Model of Robust Regression with Parametric and Nonparametric Methods

Dr. Nadia H. AL – Noor* and Asmaa A. Mohammad**
*Department of Mathematics \ College of Science\ Al-Mustansiriya University-Iraq
**Department of Mathematics \ College of Science for Women, University of Baghdad-Iraq
Corresponding E-mail:nadialnoor@yahoo.com

Abstract
In the present work, we evaluate the performance of the classical parametric estimation method "ordinary least squares" with the classical nonparametric estimation methods, some robust estimation methods and two suggested methods for conditions in which varying degrees and directions of outliers are presented in the observed data. The study addresses the problem via computer simulation methods. In order to cover the effects of various situations of outliers on the simple linear regression model, samples were classified into four cases (no outliers, outliers in the *X*-direction, outliers in the *Y*-direction and outliers in the *XY*-direction) and the percentages of outliers are varied between 10%, 20% and 30%. The performances of estimators are evaluated in respect to their mean squares error and relative mean squares error.
Keywords: Simple Linear Regression model; Ordinary Least Squares Method; Nonparametric Regression; Robust Regression; Least Absolute Deviations Regression; M-Estimation Regression; Trimmed Least Squares Regression.

1. Introduction

The simple linear regression model is expressed as:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (1)$$

Where: Y is called response variable or dependent variable; X is called predictor variable, regressor variable or independent variable, and ε is called prediction error or residual. The symbols β_0 and β_1 are called intercept and slope respectively which they represents the linear regression unknown parameters or coefficients.

The process of estimating the parameters of regression model is still one of important subjects despite of large number of papers and studies written in this subject which differ in techniques followed in the process of estimation. The ordinary least squares (OLS) method is the most popular classical parametric regression technique in statistics and it is often used to estimate the parameters of a model because of nice property and ease of computation. According to Gauss-Marcov theorem, the OLS estimators, in the class of unbiased linear estimators, have minimum variance i.e. they are best linear unbiased estimator (BLUE)[10]. Nonetheless, the OLS estimates are easily affected by the presence of outliers, "outliers are observations which are markedly different from the bulk of the data or from the pattern set by the majority of the observations. In a regression problem, observations corresponding to excessively large residuals are treated as outliers[18]", and will produce inaccurate estimates. The breakdown point of the OLS estimator is 0% which implies that it can be easily affected by a single outlier. So alternative methods such as nonparametric and robust methods should be put forward which are less affected by the outliers. However, most robust methods are relatively difficult and computationally complicated. As an alternative to OLS, least absolute deviations regression (LAD or L1) has been proposed by Boscovich in 1757, then Edgeworth in 1887. LAD regression is the first step toward a more robust regression [22][26]. The next direct step to obtain robust regression was the use of M-estimators. The class of M-estimators was defined by Huber (1964, 1968) for the location model and extended by him to the regression model in (1973) [12] as an alternative robust regression estimator to the least squares. This method based on the idea of replacing the squared residual in OLS by another symmetric function, ρ, of the residuals [13]. Rousseeuw and Yohai (1984) [24] introduced the Trimmed Least Squares (TLS) regression which is a highly robust method for fitting a linear regression model. The TLS estimator minimizes the sum of the (h) smallest squared residuals. Alma (2011) [1] compare some robust regression methods such that TLS and M-estimate against OLS regression estimation method in terms of the determination of coefficient. Bai (2012) [3] review various robust regression methods including "M-estimate and TLS estimate" and compare between them based on their robustness and efficiency through a simulation study where n=20,100. In other side, Theil (1950) [27] introduced a nonparametric procedure which is expected to perform well without regard to the distribution of the error terms. This procedure is based on ranks and uses the median as robust measures rather than using the mean as in OLS. Mood and Brown (1950) [19] proposed to estimate the intercept and slope simultaneously from two equations depending upon divide the observations for two groups according to the median of the variable

(X). Conover (1980) [5] calculate the estimate of the intercept by used the median of the response variables, estimated Thiel's slope and the median of the explanatory variables. Hussain and Sprent (1983) [14] presented a simulation study in which they compared the OLS regression estimator against the Theil pairwise median and weighted Theil estimators in a study using 100 replications per condition. Hussain and Sprent characterized the data modeled in their study as typical data patterns that might result from contamination due to outliers. Contaminated data sets were generated using a mixture model in which each error term is either a random observation from a unit normal distribution [N(0,1)] or an observation from a normal distribution with a larger variance [N(0, k^2), k >1]. Jajo (1989) [15] carried a simulation study to compare the estimators that obtained from (Thiel, Mood-Brown, M-estimation and Adaptive M-estimation) with the estimators that obtained from least squares of the simple linear regression model in the presence of outliers. Mutan (2004) [20] introduced a Monte Carlo simulation study to comparing regression techniques including (ordinary least squares, , least absolute deviations, trimmed least squares, Theil and weighted Theil) for the simple linear regression model when the distribution of the error terms is Generalized Logistic. Meenai and Yasmeen (2008) [17] applied nonparametric regression methods to some real and simulated data.

In the present work, we evaluate the performance of the classical nonparametric estimation methods, some robust estimation methods "least absolute deviations, M-estimation and trimmed least squares" and two suggested methods "depending upon nonparametric and M-estimation" with the OLS estimation method for conditions in which varying degrees and directions of outliers are presented in the observed data. The study addresses the problem via computer simulation methods. In order to cover the effects of various situations of outliers on the simple linear regression model, samples were classified into four cases (no outliers, outliers in the *X*-direction, outliers in the *Y*-direction "error distributed as contaminated normal", and outliers in the *XY*-direction) and the percentages of outliers are varied between 10%, 20% and 30% . The performances of estimators are evaluated in respect to their mean squares error and relative mean squares error.

2. *Classical Estimation Method for Regression Parameters [16]*

The most well-known classical parametric method of estimating the regression parameters is to use a least square error (LSE) approach. The basic idea of ordinary least squares is to optimize the fit by minimizing the total sum of the squares of the errors (deviations) between the observed values y_i and the estimated values $\hat{\beta}_0 + \hat{\beta}_1 x_i$:

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (2)$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of β_0 and β_1, respectively. The least squares estimators of β_0 and β_1, $\hat{\beta}_0$ and $\hat{\beta}_1$ are:

$$\hat{\beta}_1^{OLS} = \frac{\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)/n}{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2/n} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (3)$$

$$\hat{\beta}_0^{OLS} = \bar{y} - \hat{\beta}_1 \bar{x} \quad (4)$$

Where: $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$

3. *Alternative Estimation Methods for Regression Parameters*

3.1 Nonparametric Regression [5][11][14][15][21][27]

The OLS regression method described above assume normally distributed error terms in the regression model. In distinction, classical nonparametric methods to linear regression typically employ parameter estimation methods that are regarded as distribution free. Since nonparametric regression procedures are developed without relying on the assumption of normality of error distributions, the only presupposition behind such procedures is that the errors of prediction are independently and identically distributed (i.i.d.). Many nonparametric procedures are based on using the ranks of the observed data rather than the observed data themselves. The robust estimate of slope for nonparametric fitted line was first described by Theil (1950). He proposed two methods, namely, the complete and the incomplete method. Assumed that all the x_i's are distinct, and lose no generality that the x_i's are arranged in ascending order. The complete Theil slope estimate is computed by comparing each data pair to all others in a pairwise fashion. A data set of n (X,Y) pairs will result in $N = \binom{n}{2} = \frac{n(n-1)}{2}$ pairwise comparisons. For each of these comparisons a slope $\Delta Y/\Delta X$ is computed. The median of all possible pairwise slopes is taken as the nonparametric Thiel's slope estimate, $\hat{\beta}_1^{Thiel}$, Where:

$$S_{ij} = \frac{\Delta Y}{\Delta X} = \frac{y_j - y_i}{x_j - x_i} \quad ; x_i \neq x_j \ , 1 \leq i < j \leq n \tag{5}$$

$$\hat{\beta}_1^{Thiel} = median\left(S_{ij}\right) ; 1 \leq i < j \leq n \tag{6}$$

For incomplete method, Theil suggested using only a subset of all S_{ij}, and took as estimator of β_1 the median of the subset ($S_{i,i+n^*}$); where:

$$S_{i,i+n^*} = \frac{y_{i+n^*} - y_i}{x_{i+n^*} - x_i} \quad ; i = 1,2, \dots, n^* \tag{7}$$

If n is even then $n^* = n/2$. If n is odd, the observation with rank (n+1)/2 is not used. The incomplete Theil's slope estimator is:

$$\hat{\beta}_1^{Thiel*} = median\left(S_{i,i+n^*}\right);\ i = 1,2, \dots, n^* \tag{8}$$

For estimation the intercept parameter, Theil's intercept estimate, $\hat{\beta}_0^{TH}$, is defined as:

$$\hat{\beta}_0^{TH} = median\left(y_i - \hat{\beta}_1^{TH} x_i\right) \ ; i = 1,2, \dots, n \tag{9}$$

Where $\hat{\beta}_1^{TH}$ is the estimate of β_1 according to the complete or the incomplete Theil's slope estimator.
Other estimators of intercept have been suggested. Conover suggested estimating β_0 by using the formula:

$$\hat{\beta}_0^{CON} = median\ (y_i) - \hat{\beta}_1^{TH}.median\ (x_i) \tag{10}$$

This formula "Conover's estimator" assures that the fitted line goes through the point (X_{median} ,Y_{median}). This is analogous to OLS, where the fitted line always goes through the point $(\bar{x}, \bar{y})$.

3.2 Robust Regression

Any robust method must be reasonably efficient when compared to the least squares estimators; if the underlying distribution of errors are independent normal, and substantially more efficient than least squares estimators, when there are outlying observations. There are various robust methods for estimation the regression parameters. The main focus of this subsection is to least absolute deviations regression, M-estimation and trimmed least squares regression which are the most popular robust regression coefficients with outliers.

3.2.1 Least Absolute Deviations Regression [4][8][20][25]

The least absolute deviations regression (LAD regression) is one of the principal alternatives to the ordinary least squares method when one seeks to estimate regression parameters.

The goal of the LAD regression is to provide a robust estimator which is minimized the sum of the absolute residuals.

$$min \sum_{i=1}^{n} |r_i| \tag{11}$$

The LAD procedure was developed to reduce the influence of *Y*-outliers in the OLS. The *Y*-outliers have less impact on the LAD results, because it does not square the residuals, and then the outliers are not given as much weight as in OLS procedure. However, LAD regression estimator is just as vulnerable as least squares estimates to high leverage outliers (X-outliers). In fact, LAD estimate have low breakdown point (BP is 1/n or 0%). Although the concept of LAD is not more difficult than the concept of the OLS estimation, calculation of the LAD estimates is more troublesome. Since there are no exact formulas for LAD estimates, an algorithm is used. Birkes and Dodge (1993) explain this algorithm for the simple linear regression model. It is known that LAD regression line passes through two of the data points. Therefore, the algorithm begins with one of the data points, denoted by (x_0, y_0), and tries to find the best line passing through it. The procedure for finding the best line among all lines passing through a given data point (x_0, y_0) is describe below.
For each data point (x_i, y_i), the slope of the line passing through the two points (x_0, y_0) and (x_i, y_i) is calculated and it is equal to the $(y_i$-$y_0)$/(x_i-x_0). If $x_i = x_0$ for some i, the slope is not defined. The data points are re-indexed in such a way that: $(y_1 - y_0)/(x_1 - x_0) \leq (y_2 - y_0)/(x_2 - x_0) \leq \dots \leq (y_n - y_0)/(x_n - x_0)$
Now, the searched point $(x_j$,$y_j)$ is determined by the index j for which.

$$\left.\begin{aligned} &|x_1 - x_0| + \dots + |x_{j-1} - x_0| < \tfrac{1}{2}\mathrm{T} \\ &|x_1 - x_0| + \dots + |x_{j-1} - x_0| + |x_j - x_0| > \tfrac{1}{2}\mathrm{T} \end{aligned}\right] \tag{12}$$

Where $\mathrm{T} = \sum_{i=1}^{n} |x_i - x_0|$.
This conditions guarantee that $\hat{\beta}_1$minimizes the quantity $\sum_{i=1}^{n}\left|(y_i - y_0) - \hat{\beta}_1(x_i - x_0)\right|$

Analogously to $\sum|r_i|$ for the regression lines passing through (x_0,y_0).The $\hat{\beta}_0$ is computed in such a way that the regression line crosses (x_0,y_0). So, the best line passing through (x_0,y_0) is the line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ where:

$$\hat{\beta}_1^{LAD} = \frac{y_j - y_0}{x_j - x_0} \quad (13)$$

$$\hat{\beta}_0^{LAD} = y_0 - \hat{\beta}_1^{LAD} x_0 \quad (14)$$

We can equally verify that it passes through the data point (x_j,y_j).We just have to rename the point (x_j,y_j) by (x_1,y_1) and restart.

***3.2.2 M-Estimation Regression* [1][3][10][12]**

The most common general method of robust regression is M-estimation, introduced by Huber (1973). The M in M-estimates stands for "maximum likelihood type". That is because M-estimation is a generalization of maximum likelihood estimates (MLE). The goal of M-estimation is minimized a sum of less rapidly increasing functions of the residuals, $\sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right)$ where s is an estimate of scale which can be estimated by using the formula:

$$s = \frac{median|r_i - median\,(r_i)|}{0.6745} \quad (15)$$

A reasonable ρ should satisfy the following properties: $\rho(r) \geq 0; \rho(r) = \rho(-r); \rho(0) = 0; \rho(r_i) \geq \rho(r_j)\ for\ |r_i| \geq |r_j|$

M-estimators are robust to outliers in the response variable with high efficiency. However, M-estimators are just as vulnerable as least squares estimates to high leverage outliers. In fact, the BP (breakdown point) of M-estimates is $1/n$ or 0%. Suppose simple linear regression model, the M-estimator minimizes the objective function:

$$\sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \beta_0 - \beta_1 x_i}{s}\right) = \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{s}\right) = \sum_{i=1}^{n} \rho(u_i) \quad (16)$$

Where $u_i = \frac{r_i(\beta)}{s}$ are called standardized residuals. Let $\psi(u) = \acute{\rho}(u)$

Differentiating (16) with respect to β and setting the partial derivatives to zero, we get the normal equations:

$$\left.\begin{aligned} \sum_{i=1}^{n} \psi\left(\frac{r_i(\beta)}{s}\right) &= 0 \\ \sum_{i=1}^{n} \psi\left(\frac{r_i(\beta)}{s}\right) \mathrm{x}_i &= 0 \end{aligned}\right] \quad (17)$$

To solve (17) we define the weight function $W(x) = \frac{\psi(x)}{x}$; if $x \neq 0$ and $W(x) = \acute{\psi}(0)$; if $x = 0$. let $w_i = W(u_i)$.

Then equations (17) can be written as

$$\left.\begin{aligned} \sum_{i=1}^{n} w_i\,(y_i - \beta_0 - \beta_1 x_i) &= 0 \\ \sum_{i=1}^{n} w_i\,(y_i - \beta_0 - \beta_1 x_i) x_i &= 0 \end{aligned}\right] \quad (18)$$

Solving the estimating equations[1] (18) is a weighted least squares problem, minimizing $\sum_{i=1}^{n} w_i^2 u_i^2$. The weights, however, depend upon the residuals, the residuals depend upon the estimated coefficients, and the estimated coefficients depend upon the weights. An iterative solution (called iteratively reweighted least squares) is therefore required. So, the solution of (18) can be found by iterating between w_i and β :

1. Select an initial estimates $\hat{\beta}_0^{(0)}$ and $\hat{\beta}_1^{(0)}$, such as the least squares estimates.
2. At each iteration t, calculate standardized residuals $u_i^{(t-1)}$ and associated weights $w_i^{(t-1)} = W(u_i^{(t-1)})$ from the previous iteration.

1 Newton-Raphson and Iteratively Reweighted Least Squares (IRLS) are the two methods to solve the M-estimates nonlinear normal equations. IRLS is the most widely used in practice and we considered for this study.

3. Solve for new weighted least squares estimates $\hat{\beta}_1^{(t)}, \hat{\beta}_0^{(t)}$.

$$\hat{\beta}_1^{(t)} = \frac{\left(\sum_{i=1}^{n} w_i^{(t-1)}\right)\left(\sum_{i=1}^{n} w_i^{(t-1)} x_i y_i\right) - \left(\sum_{i=1}^{n} w_i^{(t-1)} x_i\right)\left(\sum_{i=1}^{n} w_i^{(t-1)} y_i\right)}{\left(\sum_{i=1}^{n} w_i^{(t-1)}\right)\left(\sum_{i=1}^{n} w_i^{(t-1)} x_i^2\right) - \left(\sum_{i=1}^{n} w_i^{(t-1)} x_i\right)^2} \tag{19}$$

$$\hat{\beta}_0^{(t)} = \frac{\left(\sum_{i=1}^{n} w_i^{(t-1)} x_i^2\right)\left(\sum_{i=1}^{n} w_i^{(t-1)} y_i\right) - \left(\sum_{i=1}^{n} w_i^{(t-1)} x_i\right)\left(\sum_{i=1}^{n} w_i^{(t-1)} x_i y_i\right)}{\left(\sum_{i=1}^{n} w_i^{(t-1)}\right)\left(\sum_{i=1}^{n} w_i^{(t-1)} x_i^2\right) - \left(\sum_{i=1}^{n} w_i^{(t-1)} x_i\right)^2} \tag{20}$$

Also, we can find $\hat{\beta}_0^{(t)}$ *as*:

$$\hat{\beta}_0^{(t)} = \frac{\sum_{i=1}^{n} w_i^{(t-1)} y_i}{\sum_{i=1}^{n} w_i^{(t-1)}} - \hat{\beta}_1^{(t)} \frac{\sum_{i=1}^{n} w_i^{(t-1)} x_i}{\sum_{i=1}^{n} w_i^{(t-1)}} \tag{21}$$

4. Repeat step 2 and step 3 until the estimated coefficients converge. The iteration process continues until some convergence criterion is satisfied, $\left|\hat{\beta}^{(t)} - \hat{\beta}^{(t-1)}\right| \cong 0$.

Several choices of ρ have been proposed by various authors. Two of these are presented in table (1) together with the corresponding derivatives (ψ) and the resulting weights (w).

Table (1): Different ρ functions, together with the corresponding derivatives ψ and the resulting weights w

Type	$\rho(r_i)$	$\psi(r_i)$	$w(r_i)$
Huber	$\begin{cases} \frac{1}{2} r_i^2 & ; \lvert r_i \rvert \le c \\ c\left(\lvert r_i \rvert - \frac{1}{2}c\right) & ; \lvert r_i \rvert > c \end{cases}$	$\begin{cases} r_i & ; \lvert r_i \rvert \le c \\ c\, sign(r_i); & \lvert r_i \rvert > c \end{cases}$	$\begin{cases} 1 & ; \lvert r_i \rvert \le c \\ \frac{c}{\lvert r_i \rvert} & ; \lvert r_i \rvert > c \end{cases}$
	$c =$ 1.345, 1.5, 1.7, 2.08		
Welsch	$\frac{c^2}{2}\left(1 - e^{-\left(\frac{r_i}{c}\right)^2}\right); \lvert r_i \rvert < \infty$	$r_i\, e^{-\left(\frac{r_i}{c}\right)^2} \quad ; \lvert r_i \rvert < \infty$	$e^{-\left(\frac{r_i}{c}\right)^2} \quad ; \lvert r_i \rvert < \infty$
	c = 2.4, 2.985		

3.2.3 Trimmed Least Squares Regression [3][23][24]

Rousseeuw and Yohai (1984) proposed the trimmed least squares (TLS) estimator regression. Extending from the trimmed mean, TLS regression minimizes the *h* out of *n* ordered squared residuals. So, the objective function is minimize the sum of the smallest *h* of the squared residuals and is defined as:

$$min \sum_{i=1}^{h} r_{(i)}^2 \tag{22}$$

where $r_{(i)}^2$ represents the i^{th} ordered squared residuals $r_{(1)}^2 \le r_{(2)}^2 \le \cdots \le r_{(n)}^2$ and h is called the trimming constant which has to satisfy $\frac{n}{2} < h < n$. This constant, h , determines the breakdown point of the TLS estimator. Using $h = [(n / 2) + 1]$ ensures that the estimator has a breakdown point equal to 50%. When $h = n$, TLS is exactly equivalent to OLS estimator whose breakdown point is 0%. Rousseeuw and Leroy (1987) recommended $h = [n\,(1- \alpha) + 1]$ where α is the trimmed percentage. This estimator is attractive because can be selected to prevent some of the poor results other 50% breakdown estimator show. TLS can be fairly efficient if the number of trimmed observations is close to the number of outliers because OLS is used to estimate parameters from the remaining *h* observations.

4. Suggested Estimators

4.1 First Suggested Estimator: in this estimator, we suggest to modifying Thiel estimator (complete and incomplete method). Thiel suggest using the median as a robust estimator of location instead of the mean in OLS. So, we suggest using the Gastwirth's estimator instead of median in Thiel estimator in order to not exclude too much of the information from the regression. Gastwirth's location estimator is a weighted sum of three order statistics. It is based on median with two ordered observations and therefore it contains information regarding the sample more than the median. The formula Gastwirth's location estimator is [9]:

$$\mathrm{GAS} = 0.3\, \mathrm{x}_{\left[\frac{n}{3}+1\right]} + 0.4\ \mathrm{median} + 0.3\, \mathrm{x}_{\left(n-\left[\frac{n}{3}\right]\right)} \tag{23}$$

Where: $\left[\frac{n}{3}+1\right]$: The integer part of the real number $\left(\frac{n}{3}+1\right)$ and $\left[\frac{n}{3}\right]$: The integer part of the real number $\left(\frac{n}{3}\right)$.

4.2 Second Suggested Estimator: in this estimator, we suggest to use the following function as M-estimator which satisfies the proprieties of $\boldsymbol{\rho}$ function.

$$\rho(r_i) = \frac{c}{18}\log\left(1+\left(\frac{3r_i}{c}\right)^2\right) \qquad ; |r_i| < \infty \ , c = 9 \qquad (24)$$

The $\boldsymbol{\psi}$ function will be as follow:

$$\psi(r_i) = \frac{r_i/c}{1+\left(\frac{3r_i}{c}\right)^2} \qquad ; |r_i| < \infty \ , c = 9 \qquad (25)$$

5. Simulation Study

In this section we introduced the simulation study which has been carried out to illustrate the robustness of the estimators under different cases. Simulation was used to compare the mean squares error (MSE) and relative mean squares error (RMSE) of the estimates of regression coefficients and model by using the ordinary least squares (OLS); least absolute deviation (LAD); nonparametric estimators contains "complete Thiel's estimator (CTH) and incomplete Thiel's estimator (ITH) with Conover's estimator for intercept"; suggested nonparametric estimator contains "complete Gastwirth's estimator (CGAS) and incomplete Gastwirth's estimator (IGAS) with Conover's estimator for intercept"; M-estimators "Huber's M-estimators with c=1.345 (H-M), Welsch's M-estimators with c=2.4 (W-M) and suggested M-estimator (SU-M)" and trimmed least squares (TLS) with proportion of trimmed (α) equal to (10%, 20%, 30% and 40%). The data sets are generated from the simple linear regression model as: $y_i = 1 + 3x_i + \varepsilon_i$ which means that the true value of regression parameters are $\beta_0 = 1$ and $\beta_1 = 3$. Since the parameters known, a detailed comparison can be made. The process was repeated 1000 times to obtain 1000 independent samples of *Y* and *X* of size *n*. The sample sizes varied from small (10), to medium (30) and large (50). In order to cover the effects of various situations on the regression coefficients and model, samples were classified into four cases, three of them where contaminated with outliers. In addition, three percentages of outliers (δ) were considered, δ =10%, 20% and 30%. We treated with normal and contaminated normal distribution. The simulation programs were written using Visual Basic6 programming language.

Case (1) No-outliers "Normal Case":

- Generate errors, $\varepsilon_i \sim N(0,1)$; $i = 1,2,\dots,n$.
- Generate the values of independent variable, $x_i \sim N(0,100)$; $i = 1,2,\dots,n$.
- Compute the y_i values.

Case (2) X-outliers:

- Generate errors, $\varepsilon_i \sim N(0,1)$; $i = 1,2,\dots,n$.
- Generate the values of independent variable with no X-outliers, $x_i \sim N(0,100)$; $i = 1,2,\dots,n\,(1-\delta)$.
- Generate (n δ) of X-outliers for the values of independent variable , $x_i \sim N(100,100)$; $i = n\,(1-\delta)+1, n\,(1-\delta)+2,\dots,n$.
- Compute the y_i values.

Case (3) Y-outliers:

- Generate the values with no Y-outliers using errors, $\varepsilon_i \sim N(0,1)$; $i = 1,2,\dots,n\,(1-\delta)$.
- Generate the values with Y-outliers using errors, $\varepsilon_i \sim N(0,50)$; $i = n\,(1-\delta)+1, n\,(1-\delta)+2,\dots,n$.
- Generate the values of independent variable, $x_i \sim N(0,100)$; $i = 1,2,\dots,n$.
- Compute the y_i values.

Case (4) XY-outliers:

- Generate the values with no Y-outliers using errors, $\varepsilon_i \sim N(0,1)$; $i = 1,2,\dots,n\,(1-\delta)$.
- Generate the values with Y-outliers using errors, $\varepsilon_i \sim N(0,50)$; $i = n\,(1-\delta)+1, n\,(1-\delta)+2,\dots,n$.
- Generate the values of independent variable with no X-outliers, $x_i \sim N(0,100)$; $i = 1,2,\dots,n\,(1-\delta)$.
- Generate (n δ) of X-outliers for the values of independent variable , $x_i \sim N(100,100)$; $i = n\,(1-\delta)+1, n\,(1-\delta)+2,\dots,n$.
- Compute the y_i values.

For each case, random samples of size *n* were chosen and from each sample thus obtained, MSE and RMSE using OLS, LAD, CTH, ITH, CGAS, IGAS, H-M, W-M, SU-M, TLS10%, TLS 20%, TLS 30% and TLS 40% were found and compared. MSE can be a useful measure of the quality of parameter estimation and is computed as:

$$MSE(\hat{\beta}) = Var(\hat{\beta}) + [Bias(\hat{\beta})]^2 \quad (26)$$

$$Bias\ (\hat{\beta}) = \bar{\beta} - \beta\ ;\ Var\ (\hat{\beta}) = \frac{\sum_{I=1}^{R}\left(\hat{\beta}(I) - \bar{\beta}\right)^2}{R-1}\ ;\ \bar{\beta} = \frac{\sum_{I=1}^{R}\hat{\beta}(I)}{R}; R = 1000$$

$$MSE(\hat{Y}) = \frac{\sum_{I=1}^{R} MSE(\hat{Y}(I))}{R} \quad (27)$$

$$MSE(\hat{Y}(I)) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$$

A relative mean squares error has also been used as a measure of the quality of parameter estimation. We computed RMSE as:

$$RMSE(\hat{\beta}) = \frac{MSE(\hat{\beta}^{OLS}) - MSE(\hat{\beta}^{other\ method})}{MSE(\hat{\beta}^{OLS})} \quad (28)$$

$$RMSE(\hat{Y}) = \frac{MSE(\hat{Y}^{OLS}) - MSE(\hat{Y}^{other\ method})}{MSE(\hat{Y}^{OLS})} \quad (29)$$

The formulation (28) is useful for comparing estimator performance and is interpreted as a proportionate (or percent) change from baseline, using the OLS estimator MSE within a given data condition as a baseline value [21]. Positive values of RMSE refer to the proportional reduction in the MSE of a given estimator with respect to OLS estimation. Hence, RMSE is interpreted as a relative measure of performance above and beyond that of the OLS estimator.

6. Conclusions and Recommendations

Based on simulation results that have been shown in tables (2)…(5), the following conclusions could be reached:

Under ideal conditions (unit normal error distribution, no contamination) "normal case", table (2) note the following:

- OLS indicates the best performance (as expected) for all sample sizes. The decline in the performance of the rest of the estimation methods compare to the performance of ordinary least squares "which can be seen through the negative values for the RMSE" it is the only sacrifice paid by those methods in anticipation of the existence of outliers.
- Proposed method "SU-M" provided the second best performance of the estimates for all sample sizes, as well as provided a performance equal to the performance of OLS in estimating the slope with a sample size equal to 30 and 50, followed by the performance of both H-M and W-M respectively. Consequently, the method of M-estimations surpassed the performance of the alternative methods for OLS.
- In general, the MSE values of estimating the intercept are greater than the corresponding MSE values of estimating the slope. So, the results for intercept estimator need more consideration.
- The use of GAS estimator instead of median in Thiel method reduced inflation in MSE values of model as compared to OLS. From the value of RMSE we can see the reduction was between (22%-28%) for all sample sizes in complete method whereas was between (20%-26%) for $n = 30,50$ in incomplete method.
- As the sample size increases, the value of MSE decreases.
- LAD introduced better performance comparing with nonparametric estimators in estimating intercept and model.

Under contamination cases, tables (3), (4) and (5) note the following:

- Ordinary Least squares recorded a decline in performance when outlier exists while most of the other estimation methods are recorded good performances depending on the percentage and direction of contaminations.
- In general, TLS indicates the best performance for all sample sizes depending on the proportion of trimmed. TLS can be fairly efficient when the number of trimmed observations is close to the number of outliers because OLS is used to estimate parameters from the remaining *h* observations.

- ❖ The MSE values indicate that the degree of sensitivity of all methods, except the TLS in some situations, to the existence of outliers in *Y*–direction was small compared with the degree of sensitivity to the existence of outliers in the *X*–direction and *XY*–direction.
- ❖ LAD and M-estimators are very sensitive to the presence of outliers in *X*- direction and *XY*- direction. In addition, the negative values of RMSE of LAD and M-estimators in some results indicate that these methods are more affected by outliers comparing with OLS. Also, LAD estimators are more sensitive to outliers comparing with M-estimators especially for estimating intercept and model. So, LAD and M-estimators are not robust estimators against those directions, but they are robust estimators against outliers in *Y*-direction.
- ❖ Nonparametric estimators introduced better performance in the presence of outliers in *X*- direction and *XY*- direction comparing with OLS, LAD and M-estimators especially for estimating slope and model.
- ❖ Although the performance of nonparametric estimators are better than OLS in presence of outliers in *X*-direction and *XY*- direction, it seems less better in estimating intercept when we have no outliers, thus those estimators is not robust for estimation intercept according to criterion of a robust methods that is any robust method must be reasonably efficient when compared to the least squares estimators; if the underlying distribution of errors are independent normal, and substantially more efficient than least squares estimators, when there are outlying observations.
- ❖ The use of GAS estimator instead of median in Thiel method improves the performance of this method when outliers appear in *Y*-direction. Also this estimator improves the performance of this method in some cases when outliers appear in *X*-direction and *XY*-direction and the most improvements get when it is used in an incomplete method especially for estimating intercept and model with 10% percentage of contamination and for estimating slop and model with 30% percentage of contamination.
- ❖ In general, the MSE values decrease when the sample sizes increase while the MSE values increase as the proportion of contaminations "outliers" increases.

Now, after pointing to the conclusions that were obtained in the present work, the following Recommendations for future work are relevant:

- ❖ The poor performance of OLS estimators with the presence of outliers confirms our need for alternative methods. Therefore, before analyzing the data, we should first check the presence of outliers and then construct the necessary tests whether to see the underlying assumptions are satisfied. After that, we should conduct the appropriate estimation techniques.
- ❖ Choosing a nonparametric method, especially to estimate slope and model, or choosing a trimmed method when the outliers appear in *X*- direction or *XY*-direction.
- ❖ Choosing M-estimation and LAD method, or choosing a trimmed method when the outliers are appearing in *Y*-direction.
- ❖ When the outliers appear in *X*-direction or *XY*-direction, choose RMSE or mean absolute error (MAE) as criteria for comparing between methods to avoid dealing with the large values of MSE.

References

[1] Alma, Ö. G. (2011). "Comparison of Robust Regression Methods in Linear Regression".Int. J. Contemp. Math. Sciences, Vol. 6, No. 9, pp. 409- 421.

[2] Amphanthong, P. (2009). "Estimation of Regression Coefficients with Outliers". A Dissertation Submitted to the School of Applied Statistics, National Institute of Development Administration in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics.

[3] Bai , X. (2012). "Robust Linear Regression". A report submitted in partial fulfillment of the requirements for the degree Master of Science Department of Statistics College of Arts and Sciences Kansas State University, Manhattan, Kansas.

[4] Birkes, D. and Dodge, Y. (1993). "Alternative Methods of Regression". John Wiley & Sons Publication, New York.

[5] Conover, W.L. (1980) "Practical nonparametric statistics". Second edition, John Wiley & Sons Publication, New York.

[6] Dietz, E.J. (1986). "On estimating a slope and intercept in a nonparametric statistics course". Institute of Statistics Mimeograph Series No. 1689R, North Carolina State University.

[7] Dietz, E. J. (1987) "A Comparison of Robust Estimators in Simple Linear Regression" Communications in Statistics - Simulation and Computation, Vol. 16, Issue 4, pp. 1209-1227.

[8] Dodge, Y. (2008). "The Concise Encyclopedia of Statistics". Springer Science & Business Media.

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

[10] Gulasirima, R. (2005). "Robustifying Regression Models". A Dissertation Presented to the School of Applied Statistics National Institute of Development Administration in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics.

[11] Helsel D.R., and Hirsch, R.M. (2002). "Statistical methods in water resources - Hydrologic analysis and interpretation: Techniques of Water-Resources Investigations of the U.S. Geological Survey", chap. A3, book 4.

[12] Huber, P. J. (1973). "Robust regression: Asymptotics, conjectures and Monte Carlo". Ann. Stat., Vol. 1, pp. 799-821.

[13] Huber, P. J. (1981)."Robust Statistics". John Wiley & Sons Publication, New York.

[14] Hussain, S. S., and Sprent, P. (1983). Nonparametric Regression. Journal of the Royal Statistical Society, Series A, Vol. 146, pp. 182-191.

[15] Jajo, N. K. (1989). "Robust estimators in linear regression model". A Thesis Submitted to the Second Education College\ Ibn Al-Haitham\ Baghdad University in partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

[16] Marques de Sá, J. P. (2007). "Applied Statistics Using SPSS, STATISTICA, MATLAB and R". Second Edition, Springer-Verlag Berlin Heidelberg, New York.

[17] Meenai, Y. A. and Yasmeen, F. (2008). "Nonparametric Regression Analysis". Proceedings of 8th Islamic Countries Conference of Statistical Sciences, Vol. 12, PP. 415-424, Lahore-Pakistan.

[18] Midi, H., Uraibi , H. S. and Talib , B. A. (2009)." Dynamic Robust Bootstrap Method Based on LTS Estimators". European Journal of Scientific Research, Vol.32, No.3, pp. 277-287

[19] Mood, A. M. (1950). "Introduction to the theory of statistics". McGraw-Hill, New York.

[20] Mutan, O. C. (2004). "Comparison of Regression Techniques Via Monte Carlo Simulation". A Thesis Submitted to the Graduate School of Natural and Applied Sciences of Middle East Technical University in partial fulfillment of the requirements for the degree of Master of Science in Statistics.

[21] Nevitt, J. and Tam, H.P. (1998). "A comparison of robust and nonparametric estimators under the simple linear regression model: Multiple linear regression viewpoints, Vol. 25, pp. 54–69.

[22] Ronchetti, E.M. (1987) " Statistical Data Analysis Based on the L1-Norm and Related Methods". North-Holland, Amsterdam.

[23] Rousseeuw, P. J. and Leroy, A. M. (1987) "Robust Regression and Outlier Detection". John Wiley & Sons Publication, New York.

[24] Rousseeuw, P.J. and Yohai, V. (1984). "Robust regression by means of S-estimators, Lecture Notes in Statistics No.26, pp. 256-272, Springer Verilog, New York.

[25] Seber, G. A. and Lee, A. J. (2003). "Linear regression analysis". Second Edition, John Wiley & Sons Publication, New York.

[26] Stigler, S. M. (1986). "The History of Statistics: The Measurement of Uncertainty before 1900". Harvard University Press, Cambridge.

[27] Theil, H. (1950). "A rank - invariant method of linear and polynomial regression analysis". Indagationes Mathematicae, Vol. 12, pp. 85-91.

Table (2) MSE and RMSE results for estimating intercept, slope and model in Normal case "No outliers"

Methods	MSE									RMSE								
	n=10			n=30			n=50			n=10			n=30			n=50		
	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model
OLS	0.10922	0.00142	1.01042	0.03415	0.00038	0.99886	0.02004	0.00022	1.00518	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD	0.16302	0.00240	1.14429	0.05444	0.00057	1.03926	0.03085	0.00033	1.02849	-0.49258	-0.69014	-0.13249	-0.59414	-0.50000	-0.04045	-0.53942	-0.50000	-0.02319
CTH	0.48979	0.00186	1.52483	0.44160	0.00042	1.43921	0.41883	0.00025	1.42273	-3.48444	-0.30986	-0.50911	-11.93119	-0.10526	-0.44085	-19.89970	-0.13636	-0.41540
CGAS	0.26369	0.00193	1.23762	0.20078	0.00042	1.18333	0.19919	0.00025	1.18660	-1.41430	-0.35915	-0.22486	-4.87936	-0.10526	-0.18468	-8.93962	-0.13636	-0.18049
ITH	0.62606	0.01148	2.53317	0.45046	0.00203	1.60231	0.42230	0.00117	1.51406	-4.73210	-7.08451	-1.50705	-12.19063	-4.34211	-0.60414	-20.07285	-4.31818	-0.50626
IGAS	0.52406	0.04527	5.85065	0.20651	0.00199	1.34222	0.20173	0.00123	1.28272	-3.79821	-30.88028	-4.79031	-5.04714	-4.23684	-0.34375	-9.06637	-4.59091	-0.27611
H-M	0.11782	0.00148	1.02497	0.03672	0.00039	1.00294	0.02130	0.00023	1.00743	-0.07874	-0.04225	-0.01440	-0.07526	-0.02632	-0.00408	-0.06287	-0.04545	-0.00224
W-M	0.13328	0.00169	1.06097	0.04018	0.00041	1.00850	0.02272	0.00025	1.01018	-0.22029	-0.19014	-0.05003	-0.17657	-0.07895	-0.00965	-0.13373	-0.13636	-0.00497
SU-M	0.11426	0.00146	1.02040	0.03578	0.00038	1.00119	0.02072	0.00022	1.00644	-0.04615	-0.02817	-0.00988	-0.04773	0.00000	-0.00233	-0.03393	0.00000	-0.00125
TLS10%	0.10922	0.00142	1.01042	0.04392	0.00049	1.01909	0.02662	0.00028	1.01750	0.00000	0.00000	0.00000	-0.28609	-0.28947	-0.02025	-0.32834	-0.27273	-0.01226
TLS20%	0.15389	0.00207	1.11416	0.04979	0.00049	1.02678	0.02788	0.00029	1.01987	-0.40899	-0.45775	-0.10267	-0.45798	-0.28947	-0.02795	-0.39122	-0.31818	-0.01461
TLS30%	0.15471	0.00209	1.11557	0.04844	0.00049	1.02656	0.02818	0.00030	1.02010	-0.41650	-0.47183	-0.10407	-0.41845	-0.28947	-0.02773	-0.40619	-0.36364	-0.01484
TLS40%	0.15880	0.00215	1.12772	0.04855	0.00046	1.02324	0.02665	0.00028	1.01851	-0.45395	-0.51408	-0.11609	-0.42167	-0.21053	-0.02441	-0.32984	-0.27273	-0.01326

Table (3) MSE and RMSE results for estimating intercept, slope and model in X-outliers

Methods	Cont.	MSE									RMSE								
		n=10			n=30			n=50			n=10			n=30			n=50		
		$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model
OLS	10%	92.14852	7.77139	881.13670	33.85944	7.44083	773.48539	22.72030	7.36628	761.89726	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		142.95472	8.61501	988.32058	47.03417	7.50494	793.17882	30.55525	7.15163	751.07685	-0.55135	-0.10855	-0.12164	-0.38910	-0.00862	-0.02546	-0.34484	0.02914	0.01420
CTH		41.71366	0.00362	53.44327	28.80270	0.00076	32.01310	24.51816	0.00029	26.76118	0.54732	0.99953	0.93935	0.14935	0.99990	0.95861	-0.07913	0.99996	0.96488
CGAS		34.39636	0.00679	44.46512	25.00387	0.00237	28.08752	23.38522	0.00168	25.70860	0.62673	0.99913	0.94954	0.26154	0.99968	0.96369	-0.02927	0.99977	0.96626
ITH		41.52151	0.05229	56.84484	28.75252	0.00591	32.34772	24.47210	0.00328	26.93074	0.54941	0.99327	0.93549	0.15083	0.99921	0.95818	-0.07710	0.99955	0.96465
IGAS		34.62179	0.11031	53.40231	24.94683	0.00903	28.56069	23.20293	0.00822	26.02195	0.62428	0.98581	0.93939	0.26322	0.99879	0.96308	-0.02124	0.99888	0.96585
H-M		100.84701	7.97210	914.30566	34.60781	7.40867	770.81151	22.47252	7.30531	755.26305	-0.09440	-0.02583	-0.03764	-0.02210	0.00432	0.00346	0.01091	0.00828	0.00871
W-M		116.68310	8.18073	955.14990	35.44255	7.27480	756.90503	22.34606	7.14704	736.82034	-0.26625	-0.05267	-0.08400	-0.04676	0.02231	0.02144	0.01647	0.02976	0.03291
SU-M		98.41461	7.96376	908.76323	34.12612	7.42166	771.69564	22.47433	7.32628	757.40965	-0.06800	-0.02475	-0.03135	-0.00788	0.00258	0.00231	0.01083	0.00543	0.00589
TLS10%		92.14852	7.77139	881.13670	24.74188	5.64394	575.47523	12.66381	4.28680	434.99055	0.00000	0.00000	0.00000	0.26928	0.24149	0.25600	0.44262	0.41805	0.42907
TLS20%		0.12523	0.00168	1.05061	0.04727	0.00052	1.02652	0.02889	0.00030	1.02135	0.99864	0.99978	0.99881	0.99860	0.99993	0.99867	0.99873	0.99996	0.99866
TLS30%		0.16457	0.00227	1.14391	0.05106	0.00052	1.03225	0.02942	0.00031	1.02339	0.99821	0.99971	0.99870	0.99849	0.99993	0.99867	0.99871	0.99996	0.99866
TLS40%		0.16577	0.00234	1.14390	0.04902	0.00051	1.02956	0.02842	0.00030	1.02149	0.99820	0.99970	0.99870	0.99855	0.99993	0.99867	0.99875	0.99996	0.99866
OLS	20%	112.98425	8.25725	962.99030	42.34553	8.06222	846.88126	28.68101	8.03100	837.85766	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		167.03947	7.96122	1007.79403	62.16646	7.91205	851.94579	38.21483	8.06243	849.51701	-0.47843	0.03585	-0.04653	-0.46808	0.01863	-0.00598	-0.33241	-0.00391	-0.01392
CTH		140.53079	0.02580	178.82824	111.57893	0.01076	121.59829	105.58343	0.00999	112.10313	-0.24381	0.99688	0.81430	-1.63496	0.99867	0.85642	-2.68130	0.99876	0.86620
CGAS		82.02756	0.67273	169.92043	71.30200	0.54560	130.89948	73.77775	0.49494	127.32147	0.27399	0.91853	0.82355	-0.68381	0.93233	0.84543	-1.57236	0.93837	0.84804
ITH		134.83417	0.47755	205.57820	102.05895	0.18581	128.43478	98.41654	0.08961	112.28821	-0.19339	0.94217	0.78652	-1.41015	0.97695	0.84834	-2.43142	0.98884	0.86598
IGAS		76.74407	1.05676	199.45617	63.02252	0.82290	149.64999	62.80647	0.80117	147.04104	0.32075	0.87202	0.79288	-0.48829	0.89793	0.82329	-1.18983	0.90024	0.82450
H-M		122.28	8.252	973.720	43.490	8.063	849.12	28.703	8.028	838.02	-	0.000	-	-	-	-	-	0.000	-

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.5, 2013

		656	56	75	99	59	523	91	02	271	0.08233	57	0.01114	0.02705	0.00017	0.00265	0.00080	37	0.00020
W-M		137.06979	8.20760	989.39269	45.27453	8.06265	851.44899	29.30285	8.02707	838.74293	-0.21318	0.00601	-0.02742	-0.06917	-0.00005	-0.00539	-0.02168	0.00049	-0.00106
SU-M		119.47043	8.24900	970.37472	42.90259	8.06210	848.07987	28.58646	8.02900	837.81485	-0.05741	0.00100	-0.00767	-0.01316	0.00001	-0.00142	0.00330	0.00025	0.00005
TLS10%		112.98425	8.25725	962.99030	39.51743	7.79656	815.53271	25.60746	7.66649	796.69147	0.00000	0.00000	0.00000	0.06679	0.03295	0.03702	0.10716	0.04539	0.04913
TLS20%		106.16258	7.93173	915.04663	28.84589	5.89851	604.94071	14.58487	4.60907	469.98506	0.06038	0.03942	0.04979	0.31880	0.26838	0.28568	0.49148	0.42609	0.43906
TLS30%		0.14520	0.00209	1.11082	0.05104	0.00056	1.03401	0.03247	0.00032	1.02612	0.99871	0.99975	0.99885	0.99879	0.99993	0.99878	0.99887	0.99996	0.99878
TLS40%		0.17873	0.00248	1.18095	0.05462	0.00054	1.03643	0.03171	0.00032	1.02602	0.99842	0.99970	0.99877	0.99871	0.99993	0.99878	0.99889	0.99996	0.99878
OLS		127.68814	8.39148	1000.71126	51.27380	8.27527	879.60999	36.79886	8.23236	867.01551	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		195.77108	8.89689	1132.21602	68.58464	8.35473	907.45603	49.59132	8.20029	878.72903	-0.53320	-0.06023	-0.13141	-0.33762	-0.00960	-0.03166	-0.34763	0.00390	-0.01351
CTH		283.83580	0.76709	420.08288	287.38623	0.07950	316.27291	270.31273	0.06125	288.92922	-1.22288	0.90859	0.58022	-4.60493	0.99039	0.64044	-6.34568	0.99256	0.66675
CGAS		199.41562	1.29984	375.48329	202.53733	0.88383	305.74674	228.14737	0.85378	325.27757	-0.56174	0.84510	0.62478	-2.95011	0.89320	0.65241	-5.19985	0.89629	0.62483
ITH		105.86162	7.36218	874.27762	40.50964	7.11848	753.48888	24.89534	7.06094	736.68298	0.17094	0.12266	0.12634	0.20993	0.13979	0.14338	0.32348	0.14229	0.15032
IGAS		95.35216	4.27891	554.15977	57.40000	4.52308	510.71883	58.60865	4.12733	478.34122	0.25324	0.49009	0.44623	-0.11948	0.45342	0.41938	-0.59268	0.49865	0.44829
H-M	30 %	137.21961	8.36506	1009.50396	53.61668	8.27498	882.17463	36.94705	8.23448	867.75962	-0.07465	0.00315	-0.00879	-0.04569	0.00004	-0.00292	-0.00403	-0.00026	-0.00086
W-M		156.62945	8.37659	1036.18068	56.39295	8.27780	885.55999	37.80307	8.23498	868.88008	-0.22666	0.00177	-0.03544	-0.09984	-0.00031	-0.00676	-0.02729	-0.00032	-0.00215
SU-M		134.85886	8.36503	1006.27900	52.62260	8.27310	880.98021	36.74888	8.23220	867.16309	-0.05616	0.00315	-0.00556	-0.02631	0.00026	-0.00156	0.00136	0.00002	-0.00017
TLS10%		127.68814	8.39148	1000.71126	48.95759	8.18363	867.13673	33.82941	8.10130	850.13129	0.00000	0.00000	0.00000	0.04517	0.01107	0.01418	0.08069	0.01592	0.01947
TLS20%		123.56879	8.31673	986.56326	44.71163	7.87599	830.46767	29.76339	7.74000	807.98511	0.03226	0.00891	0.01414	0.12798	0.04825	0.05587	0.19119	0.05981	0.06808
TLS30%		117.26398	8.01869	938.40186	33.48191	6.17508	638.22216	17.90232	4.93872	506.77577	0.08164	0.04442	0.06227	0.34700	0.25379	0.27443	0.51351	0.40008	0.41549
TLS40%		0.16697	0.00238	1.17131	0.05524	0.00061	1.04423	0.03579	0.00035	1.03276	0.99869	0.99972	0.99883	0.99892	0.99993	0.99881	0.99903	0.99996	0.99881

Table (4) MSE and RMSE results for estimating intercept, slope and model in Y-outliers

Methods	Cont.	*MSE*									*RMSE*								
		n=10			*n=30*			*n=50*			*n=10*			*n=30*			*n=50*		
		$\widehat{\beta}_0$	$\widehat{\beta}_1$	*Model*	$\widehat{\beta}_0$	$\widehat{\beta}_1$	*Model*	$\widehat{\beta}_0$	$\widehat{\beta}_1$	*Model*	$\widehat{\beta}_0$	$\widehat{\beta}_1$	*Model*	$\widehat{\beta}_0$	$\widehat{\beta}_1$	*Model*	$\widehat{\beta}_0$	$\widehat{\beta}_1$	*Model*
OLS		0.71939	0.00849	1.99194	0.18903	0.00140	1.24560	0.10963	0.00053	1.15527	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		0.19661	0.00233	1.19259	0.06365	0.00010	1.04978	0.03662	0.00041	1.03375	0.72670	0.72556	0.40129	0.66328	0.92857	0.15721	0.66597	0.22642	0.10519
CTH		2.10039	0.00223	3.45415	1.71485	0.00061	2.75049	1.20374	0.00035	2.22141	-1.91968	0.73734	-0.73406	-8.07184	0.56429	-1.20816	-9.98002	0.33962	-0.92285
CGAS		1.25016	0.00286	2.39309	0.79806	0.00069	1.78167	0.52011	0.00039	1.50252	-0.73781	0.66313	-0.20139	-3.22187	0.50714	-0.43037	-3.74423	0.26415	-0.30058
ITH		2.31557	0.02085	5.25097	1.72771	0.00253	3.01769	1.21025	0.00093	2.35407	-2.21880	-1.45583	-1.63611	-8.13987	-0.80714	-1.42268	-10.03941	-0.75472	-1.03768
IGAS		1.96417	0.02092	4.56526	0.80973	0.00277	2.06392	0.52312	0.00113	1.65883	-1.73033	-1.46408	-1.29187	-3.28361	-0.97857	-0.65697	-3.77169	-1.13208	-0.43588
H-M	10 %	0.16904	0.00184	1.09828	0.05009	0.00005	1.01343	0.02845	0.00032	1.01260	0.76502	0.78327	0.44864	0.73502	0.96429	0.18639	0.74049	0.39623	0.12349
W-M		0.15844	0.00166	1.13540	0.04570	0.00004	1.01947	0.02653	0.00029	1.01642	0.77976	0.80448	0.43000	0.75824	0.97143	0.18154	0.75800	0.45283	0.12019
SU-M		0.16401	0.00187	1.11355	0.04966	0.00005	1.01489	0.02785	0.00031	1.01321	0.77202	0.77974	0.44097	0.73729	0.96429	0.18522	0.74596	0.41509	0.12297
TLS10%		0.71939	0.00849	1.99194	0.05175	0.00055	1.02770	0.02623	0.00029	1.01672	0.00000	0.00000	0.00000	0.72623	0.60714	0.17494	0.76074	0.45283	0.11993
TLS20%		0.13241	0.00176	1.07408	0.04869	0.00052	1.02819	0.02921	0.00030	1.02152	0.81594	0.79270	0.46079	0.74242	0.62857	0.17454	0.73356	0.43396	0.11577
TLS30%		0.16429	0.00226	1.13973	0.05049	0.00053	1.03144	0.02978	0.00031	1.02310	0.77163	0.73380	0.42783	0.73290	0.62143	0.17193	0.72836	0.41509	0.11441
TLS40%		0.16469	0.00232	1.14446	0.04811	0.00050	1.02851	0.02826	0.00030	1.02110	0.77107	0.72674	0.42545	0.74549	0.64286	0.17429	0.74222	0.43396	0.11614
OLS		1.24561	0.01631	2.96421	0.34695	0.00321	1.48916	0.20934	0.00156	1.31311	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD	20 %	0.27746	0.00402	1.34070	0.07898	0.00018	1.06509	0.04403	0.00002	1.04256	0.77725	0.75353	0.54770	0.77236	0.94393	0.28477	0.78967	0.98718	0.20604
CTH		4.38498	0.00404	6.18636	2.69678	0.00089	3.78531	1.94557	0.00052	2.97954	-2.52035	0.75230	-1.08702	-6.77282	0.72274	-1.54191	-8.29383	0.66667	-1.26907
CGA		2.41917	0.00645	3.87849	1.26658	0.00120	2.28543	0.83765	0.00071	1.82992	-0.94216	0.60454	-0.30844	-2.65061	0.62617	-0.53471	-3.00139	0.54487	-0.39358

S																			
ITH		4.77556	0.04240	10.16990	2.74584	0.00515	4.32012	1.95733	0.00209	3.21397	-2.83391	-1.59963	-2.43090	-6.91422	-0.60436	-1.90104	-8.35000	-0.33974	-1.44760
IGAS		2.27029	0.06305	7.33323	1.28866	0.00615	2.86091	0.84595	0.00325	2.15007	-0.82263	-2.86573	-1.47392	-2.71425	-0.91589	-0.92116	-3.04103	-1.08333	-0.63739
H-M		0.31072	0.00436	1.33567	0.07100	0.00019	1.03575	0.04020	0.00003	1.02504	0.75055	0.73268	0.54940	0.79536	0.94081	0.30447	0.80797	0.98077	0.21938
W-M		0.21418	0.00307	1.27801	0.05410	0.00005	1.03267	0.03280	0.00001	1.02542	0.82805	0.81177	0.56885	0.84407	0.98442	0.30654	0.84332	0.99359	0.21909
SU-M		0.31339	0.00467	1.37790	0.07164	0.00022	1.04223	0.04062	0.00002	1.02857	0.74840	0.71367	0.53515	0.79351	0.93146	0.30012	0.80596	0.98718	0.21669
TLS10%		1.24561	0.01631	2.96421	0.13142	0.00139	1.15600	0.06298	0.00070	1.07290	0.00000	0.00000	0.00000	0.62121	0.56698	0.22372	0.69915	0.55128	0.18293
TLS20%		0.37974	0.00591	1.56266	0.05013	0.00057	1.03253	0.03116	0.00033	1.02486	0.69514	0.63765	0.47282	0.85551	0.82243	0.30664	0.85115	0.78846	0.21952
TLS30%		0.15605	0.00219	1.13857	0.05316	0.00055	1.03555	0.03172	0.00031	1.02547	0.87472	0.86573	0.61589	0.84678	0.82866	0.30461	0.84848	0.80128	0.21905
TLS40%		0.17510	0.00246	1.16908	0.05208	0.00056	1.03682	0.03133	0.00032	1.02563	0.85943	0.84917	0.60560	0.84989	0.82555	0.30376	0.85034	0.79487	0.21893
OLS		1.81333	0.02272	3.90077	0.52327	0.00537	1.77948	0.30808	0.00283	1.46887	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		0.44183	0.00784	1.70824	0.10259	0.00043	1.09578	0.05685	0.00008	1.05652	0.75634	0.65493	0.56208	0.80394	0.91993	0.38421	0.81547	0.97173	0.28073
CTH		6.51525	0.00839	8.89114	3.82712	0.00148	5.01438	2.47687	0.00079	3.53595	-2.59298	0.63072	-1.27933	-6.31385	0.72439	-1.81789	-7.03970	0.72085	-1.40726
CGAS		3.59857	0.01202	5.54886	1.78310	0.00235	2.88852	1.12133	0.00129	2.14221	-0.98451	0.47095	-0.42250	-2.40761	0.56238	-0.62324	-2.63974	0.54417	-0.45841
ITH		7.17127	0.09187	18.02163	3.92555	0.01153	6.08280	2.48396	0.00490	3.99974	-2.95475	-3.04357	-3.62002	-6.50196	-1.14711	-2.41830	-7.06271	-0.73145	-1.72300
IGAS		5.68709	0.10821	16.06516	1.84011	0.01309	4.01477	1.13824	0.00773	2.82842	-2.13627	-3.76276	-3.11846	-2.51656	-1.43762	-1.25615	-2.69462	-1.73145	-0.92558
H-M	30 %	0.57779	0.01020	1.87788	0.11520	0.00073	1.09962	0.06054	0.00016	1.05229	0.68137	0.55106	0.51859	0.77985	0.86406	0.38206	0.80349	0.94346	0.28361
W-M		0.37819	0.00617	1.60943	0.07639	0.00028	1.07039	0.04481	0.00005	1.04372	0.79144	0.72843	0.58741	0.85401	0.94786	0.39848	0.85455	0.98233	0.28944
SU-M		0.62472	0.01083	1.97187	0.12363	0.00088	1.12225	0.06445	0.00023	1.06383	0.65548	0.52333	0.49449	0.76374	0.83613	0.36934	0.79080	0.91873	0.27575
TLS10%		1.81333	0.02272	3.90077	0.25322	0.00268	1.34940	0.12663	0.00136	1.17732	0.00000	0.00000	0.00000	0.51608	0.50093	0.24169	0.58897	0.51943	0.19849
TLS20%		0.76972	0.01015	2.21349	0.09292	0.00099	1.10261	0.05210	0.00055	1.05700	0.57552	0.55326	0.43255	0.82242	0.81564	0.38038	0.83089	0.80565	0.28040
TLS30%		0.30647	0.00490	1.45468	0.05898	0.00063	1.04510	0.03498	0.00037	1.03341	0.83099	0.78433	0.62708	0.88729	0.88268	0.41269	0.88646	0.86926	0.29646
TLS40%		0.17569	0.00241	1.19141	0.05542	0.00060	1.04489	0.03443	0.00034	1.03092	0.90311	0.89393	0.69457	0.89409	0.88827	0.41281	0.88824	0.87986	0.29815

Table (5) MSE and RMSE results for estimating intercept, slope and model in XY-outliers

Methods	Cont.	MSE									RMSE								
		n=10			n=30			n=50			n=10			n=30			n=50		
		$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Model
OLS		92.04049	7.77059	880.90433	33.93830	7.43896	773.16320	22.71535	7.36523	761.84821	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		142.82002	8.61155	987.64781	46.86974	7.46566	788.75399	30.52518	7.11496	747.23289	-0.55171	-0.10822	-0.12117	-0.38103	-0.00359	-0.02016	-0.34381	0.03398	0.01918
CTH		41.46591	0.00362	53.03090	28.84611	0.00195	32.00093	24.41959	0.00132	26.63442	0.54948	0.99953	0.93980	0.15004	0.99974	0.95861	-0.07503	0.99982	0.96504
CGAS		34.56729	0.00679	44.53948	25.10213	0.00355	28.14972	23.18144	0.00319	25.47831	0.62443	0.99913	0.94944	0.26036	0.99952	0.96359	-0.02052	0.99957	0.96656
ITH		41.49638	0.05124	56.54779	28.79371	0.00591	32.33968	24.37305	0.00328	26.79981	0.54915	0.99341	0.93581	0.15159	0.99921	0.95817	-0.07298	0.99955	0.96482
IGAS	10 %	34.87539	0.11025	53.51295	25.05552	0.00903	28.62422	23.01355	0.00822	25.79443	0.62109	0.98581	0.93925	0.26173	0.99879	0.96298	-0.01313	0.99888	0.96614
H-M		100.63634	7.97149	914.15092	34.58313	7.38816	768.39006	22.39272	7.28738	753.34853	-0.09339	-0.02585	-0.03774	-0.01900	0.00683	0.00617	0.01420	0.01057	0.01116
W-M		116.28996	8.17873	954.56846	35.27835	7.22388	750.63749	21.94021	7.03813	724.91953	-0.26347	-0.05252	-0.08362	-0.03948	0.02891	0.02913	0.03412	0.04441	0.04847
SU-M		98.23576	7.96333	908.69623	34.13698	7.40587	769.79875	22.41553	7.31204	755.90321	-0.06731	-0.02480	-0.03155	-0.00585	0.00445	0.00435	0.01320	0.00722	0.00780
TLS10%		92.04049	7.77059	880.90433	24.17195	5.45501	556.67062	12.06406	4.08871	414.97693	0.00000	0.00000	0.00000	0.28777	0.26670	0.28001	0.46890	0.44486	0.45530
TLS20%		0.12523	0.00168	1.05061	0.04727	0.00052	1.02652	0.02889	0.00030	1.02135	0.99864	0.99978	0.99881	0.99861	0.99993	0.99867	0.99873	0.99996	0.99866

TLS30%		0.16457	0.00227	1.14391	0.05106	0.00052	1.02925	0.02942	0.00031	1.02339	0.99821	0.99971	0.99870	0.99850	0.99993	0.99867	0.99870	0.99996	0.99866
TLS40%		0.16577	0.00234	1.14390	0.04902	0.00051	1.02956	0.02842	0.00030	1.02149	0.99820	0.99970	0.99870	0.99856	0.99993	0.99867	0.99875	0.99996	0.99866
OLS	20%	112.77503	8.25371	963.06454	42.37761	8.05852	846.43984	28.65975	8.02745	837.54585	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		166.73751	7.94649	1006.75204	61.78185	7.89547	849.83400	37.97239	8.04644	847.72111	-0.47850	0.03722	-0.04536	-0.45789	0.02023	-0.00401	-0.32494	-0.00237	-0.01215
CTH		139.99546	0.03106	178.17154	110.36348	0.01255	120.36321	104.17848	0.01123	110.73622	-0.24137	0.99624	0.81500	-1.60429	0.99844	0.85780	-2.63501	0.99860	0.86778
CGAS		82.70022	0.68038	171.04904	70.24312	0.56919	132.22269	71.05673	0.53716	128.76433	0.26668	0.91757	0.82239	-0.65755	0.92937	0.84379	-1.47932	0.93308	0.84626
ITH		135.04957	0.47345	205.95563	101.35361	0.18561	127.59519	97.53820	0.08959	111.32251	-0.19751	0.94264	0.78615	-1.39168	0.97697	0.84926	-2.40332	0.98884	0.86708
IGAS		77.71810	1.05393	200.15723	63.15645	0.81908	149.42471	62.27748	0.79795	146.09923	0.31086	0.87231	0.79217	-0.49033	0.89836	0.82347	-1.17299	0.90060	0.82556
H-M		122.01374	8.23766	972.63403	43.52678	8.04997	847.60180	28.57509	8.01483	836.60179	-0.08192	0.00194	-0.00994	-0.02712	0.00106	-0.00137	0.00295	0.00157	0.00113
W-M		137.51693	8.18704	988.09590	45.38417	8.04112	849.13212	29.08051	8.00712	836.54617	-0.21939	0.00808	-0.02599	-0.07095	0.00216	-0.00318	-0.01468	0.00253	0.00119
SU-M		119.39188	8.23854	969.99363	42.88482	8.05106	846.84890	28.45965	8.01834	836.67153	-0.05867	0.00184	-0.00719	-0.01197	0.00093	-0.00048	0.00698	0.00113	0.00104
TLS10%		112.77503	8.25371	963.06454	38.84923	7.64031	799.08631	25.09340	7.48600	777.90494	0.00000	0.00000	0.00000	0.08326	0.05190	0.05594	0.12444	0.06745	0.07121
TLS20%		103.87399	7.73738	892.62212	27.17352	5.59538	573.61776	13.70942	4.33090	441.47918	0.07893	0.06256	0.07314	0.35878	0.30566	0.32232	0.52165	0.46049	0.47289
TLS30%		0.14520	0.00209	1.11082	0.05104	0.00056	1.04401	0.03147	0.00032	1.02612	0.99871	0.99975	0.99885	0.99880	0.99993	0.99877	0.99890	0.99996	0.99877
TLS40%		0.17873	0.00248	1.18095	0.05462	0.00054	1.03864	0.03171	0.00032	1.02670	0.99842	0.99970	0.99877	0.99871	0.99993	0.99877	0.99889	0.99996	0.99877
OLS	30%	127.58125	8.38863	1000.68538	51.30234	8.27313	879.25906	36.79661	8.23023	866.87473	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAD		195.62092	8.88976	1131.62629	68.10293	8.33614	905.33276	49.37628	8.18894	877.42982	-0.53330	-0.05974	-0.13085	-0.32748	-0.00762	-0.02965	-0.34187	0.00502	-0.01218
CTH		242.43311	1.47948	431.08093	258.31211	0.19724	295.95882	245.55224	0.14272	270.96148	-0.90023	0.82363	0.56921	-4.03509	0.97616	0.66340	-5.67323	0.98266	0.68743
CGAS		180.33850	1.64453	382.67620	190.47354	1.01165	305.01228	215.18578	0.95815	322.04720	-0.41352	0.80396	0.61759	-2.71277	0.87772	0.65310	-4.84798	0.88358	0.62850
ITH		107.59932	7.33471	873.71298	40.77348	7.07379	749.38838	25.16128	7.01429	732.01484	0.15662	0.12564	0.12689	0.20523	0.14497	0.14770	0.31621	0.14774	0.15557
IGAS		96.15435	4.26860	553.14160	57.40188	4.50734	509.24734	58.11605	4.10988	476.23910	0.24633	0.49114	0.44724	-0.11889	0.45518	0.42082	-0.57939	0.50064	0.45063
H-M		136.88045	8.35275	1007.82410	53.63145	8.26437	880.78174	36.84258	8.22601	866.87072	-0.07289	0.00428	-0.00713	-0.04540	0.00106	-0.00173	-0.00125	0.00051	0.00000
W-M		153.56852	8.35918	1029.68486	56.39445	8.26344	883.76420	37.57574	8.22398	867.57333	-0.20369	0.00351	-0.02898	-0.09926	0.00117	-0.00512	-0.02117	0.00076	-0.00081
SU-M		134.62225	8.35699	1004.95095	52.68257	8.26538	879.95936	36.62526	8.22580	866.46767	-0.05519	0.00377	-0.00426	-0.02690	0.00094	-0.00080	0.00466	0.00054	0.00047
TLS10%		127.58125	8.38863	1000.68538	48.21096	8.06464	854.29603	33.24224	7.96301	835.66820	0.00000	0.00000	0.00000	0.06026	0.02520	0.02839	0.09660	0.03247	0.03600
TLS20%		121.40988	8.18474	972.65609	43.28637	7.61905	803.28910	28.62855	7.44984	777.70839	0.04837	0.02431	0.02801	0.15625	0.07906	0.08640	0.22198	0.09482	0.10286
TLS30%		114.20717	7.74069	908.71953	31.63137	5.84757	604.41014	16.73457	4.60937	473.25086	0.10483	0.07724	0.09190	0.38343	0.29319	0.31259	0.54521	0.43995	0.45407
TLS40%		0.16697	0.00238	1.17131	0.05524	0.00061	1.04423	0.03579	0.00035	1.03276	0.99869	0.99972	0.99883	0.99892	0.99993	0.99881	0.99903	0.99996	0.99881