Pak.j.stat.oper.res. Vol.17 No. 3 2021 pp 997-1014 DOI: http://dx.doi.org/10.18187/pjsor.v17i4.3546

Pakistan Journal of Statistics and Operation Research

A Redescending M-Estimator for Detection and Deletion of Outliers in Regression Analysis

Stella Ebele Anekwe[1*], Sidney Iheanyi Onyeagu[2]

*Corresponding Author

1. Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria, stellaanekwe@gmail.com
2. Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria, si.onyeagu@unizik.edu.ng

Abstract

Outliers in a statistical analysis strongly affect the performance of the ordinary least squares, such outliers need to be detected and extreme outliers deleted. This paper is aimed at proposing a redescending M-estimator, which is more efficient and robust, compared to other existing redescending M-estimators. The proposed method is applied to real life data to verify its effectiveness in detecting and deleting of outliers. The Monte Carlo simulation method is also used to investigate the performance of the newly proposed method. The results from the real life data and the Monte Carlo simulation method show that the proposed method is effective in the detection and deletion of extreme outliers compared to other existing redescending M-estimators.

Key Words: Efficiency; M-estimators; Redescending M- estimators; Outliers; Robustness.

Mathematical Subject Classification:

1. Introduction

The standard multiple regression model in matrix notation is given as

$$Y = X\beta + \varepsilon \tag{1}$$

where $Y = (y_1, y_2, \ldots, y_n)'$ is $n \times 1$ vector of n observations, X is $n \times k$ matrix of n observations on each of the k explanatory variables, $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a $n \times 1$ vector of random error components. According to Sokal and Rohlf (2012), Ordinary Least Squares (OLS) regression fit a line to bivariate data such that the (squared) vertical distance from each data point to the line is minimized across all data points.

OLS estimates are obtained by minimizing the sum of squared error (SSE) given as

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) \tag{2}$$

Some of the assumptions of Ordinary Least Squares are: $E(\varepsilon) = 0, E(\varepsilon\varepsilon') = \sigma^2 1_n, X$ is a non–stochastic matrix and $\varepsilon \sim N(0, \sigma^2 1_n)$.

In the context of outlier detection, many researchers developed various methods. Aggarwal and Yu (2001) discovered a new technique for detecting outliers associated to very high dimensional data sets. Nguyena and Welch (2010) studied outlier detection and proposed a new trimmed square approximation. Hadi and Simonoff (1993) introduced two test procedures for the detection of multiple outliers in a linear model. Maronna et al. (2006) suggested a more reliable procedure, the graphical procedure, which uses normal Q-Q plots for the detection of outliers. Armin (2008) proposed a method for detection of outliers using Dixon's test statistics. Carling (2000) introduced the median rule for identification of outliers through studying the relationship between target outlier percentage and Generalized Lambda Distributions (GLDs). Manoj and Kaliyaperumal (2013) compared the performances of five outlier detection methods (Grubbs test, Dixon test, Hampel, Quartile method and Generalized ESD). Zhang et al. (2015) proposed an enhanced Monte Carlo outlier detection method by establishing cross-prediction models based on normal samples and analysing the distribution of prediction errors for dubious samples. Other authors who studied detection of outliers include: Tukey (1977), Atkinson (1994), Becker and Gather (1999) and Carling (2000).

Robust regression is use for improving the results of the least square estimates in the presence of outliers. Some methods of robust regression include those of: Huber (1964) who discovered M-estimators which are the generalization of the Maximum Likelihood Estimators (MLE). Rousseeuw (1982) discovered the Least Median of Squares estimators (LMS). Rousseeuw (1983) also proposed the Least Trimmed Squares (LTS) estimators. Some Redescending M-estimators for detection and deletion of outliers are also given in: Andrew et al. (1972), Beaton and Tukey (1974), Hampel et al. (1986) and Alamgir et al. (2013).

This article is aimed at proposing a Redescending M-estimator (that will be differentiable and continuous) which includes the objective function (ρ-function), the corresponding influence function (ψ-function) and weight function (w-function). Secondly, to compare the proposed Redescending M-estimator with some existing Mestimators and Redescending M-estimators in terms of efficiency and robustness

2. Review of M-estimators

M-estimators are robust estimators developed to give less weight to the observations that are outliers. It was introduced by Huber (1964) and can be regarded as a generalization of Maximum Likelihood Estimation; hence the "M". The Maximum Likelihood Estimator (MLE) is a method of estimating the parameters of a model by maximizing the model's likelihood function.

Considering the linear model in equation (1), the fitted model is

$$\hat{y}_i = \beta_0 + \beta_i x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} = X_i' \hat{\beta} \quad (3)$$

where $\hat{y}_i$ is the vector of predicted or estimated value of y and p is the number of explanatory variables.

The residuals, r_i,are given by

$$\begin{aligned} r_i &= y_i - \hat{y}_i \\ &= y_i - X_i' \hat{\beta} \end{aligned} \quad (4)$$

To obtain the parameter β in MLE, we minimize the negative log function given as

$$\hat{\beta}_{MLE} = minimize \ \sum_{i=1}^{n} [-log\ f\ (y_i\ ;\beta)] \quad (5)$$

while Ordinary Least Squares (OLS) minimizes the residual sum of squares, that is,

$$\hat{\beta}_{OLS} = minimize \ \sum_{i=1}^{n} r_i^2 \quad (6)$$

Replacing the squared error term in equation (6) by $\rho(\mathrm{r})$, M-estimator is given as

$$\hat{\beta}_{M-estimator} = minimize \ \sum_{i=1}^{n} \rho(\mathrm{r}) \quad (7)$$

where $\rho(r)$ is the objective function of an M-estimator.

Standardizing the residuals, $\mathrm{r_i}$, equation (7) can also be written as

$$\hat{\beta}_{M-estimator} = minimize \ \sum_{i=1}^{n} \rho\left(\frac{\mathrm{r_i}}{\hat{\sigma}}\right) \quad (8)$$

where $\hat{\sigma}$ is the scale parameter given as

$$\hat{\sigma} = \frac{MAD}{0.674} \quad (9)$$

and MAD is the Median Absolute Deviation given as

$$\mathrm{MAD} = median(|r_i - median(r_i)|) \quad (10)$$

Since standard deviation is not resistant to outliers, the Median Absolute Deviation (MAD) is used as a measure of spread in robust regression.

The Objective function of an M-estimator defines the probability distribution of the M-estimator. The properties of the objective function include;$\rho(0) = 0$, $\rho(r_i) \geq 0$, $\rho(r_i) = \rho(-r_i)$, $\rho(r_i) \leq \rho(r_j)$ for 0< $r_i < r_j$ and lastly, $\rho(r)$ should be continuous and diffentiable.

The Influence function describes the sensitivity of the overall estimate on the outlying data. Hampel (1974) disclosed that the robustness of an estimator is measured by its influence function. The derivative of equation (8) with respect to the regression coefficient β gives rise to the influence function (ψ-function), that is,

$$\frac{d[\sum_{i=1}^{n} \rho(\mathrm{r_i})]}{d\beta} = \Sigma\psi(r)\, X_i$$

$$\psi(r) = \Sigma\psi\left(\frac{\mathrm{y_i} - \mathrm{X_i'}\hat{\beta}}{\hat{\sigma}}\right) X_i \tag{11}$$

where $\psi(r)$ is defined as the influence function and $\hat{\sigma}$ is the scale parameter.

Draper and Smith (1998) defined the weighted function, w_i, as

$$w_i = \frac{\psi\left(\frac{\mathrm{y_i} - \mathrm{X_i'}\hat{\beta}}{\hat{\sigma}}\right)}{\left(\frac{\mathrm{y_i} - \mathrm{X_i'}\hat{\beta}}{\hat{\sigma}}\right)} \tag{12}$$

2.1 Huber M-estimator

Huber (1964) proposed the Huber M-estimator and its influence function, $\psi(r)$, is

$$\psi(r) = \begin{cases} -c & ; r < -c \\ r & ; -c \leq r \leq c \\ c & ; r > c \end{cases} \tag{13}$$

where c is arbitrary value known as tuning constant and r are the residuals scaled over Median Absolute Deviation (MAD). Its influence function is non-decreasing function with a tuning constant c = 1.345 which yields 95% efficiency on a normal distribution (the tuning constant c, determines the degree of robustness in M-estimators). Huber estimator is not robust when the outliers present in the data are in x-direction (leverage points).

Redescending M-estimators are estimators with ψ-functions redescending to zero. They are those M-estimators that reject extreme outliers. Some of these estimators discussed in the literature are:

2.2 Hampel M-estimator

Hampel's three–part redescending M-estimator was proposed by Hampel et al. (1986) in the Princeton Robustness study. It has three tuning constants a, b and c. Its ψ-function is given as

$$\psi(r) = \begin{cases} r & ; \text{if } |r| \leq a \\ a \text{ sign } (r) & ; \text{ if } a < |r| \leq b \\ \frac{(c-|r|)}{(c-b)} a \text{ sign } (r) & ; \text{ if } b < |r| \leq c \\ 0 & ; \text{ if } |r| > c \end{cases} \tag{14}$$

where a, b, c are positive constants and $0 < a \leq b < c < \infty$ and r are the residuals scaled over Median Absolute Deviation MAD. The drawback of this estimator is that, its influence function is non-differentiable.

2.3 Tukey's Biweight M-estimator

Beaten and Tukey (1974) proposed Tukey's biweight M-estimator and its ψ-function is given as

$$\psi(r) = \begin{cases} r\{1 - (\frac{r}{c})^2\}^2 & ; |r| \le c \\ 0 & ; \text{otherwise} \end{cases} \quad (15)$$

where c is arbitrary value known as tuning constant and r are the residuals scaled over MAD. For Tukey's biweight, c = 4.685 gives 95% efficiency on normal distribution. The performance of Tukey's biweight estimator was encouraging, that is, the influence function is differentiable and smooth when compared to the methods proposed by Huber (1964) and Hampel et al. (1986).

2.4 Alarm M-estimator

Alamgir et al. (2013) proposed the Alarm's Redescending M-estimator for robust regression and outlier detection. Its ψ -function is given as

$$\psi(r) = \begin{cases} \frac{16r\,(e^{-(r/c)^2}}{(1+e^{-(r/c)^4}} & ; |r| \le c \\ 0 & ; |r| > c \end{cases} \quad (16)$$

where c is the tuning constant and r are the residuals scaled over MAD.

The Alarm estimator was based on the modified tangent hyperbolic (tan h) type weight function. Its Mean Square Errors (MSE) are the smallest when compared with that of Huber (1964), Beaton and Tukey (1974) and Hampel et al. (1986), yielding efficient results. For Alarm M-estimator, c = 3 gives approximately 95% efficiency at normal distribution.

3. Methodology

3.1 The Proposed Estimator

In this section, a new Influence function, $\psi(r)$, is proposed (based on modified Tukey's biweight ψ-function) with 95% efficiency at normal distribution, we introduced a function $g(r)$

$$g(r) = \left(1 + \left(\frac{r}{c}\right)^2\right)^2 \quad (17)$$

$g(r)$ is a smooth and differentiable function for all r, r are the residuals scaled over Median Absolute Deviation (MAD) and c is the tuning constant.

In addition, we multiply the function, $g(r)$, by the Tukey's biweight ψ-function resulting in the proposed influence function, $\psi(r)$, given as

$$\psi(r) = \begin{cases} r\left(1 - \left(\frac{r}{c}\right)^2\right)^2 \left(1 + \left(\frac{r}{c}\right)^2\right)^2; & |r| < c \\ 0 & ; \quad |r| \ge c \end{cases} \quad (18)$$

where c is the tuning constant for the ith observation and the variable r are the residuals scaled over MAD.

By integrating the $\psi(r)$ with respect to r, we obtain the corresponding objective function, $\rho(r)$, given as

$$\rho(r) = \begin{cases} \frac{r^6}{c^4} + \frac{r^{10}}{2c^8} - \frac{2r^6}{c^4} + \frac{r^2}{2} - \frac{2r^6(3r^4-5c^4)}{15c^8}; & |r| \le c \\ \frac{4c^2}{15} \quad ; & |r| > c \end{cases} \quad (19)$$

where c is the tuning constant for the ith observation and the variable, r, are the residuals scaled over MAD.

Derivation of equation (19)

$$\rho(r) = \int \psi(r)dr \quad (20)$$

where $\psi(r)$ and $\rho(r)$ are influence and objective functions respectively, r are the residuals scaled over MAD (Median Absolute Deviation) while c is the tuning constant.

Given: $\psi(r) = r\left(1-\left(\frac{r}{c}\right)^2\right)^2\left(1+\left(\frac{r}{c}\right)^2\right)^2$

Using the identity;

$a^2 - b^2 = (a+b)(a-b)$

Squaring both sides;

$(a^2 - b^2)^2 = \{(a+b)(a-b)\}^2$

Where a = 1 and b = $\left(\frac{r}{c}\right)^2$

$\Rightarrow \psi(r) = r\left(1-\left(\frac{r}{c}\right)^4\right)^2$

and

$$\rho(r) = \int \psi(r)dr$$

$$= \int r\left(1-\left(\frac{r}{c}\right)^4\right)^2 dr$$

Using integration by parts;

Let

u = $\left(1-\left(\frac{r}{c}\right)^4\right)^2$, du = - $\frac{8r^3(c^4-r^4)}{c^8}\,dr$

and

$dv = rdr, \quad v = \frac{r^2}{2}$

$$\int\left(1-\left(\frac{r}{c}\right)^4\right)^2 rdr = \left(1-\left(\frac{r}{c}\right)^4\right)^2\left(\frac{r^2}{2}\right) + \int\left(\frac{r^2}{2}\right)\left(\frac{-8r^3(c^4-r^4)}{c^8}\right)dr$$

$$= \frac{r^2(c^4-r^4)}{2c^8} + 2\,\frac{(5c^4r^6)-3r^{10}}{15c^8}$$

$$= \frac{r^{10}}{2c^8} - \frac{r^6}{c^4} + \frac{r^2}{2} - \frac{2r^6(3r^4-5c^4)}{15c^8} \quad (21)$$

For the second part of $\rho(r)$, we use the same argument in Beaton and Tukey (1974) by substituting r for c in equation (21).

$$\int\left(1-\left(\frac{r}{c}\right)^4\right)^2 rdr = \frac{c^{10}}{2c^8} - \frac{c^6}{c^4} + \frac{c^2}{2} - \frac{2c^6(3c^4-5c^4)}{15c^8}$$

$$= \frac{4c^2}{15} \quad (22)$$

The proposed $\rho(r)$ satisfies the standard properties of the objective function of an M-estimator.

Dividing the proposed $\psi(r)$ by r gives the weight function, $w(r)$, as follows:

$$w(r)=\begin{cases}\left(1-\left(\frac{r}{c}\right)^2\right)^2\left(1+\left(\frac{r}{c}\right)^2\right)^2; & |r|<c\\ 0 & ;\ |r|\geq c\end{cases} \tag{23}$$

Graphs of the proposed objective, influence and weight functions are shown below:

Figure 1 : Graph of the Proposed Influence Function

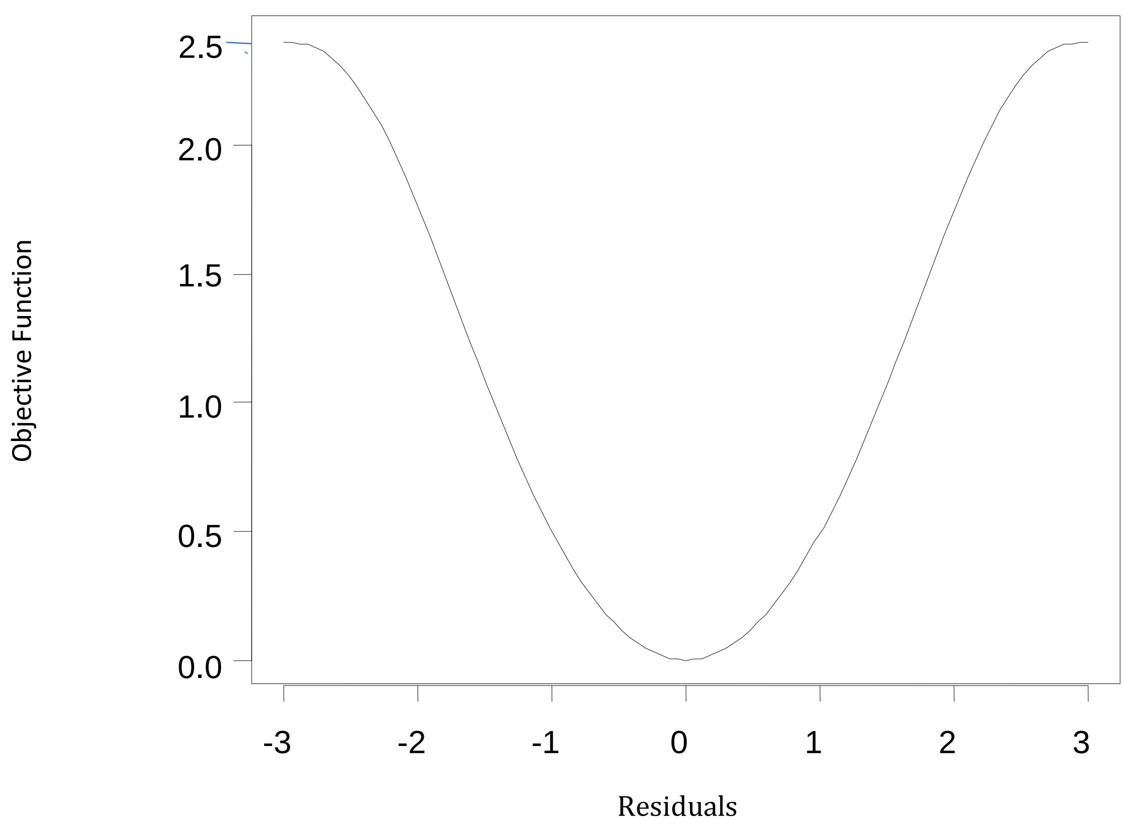
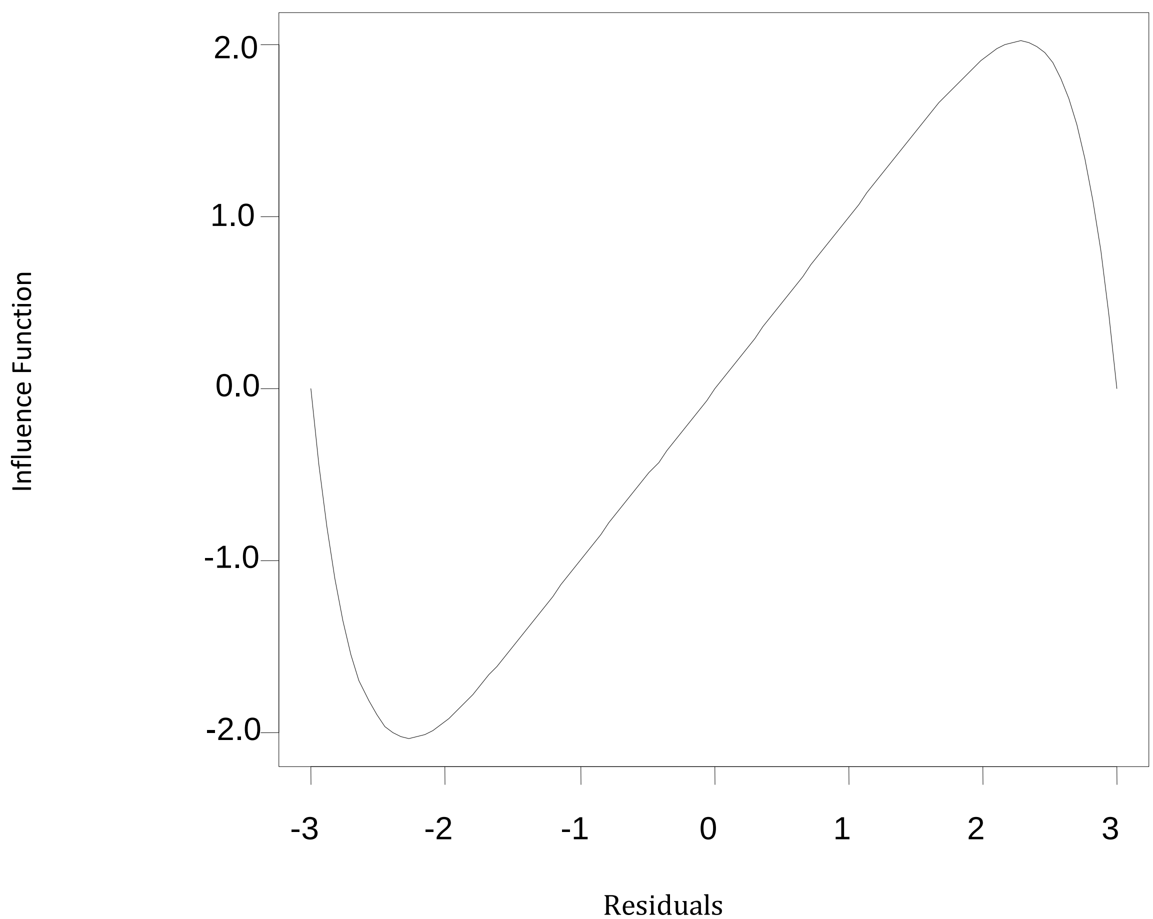


Figure 1 shows that the values of the objective function are non- negative. Secondly, the graph indicates that the objective function is symmetric, differentiable and a continuous function.

Pak.j.stat.oper.res. Vol.17 No. 3 2021 pp 997-1014 DOI: http://dx.doi.org/10.18187/pjsor.v17i4.3546

Figure 2 : Graph of the Proposed Influence Function



In Figure 2, the proposed estimator redescends to zero by assigning zero influence to extreme outliers thereby rejecting them.

Figure 3: Graph of the Proposed Weight Function

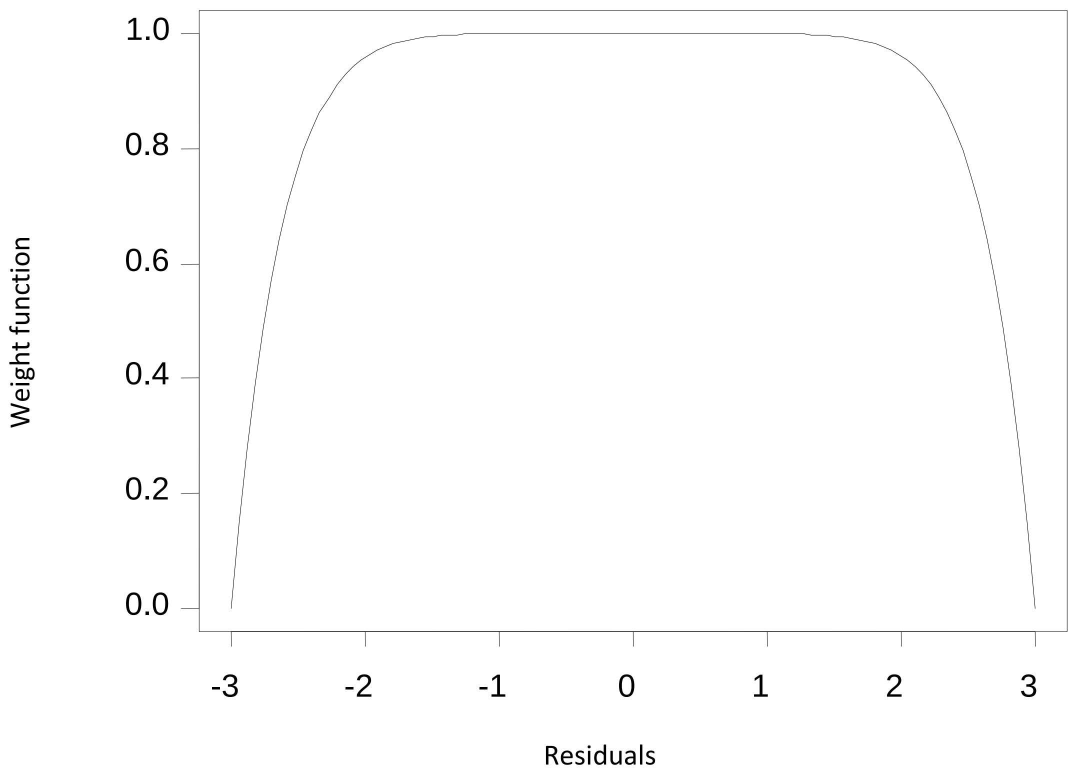


Figure 3 indicates that, the good observations are assigned bigger weights while outliers have the smaller weights. In addition, extreme outliers are assigned zero weights, which implies that they are eliminated from the dataset.

3.2 Simulation Design

Monte Carlo simulation method is used to generate random data from different probability distributions. The purpose of the simulation study is to determine the extent our estimates differ from their true values (robustness). We took the true parameters to be 1, 2, and 5 for β_0, β_1, and β_2 respectively. Each simulation case was replicated $M = 1000$ times. The estimates of each estimator were calculated in each of the iteration and the mean of the M replicated estimates given by

$$\hat{\beta}_j = \frac{\sum_{i=1}^{M} \hat{\beta}_{ji}}{M} \qquad for\ j = 0, 1, 2, \cdots, p \tag{24}$$

was recorded for each estimator.

For comparison, the parameters estimates of the Mean Square Error (MSE) and the absolute bias (BIAS) of the OLS, Huber (1964), Hampel et al. (1986), Beaton and Tukey (1974) and Alamgir et al. (2013) alongside the proposed Redescending M-estimators are computed.

Robustness of an estimator is measured using absolute bias given as

$$AbsBias(\hat{\beta}_j) = |\beta_j - \hat{\beta}_j| \qquad for\ j = 0, 1, 2, \cdots, p \tag{25}$$

Efficiency of an estimator is measured using the MSE (mean square error) defined as

$$MSE(\hat{\beta}_j) = \frac{\sum_{i=1}^{M} (\beta_j - \hat{\beta}_{ji})^2}{M} \qquad for\ j = 0, 1, 2, \cdots, p \tag{26}$$

and the variance of the estimator is defined as

$$Var(\hat{\beta}_j) = MSE(\hat{\beta}_j) - \left(Bias(\hat{\beta}_j)\right)^2 \qquad for\ j = 0, 1, 2, \cdots, p \tag{27}$$

The estimator with lowest MSE is the most efficient; the smaller the MSE the more efficient is the estimator.
Simulated data were generated (including percentage mixtures of contaminated and uncontaminated data) in both simple and multiple regressions, using two sample sizes, $n = 20$ and 200.
The percentages of outliers considered in the simulation study were as follows:
For the x- axis, we chose contamination at 20%, and 30%.
For the y- axis, we chose contamination at 20%, 30% and 40%.

3.3 Algorithm for the Simulation Studies

1. Compute the initial estimates using the Least Median Squares (LMS).

2. Obtain the corresponding residuals from our initial estimates.

3. Compute the corresponding weights based on the proposed weight function.

4. Calculate the new estimates of the regression coefficients using weighted least squares.

5. Repeat step 2 to 5 until convergence.

3.4 Choice of the Tuning Constant c

A simulation study was done to determine the choice of the tuning constant. A tuning constant of c = 3 gives the best result for estimating the true parameters, detecting and delecting of outliers.

4. Results

The Simulated results for the proposed estimator and that of OLS, Huber, Hampel, Bisquare (Biweight) and Alarm estimators are discussed as follows:

4.1 Discussion of Simulated results for data without outlier

Tables 1 and 7 present detailed stimulated results for uncontaminated data from simple and multiple regressions respectively. The OLS having the least MSE, outperformed the Huber, Hampel, Bisquare, Alarm and the proposed estimators, that is, the most efficient estimator. Similarly, the OLS, Huber, Hampel, Bisquare, Alarms and the proposed estimators are all closer to their true parameters estimates (robustness).

4.2 Discussion of stimulated results for data with outliers in the x- direction (leverage points)

Table 2 presents the stimulated result for 20% outliers in the x-direction in a simple regression model. The proposed and Alarm estimators are more efficient and robust compared to OLS, Huber, Hampel, and Bisquare estimators. Moreover, outliers strongly affect the slopes of OLS, Huber, Hampel, and Bisquare estimators.

With the increase of the percentage of outliers in the x-direction in a simple regression model to 30% as shown in table 3, all the estimators performed badly for both MSE and BIAS. For comparison, the proposed and Alarms estimators are more efficient and robust compared to OLS, Huber, Hampel, and Bisquare estimators.

Table 8 presents the result for 20% outliers in the x-direction in a multiple regression analysis. The proposed, Hampel and Alarm estimators are more efficient and robust compared to OLS, Huber, and Bisquare estimators. Moreover, outliers strongly affect the slopes of the OLS, Huber, and Bisquare estimators.

Based on data generated from 30% outliers in x- direction in a multiple regression, shown in table 9, the result indicates that the Alarm estimator is the most efficient having the smallest MSE among others while Hampel estimator is the second. The third most efficient is the proposed estimator followed by the remaining three estimators (OLS, Huber, and Bisquare estimators). Furthermore, outliers strongly affect the slopes of all the estimators (The proposed, Hampel, Alarm, OLS, Huber, and Bisquare estimators). All the estimators performed badly in this category.

4.3 Discussion of stimulated results for data with outliers at the response, that is, in the y-direction

Furthermore, the results from tables 10 and 4 (20% outliers in the y-direction for simple and multiple regressions respectively), indicate that the proposed estimator is the most efficient and robust, followed closely by the Bisquare estimator. The Alarm, Hampel and Huber estimators follow thereafter, while OLS is the least.

Tables 5 and 11, present the results for 30% outliers in the y-direction for simple and multiple regressions respectively. The proposed estimator competes favourably with the Bisquare estimator as the most efficient estimator but Bisquare estimator gets better with an increase in the sample size. The least efficient is the OLS followed by the Huber, then, the Hampel's estimator. Nevertheless, the proposed and Bisquare estimators are more robust compared to the Hampel, Alarm, OLS and Huber estimators.

Nevertheless, in tables 6 and 12 (results for 40% outliers in the y-direction for simple and multiple regressions respectively), the proposed and Bisquare estimators are also more efficient and robust compared to the Hampel, Alarm, OLS and Huber estimators.

Table 1: Simulated MSE and BIAS on Simple Regression for data with no outlier.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	0.0033	0.0020	0.0017	0.0031	0.0018	0.0031
	β_0	MSE	0.0509	0.0547	0.0565	0.0521	0.0540	0.0604
20	β_1	BIAS	0.0071	0.0063	0.0060	0.0072	0.0086	0.0173
	β_1	MSE	0.1526	0.1586	0.1630	0.1542	0.1585	0.1866
200	β_0	BIAS	0.0027	0.0036	0.0038	0.0032	0.0033	0.0040
	β_0	MSE	0.0049	0.0051	0.0051	0.0049	0.0049	0.0052
200	β_1	BIAS	0.0047	0.0050	0.0051	0.0052	0.0052	0.0062
	β_1	MSE	0.0147	0.0155	0.0156	0.0148	0.0149	0.0156

Table 2: Simulated MSE and BIAS on Simple Regression for 20% outliers in x-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	0.0107	0.0112	0.0117	0.0103	0.0006	0.0002
	β_0	MSE	0.1391	0.1585	0.1625	0.1436	0.1052	0.1120
20	β_1	BIAS	1.9871	1.9867	1.9860	1.9869	0.0702	0.6573
	β_1	MSE	3.9516	3.9504	3.9479	3.9511	1.6876	1.5633
200	β_0	BIAS	0.0377	0.0403	0.0400	0.0382	0.0024	0.0021
	β_0	MSE	0.0150	0.0169	0.0168	0.0153	0.0061	0.0062

200	β_1	BIAS	1.9841	1.9831	1.9831	1.9839	0.0285	0.0391
	β_1	MSE	3.9370	3.9329	3.9328	3.9360	0.0699	0.0932

Table 3: Simulated MSE and BIAS on Simple Regression for 30% outliers in x-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	0.0300	0.0302	0.0288	0.0286	0.0066	0.0090
	β_0	MSE	0.1563	0.1756	0.1840	0.1633	0.1462	0.1653
20	β_1	BIAS	1.9899	1.9898	1.9896	1.9899	1.2578	1.2324
	β_1	MSE	3.9620	3.9619	3.9613	3.9621	2.7002	2.6389
200	β_0	BIAS	0.0334	0.0353	0.0354	0.0339	0.0102	0.0150
	β_0	MSE	0.0168	0.0189	0.0188	0.0172	0.0140	0.0122
200	β_1	BIAS	1.9911	1.9906	1.9906	1.9910	0.6651	0.9326
	β_1	MSE	3.9646	3.9628	3.9628	3.9642	1.3739	1.9144

Table 4: Simulated MSE and BIAS on Simple Regression for 20% Outliers in y-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	1.8531	0.3405	0.0336	0.1299	0.0527	0.0315
	β_0	MSE	4.1726	0.2057	0.0752	0.1340	0.0819	0.0772
20	β_1	BIAS	0.8329	0.1530	0.0035	0.0436	0.0079	0.0020
	β_1	MSE	2.0266	0.2214	0.1813	0.2103	0.1942	0.1883
200	β_0	BIAS	2.0044	0.3597	0.0319	0.1083	0.0523	0.0250
	β_0	MSE	4.0958	0.1393	0.0084	0.0218	0.0110	0.0081
200	β_1	BIAS	0.1453	0.0483	0.0173	0.0259	0.0198	0.0160
	β_1	MSE	0.3063	0.0308	0.0218	0.0292	0.0245	0.0216

Table 5: Simulated MSE and BIAS on Simple Regression for 30% Outliers in y-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	2.9036	0.9170	0.1008	1.2486	0.1619	0.0860

	β_0	MSE	9.5809	1.2896	0.1452	3.6198	0.1700	0.1172
20	β_1	BIAS	0.3581	0.1170	0.0045	0.1599	0.0091	0.0027
	β_1	MSE	1.7931	0.2551	0.1970	0.4807	0.2225	0.2072
200	β_0	BIAS	3.0142	0.8234	0.0773	0.4167	0.1534	0.0743
	β_0	MSE	9.2003	0.7147	0.0164	0.3854	0.0387	0.0172
200	β_1	BIAS	0.2589	0.1076	0.0218	0.0702	0.0304	0.0216
	β_1	MSE	0.4556	0.0623	0.0304	0.0702	0.0434	0.0331

Table 6: Simulated MSE and BIAS on Simple Regression for 40% Outliers in y-axis**.**

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	3.7895	2.3770	1.2854	2.9863	0.7094	0.3102
	β_0	MSE	15.8511	7.5439	4.4558	11.8804	1.4118	0.4626
20	β_1	BIAS	1.3977	1.0138	0.6583	1.1816	0.3185	0.1331
	β_1	MSE	4.4199	2.1769	1.6292	3.2272	0.7103	0.3782
200	β_0	BIAS	4.0119	2.2093	0.2309	3.2384	0.5412	0.2731
	β_0	MSE	16.2480	5.1134	0.0851	10.9361	0.3561	0.1095
200	β_1	BIAS	0.0871	0.0731	0.0241	0.0852	0.0302	0.0268
	β_1	MSE	0.5570	0.2131	0.0606	0.4004	0.1229	0.0934

Table 7: Simulated MSE and BIAS on Multiple Regression for Data with no Outliers.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	0.0100	0.0100	0.0098	0.0096	0.0092	0.0151
	β_0	MSE	0.0520	0.0556	0.0576	0.0529	0.0549	0.0677
20	β_1	BIAS	0.0006	0.0021	0.0048	0.0006	0.0021	0.0041
	β_1	MSE	0.1537	0.1660	0.1771	0.1576	0.1670	0.2225
20	β_2	BIAS	0.0085	0.0105	0.0113	0.0093	0.0111	0.0130
	β_2	MSE	0.0462	0.0471	0.0499	0.0460	0.0487	0.0622

200	β_0	BIAS	0.0014	0.0011	0.0014	0.0016	0.0016	0.0016
	β_0	MSE	0.0050	0.0052	0.0053	0.0050	0.0050	0.0052
200	β_1	BIAS	0.0037	0.0027	0.0028	0.0034	0.0034	0.0031
	β_1	MSE	0.0144	0.0149	0.0150	0.0145	0.0145	0.0151
200	β_2	BIAS	0.0016	0.0007	0.0008	0.0013	0.0013	0.0009
	β_2	MSE	0.0033	0.0035	0.0035	0.0034	0.0034	0.0035

Table 8: Simulated MSE and BIAS on Multiple Regression for 20% Outliers in x-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	0.0412	0.0400	0.0377	0.0078	0.0095	0.0082
	β_0	MSE	0.1631	0.1814	0.1900	0.1150	0.1266	0.1484
20	β_1	BIAS	1.9871	1.9872	1.9871	0.8225	0.8270	0.8325
	β_1	MSE	3.9521	3.9528	3.9527	1.9240	1.9260	1.9626
20	β_2	BIAS	0.0206	0.0205	0.0197	0.0175	0.0174	0.0180
	β_2	MSE	0.1053	0.1145	0.1198	0.0778	0.0826	0.1033
200	β_0	BIAS	0.5143	0.5684	0.5684	0.1693	0.1349	0.1593
	β_0	MSE	0.0162	0.0182	0.0180	0.0073	0.0067	0.0068
200	β_1	BIAS	2.0000	2.0000	2.0000	0.9744	0.5278	0.4601
	β_1	MSE	3.9467	3.9433	3.9434	0.1702	0.1042	0.0961
200	β_2	BIAS	2.4397	0.0485	0.0623	0.1722	0.0896	0.0846
	β_2	MSE	0.0097	0.0106	0.0106	0.0049	0.0049	0.0049

Table 9: Simulated MSE and BIAS on Multiple Regression for 30% Outliers in x-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	0.0430	0.0461	0.0479	0.0364	0.0439	0.0471
	β_0	MSE	0.1773	0.1983	0.2099	0.1667	0.1809	0.2113
20	β_1	BIAS	1.9904	1.9903	1.9901	1.4550	1.4457	1.3603
	β_1	MSE	3.9642	3.9640	3.9636	3.0673	3.0473	2.8826
20	β_2	BIAS	0.0081	0.0082	0.0088	0.0079	0.0055	0.0098

	β_2	MSE	0.1017	0.1116	0.1168	0.0970	0.1056	0.1303
200	β_0	BIAS	0.0453	0.0480	0.0482	0.0265	0.0233	0.0299
	β_0	MSE	0.0183	0.0197	0.0198	0.0129	0.0122	0.0144
200	β_1	BIAS	1.9910	1.9895	1.9895	0.9721	0.8087	1.0685
	β_1	MSE	3.9609	3.9583	3.9584	1.9867	1.6560	2.1805
200	β_2	BIAS	0.0031	0.0030	0.0033	0.0008	0.0016	0.0023
	β_2	MSE	0.0087	0.0095	0.0095	0.0071	0.0071	0.0075

Table 10: Simulated MSE and BIAS on Multiple Regression for 20% Outliers in y-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	1.7370	0.3018	0.0149	0.0948	0.0365	0.0123
	β_0	MSE	3.7599	0.1805	0.0709	0.0985	0.0754	0.0758
20	β_1	BIAS	0.6647	0.0782	0.0109	0.0083	0.0084	0.0108
	β_1	MSE	2.1465	0.2271	0.2026	0.2269	0.0280	0.2201
20	β_2	BIAS	0.5552	0.1733	0.0425	0.0880	0.0507	0.0329
	β_2	MSE	1.4471	0.1305	0.0852	0.1316	0.1000	0.1000
200	β_0	BIAS	2.0080	0.3556	0.0279	0.1044	0.0486	0.0218
	β_0	MSE	4.1176	0.1366	0.0079	0.0202	0.0103	0.0077
200	β_1	BIAS	0.0188	0.0301	0.0026	0.0021	0.0071	0.0033
	β_1	MSE	0.3028	0.0281	0.0233	0.0282	0.0257	0.0238
200	β_2	BIAS	0.1287	0.0286	0.0034	0.0101	0.0059	0.0039
	β_2	MSE	0.0681	0.0063	0.0047	0.0058	0.0050	0.0047

Table 11: Simulated MSE and BIAS on Multiple Regression for 30% Outliers in y-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	2.6733	1.0070	0.2159	0.5721	0.1781	0.0764
	β_0	MSE	8.3023	1.6545	0.4679	1.2216	0.2081	0.1200

20	β_1	BIAS	0.0159	0.1784	0.0952	0.0754	0.0576	0.0345
	β_1	MSE	2.2758	0.4384	0.3291	0.4289	0.2970	0.2803
20	β_2	BIAS	1.0262	0.6365	0.2144	0.3277	0.1498	0.0804
	β_2	MSE	2.5571	0.8740	0.4153	0.5396	0.2389	0.1888
200	β_0	BIAS	3.0213	0.8221	0.0765	0.3375	0.1528	0.0729
	β_0	MSE	9.2560	0.7156	0.0157	0.1739	0.0381	0.0166
200	β_1	BIAS	0.2577	0.0871	0.0046	0.0304	0.0107	0.0057
	β_1	MSE	0.4520	0.0563	0.0314	0.0526	0.0409	0.0355
200	β_2	BIAS	0.1929	0.0689	0.0059	0.0282	0.0119	0.0044
	β_2	MSE	0.1234	0.0158	0.0068	0.0129	0.0091	0.0076

Table 12: Simulated MSE and BIAS on Multiple Regression for 40% Outliers in y-axis.

Sample Size	Beta	Criteria	OLS	Huber	Bisquare	Hampel	Alarm	Proposed
20	β_0	BIAS	3.9455	2.6372	1.8229	2.3304	0.9315	0.4042
	β_0	MSE	17.3861	9.3040	6.6230	9.5266	2.3911	0.8166
20	β_1	BIAS	1.6181	1.5173	1.4029	1.1184	0.5720	0.2760
	β_1	MSE	6.5166	5.7200	5.8235	4.4222	3.0762	1.7525
20	β_2	BIAS	0.6502	0.8344	0.8935	0.5206	0.3212	0.1689
	β_2	MSE	3.0760	2.8364	3.6828	2.3742	2.1760	1.4693
200	β_0	BIAS	4.0250	2.2330	0.2428	3.0746	0.5587	0.2715
	β_0	MSE	16.3720	5.2455	0.1187	10.3992	0.3834	0.1057
200	β_1	BIAS	0.0841	0.0420	0.0039	0.0646	0.0018	0.0060
	β_1	MSE	0.5241	0.2049	0.0617	0.3634	0.1239	0.0871
200	β_2	BIAS	0.3261	0.2437	0.0313	0.2937	0.0695	0.0340
	β_2	MSE	0.2172	0.0999	0.0148	0.1646	0.0297	0.0186

4.4 Data Analysis

In this section, we applied the proposed estimator to real-life data to verify its effectiveness in detecting and deleting of outliers. These datasets had been extensively used by other researchers in the area of robust regression.

Example 1: Telephone-Call Data (Simple Regression Case)

This is a real regression data with few outliers in y-direction. The data set is taken from Belgium Statistical Survey (Rousseeuw and Leroy, (1987)). The data contains 24 data points and 2 variables. The dependent variable is the number of telephone calls made from Belgium and the independent variable is the year.

Table 13: Estimates of the Model Parameters for Telephone Calls Data

Parameter	OLS	Huber	Hampel	Biweight	Alarm	Proposed
β_0	-260.059	-99.905	-52.389	-52.348	-52.454	-52.456
β_1	5.041	1.987	1.101	1.100	1.102	1.102
Data points used	24	24	18	17	17	17
Residual standard error	56.22	19.51	1.62	1.24	1.38	1.39

The summary of the results for estimates of the model parameters for Telephone Calls Data for the estimators are presented in Table13. The Biweight, Alarm, Hampel and the proposed estimators, performed better than OLS and Huber estimators. In addition, OLS and Huber estimators used all the data in the analysis while Alarm, Biweight and the proposed method detected and deleted 7 outliers in the robust fit.

Example 2: The Hawkins, Bradu, and Kass data (Multiple Regression Case)

The Hawkins et al. (1984) (Rousseeuw and Leroy, (1987)) generated artificial data for testing the performance of robust estimators. The data contains 75 observations in four dimensions (one response and three explanatory variables). The first 10 observations are bad leverage points, and the next four points are good leverage points (i.e., their xi are outlying, but the corresponding yi fit the model quite well).

Table 14: Estimates of the model parameters for Hawkins, Bradu and Kass data.

Parameter	OLS	Huber	Hampel	Biweight	Alarm	Proposed
β_0	-0.388	-0.776	-0.181	-0.946	-0.181	-0.181
β_1	0.239	0.167	0.081	0.145	0.082	0.081
β_2	-0.335	0.007	0.040	0.197	0.040	0.040
β_3	0.383	0.274	-0.052	0.180	-0.052	-0.052
Data points used	75	75	65	71	65	65
Residual standard error	2.25	1.13	0.77	0.63	0.56	0.56

The summary of the results for estimates of the model parameters for Hawkins, Bradu and Kass data for the estimators are presented in Table 14. With smaller Residual Standard Error (RSE), the Alarm, Hampel, Biweight and the proposed estimators, performed better than OLS and Huber estimators. In addition, OLS and Huber used all the data in the analysis while Alarm, Hampel and the proposed method detected and deleted 10 outliers in the robust fit. The Biweight estimator detected and deleted 4 outliers in the analysis.

5. Conclusion

A redescending M-estimator was proposed and the graphs of its objective, influence and weight functions satisfied the various properties of these functions. Simulation studies were done to ascertain the effectiveness of the proposed redescending M-estimator and for comparison with other existing methods. Mean square error (MSE) and BIAS were used for comparison under two different sample sizes.

From the Stimulated results, it was obvious that Ordinary least squares estimator outperformed other estimators in an uncontaminated data (clean data). On the other hand, all the estimators performed very well when outliers are in the

y-direction but the proposed estimator tops the list as the most efficient and robust estimator while the Biweight, Alarm and Hampel estimators followed closely. Consequently, when outliers are in the leverage points, the proposed and Alarm estimators take the lead as the most efficient and robust estimators among others.

In addition, robust regression analysis was fitted using the Telephone call data and the Hawkins, Bradu and Kass data to illustrate the ability of the proposed estimator to detect and delete outliers and to compare with the existing ones. The results from the two robust fits showed that the proposed method can successively detect and delete outliers and for comparison, the proposed estimator alongside the Alarm, Hampel and Biweight (only when outliers are in the response) estimators showed great resistance to outliers.

6. Future Research

This work can be extended in future to handle outliers effectively on x-axis with higher percentages, that is, 30% and 40%.

References

1. Aggarwal, C. & Yu, P. (2001). Outlier Detection for High Dimensional Data. In Proceedings of the ACM SIGMOD International Conference on Management of Data. ACM Press, 37-46.
2. Alamgir, A. A., Khan, S.A, Khan, D.M. & Khalil, U. (2013). A New Efficient Redescending M-estimator, Alamgir Redesending M-estimator. Research Journal of Recent Sciences, 2(8). 79 - 91.
3. Andrew, D.F., Bickel, F.R. Hampel, P.J. Huber, Tukey J.W. & Rogers W.H. (1972). Robust Estimates of Location Survey and Advances, Princeton University Press, Princeton, NJ.
4. Atkinson, A. (1994). Fast Very Robust Methods for the Detection of Multiple Outliers, Journal of American Statistical Association, 89, 1329-1339.
5. Armin, B. (2008). One-sided and two sided critical values for Dixon's Outlier Test for sample sizes up to n=30,Economic Quality Control, Vol. 23, No 1, 5-13.
6. Beaton, A.E. & Tukey, J.W. (1974). The Fitting of Power Series, Meaning Polynomials, Illustrated on Band Spectroscopic Data. Tecnometrics, 16, (2), 147-185.
7. Becker, C. & Gather, U. (1999). The Masking Breakdown Point of Multivariate Outlier Identification Rule, Journal of the American Statistical Association, 94, 949-955.
8. Carling, K. (2000). Resistant Outlier Rules and Non-Gaussian Case, Computational Statistics and Analysis, Vol. 33, 249 – 258.
9. Draper, N.R. & Smith H. (1998). Applied Regression Analysis, Third Edition, John Wiley and Sons, New York.
10. Hadi, A.S. & Simonoff J.S. (1993). Procedures for Identification of Multiple Outliers in Linear Models, Journal of the American Statistical Association, 88(424), 1264- 1272.
11. Hampel, F.R. (1974). The Influence Curve and Its Role in Robust Estimation, Journal of the American Statistical Association, Vol. 69, No 346, pp. 383-393.
12. Hampel, F.R. (1997). Some Additional Notes on the 'Princeton Robustness Year,' In: The Practice of Data Analysis: Essays in Honour of Tukey J.W. eds Brillinger, D.R., and Ferholz, L.T. Princeton: Princeton University press, 133-153.
13. Hampel, F.R., Ronchetti, E.M. Rousseeuw, P.J. & Stahel, W.A. (1986). Robust Statistics. The Approach Based on Influence Functions. New York: John Wiley.
14. Hawkins, D.M., Bradu, D. & Kass, G. V. (1984). Location of several outliers in multiple regression data using elemental sets. Technometrics 26, 197-208.
15. Huber, P.J. (1964). Robust Estimation of Location Parameter. The Annals of Mathematical Statistics, 35, 73 – 101.
16. Manoj, K. & Kaliyaperumal, S.K. (2013). Comparison of Methods for Detecting Outliers, International Journal of Scientific and Engineering Research. Vol 4. Issue 9.
17. Maronna, R.A., Martin, R.A. & Yohai, V.J. (2006). Robust Statistics. Theory and Methods, Wiley, New York.
18. Nguyena T.D. & Welch, R. (2010). Outlier Detection and Least Trimmed Squares Approximation using Semi-definite Programming, Comput Stat Data, 54: 3212-3226.
19. Rousseeuw, P.J. (1982). Least Median of Squares Regression, Research Report No. 178, Centre for Statistics and Operations Research, VUB Brussels.

20. Rousseeuw, P.J. (1983). Multivariate Estimation with High Breakdown point, Research Report No. 192, Centre for Statistics and Operations Research, VUB Brussels.
21. Rousseeuw, P.J. & Leroy, A.M. (1987). Robust Regression and Outlier Detection, Wiley-Interscience, New York.
22. Rousseeuw, P.J. & Yohai, V. (1984). "Robust Regression by Means of S-Estimators", in Robust and Nonlinear Time Series, edited by J. Franke, W. Hardle, and R.D. Martin, Lecture Notes in Statistics 26, Springer Verlag, New York, 256-274.
23. Sokal, R.R & Rohlf, F. J. (2012). Biometry, 4th Edition. W.H. Freeman and Co, New York.
24. Tukey, J. W. (1977), Exploratory Data Analysis. Reading, MA: Addison–Wesley.
25. Zhang, L., Li, P., Mao, J., Ma, F., Ding, X. & Zhang, Q. (2015). An Enhanced Monte Carlo Outlier Detection Method, J Comput Chem. (2015) 36(25), 1902-6. DOI: 10.1002/jcc.24026. Epub 2015 Jul 31.