

## Research Article

# A New Efficient Redescending M-Estimator for Robust Fitting of Linear Regression Models in the Presence of Outliers

Dost Muhammad Khan <sup>1</sup>, Muhammad Ali,<sup>1</sup> Zubair Ahmad <sup>2</sup>, Sadaf Manzoor,<sup>3</sup>  
and Sundus Hussain<sup>4</sup>

<sup>1</sup>Department of Statistics, Abdul Wali Khan University Mardan, Mardan, KP, Pakistan

<sup>2</sup>Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran

<sup>3</sup>Department of Statistics, Islamia College Peshawar, Peshawar, KP, Pakistan

<sup>4</sup>Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, KP, Pakistan

Correspondence should be addressed to Zubair Ahmad; [zubair@stu.yazd.ac.ir](mailto:zubair@stu.yazd.ac.ir)

Received 23 July 2021; Revised 4 October 2021; Accepted 8 November 2021; Published 22 November 2021

Academic Editor: Ishfaq Ahmad

Copyright © 2021 Dost Muhammad Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Robust regression is an important iterative procedure that seeks analyzing data sets that are contaminated with outliers and unusual observations and reducing their impact over regression coefficients. Robust estimation methods have been introduced to deal with the problem of outliers and provide efficient and stable estimates in their presence. Various robust estimators have been developed in the literature to restrict the unbounded influence of the outliers or leverage points on the model estimates. Here, a new redescending M-estimator is proposed using a novel objective function with the prime focus on getting highly robust and efficient estimates that give promising results. It is evident from the results that, for normal and clean data, the proposed estimator is almost as efficient as ordinary least square method and, however, becomes highly resistant to outliers when it is used for contaminated datasets. The simulation study is being carried out to assess the performance of the proposed redescending M-estimator over different data generation scenarios including normal, *t*-distribution, and double exponential distributions with different levels of outliers' contamination, and the results are compared with the existing redescending M-estimators, e.g., Huber, Tukey Biweight, Hampel, and Andrew-Sign function. The performance of the proposed estimators was also checked using real-life data applications of the estimators and found that the proposed estimators give promising results as compared to the existing estimators.

## 1. Introduction

Robust regression is an alternative method to the ordinary least square (OLS) regression when the basic assumptions are violated by the nature of the data. The method of OLS requires several assumptions to fit a regression line efficiently, but it produces very poor estimates of the regression coefficients when the same assumptions are not fulfilled, and consequently, the residuals become very large leading to inflated standard errors, which can seriously distort model predictions. Thus, the confidence interval becomes wider, and the estimates of the regression coefficients are no longer asymptotically consistent. Observations that bias the

estimates of the parameters are called bad leverage points, whereas observations that lie along the predicted line are called good leverage points.

During estimation of the regression coefficients, if assumptions of the OLS are violated, then this problem is often fixed by using the transformation techniques. The transformed variables sometimes eliminate the effect of the influential outliers, which can affect the significance of the regression coefficients that can also lead to incorrect predictions. Under these situations, robust regression methods that are resistant to outliers become the only best choice left to fit the regression line and find efficient estimates of the regression coefficients.

A regression model is generally defined as

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

where the dependent variable  $Y$  and the vector of true residuals  $\varepsilon$  are  $n \times 1$ , and the design matrix  $X$  is  $n \times p$ . Write  $\hat{\beta}$  for an estimate of  $\beta$ , and

$$e = Y - X\hat{\beta}, \quad (2)$$

for the corresponding fitted residuals.

The least squares method is commonly used for parameter estimation of the regression model under certain assumptions; for example, the errors term may follow normal distribution with zero mean and constant variance, i.e.,  $e_i \sim N(0, \delta^2)$ . The basic notion of the least square principle for parameter estimation is to minimize the sum of squared residuals. For normal data, the least square method estimation is defined by

$$\min \sum_{i=1}^n e_i^2, \quad (3)$$

where  $e_i = y_i - x_i^T \beta$  represents the vectors of residuals. However, this estimator is very sensitive to bad observations and to even a small departure of the data points from normality. Therefore, statisticians and data-scientists have been working to introduce new robust methods that are highly resistant to outliers.

While numerous robust methods have been proposed in the literature that are resistant to outliers. For example, some of the popular robust regression estimators are presented here: L1-estimator, M-estimator, R-estimators, GM-estimator, RM-estimator, S-estimator, LMS and LTS, MM;  $\tau$ -estimator and redescending M-estimators [1, 2]. However, there is no single method that can outperform all the methods in all outlying scenarios. One method may be good and perform well for one outlier scenario but may not be good for other outlying situation.

Keeping in view the weaker performance of the available robust estimators, an attempt has been made in this paper to introduce a novel redescending M-estimator that outperforms the existing robust methods in terms of efficiency and robustness, which gives promising results. The main contribution in this paper is to propose a new redescending M-estimator that has the potential to perform well in almost all kinds of contaminated datasets without compromising on its efficiency.

## 2. Literature Review

In this section, we present and discuss all those robust methods that have been introduced in the literature by different researchers, with the main objective being coping with the problem of outliers that may be vertical outliers or leverage points. The details of these methods are as follows.

**2.1. L-Estimators.** L-estimators are computed by using some linear combinations of order statistics. The first type of this estimator is called L-estimation method. It is more resistant

than the classical ordinary least square and is known as least absolute deviation (LAD) regression, which is also called  $L_1$ -regression, because it minimizes the sum of absolute deviations. Least absolute value is a very simple and easy procedure towards bounded influence of robust regression. The method of least square regression, which minimizes the sum of square of regression, also fits the definition of  $L$ -estimators and is sometimes called as  $L_2$ -norm. Other types of  $L$ -estimators are least median of square (LMS) and least trimmed square (LTS). However, a brief discussion on  $L$ -estimators is given as follows.

**2.2. Least Absolute Deviation (LAD) Estimator.** The LAD estimator was introduced by Edgeworth in 1887 [3], which is more resistant to outliers, specifically when they appear in the direction of dependent variable  $y$ . The estimates of the regression coefficients can be obtained by minimizing the sum of absolute value of the residuals. The mathematical form of the least absolute value regression is defined as

$$\min \sum_{i=1}^n |e_i| = \min \sum_{i=1}^n \left| y_i - \sum x_{ij} \beta_j \right|. \quad (4)$$

The objective function of the LAD estimator denoted by  $\rho(r)$  is a more general form of the quantile regression and can be written as

$$\sum_{i=1}^n \rho_\omega(e_i), \quad (5)$$

where

$$\rho_\omega(e_i) = \begin{cases} \omega e_i & \text{if } e_i \geq 0 \\ (\omega - 1)e_i & \text{if } e_i < 0 \end{cases}, \quad (6)$$

and  $\omega$  is the quantile, which needs to be estimated. More details on the topic of quantile regression and its application are given in Koenker and Basset [4] and Koenker & D'orey [5].

Although *LAD* is less affected by outliers in the direction of response variable  $y$  and performs better than OLS, yet when unusual values appear in the direction of explanatory variable  $x$  (called leverage points), the method of *LAD* produces very poor estimates of the regression coefficients as its breakdown point becomes zero, i.e.,  $BDP = 0$ . The *LAD* estimates are not efficient in the case of mean, using the assumption that  $y \sim N(\mu, \delta^2)$  and the sampling variance of  $y$  for ordinary least square (OLS) is  $\delta^2/n$ , and for *LAD* it is  $(\pi/2) = 1.57$  times greater than that of the variance of the OLS, i.e.,  $\text{Var}(LAD) = (\pi\delta^2/2n)$ . The low breakdown points in case of outliers in the direction of explanatory variable  $x$  and low efficiency of *LAD* are compared to the ordinary least square (OLS), which makes the *LAD* regression estimators less attractive than other robust regression techniques [6].

**2.3. Least Median of Squares (LMS) Regression.** Least median of square (LMS) was introduced by Rousseeuw and Yohai [7], which is based on the median of square residuals instead

of sum of square residuals used in the ordinary least squares. The mathematical form of the least median square is given as

$$\widehat{\beta}_{\text{LMS}} = \min(r^2)_{h:n}, \quad (7)$$

where  $(r^2)_{1:n} \leq (r^2)_{2:n} \leq \dots \leq (r^2)_{n:n}$  are the ordered squared residuals and where the optimal coverage is chosen as  $h = [(n + p + 1)/2]$ . Here,  $p$  is the number of regression coefficients to be estimated including the intercept term. The LMS procedure performs well on the majority of the data points. One formal technique for expressing and quantifying this feature depends upon the breakdown point of an estimator. Usually, the breakdown point (BD) of an estimator is the minimum fraction of change in the data points that can make an estimator large. The breakdown point of the location estimate, which is called mean, is zero percent, and even a single outlier out of  $n$  can bring arbitrarily large changes. Similarly, the breakdown point for median is 50%. It can be seen that these figures persist for the regression; i.e., the breakdown point for OLS is 0%, and that for the LMS regression is 50%. Since the breakdown point of LMS estimator is 50%, therefore, it has certain disadvantages that limit its use. The maximum relative efficiency of LMS is 37% [8]; because of the convergence rate of  $n^{-1/3}$  the influence function is not well defined [7]. Despite these disadvantages, it can be seen that LMS estimates are significant and play an important role in the calculation of more efficient estimators called MM-estimators by giving them initial estimates of the residual.

**2.4. Least Trimmed Square (LTS) Estimator.** Least trimmed square (LTS) estimator was proposed by Rousseeuw in 1984, which is based on the idea of trimmed sum of squared residuals that allows some observations to have potentially large residuals. The principle of LTS estimation is to minimize the trimmed sum of squared residuals. It is generally defined by the following mathematical expression:

$$\min \sum_{i=1}^w e_{(i)}^2, \quad (8)$$

where  $w = [n(1 - \lambda) + 1]$  is the subset of observations having smallest residuals with the minimum least squares objective function, and  $\lambda$  is the proportion of trim values. It is equivalent to searching for finding  $w$  observations having smallest residuals and applying LS fit to these  $w$  observations. For the value of  $w = (n/2) + 1$  the breakdown point for LTS estimator is 50%, i.e., BDP = 0.50 [9].

Even though the LTS estimator has more resistance to outliers, it suffers from the problem of very low relative efficiency [10]. Because of the low efficiency of the LTS estimator, it is not a desirable one. The importance of the LTS estimator cannot be denied, because it plays a significant role in the calculation of the other robust estimators like GM-estimators, which were introduced by Coakley and Hettmansperger [11]; and the residuals obtained from LTS method can be used successfully in outlier diagnostic plots.

### 3. Huber M-Estimators

In order to restrict the influence of outliers in a regression problem, the  $M$  estimator was introduced by Huber [12], which is obtained by minimizing a less rapidly increasing function of residuals instead of sum of squared residuals. The main objective behind its development was to make a trade-off between the efficiency of OLS and robustness of LAD estimators.  $M$ -estimator is nothing but a generalization of the OLS and LAD estimators. The method of  $M$ -estimation is based on the minimization of some function of the residuals. The robustness of the  $M$ -estimation depends upon the choice of the weight function.

If the assumptions, like linearity, homoscedasticity, and no autocorrelation between error terms of the classical regression, are satisfied, then the OLS estimator of  $\beta$  can be obtained by minimizing the sum of square of error by using the method of maximum likelihood estimation, i.e.,

$$\min \sum_{i=1}^n (r_i)^2. \quad (9)$$

The criterion function for  $M$ -estimator based on the minimization of the sum of square of residuals is replaced with a less rapidly increasing function of residuals, say,  $\rho(r_i)$ , i.e.,

$$\min \sum_{i=1}^n \rho(r_i), \quad (10)$$

where the function  $\rho(r_i)$  is called objective function, which possesses the following properties:

- (i)  $\rho(r_i) \geq 0$
- (ii)  $\rho(0) = 0$
- (iii)  $\rho(r_i) = \rho(-r_i)$
- (iv)  $\rho(r_i) \geq \rho(r_j)$  for  $|r_i| \geq |r_j|$
- (v)  $\rho$  is continuous ( $\rho$  is differentiable)

The  $M$ -estimator is not necessarily scale invariant (i.e., if the errors  $r_i = y_i - X_i' \beta$  are multiplied by a constant, the new solution to eq. (8) might not be the same as the old one). The  $\psi$  function is obtained by taking the derivative of the objective function with respect to the coefficients  $\beta$ , and then equating it to zero gives a system of  $k+1$  estimating equations for the coefficients, that is,

$$\sum_{i=1}^n \psi(r_i) = 0, \quad (11)$$

where  $\psi$  is the score function. The weight function is then obtained by dividing the  $\psi$ -function by the corresponding residuals, that is,  $w(r_i) = \psi(r_i)/r_i$ . Now, suppose that  $w_i = w(r_i)$ , then the above estimating equation can be written as

$$\sum_{i=1}^n w_i(r_i) x_i' = 0. \quad (12)$$

For computation of  $M$ -estimator, an iterative method is required to solve the above system of nonlinear equation. For this purpose, the most commonly used optimization

techniques is iterative reweighted least square (IRLS) method.

**3.1. Redescending M-Estimators.** Redescending M-estimator was proposed by Hampel in 1972 [13], which is also known as three-part redescending M-estimator. The objective function  $\rho(r_i)$  of the redescending estimator is bounded, and for large values of residuals, its psi-function  $\psi$  re-descends towards zero. Redescending M-estimators completely reject extreme outliers and are thus more robust than the M-estimators. Alternatively, if the derivative of  $\rho$  (objective function) redescends to zero, the estimator is said to be redescending M-estimator, i.e.,

$$\lim_{\eta \rightarrow \pm\infty} \rho'(r_i) = 0. \quad (13)$$

The breakdown point of the redescending M-estimators approximately equals 0.5 and is considered to be high breakdown robust estimators. The score function of such high breakdown point estimators smoothly redescends to zero [14]. There are many types of redescending M-estimators, some of which are discussed as follows.

**3.2. Hampel's Three-Part Function.** Hampel's three-part redescending M-estimator was introduced by Andrews et al. [15] and was considered as a first attempt towards a redescending M-estimator. The objective function of Hampel's estimator is a step function, which is also known as Hampel's three part redescending estimators. The objective, Psi, and weight function of the Hampel' estimator are as follows:

Hampel's objective function is defined by

$$\rho(r) = \begin{cases} \frac{r^2}{2}, & \text{for } |r| \leq a, \\ a|r| - \frac{1}{2}a^2, & \text{for } a < |r| \leq b, \\ ab - \frac{a^2}{2} + \frac{a(c-b)}{2} \left[ 1 - \left( \frac{c-|r|}{c-b} \right)^2 \right], & \text{for } b < |r| \leq c, \\ ab - \frac{a^2}{2} + \frac{a(c-b)}{2}, & \text{for } |r| > c. \end{cases} \quad (14)$$

The corresponding redescending  $\Psi$ -function is

$$\psi(r) = \begin{cases} r, & \text{for } |r| \leq a, \\ a \operatorname{sign}(r), & \text{for } a < |r| \leq b, \\ a \frac{c-|r|}{c-b} \operatorname{sign}(r), & \text{for } b < |r| \leq c, \\ 0, & \text{for } |r| > c. \end{cases} \quad (15)$$

The weight function is

$$w(r) = \begin{cases} 1, & \text{for } |r| \leq a, \\ \frac{a}{|r|}, & \text{for } a < |r| \leq b, \\ \frac{a(c-|r|)}{(c-b)|r|}, & \text{for } b < |r| \leq c, \\ 0, & \text{for } |r| > c, \end{cases} \quad (16)$$

where the parameters  $a$ ,  $b$ , and  $c$  are constants, and their ranges vary from  $0 < a \leq b < c < \infty$ . The performance of Hampel estimator is efficient as shown in Figure 1. The score function is differentiable. However, a smooth  $\psi$ -function was required that led to the development of Andrews' sine function.

**3.3. Andrew's Sine Function.** Andrews et al. [15] introduced a new redescending M-estimator popularly known as Andrews-sine function. This estimator is more resistant to outliers and smoothly redescending as can be observed from Figure 2. The objective function, Psi, and weight function of the Andrews-sine estimator are as follows:

$$\rho(r) = \begin{cases} c^2 \left[ 1 - \cos\left(\frac{r}{c}\right) \right], & \text{for } |r| \leq c\pi, \\ 2c^2, & \text{for } |r| > c\pi. \end{cases} \quad (17)$$

The corresponding redescending  $\Psi$ -function is

$$\psi(r) = \begin{cases} c \sin\left(\frac{r}{c}\right), & \text{for } |r| \leq c\pi, \\ 0, & \text{for } |r| > c\pi. \end{cases} \quad (18)$$

The weight function is

$$w(r) = \begin{cases} \frac{c}{r} \sin\left(\frac{r}{c}\right), & \text{for } |r| \leq c\pi, \\ 0, & \text{for } |r| > c\pi. \end{cases} \quad (19)$$

We graph the  $\rho$ -function,  $\Psi$ -function, and the weight function in Figure 2.

**3.4. Tukey's Biweight Function.** Beaton and Tukey [16] introduced a biweight function. Tukey's biweight function is smooth and has been used successfully and extensively used in a wide variety of applications. The objective function  $\rho$ ,  $\Psi$ -function, and weight function can be defined as follows:

$$\rho(r) = \begin{cases} \frac{c^2}{6} \left[ 1 - \left( 1 - \left( \frac{r}{c} \right)^2 \right)^3 \right], & \text{for } |r| \leq c, \\ \frac{c^2}{6}, & \text{for } |r| > c. \end{cases} \quad (20)$$

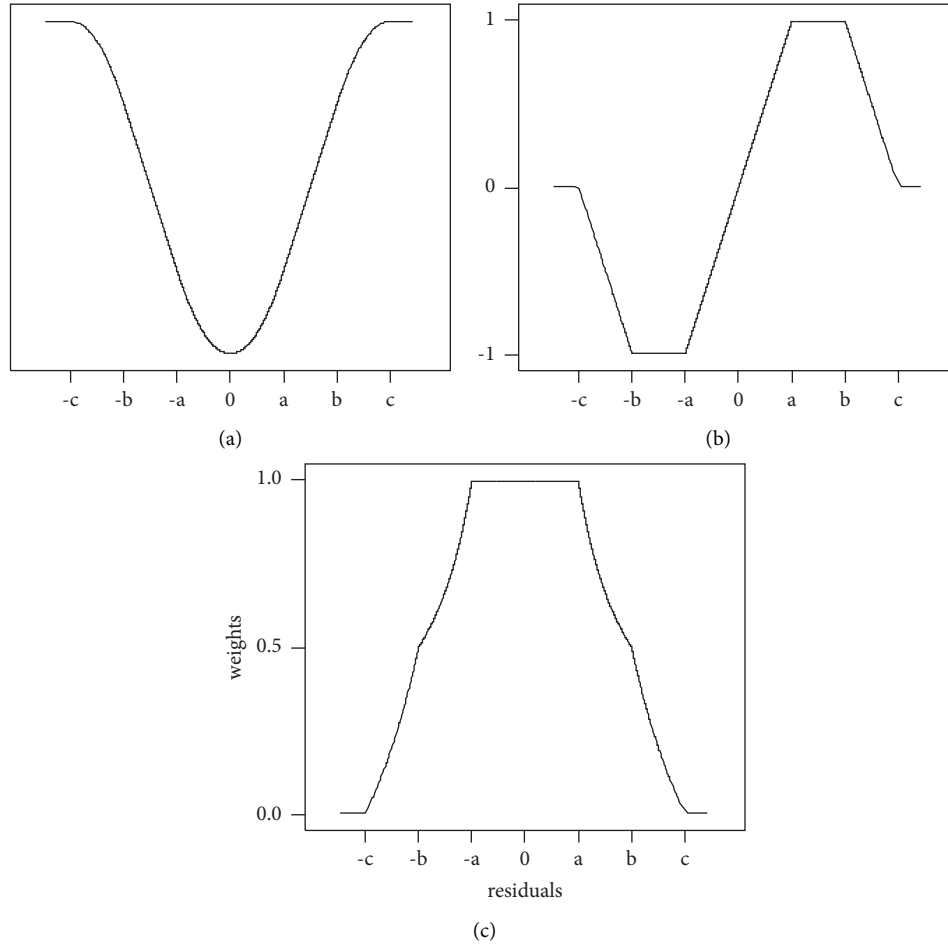


FIGURE 1: Graph of Hampel three-part function. (a) Objective function. (b)  $\Psi$ -function. (c) Weight function.

The corresponding  $\Psi$ -function is

$$\psi(r) = \begin{cases} r \left[ 1 - \left( \frac{r}{c} \right)^2 \right]^2, & \text{for } |r| \leq c, \\ 0, & \text{for } |r| > c. \end{cases} \quad (21)$$

The weight function is

$$w(r) = \begin{cases} \left( 1 - \left( \frac{r}{c} \right)^2 \right)^2, & \text{for } |r| \leq c, \\ 0, & \text{for } |r| > c. \end{cases} \quad (22)$$

We graph the  $\rho$ -function,  $\Psi$ -function, and the weight function in Figure 3.

**3.5. Iteratively Reweighted Least Squares (IRLS).** Since there is no closed form solution of robust estimator, therefore, IRLS fitting algorithm is commonly used to estimate the regression coefficients. Before finding the residuals, one needs to fit the model, which is not possible in a single step. The estimates of the regression coefficients cannot be found

until the values of the residuals are known. That is why the method of iterative reweighted least square is employed. It has the following steps:

- (i) Regression line is fitted to the data through OLS method by setting the iteration counter at  $I = 0$ , and the initial values of the estimates of the regression parameters are found.
- (ii) Initial estimates of the weights can be found from the residuals  $e_i^{(0)}$  which are extracted from fitting the OLS regression to the data.
- (iii) A weight function is then applied to the initial values of the OLS residuals to create preliminary weights,  $w(e_i^{(0)})$ .
- (iv) The weighted least square minimizes  $\sum w_i^{(1)} e_i^2$  and thus obtains  $\hat{\beta}^{(1)}$  in the first iteration at  $I = 0$ . In matrix form, with  $W$  representing the  $n \times n$  diagonal matrix of individual weights, the solution is

$$\hat{\beta}^{(1)} = (X'WX)^{-1} X'Wy. \quad (23)$$

- (v) The new weights  $w_i^{(2)}$  can be obtained from the residuals of the initial weighted least square (WLS).

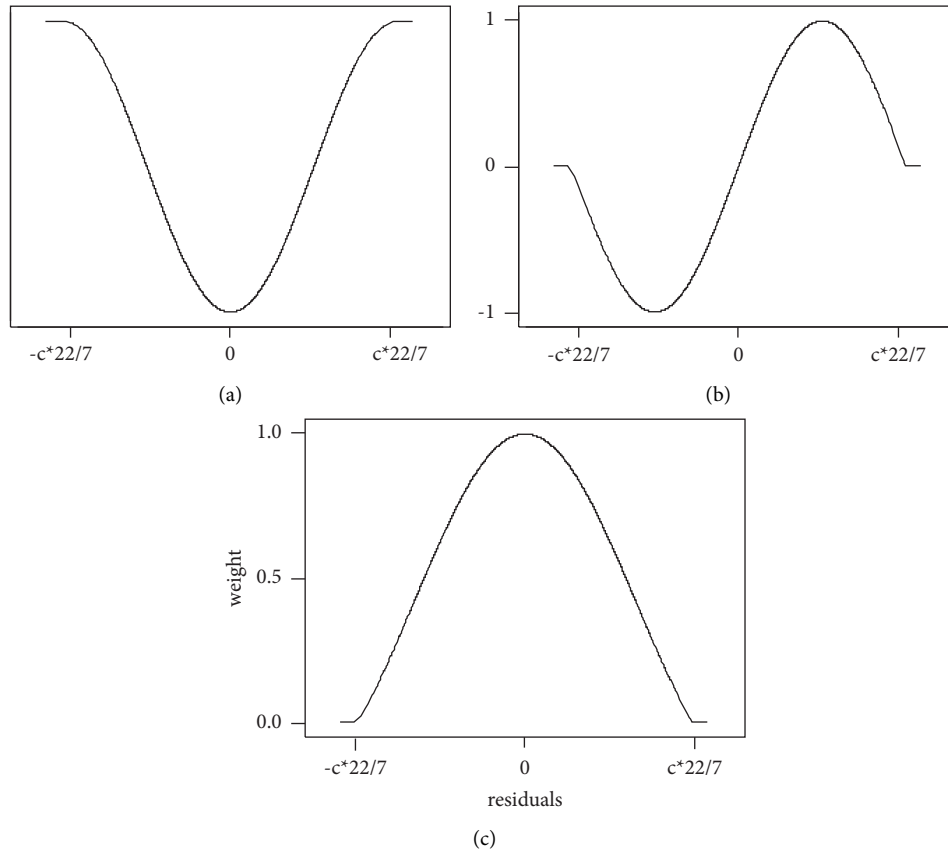


FIGURE 2: Graph of Andrew's sine function. (a) Objective function. (b)  $\Psi$ -function. (c) Weight function.

(vi) The new weights  $w_i^{(2)}$  are used in the next iteration,  $I=2$ , of weighted least square (WLS) to estimate  $\hat{\beta}^{(2)}$ .

(vii) The iteration will be stopped when estimates of  $\hat{\beta}$  stabilize from the previous iteration.

In general, at each of the  $q$  iterations, the solution is  $\hat{\beta}^{(I)} = (X'W_qX)^{-1}X'W_qy$ , where  $W_q = \text{diag}\{w_i^{(I-1)}\}$ . The procedure of iteration continues until  $\hat{\beta}^{(I)} - \hat{\beta}^{(I-1)} \cong 0$ . The iteration should be stopped, and the solution is considered to be converged if the deviation in estimates is not more than 0.01% from the preceding iteration.

M-estimators are considered to be robust in situations when the distribution of error is heavy-tailed, the variance of error does not follow the assumption of homoscedasticity, and outliers appear in the direction of dependent variable  $y$ . M-estimators also consider that the matrix  $X$  of the model is measured without error. Under these situations, M-estimators are considered to be more efficient than OLS estimates [14].

The efficiency of the M-estimates is about 95% as compared to OLS under the assumptions of classical regression model. Although M-estimators are developed in place of the OLS especially in terms of resistance and robustness to those unusual data points, which appear in the direction of  $y$ , like least absolute value (LAV) estimators, M-estimators are not totally resistant to odd observations, because they are not those outliers that are

leverage points. It is to be noted that M-estimates of location are highly robust, because they have a breakdown point (BDP) of 0.5 as well as a bounded influence function, but M-estimates for regression have these characteristics only for  $y$  but not for  $x$ , and thus their breakdown point equals zero, i.e.,  $\text{BDP} = 0$  [17]. In other words, in some situations, M-estimators perform no better than OLS [2]. The importance of M-estimators cannot be denied because of the fact that it plays a significant role in computing more robust estimates.

#### 4. Proposed Estimator

Redescending estimators aim to minimize  $\sum_{i=1}^n \rho(e_i)$  where  $\rho$  is even, smooth, differentiable, and continuous function. The choice of  $\rho$  mainly depends on keeping in view the objective of the estimators. For example, the objective function should be that function that is able to resist extreme outliers and should have bounded influence  $\psi$ -function that is zero for large residuals. Similarly, redescending estimators should have ideally differential and smooth function to maintain high efficiency along with robustness.

In this study, an attempt has been made to cover some of the drawbacks of the above mentioned redescending estimators. For this purpose, the properties of the tangent hyperbolic function were assessed, and it was found that tangent hyperbolic function satisfies all the properties associated with a good objective function as mentioned below.

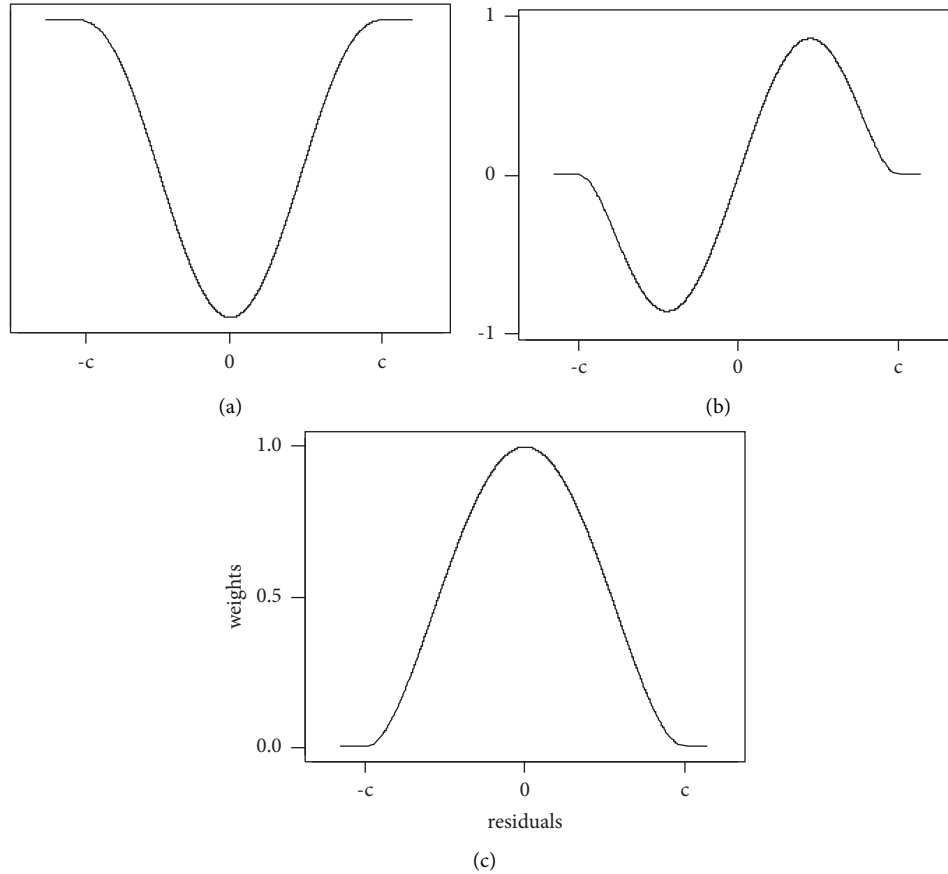


FIGURE 3: Graph of Tukey's biweight function. (a) Objective function. (b)  $\Psi$ -function. (c) Weight function.

Consider the following tangent hyperbolic function as the objective function:

$$\rho(r) = c \tan^{-1} \left[ \tanh \left( \frac{r^2}{2c} \right) \right], \quad \text{for } |r_i| \geq 0, \quad (24)$$

where  $c$  and  $r_i$  in the above defined objective function are the tuning constant and the scaled residuals over MAD. The proposed  $\rho$ -function satisfies all the properties associated with a good objective function, namely,

- (i)  $\rho(r_i) \geq 0$
- (ii)  $\rho(0) = 0$ , the objective function is zero for any residual  $r_i = 0$
- (iii)  $\rho(r_i) = \rho(-r_i)$ , the objective function should be monotonic
- (iv)  $\rho(r_i) \geq \rho(r_j)$  for  $|r_i| \geq |r_j|$
- (v)  $\rho$  is continuous ( $\rho$  is differentiable)

The corresponding score function  $\Psi$  is obtained by taking the first derivative of the proposed  $\rho$ -function, that is,

$$\psi(r_i) = \frac{r_i}{\cosh(r_i^2/c)}, \quad \text{for } |r_i| \geq 0. \quad (25)$$

Similarly, the corresponding weight function is obtained by dividing the  $\Psi$ -function by  $r_i$ , and we get

$$w(r) = \frac{1}{\cosh(r^2/c)}, \quad \text{for } |r_i| \geq 0. \quad (26)$$

The graphical representation of the objective function, score function, and weight function is presented in Figure 4.

An important characteristic of a robust estimator is the breakdown point (BDP) and efficiency, which measure the stability of a robust estimator regarding how much it is consistent against wild data points present in the dataset. A common practice to these questions is a compromise between efficiency and resistance; that is, choose an optimal estimator, which has maximum resistance with minimum loss of efficiency. In other words, avoid a robust estimator at the cost of maximum efficiency loss, nor should you choose completely nonrobust estimator with maximum efficiency, but prefer make a compromise between these two properties. The proposed  $\Psi$ -function, given in Figure 2(b), has a different mechanism as compared to Andrew's sine and Tukey's biweight estimators. The score function  $\Psi$  of the mean is a linear-straight line, and, therefore, it becomes a more efficient estimator provided that all the assumptions of normality are fulfilled. The longer and linear central section of the proposed  $\Psi$ -function behaves linearly for most of the central observations, where other smoothly redescending estimators fail to do so. This increased linearity, certainly, improves the efficiency as most of the central observations

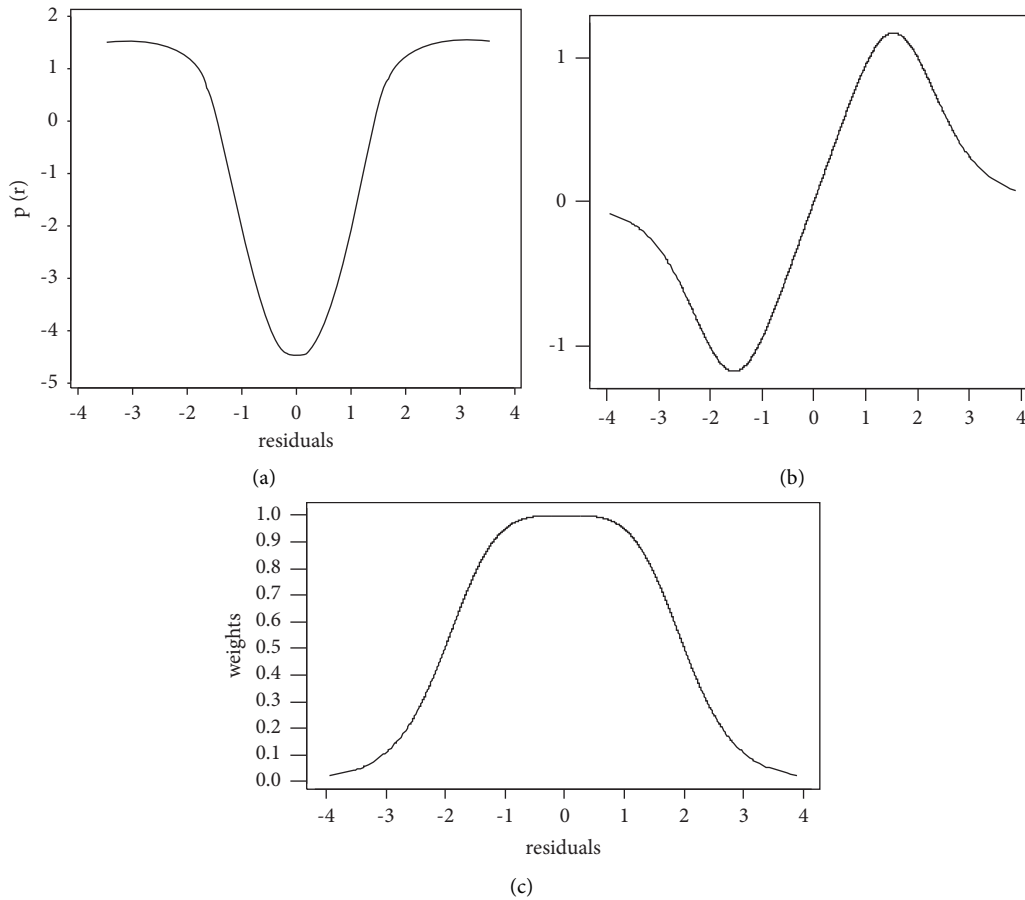


FIGURE 4: Graph of Ali function. (a) Objective function. (b) Score function. (c) Weight function.

still satisfy the normality assumption. The  $\Psi$ -function, then, redescends slowly for large values of residuals and becomes zero for values lying outside the particular band.

The most interesting point of the proposed estimator is its weight function. The weight function of the proposed estimator assigns weights nearly equal to 1 from zero to moderate size of the residuals and then weights down the residuals, and thus the outlier gets weight that is nearly equal to zero. Figure 2(c) shows graph of the proposed weight function.

## 5. Simulation Study

To assess the overall performance of the proposed estimator and the existing methods in the literature, an extensive simulation study is carried out. For comparison purposes, datasets have been simulated from different distributions and levels of contamination in the datasets. A simulation study is carried out by considering different factors that can potentially influence the model fit and the outcomes of the competing methods. For the purpose of simulation study, 1000 datasets were generated, and the proposed estimator, called Ali redescending M-estimator, is utilized along with Huber, Tukey biweight, Andrew's sine, Hampel, and OLS estimator. For comparison purposes, total mean square error is calculated for all the replications in the simulation study by using the following formula:

$$TMSE = \sqrt{\frac{\sum_{i=1}^g \sum_{j=1}^p (\hat{\beta}_{ij} - \beta_{ij})^2}{gp}} \quad (27)$$

The simulated datasets were generated from well-known distributions, of which two are heavy tailed distributions. The following simulation scenarios were considered here in this study:

- (i) Generating both  $X$  and  $Y$  from standard normal distribution, i.e.,  $X \& Y \sim N(0, I)$
- (ii) Student's  $t$ -distribution, i.e.,  $X \& Y \sim t(3)$
- (iii) Double exponential distribution, i.e.,  $X \& Y \sim L(0, I)$ .

All the datasets were either generated from the above-mentioned distributions or some contaminations were added in the form of outliers, and the results of the competing methods are presented in the following tables along with data generation scenarios and dimensions.

The three desiring properties of every robust method are bias, efficiency, and breakdown point. The results of the simulation study for the clean data set are generated from normal distribution with  $\lambda = 0\%$  level of contamination summarized in Table 1. It can be observed from the results given in Table 1 that the OLS estimator has the lowest total



TABLE 1: Total mean square error for datasets generated from normal distribution ( $X$  &  $Y \sim N(0, 1)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Proposed
$N = 30$	$P = 2$	0.301	0.301	0.301	0.301	0.301	0.301
	$P = 3$	0.339	0.350	0.342	0.369	0.369	0.342
	$P = 6$	0.523	0.547	0.550	0.741	0.731	0.527
$N = 50$	$P = 2$	0.196	0.204	0.200	0.206	0.012	0.197
	$P = 3$	0.248	0.256	0.250	0.260	0.260	0.249
	$P = 6$	0.362	0.381	0.373	0.371	0.411	0.363
$N = 100$	$P = 2$	0.138	0.144	0.141	0.144	0.144	0.108
	$P = 3$	0.175	0.180	0.175	0.180	0.180	0.170
	$P = 6$	0.242	0.258	0.253	0.261	0.261	0.207

mean square error (TMSE). The proposed Ali redescending M-estimator is a good competitor of OLS in terms of the smallest total mean square error and highest efficiency.

Table 2 presents the results obtained for the second scenario, where the datasets of different sizes and number of predictor variables were generated from t-distribution, and the total mean square errors were obtained for all the competing methods including the proposed estimator. It is evident from the results given in Table 2 that the proposed method outperforms all the competing methods in terms of smallest mean square error. Similarly, the results presented in Table 3 are based on simulated datasets generated from standard Laplace distribution, which is a heavy tailed distribution with different sample sizes and number of predictors. It can be observed from the given results that the performance of the proposed estimators based on lowest total mean square error is better than the rest of the mentioned methods. Table 4 presents the total mean square error of the proposed estimator along with competing methods based on datasets generated from normal, where 90% of the observation are normal, and 10% consist of outliers in the response as well as predictors. From the given results, it is evident that the proposed method performs well than the rest of the methods having lowest TMSE in almost all the data generation scenarios with different sample sizes and predictors. From Table 5, it is evident that the proposed methods are more efficient having lowest TMSE as compared to the competing methods for the simulated datasets, where 80% of the observations have been generated from standard normal distribution, and the remaining 20% were generated as outliers from the same distribution.

The results presented above show that the performance of the proposed method is promising for higher dimension, which means that as the sample size increases, the efficiency of the proposed method also increases, and the TMSE decreases. Similarly, when the number of predictors increases, the performance of the proposed methods becomes better, which testifies the performance of the proposed method for higher dimensional simulated datasets. From the simulations results given in Table 6, where 80% of the observations of the explanatory and response variables were simulated from standard normal and 20% from the same distribution with different location-scale parameters as mentioned, it is evident that the performance of the proposed Ali estimator is more efficient than the other competing methods with smallest total mean square error. Moreover, the proposed

TABLE 2: Total mean square error for datasets generated from  $t$ -distribution ( $X$  &  $Y \sim t$ -distribution,  $t(3)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	0.373	0.301	0.311	0.305	0.303	0.300
	$P = 3$	0.441	0.361	0.370	0.380	0.378	0.301
	$P = 6$	0.671	0.507	0.532	0.660	0.652	0.530
$N = 50$	$P = 2$	0.288	0.210	0.221	0.214	0.217	0.240
	$P = 3$	0.342	0.260	0.266	0.261	0.263	0.267
	$P = 6$	0.458	0.360	0.369	0.380	0.370	0.384
$N = 100$	$P = 2$	0.199	0.157	0.163	0.155	0.155	0.162
	$P = 3$	0.227	0.172	0.175	0.173	0.175	0.180
	$P = 6$	0.300	0.237	0.245	0.243	0.243	0.245

TABLE 3: Total mean square error for datasets generated from Laplace distribution ( $X$  &  $Y \sim L(0, 1)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	0.336	0.293	0.305	0.300	0.299	0.227
	$P = 3$	0.415	0.349	0.365	0.372	0.370	0.305
	$P = 6$	0.562	0.507	0.532	0.685	0.675	0.529
$N = 50$	$P = 2$	0.243	0.214	0.225	0.221	0.220	0.201
	$P = 3$	0.303	0.259	0.270	0.261	0.261	0.201
	$P = 6$	0.562	0.505	0.532	0.685	0.673	0.528
$N = 100$	$P = 2$	0.172	0.149	0.157	0.150	0.150	0.102
	$P = 3$	0.200	0.172	0.181	0.172	0.173	0.165
	$P = 6$	0.275	0.235	0.245	0.235	0.236	0.216

TABLE 4: Total mean square error for contaminated datasets generated from normal distribution (90%  $X$  &  $Y \sim N(0, 1)$  and 10%  $X \sim N(20, 1)$ ; and 10%  $Y \sim N(100, 1)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	4.980	4.982	0.296	0.302	0.299	0.284
	$P = 3$	3.628	3.695	0.362	0.380	0.377	0.363
	$P = 6$	2.500	3.001	0.703	0.802	0.704	0.702
$N = 50$	$P = 2$	5.103	5.103	0.222	0.222	0.207	0.201
	$P = 3$	4.002	4.003	0.300	0.302	0.302	0.300
	$P = 6$	2.370	2.123	0.411	0.421	0.408	0.401
$N = 100$	$P = 2$	4.980	4.710	0.150	0.150	0.153	0.142
	$P = 3$	3.531	4.011	0.212	0.223	0.215	0.201
	$P = 6$	2.290	2.001	0.273	0.280	0.280	0.276

estimator has shown similar performance when the rate of contamination of the outliers increased from 20% to 40% in the simulated data as shown in Tables 7–9. It is also clearly

TABLE 5: Total mean square error for contaminated datasets generated from normal distribution (80%  $X \sim N(0, 1)$  and 20%  $X \sim N(20, 1)$  and 20%  $Y \sim N(100, 1)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	5.050	5.092	0.297	0.312	0.314	0.201
	$P = 3$	3.600	3.703	0.445	0.432	0.504	0.410
	$P = 6$	2.509	2.570	1.133	1.125	1.124	1.101
	$P = 2$	4.989	5.012	0.253	0.243	0.233	0.210
$N = 50$	$P = 3$	3.598	3.624	0.330	0.330	0.335	0.290
	$P = 6$	2.382	2.426	0.490	0.505	0.505	0.454
	$P = 2$	5.053	4.985	0.155	0.159	0.159	0.153
$N = 100$	$P = 3$	3.659	3.567	0.197	0.199	0.199	0.191
	$P = 6$	2.509	2.318	0.284	0.286	0.287	0.281

TABLE 6: Total mean square error for contaminated datasets generated from normal distribution (80%  $X \sim N(0, 1)$  and 20%  $X \sim N(20, 1)$  and 20%  $Y \sim N(10, 5)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	0.545	0.537	0.324	0.317	0.325	0.312
	$P = 3$	0.625	0.515	0.490	0.490	0.489	0.457
	$P = 6$	1.036	0.695	0.656	0.771	0.766	0.623
$N = 50$	$P = 2$	0.519	0.512	0.484	0.409	0.418	0.402
	$P = 3$	0.526	0.450	0.433	0.408	0.416	0.403
	$P = 6$	0.770	0.528	0.522	0.597	0.519	0.502
$N = 100$	$P = 2$	0.514	0.505	0.499	0.413	0.417	0.401
	$P = 3$	0.625	0.397	0.399	0.350	0.355	0.307
	$P = 6$	1.035	0.385	0.390	0.360	0.363	0.349

TABLE 7: Total mean square error for datasets generated from (60%  $X \sim N(0, 1)$  and 40%  $X \sim N(20, 1)$ ; and 20%  $Y \sim N(10, 5)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	5.123	5.163	0.523	0.523	0.523	0.511
	$P = 3$	3.703	3.725	1.264	1.283	1.278	1.212
	$P = 6$	2.534	2.576	2.502	2.731	2.720	2.517
$N = 50$	$P = 2$	5.044	5.061	0.257	0.258	0.258	0.207
	$P = 3$	3.629	3.645	0.521	0.528	0.528	0.509
	$P = 6$	2.534	2.417	2.127	2.159	2.157	2.012
$N = 100$	$P = 2$	5.014	5.023	0.181	0.186	0.186	0.171
	$P = 3$	3.576	3.585	0.229	0.229	0.229	0.219
	$P = 6$	2.310	2.321	1.547	1.551	1.551	1.527

evident from the results in these tables that the performance of the proposed estimators has improved from smaller sample size to large and from lowest to high-dimensional datasets.

### 6. Applications on Real Datasets

Here, in this section, the performance of the proposed estimator will be assessed using real world datasets in order to check its performance along other competing methods.

6.1. *Brownlee’s Stack Loss Data.* The Stack loss data set is a well-known data set taken from Brownlee [18]. It consists of the examination of a plant for the oxidation of ammonia in the production of nitric acid. There are 21 observations in the

TABLE 8: Total mean square error for contaminated datasets generated from ( $X \sim N(0, 1)$ ; 80%  $Y \sim N(0, 1)$  and 20%  $Y \sim N(100, 1)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	21.410	0.615	0.314	0.317	0.317	0.307
	$P = 3$	23.81	0.727	0.398	0.405	0.405	0.390
	$P = 6$	27.764	3.278	0.608	0.649	0.649	0.599
$N = 50$	$P = 2$	20.800	0.559	0.225	0.229	0.229	0.213
	$P = 3$	21.885	0.681	0.288	0.289	0.289	0.270
	$P = 6$	24.249	0.858	0.421	0.429	0.428	0.410
$N = 100$	$P = 2$	20.467	0.540	0.158	0.160	0.160	0.150
	$P = 3$	20.883	0.567	0.200	0.205	0.205	0.197
	$P = 6$	22.150	0.664	0.282	0.288	0.288	0.271

TABLE 9: Total mean square error for datasets generated from  $X \sim N(0, 1)$ ; 60%  $Y \sim N(0, 1)$  and 40%  $Y \sim N(100, 1)$ ).

$N$	$P$	OLS	Huber	Hampel	Andrew	Tukey	Ali
$N = 30$	$P = 2$	41.297	12.45	0.356	0.356	0.356	0.345
	$P = 3$	42.357	23.84	0.466	0.467	0.467	0.445
	$P = 6$	46.583	48.64	0.755	0.752	0.752	0.728
$N = 50$	$P = 2$	40.617	5.583	0.260	0.260	0.260	0.245
	$P = 3$	41.376	14.14	0.337	0.337	0.401	0.327
	$P = 6$	43.420	34.79	0.497	0.497	0.497	0.477
$N = 100$	$P = 2$	40.38	2.297	0.183	0.183	0.183	0.161
	$P = 3$	40.60	4.247	0.233	0.233	0.269	0.211
	$P = 6$	42.637	11.335	0.330	0.334	0.334	0.310

TABLE 10: Estimates of the model parameters for Brownlee’s stackloss data.

Method	Estimate of $\beta_0$	Estimate of $\beta_1$	Estimate of $\beta_2$	Estimate of $\beta_3$	WSSR
OLS	-39.9197	0.7156	1.2953	-0.1521	178.83
Huber	-41.0265	0.8294	0.9261	-0.1278	200.765
Hampel	-40.4747	0.7411	1.2251	-0.1455	256.768
Andrew	-40.7300	0.9370	0.5350	-0.111	225.870
Tukey	-42.2853	0.9275	0.6507	-0.1123	253.426
ALI	-37.1170	0.7679	0.7951	-0.1046	212.148

dataset and three explanatory variables, that is, air flow ( $X_1$ ), cooling temperature ( $X_2$ ), and acid concentration ( $X_3$ ), and the dependent variable is stackloss ( $Y$ ).

The analysis results presented in Table 10 give estimates of the regression coefficients and their corresponding sum of squares computed by OLS and the existing robust methods including the proposed redescending M-estimator. The estimates of the regression coefficients and fitted values obtained by the proposed estimator are promising as compared to the competing robust methods as shown in the table. Similarly, the residuals computed from the model fitted through the proposed estimator are smaller than the residuals computed from the models fitted through other redescending M-estimators that lead to smaller sum of square residuals. Hence, from the simulation study, as well as real data applications, it can be concluded that the performance of the proposed estimator is better than the redescending M-estimators considered in the study.

## 7. Conclusion

The proposed redescending estimator is more continuous, smooth, and linear in the middle, and, therefore, according to Winsor's principle, it satisfies almost all the properties of a good estimator in the presence of outliers. To assess the overall performance of the proposed estimator, a simulation study was conducted in different data generation scenarios, for example, normal distribution,  $t$ -distribution, and Laplace distribution along with different contamination scenarios in the response as well as explanatory variables. It was found from the simulation results that when the dataset is normal, the performance of the proposed estimator is almost the same as that of OLS in terms of total mean square error, but it outperforms the existing robust estimators. In case of contaminated datasets (that is, when outliers are present in response as well as explanatory variable), the proposed estimator was found to be the best among all the estimators for all sample size " $n$ " and number of predictors " $p$ " with different level of contamination in the datasets. Similarly, it can also be concluded from the simulation results that the performance of the proposed estimator has improved as the sample size increases while maintaining its robustness higher as compared to the competing estimators. Moreover, the efficiency of the proposed estimator was found to be higher than the rest of the estimators. The same results were also found in the real data applications of the proposed estimator as compared to the existing estimator. Finally, it was concluded from the simulation study, as well as real data applications, that all robust estimators perform well for all levels of contamination in response, but their performance gets poor in case of leverage points. However, the proposed estimator was found to perform better than all the redescending estimators considered in this study.

## Data Availability

To replicate the results of this paper, the data are available from the corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics—Theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977.
- [2] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, Hoboken, NJ, USA, 1987.
- [3] F. Y. Edgeworth, "On observations relating to several quantities," *Hermathena*, vol. 6, no. 13, pp. 279–285, 1887.
- [4] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [5] R. Koenker and V. D'Orey, "Remark AS R92: a remark on algorithm AS 229: computing dual regression quantiles and regression rank scores," *Applied Statistics*, vol. 43, no. 2, pp. 410–414, 1994.
- [6] D. M. Khan, A. Yaqoob, S. Zubair, M. A. Khan, Z. Ahmad, and O. A. Alamri, "Applications of robust regression techniques: an econometric approach," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6525079, 9 pages, 2021.
- [7] P. Rousseeuw and V. Yohai, "Robust regression by means of S-estimators," in *Robust and Nonlinear Time Series Analysis*, pp. 256–272, Springer US, Berlin, Germany, 1984.
- [8] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.
- [9] D. M. Khan, S. Ihtesham, A. Ali, U. Khalil, S. A. Khan, and S. Manzoor, "An efficient and high breakdown estimation procedure for nonlinear regression models," *Pakistan Journal of Statistics*, vol. 33, no. 3, 2017.
- [10] A. J. Stromberg, O. Hössjer, and D. M. Hawkins, "The least trimmed differences regression estimator and alternatives," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 853–864, 2000.
- [11] C. W. Coakley and T. P. Hettmansperger, "A bounded influence, high breakdown, efficient regression estimator," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 872–880, 1993.
- [12] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [13] F. R. Hampel, "Robust estimation: a condensed partial survey," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 27, no. 2, pp. 87–104, 1973.
- [14] D. M. Khan, S. A. Khan, K. U. Alamgir, and A. Ali, "A comparative study of three improved robust regression procedures," *Pakistan Journal of Statistics*, vol. 32, no. 6, 2016.
- [15] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, NJ, USA, 1972.
- [16] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, no. 2, pp. 147–185, 1974.
- [17] S. Van Aelst, G. Willems, and R. H. Zamar, "Robust and efficient estimation of the residual scale in linear regression," *Journal of Multivariate Analysis*, vol. 116, pp. 278–296, 2013.
- [18] K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, Wiley, New York, NY, USA, pp. 491–500, 1965.