An Evaluation of the Impact of Multicollinearity on the Performance of Various Robust Regression Methods

Muhammad Ilyas^{1*} and Salahuddin²

¹Department of Statistics, University of Malakand, Chakdara, Lower Dir, KPK, Pakistan; m.ilyasstatistics2015@gmail.com ²Department of Basic Sciences and Humanities, CECOS University of IT and Emerging Sciences, Hayatabad, Peshawar, Pakistan; salahuddin_90@yahoo.com

Abstract

Objectives: To examine the performance of several regression methods comprising of Ordinary Least Square (OLS), and certain robust methods including; M-regression, Least Median of Squared (LMS), Least Trimmed Square (LTS), MM-estimation and S-estimation, under fluctuating levels of collinearity, using the criterion, Total Absolute Deviation (TAB) and Total Mean Square Error (TMSE) with some graphical tools. **Methods/Statistical Analysis**: Robust Regression methods insure good performance even in case the fundamental assumption of normality is not satisfied. The presence of multicollinearity affects the results of robust regression methods and marks them unsatisfactory. A quantitative evaluation of these techniques is provided by using the criterion, TAB and TMSE. Results are summarised by using box plot of absolute bias, along with the graphs of TAB, and TMSE. **Findings**: The results show that for minor levels of collinearity the effect is low and similar, but at greater levels of collinearity the effect is high and performance wise all the methods give quite incompatible results. It is also illustrated that greater magnitude of collinearity along with higher percentages of outliers ranks the underlying methods quite differently, resulting in MM-estimation method to be the most unpleasant. **Conclusion**: While applying any statistical method it is necessary to consider all the assumption underlying that method as well as every aspect of our data to avoid misleading results. It is illustrated that MM-estimation method although a best candidate for higher percentages of outliers alone, become the most unpleasant, by a simultaneous interruption of high level of collinearity, hence robust ridge techniques need to be adopted.

Keywords: Multicollinearity, Ordinary Least Squares, Outliers, Robust Regression Methods

1. Introduction

Regression analysis deals to model the relationship between variables approximated by some appropriate mathematical Eq. In case the regression model satisfies certain basic assumptions, the OLS estimators happen to be best linear unbiased estimates^{1,2}. The estimate is too much sensitive to violation of these assumptions and even a single contaminated observation can result in the OLS estimator to be unreliable. Researchers have been attempting for alternative estimating procedures, known as robust regression methods which are robust to outliers.

*Author for correspondence

These methods down weight outliers; preserve good efficiency, in case normality hold and are resilient to highly discrepant values. Most often used robust estimators are Huber M-estimation, Least Trimmed Square (LTS) estimator, Least Median Square (LMS) estimator, S-estimator by^{3,4}. Besides non-normality multicollinearity in regression is a condition where two or more predictor variables remain linearly related⁶⁻⁸. Exact relationship results in matrix (X'X) to be non-invertible⁹. This is a phenomenon where an infinite set of coefficients there are to minimize the totality of squared residuals¹⁰. However, usually relationships between the predictors are close to exact, decreases reliability of estimates, potentially affecting estimation, forecasting and hypothesis testing. Robust regression being a good alternative in case of violation from normality, but are suspected to lose performance when the problem of non-normality is joined by multicollinearity concurrently. In this paper efforts are carried out to evaluate the performance of various methods (OLS, M, LTS, LMS, S, and MM) under various simulation settings, specifically investigating the influence of collinearity levels on the performance of these methods.

2. Methods to be Compared

2.1 Ordinary Least Squares Method

Least square method is generally used technique to estimate the parameters in the model. In this technique, estimates of the parameters are obtained by principle which minimizes the totality of squared residuals. In case the linear regression model accomplishes the basic assumptions, OLS estimators stay best linear unbiased¹. This method provides an explicit estimate of the true values from observed data as:

$$\hat{\boldsymbol{\beta}} = \left(X'X\right)^{-1} X'Y \tag{1}$$
With $var(\hat{\boldsymbol{\beta}}) = \sigma^2 \left(X'X\right)^{-1}$

The logic behind frequent use of this method is its computational easiness, but unfortunately this method

depends upon a controlled set of assumptions, now being criticized to a greater extent for lacking robustness³.

2.2 M-estimation Method

M-estimation being a common robust regression procedure was primarily introduced by². This technique in a sense is a general form of the least squares substituting the quadratic loss function by function $\rho(.)$. The function $\rho(.)$ is symmetric, continuous having a unique minimum at zero^{4,11}. The function $\rho(.)$ may be chosen in such a way that it denotes some weighting scheme of the residuals. The set of normal Eqs to be solved is given by the system:

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) x_i = 0$$
⁽²⁾

Where $\psi(e) = \frac{d\rho(e)}{de}$, the score function and $W(e) = \frac{\psi(e)}{e}$, the weight function's being an estimate of the residuals scale. The choice of ψ to be monotone will not weight discrepant values as much as the least squares, whereas a re-descending ψ function results in a weighting scheme that assign the weights in decreasing order up to a definite distance (e.g. 3σ) and then declines the weight to zero as the remote distance is increased much. Some of the proposals for $\rho(.) \psi$ and the weight function are given in Table 1.

Table 1. ρ , ψ and weight functions of various re-descending estimators

Method	Objective function $ ho(u)$	ψ function $ ho'(u)$	Weight function W(u) = $\psi(u)$	Range of u
OLS	$\frac{u^2}{2}$	U	1	$ u \leq \infty$
LAD	u	Sign (u)	sign(u)/u	$ u \leq \infty$
Huber	$\frac{u^2}{2}$	и	1	$ u \leq c$
	$c. u -c^2/2$	c.sign(u)	^c /u	u > c
Bi-Square	$c2/6[1-\{1-(u/c)2\}3]$	u[1-(r/c)2]2	[1- (u/c)2]2	$ u \leq c$
	c2/6	0	0	u > c

Method	Objective function $\rho(u)$	ψ function $ ho'(u)$	Weight function W(u) = $\psi(u)/u$	Range of u
Andrews	c2[1-cos(u/c)]	c sin(u/c)	sin(u/c)/(u/c)	$ u \leq c\pi$
	2 c2	0	0	$ u > c\pi$
Alamgir Re-descending	$\frac{2c^2}{3} \begin{bmatrix} 2\left(1+3e^{\left(\frac{u}{c}\right)^2}\right) \\ 1-\frac{2\left(1+3e^{\left(\frac{u}{c}\right)^2}\right)}{\left(1+e^{\left(\frac{u}{c}\right)^2}\right)^3} \end{bmatrix}$	$\frac{16ue^{-2\left(\frac{u}{c}\right)^2}}{\left(1+e^{-\left(\frac{u}{c}\right)^2}\right)^4}$	$\frac{\left(4e^{-\left(\frac{u}{c}\right)^{2}}\right)^{2}}{\left(1+e^{-\left(\frac{u}{c}\right)^{2}}\right)^{4}}$	u ≤ c
M-estimator (ALARM)	$\frac{2c^2}{3} \left[1 - \frac{2(1+3e)}{(1+e)^3} \right]$	0	0	u > c

2.3 Least Median of Squares (LMS) Regression

The LMS regression was suggested by³, using the idea of minimization of the median of the squared residuals not the sum of the squared residuals. Under this procedure the estimates for the model parameters are provided by the Eq:

$$\hat{\beta}_{LMS} = Minimize \ mediam\left(e_i^2\right)$$
 (3)

LMS regression estimator attaining a high breakdown of almost 0.5 is the first Equivariant estimator. Although the (LMS) estimator is robust to outliers in y-direction as well as in x-space, it has a drawback that the efficiency of this methodises quite low as compared to the least squares in the instance of Gaussian errors. Due to this deficiency LMS estimator has a very little direct use, but is often used as an initial estimator for diagnostics purposes or in some other robust techniques⁵.

2.4 LTS Regression

The LTS regression method is an alternative robust regression technique suggested by Rousseeuw³. To elude outliers, this procedure minimizes the totality of squared residuals after the largest a squared residuals are trimmed. The LTS regression estimator is given by:

$$\hat{\beta}_{LTS} = \underbrace{\text{mimimize}}_{\beta} \sum_{i=1}^{h} \left(e_i^2 \right)_{i:n} \tag{4}$$

Where $(e^2)_{1:n} \le (e^2)_{2:n} \le \dots \le (e^2)_{n:n}$, are squared residuals arranged in order of magnitude for a value of $h = \left[\frac{(n+p+1)}{2}\right]$, this method achieves a breakdown point, (n/2 - p + 2)/n. A difficulty of the LTS method is the action rEquired for sorting the squared residuals in its objective function⁵. Several algorithms suggested in literature for this approach are simulated annealing based LTS- algorithm developed by¹⁴, 'Feasible Set Algorithm' by¹². Another algorithm called FAST-LTS, has been given by¹³, which is too fast than all the existing algorithms. The

by¹³, which is too fast than all the existing algorithms. The high statistical efficiency and faster rate of convergence of the LTS over the LMS make LTS a more appropriate nominee than LMS as an initial step for two-stage estimators such as MM-estimator and the generalized M-estimators^{13,15,16}. In¹⁷ an L-1 penalty is imposed on LTS estimator a spars estimator has been developed.

2.5 S-Estimator (S-Regression)

The S-estimator an alternative estimator possessing a high breakdown is suggested³. The S-estimator minimizes an M-estimate of the residuals scale. This method estimates the true values as:

$$\hat{\beta}_{s} = \underbrace{\text{mimimize}}_{\beta} S(e_{1}(\beta), e_{2}(\beta), \dots, e_{n}(\beta))$$
(5)

Where the dispersion of the residuals S [$e_1(\beta), e_2(\beta), \dots, e_n(\beta)$], is determining by solution to the Eq:

$$\frac{1}{n-p}\sum_{i=1}^{p}\rho\left(\frac{y_{i}-x_{i}'\hat{\beta}}{s}\right) = K$$
(6)

Where *K* is Equal to $E\varphi[\rho]$, with φ stances standard normal distribution and ρ is a continuously differentiable, and symmetric, behaving to strictly increase over [0, c] and remain constant over [0, c]. The final estimate of the scale for S-estimate of regression is given by:

$$\hat{\sigma} = S\left(e_1\left(\hat{\beta}\right), \ e_2\left(\hat{\beta}\right), \dots, \ e_n\left(\hat{\beta}\right)\right)$$
(7)

A dominant choice of ρ is:

$$\rho_{c}(S) = \begin{cases} 3\left(\frac{s}{c}\right)^{2} - 3\left(\frac{s}{c}\right)^{4} + \left(\frac{s}{c}\right)^{6} & \text{if } |S| \le c \\ 1 & \text{if } |S| > c \end{cases}$$
(8)

The S-estimator may possess a high breakdown value of 50%, if K in Eq (6) and ρ in Eq (8) satisfies:

$$\frac{K}{\rho(c)} = \frac{1}{2}$$

S-estimator possesses the properties of high breakdown and asymptotic normality. The compromise between breakdown point and efficiency is determined by choice of the tuning constants and K. In¹⁸ it is concluded that the S-estimator under Gaussian errors can achieve an efficiency of 0.33 with a breakdown of 50% has suggested fast-S, an approximating algorithm for obtaining S-estimator of regression¹⁹.

2.6 MM-Estimation

The MM-estimation recommended in⁴, a special brand of M-estimation is a multistage estimator joining the high breakdown from an initial robust estimator and high efficiency from another, bringing about high breakdown and good efficiency from another robust estimator. The calculation of MM-estimator contain, considering a high breakdown point initial estimator, compute an M-estimate of the residuals scale and obtaining an M-estimate of the true values on the basis of the M-estimate of residuals scale . The algorithm for MM-estimation procedure may be given by;

1. Choose an initial high breakdown estimate $\beta_{initial}$, consistent but not essentially efficient and calculate residuals $r_i(\hat{\beta}_{initial})$ based on the initial estimate.

2. Calculate the M-estimate of residuals scale S_m using residuals $r_i(\hat{\beta}_{initial})$, where S_m may be created as a solution to:

$$\frac{1}{n-p}\sum_{i=1}^{n}\rho_{c_0}\left(\frac{y_i - x_i'\hat{\beta}_{initial}}{s_{mm}}\right) = K$$
(9)

3. The third step use an M-estimate of coefficient $\beta_{initial}$ from the first and estimate of residuals scale S_m from the second step, gives $\hat{\beta}_{inal}$ as a solution to:

$$\sum_{i=1}^{n} x_{ij} \psi \left(\frac{y_i - x_i^t \hat{\beta}_{initial}}{S_m} \right) = 0$$

$$i = 1, \ 2...nandj = 1, \ 2...p$$
(10)

And the final scale estimate of the residuals is given by solution to Eq:

$$\frac{1}{n-p} \sum_{i=1}^{n} \rho_{c_0} \left(\frac{y_i - x_i^t \hat{\beta}_{final}}{S_{mm}} \right) = 0.5$$
(11)

For a particular value of c_0 , the tuning constant.

The MM-estimator enjoys high efficiency, a high breakdown value of (50%), unluckily may be influenced by the occurrence of high leverage observations²⁰⁻²². In²³ a robust version of ridge estimator referred to as Weighted Ridge MM-estimator (WRMM) is offered by means of weighted ridge and the MM-estimation penalized MM-estimation called (MM-lasso) by using the L-1 penalty and the mechanism of MM-estimation²⁴.

3. Simulation Studies

To compare the performance of different methods, numerous simulation options are examined. A simulation structure is implemented to allow on- normality and multicollinearity together. The particulars of various aspects considered in various settings of simulation are:

Methods evaluated: Methods assessed under various simulation settings include, OLS, M-estimation, LMS, LTS,S and MM.

Sample size: In different simulation settings we have considered, the sample size at 50,100, 150 and 200.

Number of predictor variables: for fitting models using different multicollinearity levels and fractions of outliers, the number of predictors (P) is considered at 2, 4 and 6.

Fractions of outliers: In various settings of simulation, particularly focus is on y-outliers. While judging the performance of different methods, numerous fractions of outliers; particularly 10%, 20%, 30%, and 40% outliers are generated in data sets.



Figure 1. Left column graph of Ψ function right column graph of weight function for LS, Huber, Bi-Square, and Alarm objective functions.



Figure 2. Scatterplots [(a), (b), (c) & (d)] of predictor variables $(X_1, X_2, X_3 & X_4)$ generated through scheme given in Eq (12).

To let different levels of collinearity, the values on predictors are generated by means of a methodology used in²⁵⁻²⁸. To generate explanatory variables the following mechanism is used:

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{im}, i = 1, 2, \dots, j = 1, 2, \dots, p,$$

and $m = p + 1$ (12)

 z_{ij} , is a standard normal variate and ρ is specified in such a way that the correlation coefficient between any two predictor variables is maintained at ρ^2 . The scatterplot matrices of predictors X₁, X₂, X₃ and X₄ generated by system in Eq (12) with different values of ρ are given in Figure 1-2 and correlation matrix in Table 2.

$ ho^2 = 0.98$				$\rho^2 = 0.5$					
	X ₁	X ₂	X ₃	\mathbf{X}_4		X ₁	X ₂	X ₃	\mathbf{X}_4
X	1.000	0.978	0.9824	0.9826	X ₁	1.0000	0.5057	0.5970	0.6101
X ₂	0.987	1.000	0.9817	0.9789	X ₂	0.5057	1.0000	0.5181	0.4575
X ₃	0.982	0.981	1.0000	0.9779	X ₃	0.5970	0.5181	1.0000	0.4462
X ₄	0.982	0.978	0.9779	1.0000	X ₄	0.6101	0.4575	0.4462	1.0000

(15)

Table 2. Correlation matrix for variables $(X_1, X_2, X_3 \otimes X_4)$ generated by Eq (12) with different ρ values

Model: In simulation studies the following models are used:

Model 1:
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$
 (13)

Model 2:
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

(14)
Model 3: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \varepsilon_i$

For each model the predictors are generated with scheme (12), and all β 's are used at a constant value one, i.e., $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$.

To consider different fractions of outliers different options for error terms ϵ_i 's values, specifically the following cases are observed.

Case: I $\epsilon \sim N(0, 1)$ Case: II $\epsilon \sim 0.9N(0, 1) + 0.1N(10, 1)$ Case: III $\epsilon \sim 0.8N(0, 1) + 0.2N(10, 1)$ Case: IV $\epsilon \sim 0.7N(0, 1) + 0.3N(10, 1)$ Case: V $\epsilon \sim 0.6N(0, 1) + 0.4N(10, 1)$

Assessment Measures: To assess the results of methods understudy, for all simulations of the estimate $\hat{\beta}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \hat{\beta}_{i2}, \hat{\beta}_{i3}, \dots, \hat{\beta}_{ip}), 1 \le i \le m$, with m=1000, the two criterions used are:

1. Total Mean Square Error (TMSE): For all replication, i.e., m=1000, the TMSE is obtained by the formula:

$$TMSE = \frac{\sum_{i}^{m} \sum_{j}^{p} \left(\hat{\beta}_{ij} - \beta_{ij} \right)^{2}}{m}$$
(16)

2. Total Absolute Bias (TAB):

$$TAB = \frac{\sum_{i}^{m} \sum_{j}^{p} \left| \hat{\beta}_{ij} - \beta_{ij} \right|}{m}$$
(17)

For error distribution considered in Case I to Case V, atn = 50, 100, 150, 200 and P = 2, 4, 6, though the values of TMSE vary, increases with an increase in the value of P and decreases as the sample size grow. With respect to performance for all the methods nearly similar pattern is observed for each value of n(50,100,150,200) and P (2, 4, 6). It is observed that for lower collinearity levels, the values of TMSE for all techniques are not much divergent, but at greater collinearity levels, the TMSE values for all the methods are pretty different. The graphs of TMSE for (OLS, M,LTS, LMS, MM and S) for different collinearity levels and outliers percentages for n = 200 are given in Figure 3. Figure 3 reveals that at smaller levels of multi-collinearity with increasing percentages of outliers causes the (TMSE) to increase with little differences but the (TMSE) values at higher levels of multi-collinearity increased markedly with wider differences and give different ranking of the methods. In Figure 4, a graphical analysis for (OLS, M, LMS, LTS,S and MM) is given for the second performance measure total absolute bias. The results from Figure 4 are consistent with the results in Figure 3. The results over the two performance measures, TMSE and TAB for different scenarios are given in Table 3 and 4, respectively. In Figure 5, a graphical analysis of box plots for various situations is also given.

M	ethod		Total MSE				
Rho	Rho values		0.5			0.7	
Sample size	Outlier%	P=2	P=4	P=6	P=2	P=4	P=6
OLS	0	0.01562	0.02870	0.04082	0.01818	0.03746	0.05587
	10	1.1118	1.2442	1.3742	1.1374	1.3351	1.5324
	20	4.1868	4.4176	4.6300	4.2306	4.5759	4.8941
	30	9.2401	9.5461	9.8615	9.3016	9.7489	10.2144
	40	16.2701	16.5553	16.9323	16.1487	16.7716	17.3143
М	0	0.01653	0.03008	0.04353	0.01924	0.03929	0.05962
	10	0.0613	0.0844	0.1063	0.0657	0.0990	0.1321
	20	0.3347	0.3949	0.4520	0.3431	0.4260	0.5075
	30	8.1188	8.6069	9.0588	8.1979	8.8636	9.4877
	40	16.2689	16.5535	16.9287	16.3318	16.7708	17.3127
LMS	0	0.12045	0.22525	0.3346	0.14318	0.28667	0.47401
	10	0.1120	0.2110	0.3571	0.1264	0.2794	0.4846
	20	0.1089	0.2132	0.4202	0.1278	0.2759	0.5688
	30	0.0999	0.2192	0.4984	0.1090	0.2887	0.6857
	40	0.1012	0.2434	0.7071	0.1164	0.3144	0.9643
LTS	0	0.13652	0.25097	0.36595	0.16047	0.32987	0.50627
	10	0.1222	0.2300	0.3815	0.1404	0.2921	0.5249
	20	0.1043	0.2116	0.4275	0.1196	0.2841	0.5763
	30	0.0844	0.3051	0.4881	0.0982	0.2874	0.6761
	40	0.0632	0.2221	0.6591	0.0731	0.2918	0.9336
ММ	0	0.01814	0.0312128	0.04382	0.02120	0.04123	0.06012
	10	0.0180	0.0334	0.0474	0.0209	0.0436	0.0646
	20	0.0203	0.0357	0.0560	0.0238	0.0465	0.0764
	30	0.0238	0.0433	0.0650	0.0276	0.0563	0.0889
	40	4.0261	6.8552	9.6346	4.2035	7.6005	11.0224
S	0	0.07066	0.16384	0.29674	0.07865	0.21358	0.40467
	10	0.0609	0.1611	0.3162	0.0687	0.2152	0.4327
	20	0.0583	0.1770	0.3905	0.0676	0.2322	0.5393
	30	0.0592	0.2123	0.5162	0.0706	0.2776	0.6973
	40	0.0705	0.2676	0.7710	0.0816	0.3539	1.0465

Table 3. TMSE values for sample size n = 200

M	ethod	Total MSE						
Rho	values	0.9			0.99			
Sample size	Outlier %	P=2	P=4	P=6	P=2	P=4	P=6	
OLS	0	0.03398	0.08771	0.13896	0.25705	0.78028	1.27754	
	10	1.2982	1.8496	2.4015	3.5640	8.9336	14.3019	
	20	4.5132	5.4750	6.3456	8.4940	17.8580	26.2344	
	30	9.6865	10.8984	12.1567	15.0544	26.6940	38.7711	
	40	16.7600	17.9982	19.4186	22.8454	34.8704	48.2549	
М	0	0.03602	0.09198	0.14826	0.27271	0.81808	1.36314	
	10	0.0920	0.1819	0.2736	0.4626	1.3253	2.2104	
	20	0.3975	0.6026	0.8116	1.1637	3.0351	4.9856	
	30	8.6926	10.2927	11.8542	15.5564	29.9880	44.3502	
	40	16.7598	18.0026	19.4297	22.8591	34.9442	48.4432	
LMS	0	0.25797	0.69073	1.15800	1.92192	6.07449	10.45611	
	10	0.2292	0.6413	1.2067	1.7611	5.5850	11.0966	
	20	0.2392	0.6366	1.3835	1.8283	5.7739	13.0223	
	30	0.2108	0.6963	1.6806	1.5445	6.3729	15.7386	
	40	0.2005	0.7364	2.4300	1.5514	6.7014	22.2894	
LTS	0	0.29857	0.77320	1.28011	2.25997	6.98294	11.84773	
	10	0.2620	0.6817	1.3313	2.0226	6.2598	12.1469	
	20	0.2268	0.6811	1.4549	1.6894	6.0393	13.5082	
	30	0.1927	0.7295	1.7001	1.4710	6.5908	16.1195	
	40	0.1383	0.7141	2.3973	1.0515	6.5808	22.5637	
ММ	0	0.04021	0.09756	0.14938	0.30721	0.87201	1.37167	
	10	0.0388	0.1020	0.1597	0.2918	0.9062	1.4630	
	20	0.0452	0.1080	0.1889	0.3456	0.9561	1.7290	
	30	0.0519	0.1310	0.2203	0.3935	1.1607	2.0192	
	40	5.3527	11.8497	18.6131	21.4597	70.4481	122.4721	
S	0	0.14511	0.50016	0.98509	1.19003	4.37143	9.08112	
	10	0.1288	0.5094	1.0598	0.9357	4.6377	9.7189	
	20	0.1300	0.5488	1.2785	0.9784	4.9468	11.8010	
	30	0.1358	0.6581	1.7183	0.9913	5.8713	15.8645	
	40	0.1505	0.8662	2.5929	1.1382	7.8683	23.8235	

Table 4. TAB values for sample size 200

Met	thod	Total Absolute Bias						
Rhov	values	0.5			0.7			
Sample size	Out-lier %	P = 2	P = 4	P = 6	P = 2	P = 4	P = 6	
OLS	0	0.1725	0.3023	0.4254	0.1851	0.3441	0.4958	
	10	1.3666	1.7858	2.1893	1.4116	1.9216	2.4173	
	20	2.4858	3.0276	3.5494	2.5392	3.2058	3.8523	
	30	3.5629	4.1780	4.7837	3.6283	4.3837	5.1273	
	40	4.5973	5.2070	5.8781	4.6660	5.4146	6.2386	
М	0	0.1770	0.3095	0.4397	0.1899	0.3525	0.5130	
	10	0.3456	0.5125	0.6793	0.3637	0.5656	0.7724	
	20	0.7533	1.0047	1.2613	0.7764	1.0843	1.3972	
	30	3.4327	4.1355	4.8082	3.5078	4.3668	5.1895	
	40	4.5975	5.2083	5.8815	4.6663	5.4162	6.2430	
LMS	0	0.4821	0.8445	1.2159	0.5235	0.9533	1.4462	
	10	0.4664	0.8149	1.2626	0.4921	0.9329	1.4657	
	20	0.4559	0.8222	1.3660	0.4961	0.9359	1.5907	
	30	0.4343	0.8369	1.4901	0.4590	0.9488	1.7369	
	40	0.4391	0.8844	1.7721	0.4734	0.9988	2.0534	
LTS	0	0.5104	0.8890	1.2681	0.5508	1.0146	1.4907	
	10	0.4846	0.8533	1.2963	0.5166	0.9616	1.5136	
	20	0.4450	0.8230	1.3805	0.4772	0.9461	1.5951	
	30	0.4027	0.9742	1.4679	0.4339	0.9380	1.7264	
	40	0.3469	0.8305	1.6906	0.3738	0.9429	2.0054	
ММ	0	0.1846	0.3134	0.4445	0.1991	0.3592	0.5190	
	10	0.1868	0.3262	0.4596	0.2010	0.3715	0.5356	
	20	0.1983	0.3374	0.4999	0.2142	0.3843	0.5814	
	30	0.2134	0.3717	0.5358	0.2296	0.4216	0.6248	
	40	2.5675	4.2779	5.9513	2.6666	4.6552	6.6326	
S	0	0.3656	0.7192	1.1504	0.3870	0.8232	1.3451	
	10	0.3395	0.7193	1.1907	0.3608	0.8276	1.3845	
	20	0.3355	0.7562	1.3211	0.3601	0.8639	1.5466	
	30	0.3375	0.8225	1.5233	0.3657	0.9368	1.7754	
	40	0.3672	0.9244	1.8619	0.3950	1.0649	2.1665	

Method		Total Absolute bias						
Rho values	;	0.9			0.99			
Sample size	Outlier %	P = 2	P = 4	P = 6	P = 2	P = 4	P = 6	
OLS	0	0.2464	0.5160	0.7693	0.6160	1.4641	2.2557	
	10	1.6137	2.4727	3.2997	2.8078	5.5099	8.0909	
	20	2.8075	3.9294	5.0074	4.4003	7.9189	11.2686	
	30	3.9451	5.2135	6.4536	5.8037	9.7791	13.6473	
	40	4.9917	6.2611	7.6226	6.9262	10.9315	15.1408	
М	0	0.2527	0.5294	0.7964	0.6320	1.5027	2.3348	
	10	0.4454	0.7844	1.1301	0.9345	1.9939	3.0701	
	20	0.8927	1.4053	1.9180	1.5877	3.1725	4.7478	
	30	3.8684	5.2956	6.6627	5.9826	10.4072	14.6490	
	40	4.9922	6.2639	7.6306	6.9280	10.9414	15.1687	
LMS	0	0.6907	1.4417	2.2128	1.7215	4.0641	6.4711	
	10	0.6552	1.3807	2.2666	1.6484	3.8850	6.6635	
	20	0.6642	1.3983	2.4501	1.6666	4.0076	7.2612	
	30	0.6207	1.4484	2.6895	1.5324	4.1749	7.9610	
	40	0.6093	1.4966	3.2084	1.5234	4.2670	9.4167	
LTS	0	0.7368	1.5231	2.3243	1.8401	4.3638	6.8480	
	10	0.6892	1.4320	2.3660	1.7450	4.1201	6.9573	
	20	0.6480	1.4282	2.4906	1.6004	4.0798	7.3557	
	30	0.5852	1.4607	2.6885	1.4512	4.2036	8.0262	
	40	0.5004	1.4419	3.1517	1.2575	4.2015	9.4308	
ММ	0	0.2684	0.5416	0.8047	0.6747	1.5428	2.3544	
	10	0.2675	0.5560	0.8297	0.6679	1.5741	2.4259	
	20	0.2872	0.5741	0.8985	0.7182	1.6214	2.6243	
	30	0.3080	0.6297	0.9693	0.7768	1.7802	2.8391	
	40	3.1588	6.1975	9.2492	6.1294	14.7160	23.4415	
S	0	0.5139	1.2367	2.0624	1.3421	3.4483	6.0518	
	10	0.4815	1.2448	2.1368	1.1778	3.5635	6.2528	
	20	0.4856	1.3033	2.3534	1.2200	3.7101	6.9023	
	30	0.4924	1.4112	2.7355	1.2129	4.0323	8.0522	
	40	0.5220	1.6298	3.3513	1.3180	4.6490	9.8129	



Figure 3. [(a)....(o)], n=200, Total MSE vs ρ values.



Figure 4. [(a), ..., (o)], Total absolute bias vs ρ values, sample size n=200.



Figure 5. [(a),, (o)]: Boxplot of various methods for different outliers percentages and multicollinearity levels, with a sample size n=200.

4. Conclusions

In this study efforts are carried out to compare the performance of different regression methods under the influence of varying levels of multicollinearity. A case wise discussion over performance of different methods is as under:

Case I: $\epsilon \sim N(0, 1)$: For the error distribution considered standard normal, at lower collinearity levels all the methods performance is very similar but as the level of collinearity grows they behave quite differently. Moreover it is evident that at higher collinearity levels LTS, LMS seem the poor, S the next whereas OLS, M and MM appear to perform reasonably fine.

Case II: $\epsilon \sim 0.9N(0, 1) + 0.1N(10, 1)$: In this case the error distribution is considered in this case allow 10% outliers in y-direction. In this case at lower levels of collinearity OLS appear to be markedly different, while the other methods behave fairly alike. At higher collinearity levels and 10% fraction of outliers OLS appear the worse whereas (LTS, LMS and S) the next and (M, MM) perform sensibly well with MM the finest of all.

Case III: $\epsilon \sim 0.8N(0, 1) + 0.2N(10, 1)$: For a fraction of 20% outliers OLS is markedly different at all levels while the remaining all methods perform nearly similar at lower collinearity levels but behave very differently for greater values of collinearity levels. In this case the ranking of the methods nearly similar to that in Case II.

Case IV: $\epsilon \sim 0.7N(0, 1) + 0.3N(10, 1)$: In this case at lower levels of collinearity there seem to be two categories (M-Estimation and OLS) and (LTS, LMS, S and MM), together all the methods have quite diverse behavior at higher collinearity levels. Particularly at higher levels of collinearity MM is the best with (LTS,LMS, and S) the succeeding set of best, whereas OLS and M perfomance is very poor and M-Estimation give the worse result.

Case V: $\epsilon \sim 0.6N(0, 1) + 0.4N(10, 1)$: In this case 40% outliers all together with lower to moderate collinearity levels OLS and M method results in higher values of (TMSE), MM the next method giving subsEquent higher values of (TMSE), the other methods (LTS, LMS and S) havinglow and nearly similar values. However 40% outliers considered with a high multi-collinearity level, ranks the methods quite differently. The MM method which is unsurpassed one in all cases turn out to worse among all, M and OLS forming the second set resulting in higher (TMSE) values, and the other three methods (LTS, LMS and S) generating relatively small values of (TMSE).

5. Reference

- Draper NR, Smith H. Fitting a Straight Line by Least Squares. Applied Regression Analysis; 1998. p. 15–46. https://doi.org/10.1002/9781118625590.ch1.
- Huber PJ. Robust regression: asymptotic, conjectures and Monte Carlo, The Annals of Statistics. 1973; 1(5):799–821. https://doi.org/10.1214/aos/1176342503.
- Rousseeuw P, Yohai V. Robust Regression by Means of S-Estimators. In: Franke J., Härdle W., Martin D. (eds) Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics, vol 26. Springer, New York, NY; 1984. p. 256–72. https://doi.org/10.1007/978-1-4615-7821-5_15.
- Yohai VJ. High breakdown-point and high efficiency robust estimates for regression, The Annals of Statistics. 1987; 15(2):642–56. https://doi.org/10.1214/aos/1176350366.
- Leroy AM, Rousseeuw PJ. Robust Regression and Outlier Detection. Wiley Series in Probability and Mathematical Statistics, New York: Wiley; 1987. https://doi. org/10.1002/0471725382.
- Belsley DA. A guide to using the collinearity diagnostics, Computer Science in Economics and Management. 1991; 4(1):33-50.
- Yamagata T, Orme CD. On testing sample selection bias under the multicollinearity problem, Econometric Reviews. 2005; 24(4):467–81. https://doi. org/10.1080/02770900500406132.
- Alauddin M, Nghiemb HS. Do instructional attributes pose multicollinearity problems? An empirical exploration, Economic Analysis and Policy. 2010; 40(3):351. https://doi. org/10.1016/S0313-5926(10)50034-1.
- Thomas RL. Introductory econometrics. Addison-Wesley Longman Limited; 1993.
- Stewart GW. Collinearity and least squares regression, Statistical Science. 1987; 2(1):68–84. https://doi.org/10.1214/ss/1177013439.
- Andersen R. Modern methods for robust regression, Sage. 2007; 128. https://doi.org/10.4135/9781412985109.
- Hawkins DM. The accuracy of elemental set approximations for regression, Journal of the American Statistical Association. 1993; 88(422):580–89. https://doi.org/10.1080 /01621459.1993.10476310.
- Rousseeuw PJ, Van Driessen K. Computing LTS regression for large data sets, Data Mining and Knowledge Discovery. 2006; 12(1):29–45. https://doi.org/10.1007/s10618-005-0024-4.
- Atkinson AC, Weisberg S. Simulated annealing for the detection of multiple outliers using least squares and least median of squares fitting, Institute for Mathematics and Its Applications. 1991; 33:7–20. https://doi.org/10.1007/978-1-4615-6861-2_2.

- Simpson DG, Ruppert D, Carroll RJ. On one-step GM estimates and stability of inferences in linear regression, Journal of the American Statistical Association. 1992; 87(418):439–50. https://doi.org/10.1080/01621459.1992.10 475224.
- Coakley CW, Hettmansperger TP. A bounded influence, high breakdown, efficient regression estimator, Journal of the American Statistical Association. 1993; 88(423):872–80. https://doi.org/10.1080/01621459.1993.10476352.
- Alfons A, Croux C, Gelper S. Sparse least trimmed squares regression for analyzing high-dimensional large data sets, The Annals of Applied Statistics. 2013; 7(1):226–48. https:// doi.org/10.1214/12-AOAS575.
- Hössjer O. On the optimality of S-estimators, Statistics and Probability Letters. 1992; 14(5):413–19. https://doi. org/10.1016/0167-7152(92)90103-C.
- Salibian-Barrera M, Yohai VJ. A fast algorithm for S-regression estimates, Journal of Computational and Graphical Statistics. 2006; 15(2):414–27. https://doi. org/10.1198/106186006X113629.
- 20. Ruppert D. Computing S-estimators for regression and multivariate location/dispersion, Journal of Computational and Graphical Statistics. 1992; 1(3):253–70. https://doi.org/ 10.1080/10618600.1992.10474584.
- 21. Yohai VJ, Stahel WA, Zamar RH. A procedure for robust estimation and inference in linear regression. In: Directions in robust statistics and diagnostics. Springer, New York,

NY; 1991. p. 365-74. https://doi.org/10.1007/978-1-4612-4444-8_20.

- 22. Marazzi A. Algorithms, routines, and S-functions for robust statistics. CRC Press; 1993.
- 23. Zahari SM, Zainol MS, Al-Banna MI, Ismail B. Weighted ridge mm-estimator in robust ridge regression with multicollinearity. Proceedings of Mathematical Models and Methods in Modern Science; 2012. p. 124–29.
- Darwish K, Buyuklu AH. Robust linear regression using L1-Penalized MM-Estimation for high dimensional data, American Journal of Theoretical and Applied Statistics. 2015; 4(3):78–84. https://doi.org/10.11648/j.ajtas.20150403.12.
- 25. McDonald GC, Galarneau DI. A Monte Carlo evaluation of some ridge-type estimators, Journal of the American Statistical Association. 1975; 70(350):407–16. https://doi. org/10.1080/01621459.1975.10479882.
- Wichern DW, Churchill GA. A comparison of ridge estimators, Techno Metrics. 1978; 20(3):301–11. https://doi.org/1 0.1080/00401706.1978.10489675.
- Gibbons DG. A simulation study of some ridge estimators, Journal of the American Statistical Association. 1981; 76(373):131–39. https://doi.org/10.1080/01621459.1981.10 477619.
- Kibria BG. Performance of some new ridge regression estimators, Communications in Statistics-Simulation and Computation. 2003; 32(2):419–35. https://doi.org/10.1081/ SAC-120017499.