Pak. J. Statist.
2021 Vol. 37(4), 369-391

ON SOME NEW RIDGE M-ESTIMATORS FOR LINEAR REGRESSION MODELS UNDER VARIOUS ERROR DISTRIBUTIONS

Muhammad Suhail[§], **Sohail Chand** and **Iqra Babar**
College of Statistical and Actuarial Sciences (CSAS)
University of the Punjab, Lahore, Pakistan
Email: msuhail88@aup.edu.pk
sohail.stat@pu.edu.pk
iqra.babar@ucp.edu.pk
[§] Corresponding author

ABSTRACT

Ridge regression is used to circumvent the problem of multicollinearity in the multiple linear regression models. Beside the multicollinearity, when the outliers in the y-direction are also present, then the usual ridge regression estimators gives inefficient results in terms of mean squared error (MSE). In order to mitigate such situation, ridge M-estimators are often used. Several estimators are available in literature but they do not perform well in terms of MSE when the joint problem of high multicollinearity and y-direction outliers is present. In this article, some new quantile based ridge M-estimators are proposed. The new estimators are then compared with other existing estimators through extensive Monte Carlo simulations for various error term distributions, degrees of multicollinearity and percentage of y-direction outliers. Based on simulation study with minimum MSE criterion, the new estimators outperform in many considered scenarios. Particularly, in case of high multicollinearity, y-direction outliers and heavy tailed error distributions, the proposed estimators have shown efficient results. A numerical example is also presented to support the simulation results.

KEYWORDS

Ridge regression; Multicollinearity; Outliers; M-estimators; MSE.

MSC 2010 Classification: 62J05; 62J07.

1. INTRODUCTION

Consider the matrix form of a multiple linear regression model:

$$y = X\beta + \varepsilon, \quad (1)$$

where y is the $n\times 1$ vector of response variables and ε is the $n\times 1$ vector of random errors. X is the $n\times p$ matrix whose column contains centered and standardized explanatory variables. Vector β is a vector of unknown parameters of order $p\times 1$. The ordinary least square (OLS) estimator of vector β is given as follows:

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

OLS estimators give imprecise results in the case of multicollinearity (Ertaş 2018). (Hoerl and Kennard 1970) suggested ridge regression (RR) method to overcome with multicollinearity problem and can be defined as:

$$\begin{aligned} \hat{\beta}(k) &= (X'X + kI)^{-1} X'Y \\ &= R_k \hat{\beta} \end{aligned} \tag{3}$$

where $R_k = (X'X + kI)^{-1} X'X$ and $k(>0)$ is known as the ridge or biasing parameter. OLS and RR estimators are sensitive to typical unusual observations (Maronna 2011). Robust M-estimator (ME) proposed by (Huber 1981) is used to deal with outliers. ME can be obtained by the solution of M-estimating equations $\sum \psi(e_j / s) = 0$ and $\sum \psi(e_j / s) x_j = 0$ where $e_j = Y_j - x_j' \hat{\beta}_m$, s is the scale estimator for errors, $\hat{\beta}_m$ is the ME and $\psi(.)$ is some suitably chosen function (Hampel et al. 1986).

(Silvapulle 1991) first suggested the use of ridge M-estimator (RM) to deal with multicollinearity and outliers in y-direction and is defined as:

$$\hat{\beta}_m(k) = R_k \hat{\beta}_m \tag{4}$$

Another approach based on prior information b_0 proposed by (Swindel 1976) as follows:

$$\tilde{\beta}(k, b_0) = (X'X + kI)^{-1} (X'Y + kb_0) \tag{5}$$

(Jahufer and Jianbao 2009) and (Gültay and Kaçıranlar 2015) suggested the use of prior information $b_0 = \hat{\beta}(k)$, therefore the modified RR (MRR) estimator becomes:

$$\begin{aligned} \tilde{\beta}(k) &= (X'X + kI)^{-1} (X'y + k\hat{\beta}(k)) \\ &= (I + kH_k) H_k X'X \hat{\beta} \\ &= S_k \hat{\beta}_{OLS}, \end{aligned} \tag{6}$$

where $H_k = (X'X + kI)^{-1}$ and $S_k = (I + kH_k) H_k X'X$. To make MRR robust towards outliers in y-direction, (Ertaş 2018) proposed the modified ridge M-estimator (MRM) by replacing the OLS estimator with ME in Eq. (6):

$$\tilde{\beta}_m(k) = S_k \hat{\beta}_m \tag{7}$$

The RR and RM estimators mainly depend on the good choice of ridge parameter k. Several approaches have been proposed for different models by different researchers for estimating k, see e.g., (Algamal 2018a), (Alheety and Kibria 2013), (Arashi et al. 2014), (Hoerl and Kennard 1970), (Kibria 2003), (Khalaf, Månsson, and Shukur 2013), (Lawless and Wang 1976), (Gültay and Kaçıranlar 2015), (Kibria and Banik 2016), (Ertaş 2018), (Ali et al. 2019), (Suhail and Chand 2019), (Suhail, Chand, and Kibria 2020; 2019) and very recently (Roozbeh, Arashi, and Hamzah 2020), (Suhail, Ilyas, and Ayanullah 2020) and (Yasin, Kamal, and Suhail 2020) among others. Inspired by (Ertaş 2018), we have reviewed the relevant existing estimators and proposed some new quantile based RR and RM estimators for estimating k and is defined in Section 2. The criterion for the comparison of estimators is also presented in Section 2. The Monte Carlo simulation study and its results are described in Section 3. A real life application is considered in Section 4. Section 5 presents some concluding remarks.

2. STATISTICAL METHODOLOGY

Consider the canonical form of model in Eq. (1):

$$y = Z\alpha + \varepsilon, \tag{8}$$

where $Z = XD$ and $\alpha = (\alpha_1, \alpha_2, ..., \alpha_p)' = D'\beta$, D is an orthogonal matrix that is $D'D = I$ and $Z'Z = D'X'XD = \Lambda$, $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ contains the eigen values of the $X'X$ matrix. The estimators defined in Eq. (2) to (4) can be written in canonical form as follows:

$$OLS: \quad \hat{\alpha} = \Lambda^{-1} Z'Y, \tag{9}$$

$$RR: \quad \hat{\alpha}(k) = R_k^* \hat{\alpha}, \tag{10}$$

$$RM: \quad \hat{\alpha}_m(k) = R_k^* \hat{\alpha}_m, \tag{11}$$

where $R_k^* = (\Lambda + kI)^{-1}\Lambda$, $k>0$ $\hat{\alpha}$ is the OLS and $\hat{\alpha}_m$ is the M-estimator of canonical coefficient α. Note that $\hat{\alpha} = D'\hat{\beta}$ such that $MSE(\hat{\alpha}) = MSE(\hat{\beta})$. Therefore, it suffices to consider the canonical form only. The estimators in Eq. (6) and (7) can be written as:

$$MRR: \quad \tilde{\alpha}(k) = S_k^* \hat{\alpha} \tag{12}$$

$$MRM: \quad \tilde{\alpha}_m(k) = S_k^* \hat{\alpha}_m \tag{13}$$

where $S_k^* = (I + kH_k^*)H_k^*\Lambda$ and $H_k^* = (\Lambda + kI)^{-1}$. The MSE of above estimators is defined in (Ertaş 2018). The estimator for ridge parameter k was first proposed by (Hoerl and Kennard 1970) as:

$$\hat{k}_j = \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}, \qquad j = 1, 2, ..., p \tag{14}$$

where $\hat{\sigma}^2$ is the unbiased estimator of error variance σ^2. (Hoerl and Kennard 1970) also suggested to use: $\hat{k}_{RR} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2}$. Later, (Kibria 2003) suggested to take geometric mean of $\hat{k}_j$ defined in Eq. (20) and is given below:

$$\hat{k}_{KGM} = \frac{\hat{\sigma}^2}{(\prod_{j=1}^{p} \hat{\alpha}_j^2)^{1/p}} \tag{15}$$

A ridge M-estimator for $\hat{k}_j$, suggested by (Silvapulle 1991), is given by:

$$\hat{k}_{RM} = \frac{p\hat{A}^2}{\sum_{j=1}^{p} (\hat{\alpha}_m^2)_j},$$

where $\hat{A}^2$ given by (Huber 1981) is defined as:

$$\hat{A}^2 = \frac{s^2 (n-p)^{-1} \sum_{j=1}^{p} (\psi(e_j / s))^2}{(n^{-1} \sum_{j=1}^{p} \psi'(e_j / s))^2}, \tag{16}$$

where $\psi(.)$ is defined in (Hampel et al. 1986). (Gültay and Kaçıranlar 2015) suggested the following estimator:

$$\tilde{k}_j = \frac{\sqrt{\lambda_j \hat{\alpha}_j^2 \hat{\sigma}^2 + \hat{\sigma}^4} + \hat{\sigma}^2}{\hat{\alpha}_j^2} \tag{17}$$

Modified ridge M-estimator of $\tilde{k}_j$ suggested by (Ertaş 2018) is given by:

$$\tilde{k}_{mj} = \frac{\sqrt{\lambda_j (\hat{\alpha}_m^2)_j \hat{A}^2 + \hat{A}^4} + \hat{A}^2}{(\hat{\alpha}_m^2)_j} \tag{18}$$

(Ertaş 2018) further suggested to use harmonic mean of Eq. (17) and (18) and is defined as follows:

$$\tilde{k}_{MRR} = \frac{p}{\sum_{j=1}^{p} \frac{\hat{\alpha}_j^2}{\sqrt{\lambda_j \hat{\alpha}_j^2 \hat{\sigma}^2 + \hat{\sigma}^4} + \hat{\sigma}^2}} \tag{19}$$

$$\tilde{k}_{MRM} = \frac{p}{\sum_{j=1}^{p} \frac{(\hat{\alpha}_m^2)_j}{\sqrt{\lambda_j (\hat{\alpha}_m^2)_j \hat{A}^2 + \hat{A}^4} + \hat{A}^2}} \tag{20}$$

Recently, (Suhail, Chand, and Kibria 2020) suggested quantile based ridge estimators (SQ) as: $\hat{k}_\gamma = \left\{ \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2} \right\}_\gamma$, $0 < \gamma < 1$, where γ is the quantile level. They showed that $\hat{k}_{\gamma=0.95}$ outperform when the level of multicollinearity is high or error variance is greater than one.

The Suhail quantile estimator $\hat{k}_{\gamma=0.95}$ is denoted here by SQ95. New estimators for ridge parameter are proposed in the section to follow.

2.1. New Estimators for Ridge Parameter

Following (Ali et al. 2019), (Suhail and Chand 2019) and (Suhail, Chand, and Kibria 2019; 2020), we proposed some new quantile-based ridge (QR) and ridge M- (QM) estimators and are defined as follows:

$$QR(\gamma) = \left\{\tilde{k}_{(j)}\right\}_{\gamma} = \left\{\tilde{k}_{(1)}, \tilde{k}_{(2)}, ..., \tilde{k}_{(p)}\right\}_{\gamma}, 0 < \gamma < 1 \tag{21}$$

$$QM(\gamma) = \left\{\tilde{k}_{m(j)}\right\}_{\gamma} = \left\{\tilde{k}_{m(1)}, \tilde{k}_{m(2)}, ..., \tilde{k}_{m(p)}\right\}_{\gamma}, \tag{22}$$

where $\tilde{k}_{(j)}$ and $\tilde{k}_{m(j)}$ are the ordered values of Eq. (17) and (18). The quantile probability γ mainly depends on the two major factors, i.e., multicollinearity and error variance. The condition number (CN) defined as $CN = \sqrt{\lambda_{max}/\lambda_{min}}$, where λ_{max} and λ_{min} are the maximum and minimum eigen values of the matrix $X'X$ respectively, is generally used to measure the level or strength of multicollinearity (Kibria and Banik 2016). Following (Gujarati 2009), multicollinearity is moderate, high and severe when CN is between 10 and 30, 30 and 100 and greater than 100 respectively. Signal to noise ratio (SNR) defined as $SNR = 1/\sigma^2$ is normally used to measure the strength of error variance or error standard deviation (Kibria 2003). Also, according to (Kibria 2003), error variance is low when SNR is greater than one and high for less than or equal to one respectively. Table 2.1 suggests the appropriate choice of quantile probability (γ) for varying levels of CN and SNR. For the purpose of comparison, only four values of the quantile probability are considered here and are given below:

$$\gamma = 0.25, 0.50, 0.75, 0.95 .$$

The four new QR estimators are denoted by QR25, QR50, QR75 and QR95 and QM estimators by QM25, QM50, QM75 and QM95.

Table 2.1
The Appropriate Choice of Quantile Probability (γ)

SNR	CN	γ)
SNR ≥ 1	10 < CN < 30	$0.10 \leq \gamma \leq 0.50$
	30 < CN < 100	$0.25 \leq \gamma \leq 0.75$
	CN > 100	$0.50 \leq \gamma \leq 0.95$
SNR < 1	10 < CN< 30	$0.75 \leq \gamma \leq 0.95$
	30 < CN < 100	$0.90 \leq \gamma \leq 0.95$
	CN > 100	$0.95 \leq \gamma \leq 0.99$

2.2. Performance criteria

In this study, MSE criterion is used to measure the goodness of an estimator. The estimated MSE (EMSE) of $\hat{\alpha}$ is defined as:

$$MSE(\hat{\alpha}) = E(\hat{\alpha}-\alpha)'(\hat{\alpha}-\alpha) = \frac{1}{p}\sum_{j=1}^{p}(\hat{\alpha}_j - \alpha_j)^2, \tag{23}$$

Following (Khalaf, Månsson, and Shukur 2013), the new and existing estimators are compared through a Monte Carlo simulation study in the section to follow.

3. THE MONTE CARLO SIMULATION STUDY

In this section, simulation design is presented. The simulation results are also discussed in this section.

3.1. Simulation Design

Following (McDonald and Galarneau 1975), (Huang and Yang 2014) and (Algamal 2018b), the predictors are generated by using the following equation:

$$x_{ij} = (1-r^2)^{1/2} z_{ij} + r z_{ip+1}, \; i = 1,2,...,n, \; j = 1,2,..., p, \tag{24}$$

where r^2 shows the level of multicollinearity and $z_{ij} \sim N(0,1)$ are the independent pseudo random numbers. We consider $r = 0.85$, 0.95 and 0.99. The response variable can be obtained by:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i, \qquad i = 1,2,...,n, \tag{25}$$

where ε_i is the random error. In this study, the errors of the model are generated from the following distributions.

i) Normal distribution ($N(0,1)$),
ii) Student's t distribution ($t(3)$),
iii) Chi Square distribution ($\chi^2_{(4)}$),
iv) F distribution ($F(2,7)$),
v) Gamma distribution ($Gamma(2,5)$),
vi) Cauchy distribution ($Cauchy(0,1)$),

Following (Kibria 2003), we choose $\beta_0 = 0$ and eigen vector against the largest eigen value as the coefficients $\beta_1, \beta_2, ..., \beta_p$. SNR and Skewness (Sk) is computed for each of the error distributions. The other factors considered in our study are the sample size ($n = 20$ and 50) and the number of predictors ($p = 4$ and 10). We consider two different cases for outliers in y-direction, i.e. 10% and 20% outliers in y-direction also considered by (Ertaş 2018). EMSE

The EMSE is obtained using the following equation:

$$MSE(\hat{\alpha}_j) = \frac{1}{5000}\sum_{k=1}^{5000}(\hat{\alpha}_{jk} - \alpha_j)'(\hat{\alpha}_{jk} - \alpha_j),\ j = 1, 2, ..., p. \quad (26)$$

The EMSE of the estimators are presented in Tables 3.1-3.12. R-programming is used and M-estimators are computed using the code 'rlm'.

Table 3.1
EMSEEMSE from $N(0,1)$ with SNR=1, Sk=0 and 10% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	10.097	25.836	119.855	2.352	6.989	35.781
RR	1.761	4.073	18.019	1.067	2.676	13.126
KGM	1.080	2.688	6.302	0.399	1.119	4.221
MRR	1.592	3.959	17.724	0.545	1.366	6.659
M	3.444	5.945	12.581	3.422	4.196	8.894
RM	2.448	4.367	5.770	2.930	3.178	5.078
MRM	2.364	4.213	5.242	2.862	3.053	4.735
SQ95	0.553	0.502	0.229	0.345	0.496	0.616
QR25	2.045	5.414	24.925	0.684	1.679	8.216
QR50	1.059	3.134	15.820	0.435	0.763	4.657
QR75	0.582	0.827	0.279	**0.292**	0.383	0.489
QR95	**0.494**	**0.263**	**0.200**	0.330	**0.260**	**0.127**
QM25	2.366	4.215	5.214	2.896	3.049	4.806
QM50	1.991	3.747	3.704	2.677	2.712	3.832
QM75	1.647	3.034	2.513	2.304	2.345	3.117
QM95	1.524	2.717	2.218	2.126	2.177	2.907
	$p = 10$					
OLS	51.171	161.051	852.745	4.857	15.067	78.791
RR	29.559	94.449	498.167	0.675	4.818	11.321
KGM	1.655	4.393	20.417	0.222	0.410	1.518
MRR	8.514	25.913	137.593	0.491	1.430	7.345
M	12.556	38.190	196.251	3.957	6.740	23.512
RM	6.895	18.049	84.561	2.727	3.562	8.821
MRM	6.503	16.578	77.303	2.632	3.374	8.118
SQ95	0.847	0.675	0.526	0.689	0.476	0.278
QR25	6.692	20.295	116.758	0.612	1.737	8.396
QR50	2.311	7.617	44.072	0.235	0.491	2.362
QR75	0.716	1.166	7.108	**0.219**	**0.137**	0.276
QR95	**0.836**	**0.652**	**0.503**	0.670	0.445	**0.231**
QM25	6.286	15.222	69.578	2.712	3.411	7.919
QM50	4.996	11.101	47.567	2.334	2.678	5.096
QM75	3.635	7.014	25.823	1.967	2.159	3.469
QM95	2.129	3.833	11.647	1.457	1.636	2.495

Table 3.2
EMSE from $N(0,1)$ with SNR=1, Sk=0 and 20% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	45.887	96.013	372.031	20.705	66.021	337.857
RR	18.921	44.262	194.608	13.546	43.827	225.718
KGM	4.825	20.507	8.858	2.860	6.715	20.686
MRR	14.193	29.030	104.346	5.612	17.446	89.389
M	3.491	7.528	14.881	3.738	4.908	11.658
RM	2.275	5.355	6.113	3.278	3.876	7.180
MRM	2.158	5.178	5.439	3.202	3.719	6.701
SQ95	0.950	2.987	0.951	0.567	1.139	0.400
QR25	18.620	41.371	169.876	2.601	6.643	32.519
QR50	15.303	24.901	47.355	1.146	2.845	15.432
QR75	0.952	5.708	**0.887**	0.612	0.960	0.620
QR95	**0.949**	**1.560**	0.951	**0.450**	**0.347**	**0.094**
QM25	2.153	5.159	5.779	3.257	3.782	7.050
QM50	1.726	4.576	3.588	3.019	3.370	5.359
QM75	1.411	3.642	2.005	2.572	2.894	4.180
QM95	1.312	3.229	1.721	2.359	2.671	3.870
	$p = 10$					
OLS	524.296	1784.539	9514.966	7.886	21.789	104.979
RR	203.175	663.386	3591.124	1.123	3.334	19.968
KGM	55.275	161.459	737.639	0.223	0.198	0.447
MRR	146.624	495.678	2681.651	0.435	1.025	4.619
M	168.015	531.592	2826.048	3.886	6.628	23.334
RM	75.363	216.938	1101.639	2.646	3.418	8.428
MRM	73.915	214.081	1095.144	2.557	3.246	7.789
SQ95	2.437	1.141	0.334	0.775	0.667	0.434
QR25	133.652	447.395	2398.717	0.557	1.518	7.258
QR50	39.846	146.064	820.908	**0.183**	**0.157**	0.460
QR75	12.686	53.211	321.291	0.311	0.192	**0.095**
QR95	**1.357**	**0.510**	**0.139**	0.762	0.646	0.405
QM25	70.238	195.702	983.410	2.668	3.335	7.729
QM50	51.804	149.135	769.532	2.272	2.560	4.795
QM75	25.805	64.622	321.287	1.885	1.991	3.110
QM95	10.700	26.605	97.110	1.415	1.517	2.260

Table 3.3
EMSE from $t(3)$ with SNR=0.333, Sk=undefined and 10% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	85.719	218.547	1033.655	30.035	84.852	397.209
RR	14.178	33.859	153.618	13.115	34.896	153.691
KGM	2.635	13.838	16.385	1.534	1.399	5.940
MRR	11.788	30.792	147.224	5.265	13.272	57.860
M	4.096	8.174	21.992	1.540	5.027	11.854
RM	2.637	5.564	10.065	0.978	3.683	5.959
MRM	2.522	5.326	9.042	0.874	3.536	5.465
SQ95	0.956	0.357	0.880	0.699	0.773	0.296
QR25	2.505	5.380	9.158	0.913	3.555	5.324
QR50	2.062	4.600	6.195	0.693	3.164	4.153
QR75	1.628	3.555	3.967	**0.589**	2.663	3.357
QR95	1.480	3.145	3.377	0.592	2.441	3.105
QM25	15.893	42.583	216.516	6.163	15.938	71.427
QM50	7.792	29.337	159.491	1.059	0.848	1.589
QM75	**0.924**	1.554	**0.722**	0.637	**0.690**	0.312
QM95	0.955	**0.283**	0.879	0.686	0.767	**0.265**
	$p = 10$					
OLS	593.700	1316.222	5199.137	64.410	199.413	1069.498
RR	47.454	156.323	175.883	31.274	3.065	82.288
KGM	15.811	49.555	162.071	2.016	1.828	13.152
MRR	70.358	158.705	629.555	11.090	32.763	183.197
M	16.215	38.757	160.694	5.186	10.455	41.666
RM	9.758	22.178	87.060	3.142	4.778	14.272
MRM	9.168	20.567	80.179	3.027	4.513	13.211
SQ95	**0.663**	2.438	2.128	0.833	0.773	0.321
QR25	9.258	20.573	80.014	3.005	4.263	11.742
QR50	7.616	16.425	61.824	2.560	3.306	7.496
QR75	5.686	11.636	40.075	2.146	2.678	4.696
QR95	3.141	6.173	19.535	1.537	1.909	3.085
QM25	44.167	111.561	559.616	15.008	17.626	159.797
QM50	21.992	61.025	213.681	2.278	1.245	11.170
QM75	3.567	21.090	63.230	**0.647**	**0.502**	0.811
QM95	0.677	**1.907**	**1.709**	0.823	0.759	**0.293**

Table 3.4
EMSE from $t(3)$ with SNR=0.333, Sk=undefined and 20% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p=4$					
OLS	406.842	847.965	3300.164	51.506	150.324	723.791
RR	165.810	395.069	1752.095	16.711	46.806	215.322
KGM	58.524	87.041	145.977	0.830	3.491	4.376
MRR	125.655	250.387	925.713	5.771	16.019	72.981
M	4.524	8.875	29.301	3.766	3.334	13.135
RM	2.590	4.643	12.581	3.130	1.643	7.546
MRM	2.424	4.322	11.146	3.059	1.472	7.045
SQ95	2.004	1.996	1.973	0.892	0.217	0.818
QR25	2.453	4.437	11.996	3.105	1.358	7.267
QR50	1.877	3.285	7.361	2.837	1.039	5.786
QR75	1.394	2.230	4.234	2.372	0.892	4.784
QR95	**1.267**	**1.940**	3.500	2.163	0.861	4.439
QM25	161.667	373.224	1554.276	8.000	19.140	81.686
QM50	140.188	232.034	369.420	**0.652**	1.304	12.734
QM75	5.664	4.684	3.749	0.843	0.272	**0.621**
QM95	1.975	1.980	**1.970**	0.889	**0.162**	0.817
	$p=10$					
OLS	3374.776	8398.935	37031.50 5	124.012	342.451	1701.492
RR	313.520	1005.314	3066.079	39.595	102.412	470.519
KGM	16.106	56.226	386.498	7.235	15.396	51.768
MRR	429.764	1131.333	5176.690	24.970	70.404	352.722
M	49.543	126.216	514.246	5.229	10.644	43.764
RM	37.526	91.923	360.106	3.076	4.683	14.485
MRM	35.458	85.941	335.524	2.965	4.440	13.449
SQ95	**0.990**	**0.960**	0.127	0.795	0.733	0.659
QR25	36.404	87.446	343.482	2.967	4.245	11.936
QR50	32.538	76.542	296.237	2.468	3.182	7.661
QR75	26.019	60.718	219.745	2.039	2.502	5.130
QR95	13.836	36.024	135.867	1.454	1.786	3.512
QM25	160.969	582.633	2939.663	37.190	102.690	526.937
QM50	24.391	92.257	489.761	9.837	40.570	166.424
QM75	3.890	10.367	31.394	1.524	2.067	17.249
QM95	**0.990**	**0.960**	**0.100**	**0.779**	**0.718**	**0.644**

Table 3.5

EMSE from $\chi^2_{(4)}$ with SNR=0.125, Sk=1.4142 and 10% outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	155.714	397.898	1861.134	31.398	101.018	519.467
RR	21.788	51.394	233.734	0.933	0.900	0.869
KGM	2.732	17.704	13.791	**0.876**	0.921	1.389
MRR	18.361	47.509	225.030	2.044	5.418	26.633
M	8.386	21.223	72.764	6.440	10.906	54.885
RM	2.066	7.454	15.127	3.051	3.079	18.863
MRM	2.066	7.517	14.798	3.049	3.045	18.499
SQ95	0.988	0.972	0.952	0.935	**0.263**	0.861
QR25	25.502	68.011	333.142	2.874	8.765	46.148
QR50	11.614	42.967	231.737	1.068	1.520	4.726
QR75	**0.987**	1.434	**0.880**	0.885	0.802	**0.732**
QR95	0.988	**0.346**	0.952	0.935	0.893	0.860
QM25	2.173	7.968	15.883	3.058	3.071	18.762
QM50	1.687	6.362	11.467	2.526	2.494	16.470
QM75	1.297	4.360	8.262	1.891	1.835	13.255
QM95	1.191	3.615	7.010	1.672	1.598	11.690
	$p = 10$					
OLS	731.810	1923.219	8706.801	87.851	240.622	1102.439
RR	145.522	377.095	697.737	25.561	66.323	290.513
KGM	30.663	71.827	331.938	0.940	1.443	3.182
MRR	107.188	286.465	1296.336	5.307	13.943	62.605
M	49.259	128.175	562.892	10.425	26.958	121.771
RM	15.273	38.362	154.679	2.939	5.148	18.788
MRM	15.437	37.823	148.617	2.985	5.176	18.648
SQ95	0.735	0.500	0.293	0.960	0.891	0.813
QR25	150.630	420.724	1958.621	3.144	6.695	30.786
QR50	49.485	210.572	915.412	1.099	1.515	6.158
QR75	4.266	13.266	71.953	**0.867**	**0.349**	1.113
QR95	**0.650**	**0.434**	**0.248**	0.959	0.446	**0.807**
QM25	15.760	37.450	145.595	3.037	4.871	16.704
QM50	11.549	30.405	120.375	2.348	3.686	11.751
QM75	6.659	21.931	104.871	1.727	2.784	9.607
QM95	2.373	7.959	56.468	1.186	1.574	5.216

Table 3.6
EMSE from $\chi^2_{(4)}$ with SNR=0.125, Sk=1.4142 and 20% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	727.066	1511.369	5894.879	235.138	922.841	5486.803
RR	293.920	696.205	3096.316	17.929	21.230	68.623
KGM	112.949	167.605	286.060	23.620	57.052	126.227
MRR	223.221	441.352	1637.328	47.505	175.725	1147.537
M	25.736	62.868	217.919	9.511	38.406	178.243
RM	5.483	19.521	56.839	2.824	14.645	40.208
MRM	5.106	19.018	52.621	2.792	14.249	36.702
SQ95	3.789	9.957	3.267	12.512	12.781	10.511
QR25	289.201	655.749	2731.144	23.081	65.274	529.949
QR50	245.810	397.627	680.193	17.482	26.666	104.950
QR75	13.926	10.653	7.343	14.694	17.489	13.496
QR95	3.648	**3.475**	**3.266**	13.081	14.209	**10.472**
QM25	6.789	21.984	61.762	2.868	15.770	36.488
QM50	2.677	13.211	36.189	2.095	12.397	21.820
QM75	1.139	8.068	23.588	1.570	8.728	13.700
QM95	**0.909**	6.498	19.830	**1.415**	**7.625**	12.230
	$p = 10$					
OLS	1802.574	4834.788	22943.374	276.557	779.847	3777.257
RR	621.501	1734.022	5240.296	77.968	252.761	1293.847
KGM	48.994	230.357	787.934	1.764	3.373	13.007
MRR	338.547	892.807	4178.743	30.804	89.104	434.696
M	179.265	480.015	2275.136	15.505	39.482	175.808
RM	39.600	102.901	474.728	5.909	12.165	43.205
MRM	36.136	94.826	439.791	5.935	12.004	41.921
SQ95	0.869	0.803	0.420	**0.998**	**0.467**	**0.977**
QR25	501.236	1092.262	4633.615	49.874	142.073	546.504
QR50	124.448	411.434	1611.824	2.109	6.222	71.165
QR75	2.289	75.678	308.686	1.029	1.072	2.091
QR95	**0.863**	**0.709**	**0.372**	**0.998**	0.987	**0.977**
QM25	41.454	105.195	476.155	6.288	12.274	42.431
QM50	10.699	29.256	137.819	4.922	9.332	29.129
QM75	4.047	14.657	78.051	3.362	7.027	22.846
QM95	1.326	5.107	41.225	1.653	3.349	13.368

Table 3.7

EMSE from $F(2,7)$ with SNR=0.219, Sk=6.299 and 10% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p=4$					
OLS	12.679	32.767	154.257	2.097	5.830	30.023
RR	1.859	4.321	19.390	0.349	0.723	4.178
KGM	1.059	2.476	5.506	0.285	0.289	0.728
MRR	1.861	4.583	21.464	0.290	0.558	3.017
M	3.910	6.587	16.284	3.278	3.926	7.576
RM	2.248	3.818	5.543	2.507	2.622	3.505
MRM	2.230	3.771	5.279	2.491	2.598	3.433
SQ95	0.602	0.445	0.309	0.554	0.409	0.224
QR25	2.202	5.649	26.470	**0.251**	0.557	3.131
QR50	1.144	2.827	14.161	0.293	**0.217**	0.870
QR75	0.599	0.675	**0.288**	0.428	0.293	**0.158**
QR95	**0.564**	**0.298**	0.289	0.527	0.368	0.191
QM25	2.259	3.857	5.519	2.542	2.623	3.501
QM50	1.937	3.342	4.321	2.362	2.423	3.139
QM75	1.580	2.587	3.345	2.074	2.099	2.678
QM95	1.453	2.287	3.015	1.934	1.944	2.497
	$p=10$					
OLS	105.105	285.807	1388.985	8.831	27.303	130.616
RR	50.099	126.796	522.504	2.778	8.752	18.072
KGM	5.191	10.814	36.742	0.277	0.443	1.439
MRR	20.762	56.059	266.675	0.857	2.331	9.831
M	23.486	60.556	293.804	3.848	6.567	22.064
RM	8.371	19.286	88.673	2.093	2.466	4.865
MRM	8.269	18.943	86.386	2.077	2.430	4.724
SQ95	0.608	0.490	0.300	0.713	0.516	0.309
QR25	31.185	82.418	358.068	0.864	1.877	7.789
QR50	3.858	7.613	34.095	0.366	0.527	2.129
QR75	0.755	1.226	4.408	**0.259**	**0.170**	0.357
QR95	**0.554**	**0.396**	**0.207**	0.698	0.490	**0.272**
QM25	9.099	20.263	91.218	2.147	2.440	4.537
QM50	5.773	12.262	51.628	1.877	2.031	3.220
QM75	3.339	6.161	21.260	1.649	1.779	2.692
QM95	1.697	3.046	10.097	1.284	1.366	2.048

Table 3.8
EMSE from $F(2,7)$ with SNR=0.219, Sk=6.299 and 20% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	48.454	102.564	407.219	6.417	17.926	84.653
RR	19.863	45.632	200.058	1.056	2.834	14.323
KGM	4.810	6.267	10.491	0.485	0.739	1.892
MRR	14.517	29.318	111.913	0.821	1.983	9.308
M	5.908	11.957	37.257	3.280	4.098	8.891
RM	2.672	5.146	12.132	2.435	2.526	3.476
MRM	2.512	4.875	10.947	2.402	2.473	3.334
SQ95	0.944	0.934	0.934	0.464	0.346	0.208
QR25	19.187	42.012	171.801	0.911	2.349	11.206
QR50	14.075	23.771	59.644	0.358	0.536	2.391
QR75	1.008	**0.920**	**0.880**	**0.350**	**0.269**	**0.165**
QR95	**0.940**	0.932	0.933	0.414	0.292	0.181
QM25	2.694	5.294	12.820	2.460	2.508	3.562
QM50	1.935	3.689	7.526	2.218	2.233	2.821
QM75	1.296	2.483	4.266	1.913	1.913	2.241
QM95	1.179	2.185	3.584	1.774	1.784	2.108
	$p = 10$					
OLS	282.906	877.040	4510.205	17.832	54.174	265.651
RR	135.066	411.148	2128.757	7.198	21.767	17.034
KGM	2.734	5.127	19.969	0.585	0.999	5.139
MRR	24.404	72.741	377.350	2.141	6.085	29.347
M	34.637	102.826	565.132	4.794	8.640	29.906
RM	11.686	33.918	190.333	2.859	3.932	9.209
MRM	10.890	31.558	176.624	2.840	3.867	8.902
SQ95	0.883	0.768	0.628	0.638	0.517	0.270
QR25	10.505	26.452	118.175	2.082	4.792	23.754
QR50	2.748	5.514	23.457	0.684	1.130	6.899
QR75	1.022	1.059	4.538	**0.273**	**0.227**	1.305
QR95	**0.878**	**0.758**	**0.613**	0.616	0.485	**0.199**
QM25	10.281	27.865	150.725	2.884	3.858	8.710
QM50	5.797	15.044	85.172	2.592	3.380	6.869
QM75	2.968	8.021	45.821	2.207	2.979	6.053
QM95	1.168	3.195	26.103	1.465	1.963	4.272

Table 3.9
EMSE from $Gamma(2,5)$ with SNR=0.05, Sk=0.8944
and 10% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	248.265	633.071	2983.392	62.335	181.510	881.429
RR	27.577	65.149	292.185	2.236	6.405	36.536
KGM	3.571	18.092	13.306	0.952	1.428	2.898
MRR	23.394	59.619	284.531	3.631	10.179	48.068
M	20.069	56.331	223.432	10.679	27.778	129.051
RM	2.077	13.976	39.058	2.358	4.889	20.202
MRM	2.150	14.564	39.800	2.399	5.022	20.320
SQ95	**1.019**	0.464	0.978	0.720	0.705	0.699
QR25	33.398	88.826	424.894	3.028	8.859	45.035
QR50	14.357	50.127	267.800	1.183	3.033	13.243
QR75	1.081	1.230	**0.953**	**0.624**	**0.553**	**0.503**
QR95	**1.019**	**0.438**	0.978	0.714	0.702	0.697
QM25	2.457	15.857	41.941	2.611	5.568	22.470
QM50	1.609	11.835	32.609	1.993	3.980	16.330
QM75	1.182	7.188	23.448	1.457	2.567	12.013
QM95	1.089	5.568	19.572	1.317	2.143	10.270
	$p = 10$					
OLS	1168.540	3406.239	16941.627	147.580	430.978	2186.841
RR	318.068	942.986	4579.962	58.680	165.096	659.626
KGM	13.302	22.832	130.183	4.693	12.933	59.324
MRR	117.813	342.492	1710.874	18.278	53.939	280.338
M	143.861	407.113	1998.032	37.445	109.634	542.196
RM	11.453	31.136	150.649	5.095	12.446	54.081
MRM	12.312	33.449	159.525	5.347	12.900	54.935
SQ95	0.884	0.824	0.698	0.740	0.535	0.295
QR25	145.494	379.132	1493.559	25.956	76.233	439.223
QR50	23.696	29.450	343.246	8.938	28.183	181.171
QR75	2.251	2.899	27.942	1.206	3.697	23.577
QR95	**0.881**	**0.820**	**0.692**	**0.717**	**0.494**	**0.263**
QM25	11.656	27.816	122.941	5.211	11.787	48.158
QM50	5.956	16.532	71.406	3.547	8.584	35.296
QM75	2.906	9.296	51.873	2.213	5.850	29.405
QM95	1.273	2.950	21.796	1.262	2.355	14.331

Table 3.10
EMSE from $Gamma(2,5)$ with SNR=0.05, Sk=0.8944 and 20% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	1141.673	2381.325	9284.338	48.107	141.295	691.763
RR	443.630	1052.449	4676.373	9.044	26.526	126.048
KGM	172.146	254.590	430.530	0.529	0.559	1.194
MRR	333.644	661.836	2443.815	1.689	4.719	22.356
M	121.975	285.587	1027.613	17.503	46.568	212.211
RM	20.266	72.067	256.986	3.611	8.060	32.656
MRM	19.761	71.661	243.523	3.676	8.102	31.733
SQ95	6.116	12.398	**4.773**	1.734	2.042	2.317
QR25	439.033	985.460	4084.232	1.273	3.372	17.024
QR50	361.925	585.303	1043.231	**0.492**	0.545	2.202
QR75	24.419	17.876	11.653	0.662	**0.511**	**0.314**
QR95	5.862	**5.278**	4.822	0.740	0.616	0.472
QM25	28.402	85.768	281.966	4.196	9.913	40.673
QM50	9.986	50.235	179.567	2.833	5.742	21.271
QM75	3.235	28.698	125.155	1.840	3.568	15.068
QM95	**2.049**	21.691	103.885	1.586	2.934	13.203
	$p = 10$					
OLS	6284.219	17477.885	80214.918	506.972	1430.581	6991.630
RR	1361.109	3807.030	4090.377	419.753	1170.016	1604.012
KGM	269.703	492.817	1259.215	15.401	48.344	165.133
MRR	1128.099	2990.366	13545.203	123.649	356.421	1767.799
M	3564.937	10105.408	47328.955	98.627	278.514	1340.456
RM	1110.652	3016.720	13604.477	21.799	57.974	264.119
MRM	1041.407	2841.286	12914.679	21.229	55.367	248.877
SQ95	**1.085**	0.955	0.473	0.837	**0.865**	0.913
QR25	667.635	1435.024	6412.247	70.204	258.847	1424.601
QR50	242.856	493.041	2031.817	14.250	54.859	202.987
QR75	64.387	191.114	284.075	2.882	6.527	19.202
QR95	2.948	**0.818**	**0.409**	**0.785**	0.870	**0.785**
QM25	608.145	1314.196	5170.331	25.276	66.339	304.413
QM50	404.969	1025.264	4096.643	14.086	34.875	148.333
QM75	227.505	755.314	3616.216	8.267	21.088	82.940
QM95	52.324	261.530	2036.179	3.125	9.485	50.148

Table 3.11
EMSE from $Cauchy(0,1)$ with SNR=undefined, Sk=undefined and 10% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	9963462.888	43699.306	2784294.144	10985.252	111958.408	15641509.960
RR	568824.614	7792.772	604688.872	2722.248	45707.838	217319.943
KGM	15625.643	1782.270	44556.980	589.278	10069.911	117731.823
MRR	200190.481	4821.530	436540.959	1869.369	28910.020	3268025.629
M	13.228	33.491	118.559	3.531	11.191	46.936
RM	7.670	21.040	63.190	1.640	5.870	20.637
MRM	7.131	19.505	56.463	1.464	5.529	18.892
SQ95	982.633	175.404	337.176	242.230	233.157	394.311
QR25	94164.813	5441.270	559731.349	1123.102	22569.898	590204.551
QR50	15371.507	3595.772	392923.773	443.851	12151.605	164080.156
QR75	1784.069	378.194	662.822	270.235	603.040	1301.954
QR95	965.937	158.949	325.222	242.061	172.764	396.001
QM25	7.310	20.328	58.882	1.366	5.555	19.082
QM50	5.617	16.018	41.113	1.000	4.225	11.737
QM75	4.074	11.821	27.235	0.810	3.005	6.969
QM95	**3.426**	**9.725**	**22.035**	**0.770**	**2.620**	**6.128**
	$p = 10$					
OLS	73674.854	3735834.568	1532827.656	71610.716	16753506.900	9420891.372
RR	37684.302	1314176.899	597537.196	34216.714	5615756.269	4611721.715
KGM	1879.208	83572.671	26148.689	2417.205	302227.031	104589.152
MRR	12321.974	437357.181	299200.868	15022.244	3848995.937	1035653.915
M	334.867	1053.440	4891.588	10.459	24.895	106.799
RM	195.692	624.480	2879.411	5.463	11.141	40.271
MRM	178.361	554.458	2576.144	5.148	10.285	36.558
SQ95	44.503	89.107	**50.397**	23.206	105.128	166.758
QR25	7787.762	265545.675	225094.394	14240.051	4866078.841	458732.562
QR50	2430.797	103994.914	51007.198	3729.729	2108681.715	150091.562
QR75	553.827	45973.264	9558.778	283.629	11069.155	45791.585
QR95	38.724	84.885	52.714	23.830	97.079	177.346
QM25	180.485	540.819	2583.033	5.296	10.275	35.903
QM50	132.596	371.421	1824.401	3.924	6.941	20.886
QM75	71.190	196.504	970.228	2.778	4.447	9.991
QM95	**22.390**	**75.790**	333.025	**1.678**	**2.577**	**5.159**

Table 3.12
EMSE from $Cauchy(0,1)$ with SNR=undefined, Sk=undefined and 20% Outliers in y-direction

n	20			50		
r	0.85	0.95	0.99	0.85	0.95	0.99
	$p = 4$					
OLS	9963422.607	43723.591	2784215.025	10986.980	111941.455	15642487.480
RR	568848.705	7871.724	604886.531	2682.154	45624.275	216921.613
KGM	15617.879	1782.049	44531.465	589.419	10058.527	117742.963
MRR	200170.792	4834.916	436417.641	1869.991	28904.084	3268390.598
M	22.858	58.373	217.211	5.199	12.912	48.790
RM	12.270	33.256	115.359	3.102	7.456	20.066
MRM	11.238	30.675	103.138	2.974	7.107	18.220
SQ95	981.164	174.771	337.042	242.569	231.609	394.304
QR25	94177.965	5452.237	559821.889	1123.499	22571.836	590279.454
QR50	15378.739	3593.430	392895.991	444.244	12153.089	164116.745
QR75	1782.980	377.608	662.192	270.477	600.194	1301.779
QR95	964.548	158.355	325.306	242.274	171.955	396.039
QM25	12.202	33.152	111.853	**1.750**	7.054	18.011
QM50	8.082	23.509	75.872	2.934	5.725	10.782
QM75	4.768	15.685	47.744	2.449	4.483	6.561
QM95	**3.816**	**12.502**	**38.925**	1.943	**3.965**	**5.776**
	$p = 10$					
OLS	73714.798	3736270.933	1532640.381	71644.102	16753492.210	9423100.481
RR	37184.840	1313452.852	598164.381	34299.543	5616581.637	4674821.272
KGM	1871.934	83629.087	26066.229	2423.817	302287.702	104603.177
MRR	12315.301	437515.658	298845.296	15031.502	3849009.588	1035879.544
M	486.542	1343.521	7030.242	14.739	39.010	184.904
RM	278.148	750.203	3973.216	6.805	16.289	73.489
MRM	254.625	672.303	3581.000	6.226	14.504	64.922
SQ95	43.692	88.292	**50.634**	22.918	105.124	166.669
QR25	7783.992	265551.577	224900.890	14245.938	4866386.778	458864.459
QR50	2431.760	104183.214	51129.064	3733.422	2108749.741	150161.060
QR75	551.377	45986.197	9520.995	284.468	11072.136	45771.364
QR95	38.107	**84.454**	52.710	23.448	97.070	177.214
QM25	260.475	659.780	3623.836	5.926	13.122	57.616
QM50	193.069	462.828	2564.669	4.076	8.204	33.273
QM75	96.263	230.585	1225.454	2.659	4.473	14.907
QM95	**30.874**	91.036	447.726	**1.555**	**2.347**	**7.345**

3.2. Results and Discussions

Tables 3.1-3.12 presents the EMSE values of all the estimators considered in our study for different varying factors including multicollinearity, error variances, outliers, predictors, sample size and error distributions. The most efficient estimator in terms of smallest EMSE is made bold in the tables. EMSE results are distributed and discussed below in seven scenarios according to the error distributions.

Scenario-I: Tables 3.1-3.2 show the EMSE values for the estimators when the errors of the model are distributed as normal with mean zero and standard deviation one. Increase in the level of multicollinearity (r^2) and the number of predictors (p), the EMSE of all the estimators increases in general except for the estimators SQ95 and QR95. By keeping all the other factors constant, the EMSE generally decreases when the sample size (n) increases. As the percentage of outliers increases from 10 to 20, the EMSE of the considered estimators also increases in general. The estimator SQ95 is the close competitor with our new estimator QR95 but the new estimator QR95 remain winner in many cases.

Scenario-II: Tables 3.3-3.4 show the EMSE values for the estimators when the errors of the model are distributed as student's t with degrees of freedom three. It can be seen that by increasing the considered factors except the sample size deteriorates the performance of estimators except SQ95, QR95, QM75 and QM95. In this scenario, the new estimators QM75 and QM95 mostly perform better than other estimators. Estimators SQ95, QR95 and QM50 also showed best performance in some cases.

Scenario-III: Tables 3.5-3.6 show the EMSE values for the estimators when the errors of the model are distributed as chi-square with degrees of freedom four. As compared to scenario-I and II, similar pattern is also observed in this scenario with respect to different considered factors such as multicollinearity and outliers. The estimators QR75 and QR95 showed minimum EMSE values in general. When $p = 4$, $r^2 = 0.85$ and 20% outliers are present, the estimator QM95 outperforms other estimators. **Scenario-IV:** Tables 3.7-3.8 show the EMSE values for the estimators when the errors of the model are distributed as F with degrees of freedom two and seven. It is observed from this scenario that the new estimators QR75 and QR95 exhibit smallest EMSE values as compared to all other estimators in most of the considered situations and in few cases the estimators QR25 and QR50 showed better performance.

Scenario-V: Tables 3.9-3.10 show the EMSE values for the estimators when the errors of the model are distributed as gamma with scale parameter two and shape parameter five. We conclude from these tables that the estimators SQ95, QR75, QR95 and QM95 have shown generally better performance than all other estimators in most of the considered cases. Also when 20% y-direction outliers are present with $r^2 = 0.85$ and $p = 4$, the estimator QM75 for $n = 20$ and QR50 for $n = 50$ outperform.

Scenario-VI: In this scenario, the errors of the model are distributed as Cauchy with location parameter zero and scale parameter one. The results are presented in Tables 3.11-3.12. Cauchy distribution is also called heavy tailed distribution. For heavy tailed error distribution models, the ridge M-estimators perform generally better than the OLS and ridge estimators. Also, it is observed from EMSE tables 3.11-12 that among the ridge M-estimators, our new estimators QM75 and QM95 have shown efficient performance.

Therefore, we conclude from the simulation results that the new estimators denoted by QR and QM have shown efficient results in many considered scenarios. We recommend the new estimator QM for heavy tailed error distributions such as Cauchy or Student's t-distribution and QR for the normal and other non-normal considered error distributions. In general, among the new estimators, QR75, QR95, QM75 and QM95 outperform. Estimators QR95 and SQ95 remained close competitors in most cases but QR95 was the winner. Also in some cases, for moderately high multicollinearity with $p = 4, n = 50$ and 10% outliers in y-direction are present, the new estimators QR25, QR50, QM25 and QM50 perform comparatively better than the other estimators. In general, for high multicollinearity along with y-direction outliers, we recommend the use of proposed estimator QR which shows more efficient results. Also, for heavy tailed error distributions, the new estimator QM is recommended. An empirical example to illustrate the application and benefits of the new estimators is provided in the section to follow.

4. APPLICATION

In this section, we have presented an empirical example in order to support the simulation results and illustrate the benefit of the new estimators with natural datasets. We have considered the water data taken from Pakistan council of research in water resources (PCRWR). The data is published in annual report 2014-15 and available online (PCRWR 2016). The data is collected from district Islamabad, Pakistan based on water mineral contents (mg/l). The sample size is $n = 25$. The response variable and four predictors are defined as: y: Total dissolved solids (TDS), X_1: Electric conductivity (EC), X_2: Bicarbonate (HCO_3), X_3: Hardness (Hard.), X_4: Power of hydrogen (pH). The regression model considered for this data set is given below:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \qquad (27)$$

The eigen values obtained are given as:

$$\lambda_1 = 3.6811 \text{ , } \lambda_2 = 0.1804 \text{ , } \lambda_3 = 0.1336 \text{ , } \lambda_4 = 0.0490.$$

The computed value for CN is 75.1245 which shows moderately high multicollinearity exists among the predictors. Table 4.1 shows the pairwise correlation among predictors and the response variable. In addition, estimated SNR = 5, which shows small error variance exists in the data. Furthermore, observations 7, 12 and 16 have high leverage values than others, i.e., 0.42, 0.77 and 0.61. So the percentage of y-direction outliers is 12%. Moreover, Shapiro-Wilk (SW) normality test is used to test the null hypothesis that the residuals are normal. The P-value so obtained is 0.02009 which is less than 0.05. Therefore, the null hypothesis of normality is rejected at 5% level of significance. The estimates for the regression coefficients and EMSE results are provided in Table 4.2. The estimated value of the regression coefficients of ridge M-estimators gives the actual sign and magnitude than the OLS and other ridge estimators. In general, it can be seen from the table that the ridge M-estimators are more efficient with respect to minimum EMSE than the other ridge type and OLS estimators. Moreover, the new estimators QM25 and QM50 perform better than all others.

Table 4.1
Pairwise Correlations among Predictors and Dependent Variable of Water Data from (PCRWR 2016)

Correlation Table	x1	x2	x3	x4	y
x1	1.000	0.947	0.900	-0.901	0.930
x2	0.947	1.000	0.874	-0.874	0.840
x3	0.900	0.874	1.000	-0.866	0.953
x4	-0.901	-0.874	-0.866	1.000	-0.846
y	0.930	0.840	0.953	-0.846	1.000

Table 4.2
EMSE and Estimated Regression Coefficients of Water Data from (PCRWR 2016)

Estimators	MSE	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
OLS	1.416	0.485	-0.381	0.538	0.941
RR	0.794	0.485	-0.375	0.530	0.906
KGM	0.797	0.484	-0.366	0.517	0.847
MRR	0.961	0.485	-0.369	0.521	0.813
M	0.407	0.954	-0.242	0.317	0.047
RM	0.177	0.953	-0.239	0.313	0.045
MRM	0.119	0.954	-0.225	0.294	0.035
SQ95	0.914	0.483	-0.353	0.498	0.771
QR25	0.960	0.485	-0.369	0.521	0.814
QR50	1.047	0.485	-0.359	0.506	0.733
QR75	1.157	0.485	-0.327	0.460	0.573
QR95	1.238	0.484	-0.275	0.386	0.404
QM25	**0.111**	0.954	-0.229	0.301	0.038
QM50	0.115	0.954	-0.228	0.298	0.037
QM75	0.195	0.952	-0.176	0.230	0.020
QM95	0.399	0.945	-0.108	0.140	0.010

5. CONCLUSION

In this article, we proposed some new quantile based ridge M-estimators to combat with the joint problem of y-direction outliers and multicollinearity. The random errors are generated from normal and some commonly used non-normal distributions. The new estimators are compared with existing estimators through simulation study. It is concluded from the simulation results under the MSE criterion that our suggested estimators perform generally better than other existing estimators in many considered instances. An empirical example was also used to support the simulation findings and application of the new estimators. Hope the findings of this article will be helpful for practitioners.

ACKNOWLEDGEMENTS

The authors of this article highly acknowledge the financial support received from the Higher Education Commission (HEC) Pakistan under Indigenous PhD scholarship program awarded to the first author under Grant No. 315-11394-084 (Batch-III and Phase-II).

ORCID:

Muhammad Suhail
http://orcid.org/0000-0003-1784-5272

Sohail Chand
http://orcid.org/0000-0002-4564-143X

REFERENCES

1. Algamal, Z.Y. (2018). A new method for choosing the biasing parameter in ridge estimator for generalized linear model. *Chemometrics and Intelligent Laboratory Systems*, 183, 96-101.
2. Algamal, Z.Y. (2018). Developing a ridge estimator for the gamma regression model. *Journal of Chemometrics*, 32(10), e3054.
3. Alheety, M.I. and Golam Kibria, B.M. (2013). Modified Liu-type estimator based on $(r - k)$ class estimator. *Communications in Statistics-Theory and Methods*, 42(2), 304-319.
4. Ali, S., Khan, H., Shah, I., Butt, M.M. and Suhail, M. (2019). A comparison of some new and old robust ridge regression estimators. *Communications in Statistics-Simulation and Computation*, DOI: 0.1080/03610918.2019.1597119.
5. Arashi, M., Kibria, B.G., Norouzirad, M. and Nadarajah, S. (2014). Improved preliminary test and Stein-rule Liu estimators for the ill-conditioned elliptical linear regression model. *Journal of Multivariate Analysis*, 126, 53-74.
6. Ertaş, H. (2018). A modified ridge m-estimator for linear regression model with multicollinearity and outliers. *Communications in Statistics-Simulation and Computation*, 47(4), 1240-1250.
7. Gujarati, D.N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education.
8. Gültay, B. and Kaçıranlar, S. (2015). Mean square error comparisons of the alternative estimators for the distributed lag models. *Hacettepe Journal of Mathematics and Statistics*, 44(5), 1215-1233.
9. Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986). *Robust Statis Tics*. New York, USA: J. Wiley& Sons. https://doi.org/10.1186/1742-4755-7-12.
10. Hoerl, A.E. and Robert W.K. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
11. Huang, J. and Yang, H. (2014). A two-parameter estimator in the negative binomial regression model. *Journal of Statistical Computation and Simulation*, 84(1), 124-134.
12. Huber, P.J. (1981). *Robust Statistics*. *New Jersey*. John Wiley & Sons.
13. Jahufer, A. and Jianbao, C. (2009). Assessing global influential observations in modified ridge regression. *Statistics & probability letters*, 79(4), 513-518.

14. Khalaf, G., Kristofer, M. and Shukur, G. (2013). Modified Ridge Regression Estimators. *Communications in Statistics-Theory and Methods*, 42(8), 1476-87.
15. Kibria, B.M.G. (2003). Performance of Some New Ridge Regression Estimators. *Communications in Statistics Part B: Simulation and Computation*, 32(2), 419-35. https://doi.org/10.1081/SAC-120017499.
16. Kibria, B.M. and Banik, S. (2016). Some ridge regression estimators and their performances. *Journal of Modern Applied Statistical Methods*, 15(1), 206-38. https://doi.org/10.22237/jmasm/1462075860.
17. Lawless, J.F. and Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics-theory and Methods*, 5(4), 307-323.
18. Maronna, R.A. (2011). Robust ridge regression for high-dimensional data. *Technometrics*, 53(1), 44-53.
19. McDonald, G.C. and Galarneau, D.I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350), 407-416.
20. PCRWR (2016). *Annual 2014-15 Report*. Pakistan Council of Research in Water Resources (PCRWR), Islamabad Pakistan.
21. Roozbeh, M., Arashi, M. and Hamzah, N.A. (2020). Generalized cross-validation for simultaneous optimization of tuning parameters in ridge regression. *Iranian Journal of Science and Technology, Transactions A: Science*, 44(2), 473-485.
22. Silvapulle, M.J. (1991). Robust Ridge Regression Based on an M- estimator. *Australian Journal of Statistics*, 33(3), 319-33.
23. Suhail, M. and Chand, S. (2019). Performance of Some New Ridge Regression Estimators. In *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, 1-4. IEEE.
24. Suhail, M., Chand, S. and Kibria, B.M.G. (2020). Quantile based estimation of biasing parameters in ridge regression model. *Communications in Statistics-Simulation and Computation*, 49(10), 2732-2744.
25. Suhail, M., Chand, S. and Kibria, B.M.G. (2019). Quantile-based robust ridge m-estimator for linear regression model in presence of multicollinearity and outliers. *Communications in Statistics-Simulation and Computation*, 1-13. https://doi.org/10.1080/03610918.2019.1621339.
26. Suhail, M., Ilyas, M. and Ullah, A. (2019). Modified Ridge Regression Estimators: A Simulation Study. Proceedings of *15th Islamic Countries Conference on Statistical Sciences (ICCS-15),* ISOSS, Lahore, Pakistan, Vol. 34, 119-144.
27. Swindel, B.F. (1976). Good Ridge Estimators Based on Prior Information. *Communications in Statistics-Theory and Methods*, 5(11), 1065-75.
28. Yasin, S., Kamal, S. and Suhail, M. (2020). Performance of Some New Ridge Parameters in Two-Parameter Ridge Regression Model. *Iranian Journal of Science and Technology, Transactions A: Science*, 1-15. https://doi.org/10.1007/s40995-020-01019-7.