308**	**0.308**
	RLAV	0.706	0.995	0.994	0.512	0.992	0.992	0.592	0.994	0.993
		0.532	**0.999**	**0.100**	**0.413**	**0.999**	**0.999**	**0.445**	**1.000**	**0.996**
RLAV	**LS**	52.86	0.00	0.00	197.92	0.00	0.00	62.22	0.00	0.00
		8.189	**0.00**	**0.00**	**205.15**	**0.00**	**0.00**	**8.981**	**0.00**	**0.00**
	RIDGE	2.519	0.176	0.177	3.050	0.104	0.105	2.255	0.171	0.172
		2.546	**0.313**	**0.314**	**3.392**	**0.201**	**0.202**	**2.813**	**0.308**	**0.310**
RIDGE	**LS**	20.98	0.001	0.001	64.89	0.001	0.001	27.59	0.001	0.001
		3.22	**0.001**	**0.001**	**60.48**	**0.001**	**0.001**	**3.193**	**0.001**	**0.001**

Acknowledgements:

This work was supported by the Universiti Teknologi Malaysia, under the Research University Grant (QJ130000.2526.6 H 68) and Ministry of Higher Education (MOHE) of Malaysia.

Corresponding Author:
Prof. Dr. Robiah Adnan
Department of Mathematical Sciences, Faculty of Science/Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.
E-mail: robiaha@utm.my

References

1. Hatice Samkar Ozlem Alpu. Ridge Regression Based on Some Robust Estimators. Journal of Modern Applied Statistical Methods 2010; 9(2):495-501.
2. Hoerl, A. E. and R. W. Kennard. Ridge Regression: Iterative Estimation of the Biasing Parameter. Communications in Statistics: A. Theory Methods 1970a; (5): 77-88.
3. Hoerl, A. E. and R. W. Kennard Ridge Regression: Applications to Nonorthogonal Problems. Technometrics 1970b; (12): 69-82.
4. Hoerl, A. E. and K. F. Baldwin. Ridge Regression: Some Simulations. Communications in Statistics 1975; (4): 104-123.
5. Alkhamisi, M. A. and Shukur, G. A Monte Carlo study of recent ridge parameters. Communications in Statistics, Simulation and Computation 2007;36 (3): 535-547.
6. Al-Hassan,Y M. Performance of a New Ridge Regression Estimator. Journal of the Association of Arab Universities for Basic and Applied Sciences 2010; (9): 43-50.
7. Hocking, R. R. , Speed, F. M. and Lynn, M. J. A class of biased estimators in linear regression. Technometrics 1976; 18 (4): 425-437.
8. Habshah Midi and Marina Zahari. A Simulation Study on Ridge Regression Estimators in the Presence of Outliers and Multicollinearity. Jurnal Teknologi 2007; 47 (C) :59-74.
9. Montgomery, D.C. and E.A. Peck. Introduction to Linear Regression. Analysis. Wiley, New York. 1982.
10. Askin, R. G. and Montgomery, D. C. “Augmented robust estimators.” Technometrics 1980; (22): ρ 333-341.
11. Pfaffenberger, R. C. and T. E. Dielman. A Modified Ridge Regression Estimator Using the Least Absolute Value Critrion in the Multiple Linear Regression Model. Proceedings of the American Institute for Decision Sciences. Toronto 1984; 791-793.
12. Pfaffenberger, R. C. and T. E. Dielman. A Comparison of Robust Ridge Estimators. Proceedings of the American Statistical Association Business and Economic Statistics Section. Las Vegas, Nev. 1985: 631-635.
13. Maronna, R.A. Robust Ridge Regression for High-Dimensional Data. Technometrics. 2011; 53(1): 44-53.
14. Siti Meriam Zahari, Mohammad said Zainol and Muhammad Iqbal al-Banna bin Ismail Weighted Ridge MM-Estimator in Robust Ridge Regression with Multicollinearity. Mathematical Models and Methods in Modern Science 2012. ISBN: 978-1-61804-106-7.
15. Marquardt, D. W. Generalized inverse, ridge regression, biased linear estimation, and nonlinear estimation, Technometrics 1970; (12): 591-612.
16. Marquardt, D. W. and Snee, R. D. Ridge regression in practice. Amer. Statist. 1975; (29): 3-19.
17. Gibbons, D. A Simulation Study of Some Ridge Estimators. Journal of American Statistical Association 1981; (76): 131-139.
18. Alamgir1, Amjad Ali, Sajjad Ahmad Khan, Dost Muhammad Khan and Umair Khalil, A New Efficient Redescending M- Estimator: Alamgir Redescending M- estimator. Research Journal of Recent Sciences 2013; 2 (8): 79-91.
19. Betül Kan, Özlem Alpu and Berna Yazıcı. Robust ridge and robust Liu estimator for regression based on the LTS estimator, Journal of Applied Statistics 2013; 40 (3): 644-655.
20. Hampel, F. R. Beyond location parameters: Robust concepts and methods proceedings of the 40th session of the ISI 1975; 46 (1): 375–391.
21. Rousseeuw, P.J. Least median of squares regression. J. Amer. Statist. Assoc 1984; (79): 871–880.
22. Rousseeuw and Leroy. Robust Regression and outlier detection. Wiley, New York. 1987.
23. Rousseeuw, P. J. and K. Van Driessen. Computing LTS regression for large data sets 1998.
24. Moawad El-Fallah Abed El-Salam “The Efficiency of Some Robust Ridge Regression for Sciences 2013; 7 (77): 3831 – 3841.
25. K.W. Penrose, A.G. Nelson, A.G. Fisher Generalized body composition prediction equation for men using simple measurement techniques. Medicine and Science in Sports and Exercise, 1985; (17): 189.

10/15/2014