Jurnal Teknologi, 47(C) Dis. 2007: 59–74

A SIMULATION STUDY ON RIDGE REGRESSION ESTIMATORS IN THE PRESENCE OF OUTLIERS AND MULTICOLLINEARITY

HABSHAH MIDI[1] & MARINA ZAHARI[2]

Abstract. A simulation study is used to examine the robustness of six estimators on a multiple linear regression model with combined problems of multicollinearity and non-normal errors. The performance of the six estimators, namely the Ordinary Least Squares (LS), Ridge Regression (RIDGE), Ridge Least Absolute Value (RLAV), Weighted Ridge (WRID), MM and a robust ridge regression estimator based on MM estimator (RMM) are compared. The RMM is a modification of the Ridge Regression (RIDGE) by incorporating robust MM estimator. The empirical evidence shows that RMM is the best among the six estimators for many combinations of disturbance distribution and degree of multicollinearity.

Keywords: Multicollinearity; outliers; ridge regression; robust regression

Abstrak. Satu kajian simulasi telah dijalankan untuk memeriksa keteguhan beberapa penganggar ke atas model linear regresi berganda dengan gabungan masalah multikolinearan dan ralat tak normal. Prestasi keenam-enam penganggar tersebut, seperti penganggar Kuasadua Terkecil Biasa (LS), Regresi 'Ridge' (RIDGE), Nilai Mutlak Terkecil 'Ridge' (RLAV), 'Ridge' Berpemberat (WRID), MM dan Regresi Teguh 'Ridge' berasaskan penganggar MM (RMM) dibandingkan. Penganggar RMM adalah pengubahsuaian penganggar Regresi 'Ridge' dengan menggabungkan penganggar teguh MM. Bukti empirik menunjukkan RMM adalah penganggar terbaik di kalangan enam penganggar yang dikaji bagi gabungan taburan ganguan dan paras multikolinearan.

Kata kunci: Multikolinearan; titik terpencil; regresi 'ridge'; regresi teguh

1.0 INTRODUCTION

For generation, statisticians have been relying on the Classical Ordinary Least Squares (LS) method in the multiple linear regression. It has been the most popular technique due to its optimal properties and ease of computation. The estimator β is determined by minimizing the function

$$\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 = \sum_{i=1}^{n}\left(r_i\right)^2 \qquad (1)$$

[1] Laboratory of Applied and Computational Statistics, Institute for Mathematical Research, University Putra Malaysia, 43400 UPM Serdang, Selangor Malaysia
[2] Faculty of Sciences and Mathematical Studies, University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

and the estimator of the parameter β is given by

$$\widehat{B}_{\text{LS}} = \left(X^T X\right)^{-1} X^T Y \tag{2}$$

This method gives unbiased and minimum variance among all unbiased linear estimators provided that the errors are independent and identically, normally distributed. However, in the presence of multicollinearity, the LS can result in very poor estimates. Even though LS estimates are unbiased in the presence of multicollinearity, its estimates will be imprecise with inflated standard errors. Inflated variances induced by multicollinearity are quite harmful to the use of regression as a basis for hypothesis testing, estimation and forecasting. An important test of significance based on degraded estimates would have been concluded inconclusive through a high error variance, or that a confidence interval or forecast interval is found to be large.

As an alternative to the OLS, we may turn to ridge regression estimator which may improved the precision of the regression coefficients. The ridge regression (RIDGE) technique was first proposed by Hoerl [1] and was extended further by Hoerl and Kennard [2 – 3]. Several other works on ridge regression have been proposed. Among them are Hoerl, Kennard and Baldwin [4] and Marquardt and Snee [5]. However, these techniques are not immune to the deviation from the normal assumption, that is a heavy-tailed distribution which may arise as a result of outliers. Outliers which arise from bad data points may have unduly effect on the LS and the Ridge estimates. The problem is further complicated when both outliers and multicollinearity are present in the data. In recent years, major efforts have been made to obtain reliable estimates especially in the presence of heavy-tailed error distribution and also multicollinearity. A robust method which is less influenced by the outliers and a ridge regression technique which improved the multicollinearity problem, had been given substantial consideration. Nevertheless, these methods alone cannot rectify the problem of outliers and multicollinearity. To remedy these two problems simultaneously, several robust ridge regression estimators have been put forward that are much less influenced by outliers and multicollinearity. Askin and Montgomery [7] proposed using the Weighted Ridge (WRID) and Pfaffenberger and Dielman [7 – 8] suggested combining the ridge and the Least Absolute Value (LAV) robust regression techniques. In this paper, we take the initiative to develop a more robust technique to rectify these two problems. We proposed combining the ridge regression with the highly efficient and high breakdown point estimator, namely the MM estimator. We call this modified method, the Robust Ridge Regression based on MM estimator (RMM). We expect that the modified method would be less sensitive to the presence of outliers and possess a high breakdown point since we have removed the influence of outliers by the highly robust and efficient MM estimator and also the problem of multicollinearity by ridge regression.

2.0 RIDGE REGRESSION ESTIMATORS

In cases where multicollinearity exists, the singularities present in X^TX matrix and this ill-conditioned X matrix can result in very poor estimates. The degree of multicollinearity is often indicated by conditioned number (CN) of the matrix X (or X^TX). CN is defined as the ratio of the largest singular values of X to the smallest,

$$\mathrm{CN}(X) = \frac{\lambda_{\max}}{\lambda_{\min}} \geq 1 \tag{3}$$

where λ are the eigenvalues of the matrix X^TX.

Belsley *et al.* [9] have empirically shown that weak dependencies are linked to CN around 5 to 10, whereas moderate to strong relations are linked to CN of 30 to 100. Hoerl and Kennard [2 – 3] pointed out that by adding a small constant to the diagonal of a matrix, will improve the conditioning of a matrix as this would dramatically reduced its CN. The ridge regression estimator is defined as follows

$$\hat{\beta}_{\mathrm{RID}} = \left(X^T X + kI\right)^{-1} X^T Y \tag{4}$$

where I is the $(p \times p)$ identity matrix and k is the biasing constant.

In practice, the optimal value of k is unknown. Various methods in determining k have appeared in the literature such as described by Hoerl and Kennard [3] and Gibbons [10]. The estimator of k proposed by Hoerl *et al.* [11] is given by

$$k_{\mathrm{HKB}} = \frac{p s_{\mathrm{LS}}^2}{\hat{\beta}_{\mathrm{LS}}^T \hat{\beta}_{\mathrm{LS}}} \tag{5}$$

where

$$s_{\mathrm{LS}}^2 = \frac{(Y - X\hat{\beta}_{\mathrm{LS}})^T (Y - X\hat{\beta}_{\mathrm{LS}})}{n - p} \tag{6}$$

When $k = 0$, $\hat{B}_{\mathrm{RID}} = \hat{\beta}_{\mathrm{LS}}$, when $k > 0$, $\hat{\beta}_{\mathrm{RID}}$ is biased but more stable and precise than LS estimator and when $k \to \infty$, $\hat{\beta}_{\mathrm{RID}} \to 0$. Hoerl and Kennard [2] have shown that there always exist a value $k > 0$ such that $\mathrm{MSE}_{\hat{\beta}_{\mathrm{RID}}}$ is less than $\mathrm{MSE}_{\hat{\beta}_{\mathrm{LS}}}$.

3.0 ROBUST REGRESSION ESTIMATORS

Robust regression estimators have been proven to be more reliable and efficient than least squares estimator especially when disturbances are nonnormal. "Nonnormal disturbances" are disturbance distributions that have heavy or fatter tails than the normal distribution and are prone to produce outliers. Since outliers greatly influence the estimated coefficients, standard errors and test statistics, the usual statistical procedure may be most inefficient as the precision of the estimator has been affected.

A better approach is to consider the robust procedure. This procedure fit a regression by using estimators that dampen the impact of influential points and then to detect outliers; those points lying far away from the pattern formed by the good points and have large residuals from the robust fit. Several works on robust estimation have been proposed in the literature. Among them are Edgeworth [12] who proposed the Least Absolute Values (LAV) estimator and Huber [13] who introduced M-estimators. However, none of these estimators achieves high breakdown point. Rousseeuw and Leroy [14] introduced the most robust estimator having the highest possible breakdown point, that is 50% which is known as Least Median Squares (LMS) and Least Trimmed Squares (LTS). Yohai [15] improved further the efficiency of the high breakdown estimators by introducing MM-estimators. The MM estimators are defined in three stages where the first and second stage is to achieve high breakdown point and the third stage is to aim for a high efficiency.

(i) In stage one, to compute an initial regression estimate T_0 of θ_0 which is consistent robust (but not necessarily efficient) with high breakdown point, possibly 50%.

(ii) In the second stage, compute the residuals of the initial estimate,

$$r_i\left(T_{o,n}\right) = y_i - T_{o,n}^T x_i, \quad 1 \le i \le n, \tag{7}$$

then, compute an M-estimate of errors scale $s_n = s\left(r\left(T_{o,n}\right)\right)$ (8)

(using a function ρ_0, satisfying Huber [16] M-estimation assumptions and using a constant b such that

$$\frac{b}{a} = 0.5 \tag{9}$$

where $a = \max \rho_0(u)$

(iii) The third stage is to compute an M-estimate of the regression parameters based on a proper redescending psi-function.
Let ρ_1 be another function satisfying Huber [16] M-estimation assumptions such that

$$\rho_i\left(u\right) \le \rho_0\left(u\right), \tag{10}$$

$$\sup \rho_i\left(u\right) = \sup \rho_0\left(u\right) = a \tag{11}$$

The influence function denoted as $\psi(t)$ is obtained by differentiating the objective function $\rho(t)$, that is $\psi(t) = \rho'(t)$. There are several functions of $\rho(t)$ and thus $\psi(t)$ to choose from, and in this study, the Huber influence function will be employed.

Then, the MM-estimator $T_{1,n}$ is defined as any solution of

$$\sum_{i=1}^{n} \psi_1\left(\frac{r_i(\theta)}{s_n}\right) x_i = 0 \text{ and this equation must satisfies,} \tag{12}$$

$$S\left(T_{1,n}\right) \le S\left(T_{0,n}\right) \tag{13}$$

where $S(\theta) = \sum_{i=1}^{n} \rho\left(\frac{r_i(\theta)}{s_n}\right)$ (14)

and $\rho_1(0/0)$ is defined as 0.

4.0 ROBUST RIDGE REGRESSION ESTIMATORS

Although $\hat{\beta}_{\text{RIB}}$ works best in the presence of multicollinearity, however, it is not robust when there are departures from normality for the disturbances. Hence, we need to combine this technique with some robust estimation techniques to produce robust ridge regression estimators.

Since robust and ridge regression methods are unable to deal with the outliers and multicollinearity problems simultaneously, it seems worthwhile to combine both methods. There have been some studies concerning the estimation using the robust ridge regression estimators in the literature such as Vinod and Ullah [17], Askin and Montgomery [6] who introduced Weighted Ridge (WRID) estimator, and Pfaffenberger and Dielman [7] who introduced Ridge Least Absolute Value (RLAV) estimator. A slight modification of the ridge regression technique based on MM estimator is proposed. We would expect the modified method to be more robust than the Weighted Ridge (WRID) and the Ridge Least Absolute Value (RLAV).

4.1 Weighted Ridge (WRID)

The robust ridge regressions estimator proposed by Askin and Montgomery [6] first introduced a weighted least squares estimator, written as

$$\hat{\beta}_{\text{WLS}} = \left(X^T W X\right)^{-1} X^T W Y \tag{15}$$

where W is a diagonal matrix with diagonal elements W_{ii}, the weights intended to downweight extreme observations. $\hat{\beta}_{\text{WLS}}$ can be estimated by applying OLS to the transformed observations $\sqrt{W_{ii}}\, y_i$ and $\sqrt{W_{ii}}\, x_i$. The weights can be determined using any M-estimate W function and are given by

$$W_{ii} = \frac{\psi\left(y_i - x_i^T \beta\right)}{y_i - x_i^T \beta} \tag{16}$$

The WRID estimator is computed using the formula

$$\hat{\beta}_{\text{WRID}} = (X'WX + kI)^{-1} X'WY \tag{17}$$

where k is the biasing parameter determined as in Equation (5).

2.2 Ridge Least Absolute Value (RLAV)

Pfaffenberger and Dielman [7] and Lawrence and Arthur [18] suggested robust ridge regression by combining the properties of the LAV and the ridge regression estimator referred to as RLAV. The RLAV estimator can be written by

$$\hat{\beta}_{\text{RLAV}} = \left(X^T X + k * I\right)^{-1} X^T Y \tag{18}$$

The value of k^* is determined similar to that of Hoerl *et al.* [11] in Equation (5) and (6) by replacing k with k^* as in Equation (19) and (20), respectively.

$$k^* = \frac{p s_{\text{LAV}}^2}{\hat{\beta}_{\text{LAV}}^T \hat{\beta}_{\text{LAV}}} \tag{19}$$

where,
$$s_{\text{LAV}}^2 = \frac{\left(Y - X\hat{\beta}_{\text{LAV}}\right)^T \left(Y - X\hat{\beta}_{\text{LAV}}\right)}{n - p} \tag{20}$$

p is the number of parameters and n is the sample size.

$\hat{\beta}_{\text{LAV}}$ is the LAV estimator defined as the solution to $\min_{\beta} \sum_{i=1}^{N} \left| y_i - x_i^T \beta \right|$. (21)

2.3 Ridge MM (RMM)

In this article, we proposed a slight modification of the RLAV which is defined in Equation (18). In doing so, we replaced the values of the LAV estimator in (19) with the MM estimator to determine the biasing parameter k

$$k = \frac{p s_{\text{MM}}^2}{\hat{\beta}_{\text{MM}}^T \hat{\beta}_{\text{MM}}} \tag{22}$$

Likewise, the MM estimator of β is used rather than LS estimator in computing the k and s^2 values in order to reduce the effect of outliers on the value chosen for k.

$$s_{\text{MM}}^2 = \frac{\left(Y - X\hat{\beta}_{\text{MM}}\right)^T \left(Y - X\hat{\beta}_{\text{MM}}\right)}{n - p} \tag{23}$$

The RMM estimator of the parameter β is given by

$$\hat{\beta}_{\text{RMM}} = \left(X^T X + kI\right)^{-1} X^T Y \tag{24}$$

where k is given in Equation (22).

5.0 SIMULATION STUDY

A simulation study similar to that of Lawrence and Arthur (1990) has been carried out to demonstrate the efficiency of the proposed estimator, RMM in comparison with several existing estimators. Varying degrees of multicollinearity and non-normal disturbance distributions were allowed to be present simultaneously in the simulation.

There are six estimators in the study,

(i) Ordinary Least Squares (LS)
(ii) Ridge Regression (RIDGE)
(iii) Ridge Least Absolute Value (RLAV)
(iv) Weighted Ridge (WRID)
(v) MM
(vi) Ridge MM (RMM)

The following model was used in this simulation studies;

$$y_i = 1 + x_{i1} + x_{i2} + \varepsilon_i \tag{25}$$

where

$$\beta_0 = \beta_1 = \beta_2 = 1$$

The explanatory variables x_{i1} and x_{i2} were generated as below,

$$x_{ij} = \left(1 - \rho^2\right) z_{ij} + \rho z_{ij}, \quad i = 1, \ldots n; j = 1, 2 \tag{26}$$

where z_{ij} are independent standard normal random numbers. These values were held fixed throughout the experiment once generated for a given sample size n. The sample sizes used were 25 and 50. The value of ρ^2 represents the correlation between the two explanatory variables. The chosen values were 0.0, 0.5, 0.8, 0.9, 0.95 and 0.99. The final factor was the disturbance distribution. Three disturbance distributions were examined,

(i) Standard normal distribution.
(ii) Cauchy distribution with median zero and scale parameter one.
(iii) t-Student distribution with degrees of freedom three.

The non-normal distribution, such as the Cauchy and Student-t with 3 degrees of freedom having a heavy tailed distribution which prone to produce outliers, were generated to see the effect of combined problems of multicollinearity and outliers on different estimators. The Normal, Cauchy and Student-t variates were generated from S-Plus program. All computations were written in S language and executed on Acer Pentium III 866 Mhz. The performance of the RMM is assessed by looking at some summary statistics based on 1000 Monte Carlo trials. The statistics computed are the bias, root of mean squared error (RMSE), standard error (SE), and 5 pairwise MSE ratios of the estimators. The bias and MSE are given by:

$$\text{Bias} = E(\hat{\beta}) - \beta = \bar{\beta}_j - \beta_j \tag{27}$$

where $\bar{\beta}_j = \frac{\sum_{i=1}^{m} \beta_j}{m}$, $m = 1000$

$$\text{MSE}(\hat{\beta}_j) = E[\hat{\beta} - \beta]^2 = (\hat{\beta}_j - \beta_j)^2 + \frac{1}{m}\sum_{j=1}^{n}(\hat{\beta}_j^{(k)} - \bar{\beta}_j)^2 \tag{28}$$

where j = 0, 1, 2

Therefore, the RMSE is given by [MSE $(\hat{\beta}_j)]^{1/2}$ or is given by [VAR $(\hat{\beta}_j)$ + Bias^2] $^{1/2}$. The measure of closeness was calculated as the number of times estimator A was closer than estimator B to the true parameter β.

The result for the parameter β_0 were quite similar to β_1 and β_2. Similarly, the results for ρ^2 = 0.8, 0.9 and 0.99 are consistent for all methods and therefore are not included in the results to save spaces. The values in all tables are for sample size 25 except for sample size 50 were written in bold. Tables 1, 3 and 5 show the summary statistics such as bias, root of mean squared error (RMSE) and standard error (SE) of the estimators. Tables 2, 4 and 6 show the efficiency of the estimators by looking at the MSE ratios of the estimators. Values less than one indicate that the first estimator is more efficient, while values more than one indicate that the other estimator is more efficient.

From Table 1, we can see that the RMSE of the LS is relatively smaller than the other estimators when the errors are normally distributed and multicollinearity is not present. As was to be expected, the LS give the best results in the normal situation. Likewise, the result in Table 2 is in favour of LS. We see that the Mean Squared Error (MSE) ratios of RMM to OLS is greater than 1.00 indicating that the LS is more efficient than the RMM.

However, for normal error distribution and when correlation is present in the data, RMM is better than LS, WRID and MM and it's performance is almost as good as RIDGE and RLAV. Otherwise, LS is superior. The MSE ratios in Table 2 supported the results obtained from Table 1. These ratios denote the efficiency of RMM relative to other estimators. Values less than one indicate that RMM is more

Table 1 Bias, RMSE and SE of $\hat{\beta}_1$ and $\hat{\beta}_2$ with disturbance distribution Normal (0,1)

($\hat{\beta}_1$)

	Values of ρ^2								
	0.0			**0.5**			**0.95**		
Method	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**
LS	-0.0157	0.2174	0.2169	-0.0088	0.3327	0.3326	0.0446	3.0013	3.0009
	0.0017	**0.1451**	**0.1451**	**-0.0026**	**0.2284**	**0.2284**	**-0.0507**	**2.0939**	**2.0933**
RIDGE	-0.0584	0.2216	0.2137	-0.0412	0.2959	0.2930	-0.0134	1.5378	1.5377
	-0.0191	**0.1452**	**0.1439**	**-0.0185**	**0.2153**	**0.2145**	**-0.0121**	**1.0704**	**1.0704**
RLAV	-0.0617	0.2229	0.2142	-0.0426	0.2950	0.2919	-0.0261	1.5524	1.5521
	-0.0199	**0.1455**	**0.1441**	**-0.0189**	**0.2147**	**0.2139**	**-0.0255**	**1.0707**	**1.0704**
WRID	-0.0151	0.2743	0.2739	-0.0128	0.4111	0.4109	0.0607	3.5277	3.5272
	-0.0022	**0.1879**	**0.1879**	**-0.0086**	**0.2795**	**0.2794**	**-0.0563**	**2.3338**	**2.3331**
MM	-0.0146	0.2710	0.2706	-0.0107	0.4063	0.4061	0.0424	3.6993	3.6991
	-0.0028	**0.1728**	**0.1728**	**-0.0009**	**0.2728**	**0.2728**	**-0.0445**	**2.4974**	**2.4970**
RMM	-0.0607	0.2231	0.2147	-0.0422	0.2953	0.2923	0.0190	1.5057	1.5055
	-0.0198	**0.1454**	**0.1441**	**-0.0186**	**0.2152**	**0.2144**	**-0.0102**	**1.0729**	**1.0729**

($\hat{\beta}_1$)

	Values of ρ^2								
	0.0			**0.5**			**0.95**		
Method	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**
LS	-0.0051	0.2258	0.2258	0.0091	0.3496	0.3495	-0.0574	3.0116	3.0110
	0.0069	**0.1484**	**0.1482**	**0.0094**	**0.2303**	**0.2301**	**0.0555**	**2.0927**	**2.0920**
RIDGE	-0.0476	0.2266	0.2215	-0.0250	0.3103	0.3093	0.0038	1.5543	1.5543
	-0.0140	**0.1477**	**0.1471**	**-0.0076**	**0.2152**	**0.2151**	**0.0057**	**1.0686**	**1.0686**
RLAV	-0.0511	0.2276	0.2217	-0.0266	0.3087	0.3076	0.0187	1.5638	1.5636
	-0.0149	**0.1477**	**0.1469**	**-0.0081**	**0.2147**	**0.2146**	**0.0200**	**1.0710**	**1.0708**
WRID	-0.0122	0.2679	0.2676	0.0151	0.4236	0.4233	-0.0805	3.5364	3.5355
	0.0039	**0.1822**	**0.1822**	**0.0139**	**0.2795**	**0.2792**	**0.0642**	**2.3399**	**2.3390**
MM	-0.0042	0.2630	0.2629	0.0098	0.4143	0.4142	-0.0583	3.7101	3.7096
	0.0018	**0.1710**	**0.1710**	**0.0123**	**0.2699**	**0.2696**	**0.0542**	**2.4976**	**2.4970**
RMM	-0.0503	0.2266	0.2210	-0.0261	0.3094	0.3083	-0.0479	1.5210	1.5202
	-0.0148	**0.1477**	**0.1469**	**-0.0079**	**0.2150**	**0.2149**	**0.0041**	**1.0701**	**1.0701**

efficient while values greater than one indicate that the other estimators are more efficient than RMM. From Table 2, it clearly shows that RMM is almost as efficient as RIDGE and RLAV but certainly more efficient than LS, MM and WRID when outliers do not exist and when multicollinearity is present.

Table 2 MSE ratios of 15 pairwise estimators of $\hat{\beta}_1$ and $\hat{\beta}_2$ with disturbance distribution Normal (0,1)

		($\hat{\beta}_1$)			($\hat{\beta}_2$)		
		Values of ρ^2					
Estimator 1 vs Estimator 2		**0.0**	**0.5**	**0.95**	**0.0**	**0.5**	**0.95**
RMM	LS	1.05	0.79	0.25	1.01	0.78	0.26
		1.00	**0.89**	**0.26**	**0.99**	**0.87**	**0.26**
	RID	1.01	1.00	0.96	1.00	0.99	0.96
		1.00	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	WRID	0.66	0.52	0.18	0.72	0.53	0.18
		0.60	**0.59**	**0.21**	**0.66**	**0.59**	**0.21**
	RLAV	1.00	1.00	0.94	0.99	1.00	0.95
		1.00	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	MM	0.68	0.53	0.17	0.74	0.56	0.17
		0.71	**0.62**	**0.18**	**0.75**	**0.63**	**0.18**
MM	LS	1.55	1.49	1.52	1.36	1.40	1.52
		1.42	**1.43**	**1.42**	**1.33**	**1.37**	**1.42**
	RID	1.50	1.89	5.79	1.35	1.78	5.70
		1.42	**1.61**	**5.44**	**1.34**	**1.57**	**5.46**
	WRID	0.98	0.98	1.10	0.96	0.96	1.10
		0.85	**0.95**	**1.15**	**0.88**	**0.93**	**1.14**
	RLAV	1.48	1.90	5.68	1.34	1.80	5.63
		1.41	**1.61**	**5.44**	**1.34**	**1.58**	**5.44**
RLAV	LS	1.05	0.79	0.27	1.02	0.78	0.27
		1.01	**0.88**	**0.26**	**0.99**	**0.87**	**0.26**
	RID	1.01	0.99	1.02	1.01	0.99	1.01
		1.00	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**
	WRID	0.66	0.51	0.19	0.72	0.53	0.20
		0.60	**0.59**	**0.21**	**0.66**	**0.59**	**0.21**
WRID	LS	1.59	1.53	1.38	1.41	1.47	1.38
		1.68	**1.50**	**1.24**	**1.51**	**1.47**	**1.25**
	RID	1.53	1.93	5.26	1.40	1.86	5.18
		1.67	**1.69**	**4.75**	**1.52**	**1.69**	**4.79**
RID	LS	1.04	0.79	0.26	1.01	0.79	0.27
		1.00	**0.89**	**0.26**	**0.99**	**0.87**	**0.26**

*Values less than one indicate that the first estimator (first column) is more efficient than the second estimator (second column); values greater than one indicate that the second estimator (second column) is more efficient than the first estimator (first column)

From Tables 3 and 4 (Cauchy distribution), for skewed distribution and multicollinearity is not present, MM outperforms all the other estimators. However, when both multicollinearity and outliers are present, all three robust ridge estimators, WRID, RLAV and RMM perform better, in which RMM is superior, followed

Table 3 Bias, RMSE and SE of $\hat{\beta}_1$ and $\hat{\beta}_2$ with disturbance distribution Cauchy

($\hat{\beta}_1$)

	Values of ρ^2								
	0.0			**0.5**			**0.95**		
Method	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**
LS	2.948	126.361	126.327	2.778	60.726	60.662	6.272	263.854	263.779
	0.195	**18.565**	**18.564**	**0.322**	**32.098**	**32.096**	**5.177**	**340.413**	**340.374**
RIDGE	0.662	43.493	43.488	0.779	27.448	27.437	-1.028	50.754	50.743
	-0.256	**7.316**	**7.311**	**-0.243**	**8.663**	**8.659**	**0.473**	**73.983**	**73.982**
RLAV	-0.563	0.767	0.520	-0.500	0.784	0.604	-0.265	1.879	1.861
	-0.566	**0.775**	**0.529**	**-0.528**	**0.744**	**0.524**	**-0.330**	**1.061**	**1.009**
WRID	-0.209	0.509	0.464	-0.178	0.610	0.584	0.102	3.373	3.371
	-0.284	**0.473**	**0.379**	**-0.251**	**0.478**	**0.407**	**-0.124**	**1.459**	**1.453**
MM	0.025	0.432	0.432	0.009	0.670	0.670	0.598	6.195	6.166
	-0.001	**0.279**	**0.279**	**-0.022**	**0.432**	**0.431**	**-0.241**	**3.954**	**3.947**
RMM	-0.575	0.767	0.507	-0.515	0.760	0.559	-0.207	1.539	1.525
	-0.573	**0.776**	**0.524**	**-0.532**	**0.745**	**0.521**	**-0.309**	**1.008**	**0.960**

$\hat{\beta}_2$

	Values of ρ^2								
	0.0			**0.5**			**0.95**		
Method	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**
LS	3.936	144.025	143.972	1.289	78.861	78.851	-5.348	245.959	245.901
	-0.208	**17.346**	**17.345**	**-0.619**	**36.886**	**36.881**	**-5.510**	**344.521**	**344.477**
RIDGE	0.717	41.931	41.925	-0.148	21.884	21.883	1.841	46.745	46.709
	-0.398	**6.632**	**6.620**	**-0.397**	**9.743**	**9.735**	**-0.864**	**74.819**	**74.814**
RLAV	-0.571	0.793	0.550	-0.511	0.781	0.591	-0.245	1.897	1.881
	-0.579	**0.775**	**0.514**	**-0.536**	**0.748**	**0.521**	**-0.310**	**1.077**	**1.032**
WRID	-0.248	0.528	0.467	-0.222	0.597	0.555	-0.241	3.391	3.382
	-0.282	**0.469**	**0.375**	**-0.248**	**0.472**	**0.402**	**-0.111**	**1.458**	**1.453**
MM	-0.016	0.455	0.455	-0.029	0.683	0.682	-0.597	6.214	6.185
	0.011	**0.288**	**0.288**	**0.022**	**0.443**	**0.443**	**0.238**	**3.973**	**3.966**
RMM	-0.583	0.791	0.535	-0.531	0.759	0.543	-0.308	1.563	1.532
	-0.586	**0.775**	**0.508**	**-0.545**	**0.746**	**0.509**	**-0.293**	**1.013**	**0.970**

Table 4 MSE ratios of 15 pairwise estimators of $\hat{\beta}_1$ and $\hat{\beta}_2$ with disturbance distribution Cauchy

		($\hat{\beta}_1$)			(β_2)		
		Values of ρ^2					
Estimator 1 vs Estimator 2		**0.0**	**0.5**	**0.95**	**0.0**	**0.5**	**0.95**
RMM	LS	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	RID	0.000	0.001	0.001	0.000	0.001	0.001
		0.011	**0.007**	**0.000**	**0.014**	**0.006**	**0.000**
	RLAV	1.00	0.94	0.67	1.00	0.94	0.68
		1.00	**1.00**	**0.90**	**1.00**	**0.99**	**0.88**
	WRID	2.27	1.55	0.21	2.24	1.62	0.21
		2.69	**2.43**	**0.48**	**2.73**	**2.50**	**0.48**
	MM	3.15	1.29	0.06	3.02	1.24	0.06
		7.76	**2.97**	**0.07**	**7.26**	**2.83**	**0.07**
MM	LS	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	RID	0.000	0.001	0.015	0.000	0.001	0.018
		0.001	**0.002**	**0.003**	**0.002**	**0.002**	**0.003**
	RLAV	0.32	0.73	10.87	0.33	0.76	10.73
		0.13	**0.34**	**13.87**	**0.14**	**0.35**	**13.60**
	WRID	0.72	1.21	3.37	0.74	1.31	3.36
		0.35	**0.82**	**7.35**	**0.38**	**0.88**	**7.43**
RLAV	LS	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	RID	0.000	0.001	0.001	0.000	0.001	0.002
		0.011	**0.007**	**0.000**	**0.014**	**0.006**	**0.000**
	WRID	2.27	1.65	0.31	2.25	1.71	0.31
		2.68	**2.42**	**0.53**	**2.72**	**2.51**	**0.55**
WRID	LS	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	RID	0.00	0.00	0.00	0.000	0.001	0.005
		0.00	**0.00**	**0.00**	**0.005**	**0.002**	**0.000**
RID	LS	0.12	0.20	0.04	0.08	0.08	0.04
		0.16	**0.07**	**0.05**	**0.15**	**0.07**	**0.05**

*Values less than one indicate that the first estimator (first column) is more efficient than the second estimator (second column); values greater than one indicate that the second estimator (second column) is more efficient than the first estimator (first column)

closely by RLAV. A similar conclusion can be made when employing the t-Student distribution with 3 degrees of freedom. The results obtained from Tables 5 and 6 are almost parallel to the results obtained from employing the Cauchy distribution.

Table 5 Bias, RMSE and SE of $\hat{\beta}_1$ and $\hat{\beta}_2$ with disturbance distribution t-Student (3)

($\hat{\beta}_1$)

	Values of ρ^2								
	0.0			0.5			0.95		
Method	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**
LS	-0.0012	0.3755	0.3755	0.0168	0.5811	0.5808	0.2231	5.1562	5.1514
	-0.0096	**0.2435**	**0.2433**	**-0.0153**	**0.3815**	**0.3812**	**-0.1524**	**3.4542**	**3.4508**
RIDGE	-0.0927	0.3547	0.3423	-0.0646	0.4438	0.4391	0.0909	2.5437	2.5421
	-0.0655	**0.2465**	**0.2376**	**-0.0558**	**0.3295**	**0.3248**	**-0.0480**	**1.6799**	**1.6793**
RLAV	-0.1032	0.3427	0.3268	-0.0794	0.4126	0.4049	0.1088	1.7881	1.7848
	-0.0697	**0.2426**	**0.2323**	**-0.0589**	**0.3176**	**0.3121**	**-0.0203**	**1.1649**	**1.1647**
WRID	-0.0028	0.2983	0.2983	-0.0169	0.4914	0.4911	0.1509	3.8321	3.8291
	-0.0136	**0.2027**	**0.2022**	**-0.0097**	**0.3045**	**0.3043**	**-0.0136**	**2.3593**	**2.3593**
MM	0.0036	0.2938	0.2937	-0.0069	0.5207	0.5206	0.2453	4.4349	4.4281
	-0.0097	**0.2089**	**0.2087**	**-0.0093**	**0.3133**	**0.3132**	**-0.0898**	**2.8829**	**2.8815**
RMM	-0.1053	0.3410	0.3244	-0.0832	0.4073	0.3987	0.1107	1.7224	1.7189
	-0.0701	**0.2426**	**0.2322**	**-0.0598**	**0.3161**	**0.3104**	**-0.0230**	**1.1130**	**1.1128**

($\hat{\beta}_1$)

	Values of ρ^2								
	0.0			0.5			0.95		
Method	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**	**Bias**	**RMSE**	**S.E.**
LS	-0.0007	0.3604	0.3604	0.0089	0.5766	0.5765	-0.2139	5.1522	5.1478
	0.0014	**0.2573**	**0.2573**	**0.0154**	**0.3811**	**0.3808**	**0.1560**	**3.4463**	**3.4428**
RIDGE	-0.0934	0.3466	0.3337	-0.0621	0.4548	0.4506	-0.1080	2.5361	2.5338
	-0.0541	**0.2543**	**0.2485**	**-0.0313**	**0.3252**	**0.3237**	**0.0296**	**1.6753**	**1.6751**
RLAV	-0.1038	0.3372	0.3208	-0.0756	0.4261	0.4193	-0.1373	1.7843	1.7790
	-0.0584	**0.2495**	**0.2426**	**-0.0353**	**0.3131**	**0.3111**	**-0.0036**	**1.1680**	**1.1680**
WRID	-0.0134	0.3069	0.3066	-0.0047	0.4864	0.4863	-0.1390	3.8419	3.8394
	-0.0096	**0.2073**	**0.2070**	**0.0002**	**0.3182**	**0.3182**	**0.0178**	**2.3719**	**2.3719**
MM	0.0037	0.3179	0.3178	0.0078	0.5131	0.5131	-0.2294	4.4443	4.4384
	0.0034	**0.2031**	**0.2031**	**0.0135**	**0.3114**	**0.3111**	**0.0940**	**2.8804**	**2.8789**
RMM	-0.1056	0.3362	0.3192	-0.0768	0.4198	0.4127	-0.1372	1.7123	1.7068
	-0.0589	**0.2494**	**0.2424**	**-0.0353**	**0.3117**	**0.3096**	**-0.0018**	**1.1137**	**1.1137**

From Table 5, we can see that the RMM has the smallest SE and RMSE followed closely by RLAV. In terms of efficiency, again, RMM emerged as the most efficient estimator of all by having the smallest RMSE ratios as shown in Table 6. Clearly,

Table 6 MSE ratios of 15 pairwise estimators of $\hat{\beta}_1$ and $\hat{\beta}_2$ with disturbance distribution t-Student degrees of freedom (3)

		($\hat{\beta}_1$)			($\hat{\beta}_2$)		
		Values of ρ^2					
Estimator 1 vs Estimator 2		**0.0**	**0.5**	**0.95**	**0.0**	**0.5**	**0.95**
RMM	LS	0.82	0.49	0.11	0.87	0.53	0.11
		0.99	**0.69**	**0.10**	**0.94**	**0.67**	**0.10**
	RID	0.92	0.84	0.46	0.94	0.85	0.46
		0.97	**0.92**	**0.44**	**0.96**	**0.92**	**0.44**
	RLAV	0.99	0.97	0.93	0.99	0.97	0.92
		1.00	**0.99**	**0.91**	**1.00**	**0.99**	**0.91**
	WRID	1.31	0.69	0.20	1.20	0.74	0.20
		1.43	**1.08**	**0.22**	**1.45**	**0.96**	**0.22**
	MM	1.35	0.61	0.15	1.12	0.67	0.15
		1.35	**1.02**	**0.15**	**1.51**	**1.00**	**0.15**
MM	LS	0.61	0.80	0.74	0.78	0.79	0.74
		0.74	**0.67**	**0.70**	**0.62**	**0.67**	**0.70**
	RID	0.69	1.38	3.04	0.84	1.27	3.07
		0.72	**0.90**	**2.94**	**0.64**	**0.92**	**2.96**
	RLAV	0.73	1.59	6.15	0.89	1.45	6.20
		0.74	**0.97**	**6.12**	**0.66**	**0.99**	**6.08**
	WRID	0.97	1.12	1.34	1.07	1.11	1.34
		1.06	**1.06**	**1.49**	**0.96**	**0.96**	**1.47**
RLAV	LS	0.83	0.50	0.12	0.88	0.55	0.12
		0.99	**0.69**	**0.11**	**0.94**	**0.67**	**0.11**
	RID	0.93	0.86	0.49	0.95	0.88	0.49
		0.97	**0.93**	**0.48**	**0.96**	**0.93**	**0.49**
	WRID	1.32	0.71	0.22	1.21	0.77	0.22
		1.43	**1.09**	**0.24**	**1.45**	**0.97**	**0.24**
WRID	LS	0.63	0.72	0.55	0.73	0.71	0.56
		0.69	**0.64**	**0.47**	**0.65**	**0.70**	**0.47**
	RID	0.71	1.23	2.27	0.78	1.14	2.29
		0.68	**0.85**	**1.97**	**0.66**	**0.96**	**2.00**
RID	LS	0.89	0.58	0.24	0.92	0.62	0.24
		1.03	**0.75**	**0.24**	**0.98**	**0.73**	**0.24**

*Values less than one indicate that the first estimator (first column) is more efficient than the second estimator (second column); values greater than one indicate that the second estimator (second column) is more efficient than the first estimator (first column)

RMM performs significantly better than MM and other robust ridge estimators over a wide range of combination of ρ^2 and nonnormal error distributions. The simulation results for larger samples, that is for n = 50 are consistent with the results of smaller samples. The results also indicate that the estimator for larger samples are more efficient than those of smaller samples evident by the smaller values of RMSE.

6.0 CONCLUSIONS

For large sample size, when disturbances are normal and the correlation is high, RID marginally outperforms RMM, otherwise, RMM is superior. LS and MM outperform RMM in the cases when disturbances are normal and nonnormal with no multicollinearity. However, when degree of multicollinearity is high, RMM is superior to them. The comparisons among the robust ridge estimators, RMM, RLAV and WRID show that RMM is superior to the other two estimators for many combinations of error distribution type and degree of multicollinearity. The simulation studies clearly show that RMM estimator offers the most feasible option over other estimators when both multicollinearity and outliers are present.

REFERENCES

[1] Hoerl, A. E. 1962. Application of Ridge Analysis to Regression Problems. *Chemical Engineering Progress.* 58: 54-59.

[2] Hoerl, A. E. and R. W. Kennard. 1970a. Ridge Regression:Iterative Estimation of the Biasing Parameter. *Communications in Statistics: A. Theory Methods.* 5: 77-88.

[3] Hoerl, A. E. and R. W. Kennard. 1970b. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics.* 12: 69-82.

[4] Hoerl, A. E. and K. F. Baldwin. 1975. Ridge Regression: Some Simulations. *Communications in Statistics.* 4: 104-123.

[5] Marquard, D. W. and R. D. Snee. 1975. Ridge Regression in Practice. *The American Statistician.* 29:3-20.

[6] Askin, R. G. and D. C. Montgomery. 1980. Augmented Robust Estimators. *Technometrics.* 22: 333-341.

[7] Pfaffenberger, R. C. and T. E. Dielman. 1984. 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. 791-793.

[8] Pfaffenberger, R. C. and T. E. Dielman. 1985. A Comparison of Robust Ridge Estimators. *Proceedings of the American Statistical Association Business and Economic Statistics Section.* Las Vegas, Nev., 631-635.

[9] Belsley, D. A., E. Kuh and R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Col linearity.* New York: John Wiley.

[10] Gibbons, D. 1981. A Simulation Study of Some Ridge Estimators. *Journal of American Statistical Association.* 76: 131-139.

[11] Hoerl, A. E., R. W. Kennard and K. F. Baldwin. 1975. Ridge Regression: Some Simulation. *Commun. Stat.* 4:104-123.

[12] Edgeworth, F.Y. 1887. On Observations Relating to Several Quantity. *Journal Hermathena.* 6: 279-285.

[13] Huber, P. J. 1973. Robust Regression:Asymtotics. Conjectures and Monte Carlo. *The Annals of Statistics.* 1: 799-821.

[14] Rousseeuw, P. J. and A. M. Leroy. 1987. *Robust Regression and Outlier Detection.* New York: John Wiley & Sons.

[15] Yohai, V. J. 1987. High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics.* 15 (20): 642-656.

[16] Huber, P. J. 1973. Robust Regression: Asymtotics, Conjectures and Monte Carlo. *Ann. Stat.* 1:799-82.
[17] Vinod, H. D., and A. Ullah. 1981. Recent *Advances In Regression Methods.* New York: Marcel Dekker.
[18] Lawrence, K. D. and J. L. Arthur. 1990. *Robust Regression:Analysis and Application.* New York: Marcel Dekker.