Robust Regression Methods / a Comparison Study

[1]Maha Issa Ismail; [2]Huda Abdullah Rasheed
[1,2] Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, Iraq
mmaahhaa938@gmail.com; hudamath@uomustansiriyah.edu.iq

***Article History**: Received: 16,July 2021; Accepted: 13 August 2021; Published online: 4 September. 2021.*

Abstract
The ordinary least squares method (OLS) is one of the most common methods for estimating the coefficients of linear regression models. However, it is sensitive and not robust against the existence of outliers. Therefore, several robust estimation methods have been used and then represented by M-estimation using different objective functions. In this paper, a number of alternative robust methods have been suggested that represented by using Gastwirth's location estimator instead of the mean in OLS and instead of the median in different M-estimation methods. In addition to repeating the Hubers' M-estimation method (first method) until converged results are reached. A Monte-Carlo simulation study was employed to evaluate the performance of different estimation methods depending on the MSE of regression coefficients.
Keywords: Robust regression, Outlier, M-estimation, Gastwirth's location estimator.

1. Introduction

The analysis of the linear regression models is one of the important topics in statistics, which is widely used in different applied studies. It is used for describing the relationship between the response variable and one or more independent variables. The purpose is to predict the response variable, which is used for planning and making-decisions purposes. The least squares method is the most widely used and accurate method for estimating regression models because its estimators have the best linear unbiased estimator BLUE properties. The OLS estimation method required. However, the OLS is not robust against the departure of the normality assumption of error term. Therefore, a number of some robust regression methods and suggested robust methods have been compared based on MSE's of the regression coefficients. The linear regression model is defined as:

$$Y = X\underline{\beta} + \underline{\in} \qquad (1)$$

where,

Y: The vector of dependent variable in order (n×1).

X: The matrix of one or more explanatory variables in order (n×(p+1)).

β: The vector of regression coefficients in order ((p+1),1).

$\underline{\varepsilon}$: The vector of the random errors in order (n×1).

P: The number of explanatory variables.

2. Estimation methods of regression coefficients

2.1The Ordinary least squares method (OLS)

The Ordinary least squares method (OLS) was proposed by Karl Gauss in 1794. It is one of the most important estimation methods that is widely used for estimating linear models. It has the best linear unbiased estimation (BLUE) property. The OLS method aims to minimize $\sum_{i=1}^{n} e_i^2$ where e_i the value of the residual of ith observation that defined as:

$e_i = y_i - \hat{y}_i$; i = 1, 2, …, n

In general, the OLS estimators for regression coefficients can be obtained by using the following equation:

$$\underline{\hat{\beta}} = \left(X'X\right)^{-1}X'\underline{Y} \qquad (2)$$

As a special case, the simple linear regression model that is given by:

$$Y = \beta_0 + X\beta_1 + \in \qquad (3)$$

It can be estimated by using the OLS method as follows:

$$\hat{\beta}_1 = \frac{\sum yx_i - \frac{\sum y_i \sum x_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (4)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (5)$$

where,

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ and $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

The OLS method required the data set that satisfying number of assumptions, represented by:

1. Normally distributed errors, $\varepsilon_i \sim N(0, \sigma^2)$
2. $cov\left(\varepsilon_i, \varepsilon_j\right) = E\left(\varepsilon_i, \varepsilon_j\right) = 0$, $\forall i \neq j$; i, j $= 1, 2, \dots$ n
3. The explanatory variables ($X_1, X_2, \dots, X_i$; $i = 1, 2, \dots$ p) are uncorrelated with each other, i.e, $cov\left(X_i, X_j\right) = 0$; i, j = 1, 2, …,p

where, p is the number of independent random variables.

4. The error term ε_i is uncorrelated with all explanatory variables, i.e,

$cov\left(X_i, \varepsilon_i\right) = 0$, $i \neq j$

However, the OLS is not robust against the departure of its assumptions.

3. Robust regression

The concept of robustness has been introduced at first by Box in 1953. There are several definitions of robustness; perhaps the most important definition of robustness is by Huber which states that, robustness is a resistance (insensitivity) to the small departures from the assumptions of the model. [1]

The main objective of robustness in estimating the regression models is to find a robust regression method when the data set violates one or more of the assumptions of OLS method. One of the most important assumptions is the normality assumption for the error term that can be violated by the existence of outliers.

In this paper, a comparison was made between the OLS and M-estimation method with different objective functions, in addition to some of the proposed methods to reach the most robust method with the existence of outliers in the error term with different percentages.

2.2 M-estimation method

The M-estimation is one of the most common robust regression estimation methods. It is regarded as a generalization of the maximum likelihood estimates that maximize the likelihood function $L(x_1, x_2, \dots, x_n; \theta)$.
where,

$$L(x_1, x_2, \dots, x_n; \theta) = \pi_{i=1}^{n} f(x_i; \theta)$$

The base of M-estimation is obtaining an estimator that minimize the residuals weighted function $\rho\ (e_i)$:[1, 2]

$$\min \rho\ (e_i) = \min \rho\ (y_i - \sum_{j=1}^{k} x_{ij}\beta_j\ ) ; \quad i=1, 2, \dots, n ; k = 1,2, \dots, p \qquad (6)$$

The standardization of the residuals can be obtained by dividing residuals by the scale. The solution of Equation (6) can be obtained by solving:

$$min \sum_{i=1}^{n} \rho(u_i) = min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\sigma}\right) = min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{i=1}^{n} x_{ij}\beta_j}{\sigma}\right) (7)$$

To obtain $\hat{\beta}_m$, the standard deviation of the residuals (σ) should be estimated by using a robust estimation of σ as follows:[3, 4]

$$\hat{\sigma} = \frac{med|e_i - med(e_i)|}{0.6745} = \frac{MADE}{0.6745}$$

The objective function $\rho\ (.)$ should be satisfied the following constraints:

1. $\rho\ (0) = 0$
2. $(e_i) \geq 0$ (Non-negativity.)
3. $\rho\ (e_i) = \rho\ (-e_i)$(Symmetric).
4. $\rho\ (e_i) \geq \rho\left(e_j\right) ;\ if\ |e_i| \geq \left|e_j\right|$(Monotone in $|e_i|$).

Taking the partial derivative for equation (7) with respect to the regression coefficients and equating it to zero, it results in a system of estimated equations for the model parameters, yields:

$$\sum_{i=1}^{n} x_{ij}\ \Psi\left(\frac{y_i - \sum_{i=1}^{k} x_{ij}\ \beta}{\hat{\sigma}}\right) = 0 \qquad ; \qquad j = 1, 2, \dots, p \qquad (8)$$

where, $\Psi(u_i) = \acute{\rho}(u_i)$ represents the influence function.
The solution for (8) was given by Draper and Smith [5] which is as follows:
Assume that W is a diagonal matrix represents the weight function define as:

$$w_{ii}(u_i) = \frac{\Psi(u_i)}{u_i} = \frac{\Psi\left[\frac{y_i - \sum_{i=1}^{k} x_{ij}\beta}{\hat{\sigma}}\right]}{\left[\frac{y_i - \sum_{i=1}^{k} x_{ij}\beta}{\hat{\sigma}}\right]}$$

The estimated equations for the model parameters (8), can be written in term of the weighted function as follows:

$$\sum_{i=1}^{n} x_{ij}\ w_i \left[\frac{y_i - \sum_{i=1}^{k} x_{ij}\beta}{\hat{\sigma}}\right] = 0 \qquad (9)$$

The solution of the estimated equations (9) can be obtained by reweighted OLS iteratively (IRLS) as follows:

$$\underline{\hat{B}}^{t} = (\acute{X}W^{t-1}X)^{-1}\acute{X}W^{t-1}\underline{Y}$$

The initial estimates of the regression coefficients $\underline{\hat{B}}^{0}$ are mostly represented by the OLS estimators. There is a set of M-estimators that differ among themselves in terms of accuracy depending on $\rho(u_i)$. In this paper, we use three different objective functions which are as follows:

Huber's (1964) objective function (The first objective function) [6, 7, 8]

This method is summarized in reducing the effect of the large residual values, by using the following influence function:

$$\Psi(e_i) = \max\{-c, \min(e_i, c)\}\ ;\ c = 1.5, 1.7$$

The steps of applying the M-method using Huber's (1964) objective function can be shown by the following algorithm:

1. Obtaining the initial estimations of the regression coefficients by one of the estimation methods as OLS method.
2. Calculate residual values e_i
3. Calculate the diagonal vales of weighted matrix W where,
$w_{ii} = \max\{-1.5, \min(e_i, -1.5)\}/e_i$
4. Calculate $\hat{B}^{H1}$ using the weighted least squares (WLS) method as:

$$\hat{B}^{H1} = \left(\acute{X}WX\right)^{-1}\acute{X}W\underline{Y}$$

5.Repeat steps 2-4 to obtain a convergent value of $\hat{B}^{H1}$.

Huber's objective function (The second objective function)

Huber proposed the following objective function:

$$\rho(u_i) = \begin{cases} \frac{1}{2}{u_i}^2 & ; \quad |u_i| \le c \\ c\left(|u_i| - \frac{1}{2}c\right) & ; \quad |u_i| > c \end{cases}$$

$$\psi(u_i) = \begin{cases} u_i & ; \quad |u_i| \le c \\ c\ \mathrm{sign}\ (u_i) & ; \quad |u_i| > c \end{cases}$$

$$w(u_i) = \begin{cases} 1 & ; \quad |u_i| \le c \\ c/|u_i| & ; \quad |u_i| > c \end{cases} \qquad (10)$$

c =1.345

The following algorithm shows the steps of applying the M-estimation method using Huber

1. Obtaining the initial estimations of the regression coefficients by one of the estimation methods as OLS method.
2. Calculate residual value e_i
3. Compute the median (mn) of e_i
4. Compute the median MD where MD is the median of $|e_i - mn|$
5. Estimate the scale parameter σ by compute $\hat{\sigma}$ as follows:
$\hat{\sigma} = \frac{MD}{0.6745}$, where,
7. Calculate u_i, where, $u_i = e_i/\hat{\sigma}_i$
8. Calculate the diagonal vales of weighted matrix W that defined in (10)
9. Calculate $\hat{B}^{H2}$ using weighted least squares (WLS) method as:

$$\underline{\hat{B}}_i^{H2} = \left(\acute{X}W_{i-1}X\right)^{-1}\acute{X}W_{i-1}\underline{Y}$$

10.Repeat steps 2-9 to obtain a convergent value of $\underline{\hat{B}}_i^{H2}$.

- **Tukey's Biweight objective function [9]**

Tukey proposed the Biweight objective function, it's also called the dynamic weight function, and it is given by:

$$\rho(u_i)=\begin{cases}\frac{c^2}{6}\left\{1-\left[1-(\frac{u_i}{c})^2\right]^3\right\} & ; \quad |u_i| \le c \\ \frac{c^2}{6} & ; \quad |u_i| > c\end{cases}$$

$$\psi\,(u_i)=\begin{cases}u_i\left[1-(\frac{u_i}{c})^2\right]^2 & ; \quad |u_i| \le c \\ 0 & ; \quad |u_i| > c\end{cases}$$

$$\text{w}\,(u_i)=\begin{cases}\left[1-(\frac{u_i}{c})^2\right]^2 & ; \quad |u_i| \le c \\ 0 & ; \quad |u_i| > c\end{cases} \qquad (11)$$

The constant c = 4.685 or 6. In this paper we use c = 4.685.

The following algorithm shows the steps of applying the M-robust estimation as follows:

1. Obtaining the initial estimations of the regression coefficients by one of the estimation methods as OLS method.
2. Calculate residual value e_i
3. Compute the median (mn) of e_i
4. Compute the median MD where MD is the median of $|e_i - mn|$
5. Estimate the scale parameter σ by compute $\hat{\sigma}$ as follows:

$\hat{\sigma} = \frac{MD}{0.6745}$, where,

7. Calculate u_i, where, $u_i = e_i/\hat{\sigma}$
8. Calculate the diagonal vales of weighted matrix W by using equation (11)
9. Calculate $\hat{B}'$ using weighted least squares (WLS) method as:

$$\hat{B}^{TU} = (X\acute{W}X)^{-1}\acute{X}WY$$

10.Repeat steps 2-9 to obtain a convergent values of $\hat{B}^{TU}$

2.3 The suggested robust regression methods

In this paper, different suggested estimation methods have been applied as follows:

- Using Gastwirth's location estimator (G) **[10]** instead of the mean in OLS method where Gastwirth's estimation (G) is one of the robust estimators of location that represents the L-estimator. It is a weighted sum of three order statistics, which are the one-third quantile $(Q_{\frac{1}{3}})$, the median $(Q_{\frac{1}{2}})$, and the two-third quantile $(Q_{\frac{2}{3}})$. Gastwirth's estimation can be obtained as follows:

$$G = 0.3\,Q_{\frac{1}{3}} + 0.4\,Q_{\frac{1}{2}} + 0.3\,Q_{\frac{2}{3}}$$

- Using the Gastwirth's in calculating $\hat{\sigma}$instead of the median in different M-estimation methods.

3. Simulation study

To evaluate the efficiency of different estimation methods, i.e., OLS, OLS using median (OLS-MD), OLS using Gastwirth (OLS-GA), M-estimation by Huber (1964) (Huber1), M-estimation by Huber with repeated (Huber1 with repeated), M-estimation by Huber2 (Huber2), M-estimation by Huber2 with Gastwirth (Huber2-GA), M-estimation by Tukey (Tukey), and M-estimation by Tukey with Gastwirth (Tukey-GA). The Monte-Carlo

simulation experiments have been employed with different sample sizes (n = 10, 30, 50, 75, 150) to represent small, moderate, and large sample sizes. The intercept and slop have been chosen as (β_0= 2) and ($\beta_1 = 4$). In this paper, the distribution of error term has been contaminated by outliers with three different percentages (10%, 20%, and 35%). The generated samples have been replicated (1000) times (R=1000).
The Algorithm of simulation experiments can be summarized by the following table:

Table (1): The Algorithm of Simulation Experiments

n	Regression coefficients	The distribution of X	The distribution of ∈	The distribution of Y
10 30 50 75 100	β_0= 2 $\beta_1 = 4$	$X\sim N(1,1)$	$\in \sim N(0,1)$	$Y\sim N(6,1)$
		$X\sim N(1,1)$	$90\% \in \sim N(0,1) + 10\% \in \sim N(0,50)$	$90\%\ Y\sim N(6,1) + 10\% Y\sim N(6,50)$
		$X\sim N(1,1)$	$80\% \in \sim N(0,1) + 20\% \in \sim N(0,50)$	$80\%\ Y\sim N(6,1) + 20\% Y\sim N(6,50)$
		$X\sim N(1,1)$	$65\% \in \sim N(0,1) + 35\% \in \sim N(0,50)$	$65\%\ Y\sim N(6,1) + 35\% Y\sim N(6,50)$

The comparison between the different methods was made based on the mean squares of error of the regression coefficients, where,

$$MSE\left(\underline{\hat{\beta}}\right) = \frac{1}{R}\sum_{i=1}^{R}\left[(\underline{\hat{\beta}_i} - \underline{\beta})'(\underline{\hat{\beta}_i} - \underline{\beta})\right] \qquad (14)$$

$$= \frac{1}{R}\sum_{i=1}^{R}\left[(\hat{\beta}_{0i} - \beta_0)^2 + (\hat{\beta}_{1i} - \beta_1)^2\right]$$

4. **Simulation Results**

To examine and compare the behavior of different estimation methods under different cases, the simulation experiment results represented by $E(\hat{\beta}_0), E(\hat{\beta}_1)\ and\ MSE(\hat{\beta})$ are summarized in tables 2-9. The behavior of different estimation methods will be discussed briefly according to the percentage of the existence of outliers in terms of error, as follows:

1. **In case of no outlier:**

In case of the optimal state of the data is achieved, i.e., the absence of outliers. The results have been summarized in table (3), and it indicates the following:

- The OLS method is the best compared to the other methods, followed by Huber1 without repeated, for all different cases.
- With the increasing sample size, the MSE's for all estimation methods are decreasing and converged to each other, which corresponds to the central limit theory.
- If the optimal state of the data is achieved, i.e., the absence of outliers, the OLS method is the best for all sample sizes where it has the lowest MSE's followed by a method of Huber1 without repeated in terms of accuracy of the results.
- The results indicate that OLS-MD is the worst estimate of regression coefficients and for all sample sizes because it has the largest MSE's compared to other methods of estimation.

- The results indicated that the use of the proposed OLS-GA method was better than OLS-MD as the MSE values if the GA statistic is used is lower for all sample sizes than when using the medium.
- Increasing the sample size leads to a decrease in MSE values and for all methods of assessment under study.
- We note that the expected values of different estimates of regression coefficients are closer to the default values, and this is further increased by the large sample

Table (3): The expected values and MSE's for different estimation methods when $X \sim N(1,1), \in \sim N(0,1)$

Method	Criteria	n = 10	n = 30	n = 50	n =75	n = 150
OLS	$E(\hat{\beta}_0)$	1.991136	1.999987	1.999902	2.000732	1.997645
	$E(\hat{\beta}_1)$	4.005048	3.999271	3.994431	3.995453	3.995896
	MSE $(\hat{\beta})$	**0.407965**	**0.107829**	**0.060450**	**0.041365**	**0.020301**
OLS-MD	$E(\hat{\beta}_0)$	2.031086	2.039543	2.020241	2.006474	1.997841
	$E(\hat{\beta}_1)$	3.975889	3.979554	3.981870	3.983577	3.990346
	MSE $(\hat{\beta})$	0.672467	0.264030	0.181499	0.133474	0.065672
OLS-GA	$E(\hat{\beta}_0)$	2.015306	2.018572	2.001818	2.003968	1.998586
	$E(\hat{\beta}_1)$	3.991028	3.993313	3.990563	3.992205	3.994316
	MSE $(\hat{\beta})$	0.517250	0.166212	0.101754	0.068781	0.034036
Huber1 with repeated	$E(\hat{\beta}_0)$	2.001574	2.012619	2.011480	2.012999	2.009502
	$E(\hat{\beta}_1)$	4.003432	3.999912	3.995410	3.995654	3.996550
	MSE $(\hat{\beta})$	0.415307	0.112343	0.062348	0.042571	0.020815
Huber1	$E(\hat{\beta}_0)$	1.999373	2.011064	2.010220	2.011710	2.008348
	$E(\hat{\beta}_1)$	4.004072	3.999832	3.995376	3.995685	3.996492
	MSE $(\hat{\beta})$	**0.413000**	**0.111267**	**0.061854**	**0.042286**	**0.020698**
Huber2	$E(\hat{\beta}_0)$	1.992316	1.999866	2.000166	2.001262	1.997082
	$E(\hat{\beta}_1)$	4.001040	4.000774	3.995183	3.995201	3.996568
	MSE $(\hat{\beta})$	0.431083	0.116750	0.063644	0.043772	0.021253
Huber2-GA	$E(\hat{\beta}_0)$	1.993777	1.999910	1.999962	2.001177	1.997271
	$E(\hat{\beta}_1)$	4.000976	4.000747	3.995082	3.995254	3.996505
	MSE $(\hat{\beta})$	0.429254	0.116234	0.063423	0.043514	0.021159
Tukey	$E(\hat{\beta}_0)$	1.993865	2.000785	1.999093	2.000878	1.996932
	$E(\hat{\beta}_1)$	3.997302	4.000713	3.995664	3.995478	3.996796
	MSE $(\hat{\beta})$	0.459887	0.121543	0.064452	0.043642	0.021258
Tukey-GA	$E(\hat{\beta}_0)$	1.994404	2.000060	1.999181	2.000958	1.997167
	$E(\hat{\beta}_1)$	3.998237	4.000963	3.995481	3.995394	3.996712
	MSE $(\hat{\beta})$	0.450198	0.119638	0.063879	0.043274	0.021125

2. In case of the random error variable is contaminated with outliers by 10%

The results are shown in Table (4) indicate the following important points:

- OLS method efficiency decreases due to an increase in the MSE values compared to MSE values for other methods.
- The proposed method of estimating regression coefficients Tukey-GA was the best for all sample sizes compared to other methods. Followed by the Tukey method where the MSE value is lower than other methods.

- The worst method to estimate regression coefficients was OLS-MD because of the large MSE values compared to other methods and for all sample sizes.
- The proposed OLS-GA method yielded better results than OLS-MD and for all sample sizes, i.e., the use of OLS with the robust GA statistics resulted in more accurate estimates than if it were replaced by the median.

Table (4): The expected values and MSE's for different estimation methods when $X \sim N(1,1), \in \sim 0.90\ N(0,1) + 0.10\ N(0,50)$

Method	Criteria	n = 10	n = 30	n = 50	n =75	n = 150
OLS	$E(\hat{\beta}_0)$	1.979675	1.987213	1.984935	1.984257	2.000257
	$E(\hat{\beta}_1)$	4.018295	3.998626	3.999363	3.999170	3.991270
	MSE $(\hat{\beta})$	2.779872	0.656185	0.350283	0.253894	0.120755
OLS-MD	$E(\hat{\beta}_0)$	2.040513	2.044871	2.009006	2.000808	2.000848
	$E(\hat{\beta}_1)$	3.972387	3.976157	3.985472	3.986578	3.985543
	MSE $(\hat{\beta})$	2.833915	0.778850	0.460883	0.328103	0.159826
OLS-GA	$E(\hat{\beta}_0)$	2.007765	2.019646	1.992332	1.989400	2.000907
	$E(\hat{\beta}_1)$	3.994502	3.990916	3.994292	3.995191	3.989509
	MSE $(\hat{\beta})$	2.579290	0.659048	0.364698	0.251669	0.121579
Huber1 with repeated	$E(\hat{\beta}_0)$	2.003510	2.020134	2.022118	2.022376	2.019527
	$E(\hat{\beta}_1)$	4.015652	4.004127	3.993844	3.993662	3.996811
	MSE $(\hat{\beta})$	0.732960	0.169719	0.091468	0.059628	0.029296
Huber1	$E(\hat{\beta}_0)$	2.012360	2.017587	2.019205	2.019068	2.017559
	$E(\hat{\beta}_1)$	4.015192	4.004183	3.994734	3.994269	3.996354
	MSE $(\hat{\beta})$	1.140782	0.209946	0.108219	0.071737	0.033749
Huber2	$E(\hat{\beta}_0)$	1.988968	1.996695	2.000624	2.001689	1.997510
	$E(\hat{\beta}_1)$	4.008330	4.005011	3.993163	3.992711	3.996443
	MSE $(\hat{\beta})$	0.704846	0.166988	0.090497	0.058928	0.028753
Huber2-GA	$E(\hat{\beta}_0)$	1.987982	1.997081	2.001013	2.001518	1.997536
	$E(\hat{\beta}_1)$	4.009062	4.004546	3.993043	3.992874	3.996540
	MSE $(\hat{\beta})$	0.689489	0.167131	0.090657	0.058927	0.028762
Tukey	$E(\hat{\beta}_0)$	1.993372	1.998929	2.003891	2.004868	1.997596
	$E(\hat{\beta}_1)$	4.002834	4.005271	3.991971	3.992023	3.997045
	MSE $(\hat{\beta})$	**0.614509**	**0.146300**	**0.082145**	**0.050541**	**0.025463**
Tukey-GA	$E(\hat{\beta}_0)$	1.996038	1.998586	2.004419	2.004648	1.997515
	$E(\hat{\beta}_1)$	4.002291	4.005833	3.991552	3.992144	3.997115
	MSE $(\hat{\beta})$	**0.595782**	**0.143976**	**0.081585**	**0.050293**	**0.025498**

3. In case of the random error variable is contaminated with outliers by 20%

The results shown in Table-5 indicate that:

- Decrease efficiency of OLS method due to increasing the MSE values compared to other methods' MSE values.
- The proposed method for estimating the regression coefficients Tukey-GA was the best for all sample sizes compared to other methods. Followed by the Tukey method, where MSE's are less than the other methods.
- The worst method of estimation was OLS-MD because of the increase in MSE compared to other methods and for all sample sizes.

- The proposed OLS-GA method gave better results than the OLS-MD method for all sample sizes, meaning that the use of the OLS method with the robust GA resulted in more accurate estimates than if it was replaced by the median.

Table (5): The MSE values and the expected values of the coefficients of simple linear regression when $X\sim N(1,1)$, $\in \sim 0.80\ N(0,1)+0.20\ N(0,50)$

Method	Criteria	n = 10	n = 30	n = 50	n =75	n = 150
OLS	$E(\hat{\beta}_0)$	1.988665	2.018226	1.987979	1.972414	1.992878
	$E(\hat{\beta}_1)$	4.020236	3.989466	3.988709	3.997353	3.992082
	MSE $(\hat{\beta})$	4.733328	1.213448	0.632722	0.457615	0.217800
OLS-MD	$E(\hat{\beta}_0)$	2.073517	2.066055	2.017217	1.997176	1.999512
	$E(\hat{\beta}_1)$	3.961802	3.966077	3.974208	3.983998	3.985821
	MSE $(\hat{\beta})$	4.641670	1.283283	0.724710	0.515888	0.245166
OLS-GA	$E(\hat{\beta}_0)$	2.025667	2.038355	1.997171	1.981085	1.999628
	$E(\hat{\beta}_1)$	3.987620	3.980122	3.983183	3.992910	3.989892
	MSE $(\hat{\beta})$	4.363669	1.167067	0.621823	0.433461	0.204878
Huber1 with repeated	$E(\hat{\beta}_0)$	2.041333	2.031149	2.033502	2.037112	2.030956
	$E(\hat{\beta}_1)$	3.997253	4.005606	3.993665	3.990581	3.995744
	MSE $(\hat{\beta})$	1.409876	0.245415	0.127059	0.087012	0.041135
Huber1	$E(\hat{\beta}_0)$	2.045412	2.034403	2.028835	2.029344	2.025950
	$E(\hat{\beta}_1)$	4.002264	4.002304	3.993286	3.991577	3.995482
	MSE $(\hat{\beta})$	2.179730	0.366412	0.177532	0.119792	0.053817
Huber2	$E(\hat{\beta}_0)$	2.010686	1.994745	1.998656	2.002965	1.996923
	$E(\hat{\beta}_1)$	3.993664	4.006885	3.992991	3.990148	3.995592
	MSE $(\hat{\beta})$	1.502172	0.249403	0.129091	0.087488	0.040948
Huber2-GA	$E(\hat{\beta}_0)$	2.011067	1.995600	1.999097	2.002880	1.996789
	$E(\hat{\beta}_1)$	3.992058	4.006031	3.992712	3.990356	3.995773
	MSE $(\hat{\beta})$	1.508893	0.252564	0.131244	0.088266	0.041412
Tukey	$E(\hat{\beta}_0)$	2.011084	1.991682	2.001838	2.009940	1.999145
	$E(\hat{\beta}_1)$	3.991201	4.009220	3.993316	3.989318	3.994902
	MSE $(\hat{\beta})$	**1.280112**	**0.195811**	**0.108055**	**0.068999**	**0.032170**
Tukey-GA	$E(\hat{\beta}_0)$	2.014876	1.992238	2.002695	2.010361	1.998895
	$E(\hat{\beta}_1)$	3.985702	4.009245	3.992731	3.988904	3.994932
	MSE $(\hat{\beta})$	**1.344102**	**0.197513**	**0.108717**	**0.069078**	**0.032457**

4. In case of the random error variable is contaminated with outliers by 35%

The results shown in Table-6 indicate that:

- Decrease efficiency of OLS method due to increase the MSE values compared to other methods' MSE values.
- The **Huber1 with repeated** method was the best when sample size n=10 compared to other methods. Followed by the **Tukey** method,
- The performance of **Tukey** method is the best when the sample size $n\geq 30$ compared to the other methods. Followed by the **Tukey-GA** method, where the value of MSE is less than the other methods when the sample size is n=30,75,150, while the sample size is n=50, which is **Huber1 with repeated**.

- The worst method of estimation was OLS due to the large MSE relative to the other methods when the sample size is n = 10, and the worst method of estimating OLS-MD due to the large MSE relative to the other methods when the sample size is $n \geq 30$.
- The proposed OLS-GA method gave better results than the OLS-MD method for all sample sizes, meaning that the use of the OLS method with the robust GA resulted in more accurate estimates than if it was replaced by the median.

Table (6): The MSE values and the expected values of the coefficients of simple linear regression when $X \sim N(1,1)$, $\in \sim 0.65\ N(0,1) + 0.35\ N(0,50)$

Method	Criteria	n = 10	n = 30	n = 50	n =75	n = 150
OLS	$E(\hat{\beta}_0)$	1.951986	2.025625	2.011581	1.979795	1.978178
	$E(\hat{\beta}_1)$	4.022936	3.972101	3.970678	3.992656	3.994381
	MSE $(\hat{\beta})$	6.697178	2.071986	1.009795	0.746066	0.374650
OLS-MD	$E(\hat{\beta}_0)$	2.067501	2.086262	2.041235	2.002890	1.996362
	$E(\hat{\beta}_1)$	3.956559	3.946403	3.953665	3.978452	3.987397
	MSE $(\hat{\beta})$	6.477293	2.106375	1.117663	0.791860	0.390525
OLS-GA	$E(\hat{\beta}_0)$	2.028870	2.057825	2.022497	1.989220	1.988593
	$E(\hat{\beta}_1)$	3.984689	3.960656	3.963214	3.986773	3.991603
	MSE $(\hat{\beta})$	6.189917	1.970847	0.998235	0.707440	0.352771
Huber1 with repeated	$E(\hat{\beta}_0)$	2.055077	2.055965	2.057909	2.064031	2.049506
	$E(\hat{\beta}_1)$	4.001651	4.001820	3.992231	3.987106	3.996094
	MSE $(\hat{\beta})$	**2.465477**	**0.412102**	**0.211376**	0.143942	0.067997
Huber1	$E(\hat{\beta}_0)$	2.032371	2.063637	2.055240	2.049012	2.036577
	$E(\hat{\beta}_1)$	4.009899	3.990110	3.985861	3.988248	3.996069
	MSE $(\hat{\beta})$	3.689130	0.765408	0.335332	0.227431	0.103742
Huber2	$E(\hat{\beta}_0)$	1.983756	2.003063	1.999678	2.005872	1.991942
	$E(\hat{\beta}_1)$	4.002844	3.997565	3.991778	3.986595	3.996465
	MSE $(\hat{\beta})$	3.065918	0.528523	0.249486	0.168208	0.078454
Huber2-GA	$E(\hat{\beta}_0)$	1.977841	2.003198	1.998447	2.005596	1.991218
	$E(\hat{\beta}_1)$	4.004754	3.998547	3.991767	3.986196	3.996686
	MSE $(\hat{\beta})$	3.101569	0.548470	0.267439	0.177992	0.083779
Tukey	$E(\hat{\beta}_0)$	1.986722	1.989431	1.993856	2.012499	1.997801
	$E(\hat{\beta}_1)$	4.008048	4.008096	3.998253	3.985371	3.993691
	MSE $(\hat{\beta})$	**2.954632**	**0.370322**	**0.204869**	**0.126553**	**0.056252**
Tukey-GA	$E(\hat{\beta}_0)$	2.013001	1.987057	1.992433	2.013538	1.997187
	$E(\hat{\beta}_1)$	3.979294	4.011380	3.998161	3.984782	3.994367
	MSE $(\hat{\beta})$	3.206710	**0.393554**	0.222145	**0.134493**	**0.061171**

References

[1]Huber, P. J, (1964), Robust Estimation of a Location parameter. Annals of Mathematical Statistics
35:73, 101.

[2] Almetwally, E. M. and Almongy, H. M.,Comparison Between M-estimation, S-estimation, And MM Estimation Methods of Robust Estimation with Application and Simulation, International Journal of Mathematical Archive-9 (11), 2018, 55-63

[3]Rousseeuw P.J. and Leroy A.M., 1987, Robust Regression and Outlier Detection, John Wiley, New York, 202. [24]Fox J., “Robust Regression”, An R and S-PLUS Companion to Applied Regression, http://cran.r-project.org/.

[4]Coleman D., Holland P., Kaden N., Klema V. and Peters S.C., 1980, “A System of Subroutines for Iteratively Reweighted Least Squares Computations”, ACM Trans. Math. Softw., 6, 327-336

[5] N. R. Draper and H. Smith, Applied Regression Analysis, Third Edition,
Wiley Interscience Publication, United States, 1998.

[6]Bickel P., 1983, “Robust Estimation”, in Encyclopedia of Statistical Sciences, Eds. S. Kotz, N.L. Johnson, C.B. Read, N. Balakrishnan and B. Vidakovic, Vol. 8, John Wiley & Sons, Ltd, 157-163.

[7]Huber P.J., 1981, Robust Statistics, John Wiley, New York.

[8]Ruppert D., 1983, “M-Estimators”, in Encyclopedia of Statistical Sciences, Eds. S. Kotz, N.L.

[9] Gross A.M., 1981, “Confidence Intervals for Bisquare Regression Estimates”, J. Amer. Statist. Ass., 72, 341-354

[10] Gastwirth, J. L. (1966). "On Robust Procedures". J. Amer. Statist. Assn., Vol. 61, pp. 929-948.

[11]. Nandal, N., & Nandal, N. (2019). BSCQUAL: A Measuring Instrument of Service Quality for the B-Schools . International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 04, 1574-1589