Title
Robust linear regression: A review and comparison

Journal
COMMUNICATIONS IN STATISTICS-SIMULATION AND COMPUTATION, 46(8)

ISSN
0361-0918

Authors
Yu, Chun
Yao, Weixin

Publication Date
2017

DOI
10.1080/03610918.2016.1202271

Peer reviewed

Robust Linear Regression: A Review and Comparison

Chun Yu * and Weixin Yao, †

Abstract

Ordinary least-squares (OLS) estimators for a linear model are very sensitive to unusual values in the design space or outliers among y values. Even one single atypical value may have a large effect on the parameter estimates. This article aims to review and describe some available and popular robust techniques, including some recent developed ones, and compare them in terms of breakdown point and efficiency. In addition, we also use a simulation study and a real data application to compare the performance of existing robust methods under different scenarios.

Key words: Breakdown point; Robustness; Outliers; Linear Regression.

1 Introduction

Linear regression has been one of the most important statistical data analysis tools. Given the independent and identically distributed (iid) observations $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$,

*Chun Yu is Instructor, School of Statistics, Jiangxi University of Finance and Economics, Nanchang, 330013, China. Email: chunyu@jxufe.edu.cn.

†Weixin Yao is corresponding author, Associate Professor, Department of Statistics, University of California, Riverside, California 92521, U.S.A. Email: weixin.yao@ucr.edu. Weixin Yao's research is supported by NSF grant DMS-1461677.

in order to understand how the response y_is are related to the covariates $\boldsymbol{x}_i$s, we traditionally assume the following linear regression model

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \tag{1.1}$$

where $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector, and the ε_is are i.i.d. and independent of $\boldsymbol{x}_i$ with $\mathrm{E}(\varepsilon_i \mid \boldsymbol{x}_i) = 0$. The most commonly used estimate for $\boldsymbol{\beta}$ is the ordinary least-squares (OLS) estimate which minimizes the sum of squared residuals

$$\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2. \tag{1.2}$$

However, it is well known that the OLS estimate is extremely sensitive to the outliers. A single outlier can have large effect on the OLS estimate.

In this paper, we review and describe some available robust methods. In addition, a simulation study and a real data application are used to compare different existing robust methods. The efficiency and breakdown point (Donoho and Huber, 1983) are two traditionally used important criteria to compare different robust methods. The efficiency is used to measure the relative efficiency of the robust estimate compared to the OLS estimate when the error distribution is exactly normal and there are no outliers. Breakdown point is to measure the proportion of outliers an estimate can tolerate before it goes to infinity. In this paper, finite sample breakdown point (Donoho and Huber, 1983) is used and defined as follows: Let $\mathbf{z}_i = (\boldsymbol{x}_i, y_i)$. Given any sample $\boldsymbol{z} = (\boldsymbol{z}_i, \ldots, \boldsymbol{z}_n)$, denote $T(\boldsymbol{z})$ the estimate of the parameter $\boldsymbol{\beta}$. Let $\boldsymbol{z}'$ be the corrupted sample where any m of the original points of $\boldsymbol{z}$ are replaced by arbitrary bad data. Then the finite sample breakdown point δ^* is defined as

$$\delta^*(\boldsymbol{z}, T) = \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{\boldsymbol{z}'} \| T(\boldsymbol{z}') - T(\boldsymbol{z}) \| = \infty \right\}, \tag{1.3}$$

where $\|\cdot\|$ is the Euclidean norm.

Many robust methods have been proposed to achieve high breakdown point or high efficiency or both. M-estimates (Huber, 1981) are solutions of the normal equation with appropriate weight functions. They are resistant to unusual y observations, but sensitive to high leverage points on $\mathbf{x}$. Hence the breakdown point of an M-estimate is $1/n$. R-estimates (Jackel, 1972) which minimize the sum of scores of the ranked residuals have relatively high efficiency but their breakdown points are as low as those of OLS estimates. Least Median of Squares (LMS) estimates (Siegel, 1982) which minimize the median of squared residuals, Least Trimmed Squares (LTS) estimates (Rousseeuw, 1983) which minimize the trimmed sum of squared residuals, and S-estimates (Rousseeuw and Yohai, 1984) which minimize the variance of the residuals all have high breakdown point but with low efficiency. Generalized S-estimates (GS-estimates) (Croux *et al.*, 1994) maintain high breakdown point as S-estimates and have slightly higher efficiency. MM-estimates proposed by Yohai (1987) can simultaneously attain high breakdown point and efficiencies. Mallows Generalized M-estimates (Mallows, 1975) and Schweppe Generalized M-estimates (Handschin *et al.*, 1975) downweight the high leverage points on $\mathbf{x}$ but cannot distinguish "good" and "bad" leverage points, thus resulting in a loss of efficiencies. In addition, these two estimators have low breakdown points when p, the number of explanatory variables, is large. Schweppe one-step (S1S) Generalized M-estimates (Coakley and Hettmansperger, 1993) overcome the problems of Schweppe Generalized M-estimates and are calculated in one step. They both have high breakdown points and high efficiencies. Recently, Gervini and Yohai (2002) proposed a new class of high breakdown point and high efficiency robust estimate called robust and efficient weighted least squares estimator (REWLSE). Lee *et al.* (2012) and She and Owen (2011) proposed a new class of robust methods based on the regularization of case-specific parameters for each response. They further proved that the M-estimator with Huber's ψ function is a special case of their proposed estimator. Wilcox (1996) and You (1999) provided an excellent Monte Carlo comparison of some of the mentioned robust methods. This article aims to provide a more complete review and comparison of existing robust methods including some recently developed robust methods. In addition, besides comparing regression estimates for different robust methods, we also add the comparison of their performance of outlier detection based on three criteria used by She and Owen (2011).

The rest of the paper is organized as follows. In Section 2, we review and describe some of the available robust methods. In Section 3, a simulation study and a real data application are used to compare different robust methods. Some discussions are given in Section 4.

2 Robust Regression Methods

2.1 M-Estimates

By replacing the least squares criterion (1.2) with a robust criterion, M-estimate (Huber, 1964) of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}}{\hat{\sigma}}\right), \tag{2.1}$$

where $\rho(\cdot)$ is a robust loss function and $\hat{\sigma}$ is an error scale estimate. The derivative of ρ, denoted by $\psi(\cdot) = \rho'(\cdot)$, is called the influence function. In particular, if $\rho(t) = \frac{1}{2}t^2$, then the solution is the OLS estimate. The OLS estimate is very sensitive to outliers. Rousseeuw and Yohai (1984) indicated that OLS estimates have a breakdown point (BP) of BP $= 1/n$, which tends to zero when the sample size n is getting large. Therefore, one single unusual observation can have large impact on the OLS estimate.

One of the commonly used robust influence functions is Huber's ψ function (Huber, 1981), where $\psi_c(t) = \rho'(t) = \max\{-c, \min(c,t)\}$. Huber (1981) recommends using

$c = 1.345$ in practice. This choice produces a relative efficiency of approximate 95% when the error density is normal. Another possibility for $\psi(\cdot)$ is Tukey's bisquare function $\psi_c(t) = t\{1-(t/c)^2\}_+^2$. The use of $c = 4.685$ produces 95% efficiency. Bai, Yao, and Boyer (2012) also applied Huber's ψ function and Tukey's bisquare function to provide robust fitting of mixture regression models. If $\rho(t) = |t|$, then *least absolute deviation* (LAD, also called median regression) estimates are achieved by minimizing the sum of the absolute values of the residuals

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \right|. \tag{2.2}$$

The LAD is also called L_1 estimate due to the L_1 norm used. Although LAD is more resistent than OLS to unusual y values, it is sensitive to high leverage outliers, and thus has a breakdown point of BP $= 1/n \to 0$ when the sample size n is getting large (Rousseeuw and Yohai, 1984). Moreover, LAD estimates have a low efficiency of 0.64 when the errors are normally distributed. Similar to LAD estimates, the general monotone M-estimates, i.e., M-estimates with monotone ψ functions, have a BP $= 1/n \to 0$ as n becomes infinity due to lack of immunity to high leverage outliers (Maronna *et al.*, 2006). Yao et al. (2012) proposed to use a kernel function for ρ to provide a robust and efficient estimate for nonparametric regression by data adaptively choosing the tuning parameter. Their method can be also applied to linear regression.

2.2 LMS Estimates

The LMS estimates (Siegel, 1982) are found by minimizing the median of the squared residuals

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \text{Med}\{\left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}\right)^2\}. \tag{2.3}$$

One good property of the LMS estimate is that it possesses a high breakdown point of near 0.5. However, LMS estimates do not have a well-defined influence function because of its convergence rate of $n^{-\frac{1}{3}}$ and thus have zero efficiency (Rousseeuw, 1984). Despite these limitations, the LMS estimate can be used as an initial estimate for some other high breakdown point and high efficiency robust methods.

2.3 LTS Estimates

The LTS estimate (Rousseeuw, 1983) is defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{q} r_{(i)} (\boldsymbol{\beta})^2 , \tag{2.4}$$

where $r_{(1)} (\boldsymbol{\beta})^2 \leq \cdots \leq r_{(q)} (\boldsymbol{\beta})^2$ are ordered squared residuals, $q = [n (1 - \alpha) + 1]$, and α is the proportion of trimming. Using q = $\left(\frac{n}{2}\right)$ +1 ensures that the estimator has a breakdown point of BP = 0.5, and the convergence rate of $n^{-\frac{1}{2}}$ (Rousseeuw, 1983). Although highly resistent to outliers, LTS suffers badly in terms of very low efficiency, which is about 0.08, relative to OLS estimates (Stromberg *et al.*, 2000). The reason that LTS estimates call attentions to us is that it is traditionally used as an initial estimate for some other high breakdown point and high efficiency robust methods.

2.4 S-Estimates

S-estimates (Rousseeuw and Yohai, 1984) are defined by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \hat{\sigma} \left(r_1 (\boldsymbol{\beta}) , \cdots , r_n (\boldsymbol{\beta})\right) , \tag{2.5}$$

where $r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}$ and $\hat{\sigma}(r_1(\boldsymbol{\beta}), \cdots, r_n(\boldsymbol{\beta}))$ is the scale M-estimate which is defined as the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i(\boldsymbol{\beta})}{\hat{\sigma}}\right) = \delta, \tag{2.6}$$

for any given $\boldsymbol{\beta}$, where δ is taken to be $\mathrm{E}_\Phi[\rho(r)]$. For the biweight scale, S-estimates can attain a high breakdown point of BP = 0.5 and has an asymptotic efficiency of 0.29 under the assumption of normally distributed errors (Maronna *et al.*, 2006).

2.5 Generalized S-Estimates (GS-Estimates)

Croux *et al.* (1994) proposed generalized S-estimates in an attempt to improve the low efficiency of S-estimators. Generalized S-estimates are defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} S_n(\boldsymbol{\beta}), \tag{2.7}$$

where $S_n(\boldsymbol{\beta})$ is defined as

$$S_n(\boldsymbol{\beta}) = \sup\left\{S > 0; \binom{n}{2}^{-1}\sum_{i<j}\rho\left(\frac{r_i - r_j}{S}\right) \geq k_{n,p}\right\}, \tag{2.8}$$

where $r_i = y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}$, p is the number of regression parameters, and $k_{n,p}$ is a constant which depends on n and p. Particularly, if $\rho(x) = I(|x| \geq 1)$ and $k_{n,p} = \left(\binom{n}{2} - \binom{h_p}{2} + 1\right)/\binom{n}{2}$ with $h_p = \frac{n+p+1}{2}$, generalized S-estimator yields a special case, the least quartile difference (LQD) estimator, which is defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} Q_n(r_1, \ldots, r_n), \tag{2.9}$$

where

$$Q_n = \{|r_i - r_j|; i < j\}_{\binom{h_p}{2}} \tag{2.10}$$

is the $\binom{h_p}{2}$th order statistic among the $\binom{n}{2}$ elements of the set $\{|r_i - r_j|\,; i < j\}$. Generalized S-estimates have a breakdown point as high as S-estimates but with a higher efficiency.

2.6 MM-Estimates

First proposed by Yohai (1987), MM-estimates have become increasingly popular and are one of the most commonly employed robust regression techniques. The MM-estimates can be found by a three-stage procedure. In the first stage, compute an initial consistent estimate $\hat{\boldsymbol{\beta}}_0$ with high breakdown point but possibly low normal efficiency. In the second stage, compute a robust M-estimate of scale $\hat{\sigma}$ of the residuals based on the initial estimate. In the third stage, find an M-estimate $\hat{\boldsymbol{\beta}}$ starting at $\hat{\boldsymbol{\beta}}_0$.

In practice, LMS or S-estimate with Huber or bisquare functions is typically used as the initial estimate $\hat{\boldsymbol{\beta}}_0$. Let $\rho_0(r) = \rho_1\left(r/k_0\right)$, $\rho(r) = \rho_1\left(r/k_1\right)$, and assume that each of the ρ-functions is bounded. The scale estimate $\hat{\sigma}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{r_i\left(\hat{\boldsymbol{\beta}}\right)}{\hat{\sigma}}\right) = 0.5. \tag{2.11}$$

If the ρ-function is biweight, then $k_0 = 1.56$ ensures that the estimator has the asymptotic BP $= 0.5$. Note that an M-estimate minimizes

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n}\rho\left(\frac{r_i\left(\hat{\boldsymbol{\beta}}\right)}{\hat{\sigma}}\right). \tag{2.12}$$

Let ρ satisfy $\rho \leq \rho_0$. Yohai (1987) showed that if $\hat{\boldsymbol{\beta}}$ satisfies $L(\hat{\boldsymbol{\beta}}) \leq L(\hat{\boldsymbol{\beta}}_0)$, then $\hat{\boldsymbol{\beta}}$'s BP is not less than that of $\hat{\boldsymbol{\beta}}_0$. Furthermore, the breakdown point of the MM-estimate depends only on k_0 and the asymptotic variance of the MM-estimate depends only on k_1. We can choose k_1 in order to attain the desired normal efficiency without affecting its breakdown point. In order to let $\rho \leq \rho_0$, we must have $k_1 \geq k_0$; the larger the k_1 is, the higher efficiency the MM-estimate can attain at the normal distribution.

Maronna *et al.* (2006) provides the values of k_1 with the corresponding efficiencies of the biweight ρ-function. Please see the following table for more detail.

Efficiency	0.80	0.85	0.90	0.95
k_1	3.14	3.44	3.88	4.68

However, Yohai (1987) indicates that MM-estimates with larger values of k_1 are more sensitive to outliers than the estimates corresponding to smaller values of k_1. In practice, an MM-estimate with bisquare function and efficiency 0.85 ($k_1 = 3.44$) starting from a bisquare S-estimate is recommended.

2.7 Generalized M-Estimates (GM-Estimates)

2.7.1 Mallows GM-estimate

In order to make M-estimate resistent to high leverage outliers, Mallows (1975) proposed Mallows GM-estimate that is defined by

$$\sum_{i=1}^{n} w_i \psi \left\{ \frac{r_i\left(\hat{\boldsymbol{\beta}}\right)}{\hat{\sigma}} \right\} \boldsymbol{x}_i = 0, \tag{2.13}$$

where $\psi(e) = \rho'(e)$ and $w_i = \sqrt{1 - h_i}$ with h_i being the leverage of the *ith* observation. The weight w_i ensures that the observation with high leverage receives less weight than the observation with small leverage. However, even "good" leverage points that fall in line with the pattern in the bulk of the data are down-weighted, resulting in a loss of effiency.

2.7.2 Schweppe GM-estimate

Schweppe GM-estimate (Handschin *et al.*, 1975) is defined by the solution of

$$\sum_{i=1}^{n} w_i \psi \left\{ \frac{r_i\left(\hat{\boldsymbol{\beta}}\right)}{w_i \hat{\sigma}} \right\} \boldsymbol{x}_i = 0, \tag{2.14}$$

which adjusts the leverage weights according to the size of the residual r_i. Carroll and Welsch (1988) proved that the Schweppe estimator is not consistent when the errors are asymmetric. Furthermore, the breakdown points for both Mallows and Schweppe GM-estimates are no more than $1/(p+1)$, where p is the number of unknown parameters.

2.7.3 S1S GM-estimate

Coakley and Hettmansperger (1993) proposed Schweppe one-step (S1S) estimate , which extends from the original Schweppe estimator. S1S estimator is defined as

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_0 + \left[\sum_{i=1}^{n} \psi' \left(\frac{r_i\left(\hat{\boldsymbol{\beta}}_0\right)}{\hat{\sigma} w_i} \right) \boldsymbol{x}_i \boldsymbol{x}_i' \right]^{-1} \times \sum_{i=1}^{n} \hat{\sigma} w_i \psi \left(\frac{r_i\left(\hat{\boldsymbol{\beta}}_0\right)}{\hat{\sigma} w_i} \right) \boldsymbol{x}_i, \tag{2.15}$$

where the weight w_i is defined in the same way as Schweppe's GM-estimate.

The method for S1S estimate is different from the Mallows and Schweppe GM-estimates in that once the initial estimates of the residuals and the scale of the residuals are given, final M-estimates are calculated in one step rather than iteratively. Coakley and Hettmansperger (1993) recommended to use Rousseeuw's LTS for the initial estimates of the residuals and LMS for the initial estimates of the scale and proved that the S1S estimate gives a breakdown point of BP = 0.5 and results in 0.95 efficiency compared to the OLS estimate under the Gauss-Markov assumption.

2.8 R-Estimates

The R-estimate (Jackel, 1972) minimizes the sum of some scores of the ranked residuals

$$\sum_{i=1}^{n} a_n\left(R_i\right) r_i = min, \tag{2.16}$$

where R_i represents the rank of the ith residual r_i, and $a_n\left(\cdot\right)$ is a monotone score function that satisfies

$$\sum_{i=1}^{n} a_n\left(i\right) = 0. \tag{2.17}$$

R-estimates are scale equivalent which is an advantage compared to M-estimates. However, the optimal choice of the score function is unclear. In addition, most of R-estimates have a breakdown point of BP $= 1/n \rightarrow 0$ when n is close infinity. The bounded influence R-estimator proposed by Naranjo and Hettmansperger (1994) has a fairly high efficiency when the errors have normal distribution. However, it is proved that their breakdown point is no more than 0.2.

2.9 REWLSE

Gervini and Yohai (2002) proposed a new class of robust regression method called robust and efficient weighted least squares estimator (REWLSE). REWLSE is a very attractive robust estimator due to its simultaneously attaining maximum breakdown point and full efficiency under normal errors. This new estimator is a type of weighted least squares estimator with the weights adaptively calculated from an initial robust estimator.

Consider a pair of initial robust estimates of regression parameters and scale, $\hat{\boldsymbol{\beta}}_0$ and $\hat{\sigma}$ respectively, the standardized residuals are defined as

$$r_i = \frac{y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_0}{\hat{\sigma}}.$$

A large value of $|r_i|$ would suggest that $(\boldsymbol{x}_i, y_i)$ is an outlier. Define a measure of proportion of outliers in the sample

$$d_n = \max_{i>i_0} \left\{ F^+(|r|_{(i)}) - \frac{(i-1)}{n} \right\}^+, \tag{2.18}$$

where $\{\cdot\}^+$ denotes positive part, F^+ denotes the distribution of $|X|$ when $X \sim F$, $|r|_{(1)} \leq \ldots \leq |r|_{(n)}$ are the order statistics of the standardized absolute residuals, and $i_0 = \max\left\{i : |r|_{(i)} < \eta\right\}$, where η is some large quantile of F^+. Typically $\eta = 2.5$ as chosen by Rousseeuw and Leroy (1987) and the cdf of a normal distribution is chosen for F. Thus those $\lfloor nd_n \rfloor$ observations with largest standardized absolute residuals are eliminated (here $\lfloor a \rfloor$ is the largest integer less than or equal to a).

The adaptive cut-off value is $t_n = |r|_{(i_n)}$ with $i_n = n - \lfloor nd_n \rfloor$. With this adaptive cut-off value, the adaptive weights proposed by Gervini and Yohai (2002) are

$$w_i = \begin{cases} 1 & \text{if } |r_i| < t_n \\ 0 & \text{if } |r_i| \geq t_n. \end{cases} \tag{2.19}$$

Then, the REWLSE is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}, \tag{2.20}$$

where $\boldsymbol{W} = \text{diag}(w_1, \cdots, w_n)$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$, and $\mathbf{y} = (y_1, \cdots, y_n)'$.

If the initial regression and scale estimates with BP = 0.5 are chosen, the breakdown point of the REWLSE is also 0.5. Furthermore, since the cut-off values are used in a way resulting the REWLSE is asymptotically equivalent to the OLS estimates and hence has full asymptotic efficiency under the normal-error model.

2.10 Robust Regression Based on Regularization of Case-Specific Parameters

She and Owen (2011) and Lee *et al.* (2012) proposed a new class of robust regression methods using the case-specific indicators in a mean shift model with the regularization method. A mean shift model for the linear regression is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

where $\mathbf{y} = (y_1, \cdots, y_n)^T$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T, \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^T$, and the mean shift parameter γ_i is nonzero when the ith observation is an outlier and zero, otherwise.

Due to the sparsity of γ_is, She and Owen (2011) and Lee *et al.* (2012) proposed to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by minimizing the penalized least squares using L_1 penalty:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{2}\left\{\boldsymbol{y} - (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\gamma})\right\}^T \left\{\boldsymbol{y} - (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\gamma})\right\} + \lambda \sum_{i=1}^{n} |\gamma_i|, \tag{2.21}$$

where λ is a fixed regularization parameter for $\boldsymbol{\gamma}$. Given the estimate $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\beta}}$ is the OLS estimate with $\mathbf{y}$ replaced by $\mathbf{y} - \boldsymbol{\gamma}$. For a fixed $\hat{\boldsymbol{\beta}}$, the minimizer of (2.21) is $\hat{\gamma}_i = sgn(r_i)(|\gamma_i| - \lambda)_+$, that is,

$$\hat{\gamma}_i = \begin{cases} 0 & \text{if } |r_i| \le \lambda; \\ y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} & \text{if } |r_i| > \lambda. \end{cases}$$

Therefore, the solution of (2.21) can be found by iteratively updating the above two steps. She and Owen (2011) and Lee *et al.* (2012) proved that the above estimate is in fact equivalent to the M-estimate if Huber's ψ function is used. However, their proposed robust estimates are based on different perspective and can be extended to many other likelihood based models.

Note, however, the monotone M-estimate is not resistent to the high leverage outliers. In order to overcome this problem, She and Owen (2011) further proposed to replace the L_1 penalty in (2.21) by a general penalty. The objective function is then defined by

$$L_p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{2} \{\boldsymbol{y} - (\boldsymbol{X\beta} + \boldsymbol{\gamma})\}^T \{\boldsymbol{y} - (\boldsymbol{X\beta} + \boldsymbol{\gamma})\} + \sum_{i=1}^{n} p_\lambda(|\gamma_i|), \tag{2.22}$$

where $p_\lambda(|\cdot|)$ is any penalty function which depends on the regularization parameter λ. We can find $\hat{\boldsymbol{\gamma}}$ by defining thresholding function $\Theta(\boldsymbol{\gamma}; \lambda)$ (She, 2009). She (2009) and She and Owen (2011) proved that for a specific thresholding function, we can always find the corresponding penalty function. For example, the soft, hard, and smoothly clipped absolute deviation (SCAD; Fan and Li (2001)) thresholding solutions of $\boldsymbol{\gamma}$ correspond to L_1, Hard, and SCAD penalty functions, respectively. Minimizing the equation (2.22) yields a sparse $\hat{\boldsymbol{\gamma}}$ for outlier detection and a robust estimate of $\boldsymbol{\beta}$. She and Owen (2011) showed that the proposed estimates of (2.22) with hard or SCAD penalties are equivalent to the M-estimates with certain redescending ψ functions and thus will be resistent to high leverage outliers if a high breakdown point robust estimates are used as the initial values. Yu *et al.* (2015) successfully extended this robust method to mixture models.

3 Examples

In this section, we use both simulation study and real data applications to compare different robust methods in terms of parameter estimation and outlier detection. The first statistical criterion we use to compare different estimates is the mean squared errors (MSE). The other two are robust measures: robust bias (RB) and median absolute deviation (MAD)(You, 1999). They are defined as:

$$RB_i = median(\hat{\beta}_i) - \beta_i,$$

and

$$MAD_i = median(|\hat{\beta}_i - \beta_i|),$$

where $i = 0, 1$ for Example 1 and $i = 0, 1, 2, 3$ for Example 2.

We compare the OLS estimate with seven other commonly used robust regression estimates: the M estimate using Huber's ψ function (M_H), the M estimate using Tukey's bisquare function (M_T), the S estimate, the LTS estimate, the LMS estimate, the MM estimate (using bisquare weights and $k_1 = 4.68$), and the REWLSE. Due to the limited space, we cannot include all of our reviewed methods for comparison in our simulation study, such as median regression, Mallows GM-estimate, Schweppe GM-estimate, the S1S-estimator, and R-estimates. We mainly choose some methods that are popularly used and can be found from existing R packages. *R* function *rlm* provides the implementation of M_H and M_T with stating psi function as Huber and Tukey, respectively. LMS, LTS and S are computed using *R* function *lqs* with the option specified as "lms", "lts" and "S", respectively. In these *lqs* computation procedures, resampling algorithm is used. *R* package *robust* provides the implementation of MM and REWLSE and the S-estimate is used as an initial estimate via random resampling. It is known that using the initial S estimate in two-stage algorithm of MM achieves both high efficiency and robustness (Yohai, 1987). Note that we did not include the case-specific regularization methods proposed by She and Owen (2011) and Lee *et al.* (2012) since they are essentially equivalent to M-estimators. All the computations in this article are done by R. However, one can also implement those robust regression methods using SAS. In SAS, "ROBUSTREG" procedure provides implementation of M, LTS, S and MM estimates choosing "method=" to be "m", "lts", "s", or "mm", respectively.

Example 1. We generate n samples $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ from the model

$$Y = X + \varepsilon,$$

where $X \sim N(0,1)$. In order to compare the performance of different methods, we consider the following six cases for the error density of ε:

Case I: $\varepsilon \sim N(0,1)$- standard normal distribution.

Case II: $\varepsilon \sim t_3$ - t-distribution with degrees of freedom 3.

Case III: $\varepsilon \sim t_1$ - t-distribution with degrees of freedom 1 (Cauchy distribution).

Case IV: $\varepsilon \sim 0.95N(0,1) + 0.05N(0,10^2)$ - contaminated normal mixture.

Case V: $\varepsilon \sim$ N (0,1) with 10% identical outliers in y direction (where we let the first 10% of $y's$ equal to 30).

Case VI: $\varepsilon \sim$ N (0,1) with 10% identical high leverage outliers (where we let the first 10% of $x's$ equal to 10 and their corresponding $y's$ equal to 50).

Tables 1 and 2 report MSE, RB and MAD of the parameter estimates for each estimation method with sample size $n = 20$ and 100, respectively. The number of replicates is 200. From the tables, we can see that MM and REWLSE have the overall best performance throughout most cases and they are consistent for different sample sizes. For Case I, since the error distribution is normal, the performance of each estimate mainly depends on their efficiency. In this case, the OLS has the smallest MSE and MAD which is reasonable since under normal errors OLS is the best estimate; M_H, M_T, MM, and REWLSE have similar MSE to OLS, due to their high efficiency property; LMS, LTS, and S have relative larger MSE due to their low efficiency property. If the efficiency of some robust $\hat{\beta}^R$ relative to $\hat{\beta}^{OLS}$ is defined as the ratio of MSE of $\hat{\beta}^{OLS}$ to MSE of $\hat{\beta}^R$, the efficiencies of $\hat{\beta}_0^{LMS}$ and $\hat{\beta}_0^{LTS}$ relative to $\hat{\beta}_0^{OLS}$ do not exceed 28% for both $\hat{\beta}_0$ and $\hat{\beta}_1$. The efficiencies of $\hat{\beta}_0^S$ relative to $\hat{\beta}_0^{OLS}$ are between 32.38% and 41.60%, but the best efficiency of $\hat{\beta}_1^S$ relative to $\hat{\beta}_1^{OLS}$ is 29.5%. The efficiencies of other methods are much higher. MM, REWLSE, M_H, and M_T have smaller RB than those of LMS, LTS and S estimators. For Case II, M_H, M_T, MM, and REWLSE work better than other estimates in terms of MSE and MAD. For Case III, OLS has much larger MSE than other robust estimators; M_H, M_T, MM, REWLSE and S have similar MSE, RB and MAD. For Case IV, M_H, M_T, MM, and REWLSE have smaller MSE and MAD than others. From Case V, we can see that when the data contain outliers in the y-direction, OLS is much worse than any other robust estimates; MM, REWLSE, and M_T are better than other robust estimators. The reason why M_T can perform better than M_H is that Tukey's bisquare function can completely remove the effect of large outliers while Huber's ψ function can only reduce the effect of large outliers. Finally for Case VI, since there are high leverage outliers, similar to OLS, both M_T and M_H perform poorly; MM and REWLSE work better than other robust estimates.

In order to better compare the performance of different methods, Figure 1 shows the plot of their MSE versus each case for the intercept (left side) and slope (right side) parameters for example 1 when sample size $n = 100$. Since the lines for LTS and LMS are above the other lines, S, MM, and REWLSE of the intercept and slopes outperform LTS and LMS estimates throughout all six cases. In addition, the S estimate has similar performance to MM and REWLSE when the error density of ε is Cauchy distribution. However, MM and REWLSE perform better than S-estimates in other five cases. Furthermore, the lines for MM and REWLSE almost overlap for all six cases. It shows that MM and REWLSE are the overall best approaches in robust regression.

Example 2. Samples $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ are generated from the model

$$Y = X_1 + X_2 + X_3 + \varepsilon,$$

where $X_i \sim N(0, 1), i = 1, 2, 3$ and X_i's are independent. We consider the following six cases for the error density of ε:

Case I: $\varepsilon \sim N(0,1)$- standard normal distribution.

Case II: $\varepsilon \sim t_3$ - t-distribution with degrees of freedom 3.

Case III: $\varepsilon \sim t_1$ - t-distribution with degrees of freedom 1 (Cauchy distribution).

Case IV: $\varepsilon \sim 0.95N(0,1) + 0.05N(0,10^2)$ - contaminated normal mixture.

Case V: $\varepsilon \sim N(0,1)$ with 10% identical outliers in y direction (where we let the first 10% of $y's$ equal to 30).

Case VI: $\varepsilon \sim N(0,1)$ with 10% identical high leverage outliers being $X_1 = 10$, $X_2 = 10$, $X_3 = 10$, and Y = 50.

Tables 3 - 6 show MSE, RB and MAD of the parameter estimates of each estimation method for sample size $n = 20$ and $n = 100$, respectively. Figure 2 shows the plot of their MSE versus each case for three slopes and the intercept parameters with sample size $n = 100$. The results in Example 2 tell similar stories to Example 1. In summary, MM and REWLSE have the overall best performance; OLS only works well when there are no outliers since it is very sensitive to outliers; M-estimates (M_H and M_T) work well if the outliers are in y direction but are also sensitive to the high leverage outliers.

Example 3. In order to compare the performance of outlier detection, we consider two cases: 5% and 10% high leverage outliers in the model. $n = 100$ samples $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ are generated from the model

$$Y = X_1 + X_2 + X_3 + \gamma + \varepsilon,$$

where γ is a vector, $\varepsilon \sim N(0,1)$, $X_i \sim N(0,1), i = 1,2,3$ and X_i's are independent. We modify the first O rows of predictor matrix X to be $X_1 = 10$, $X_2 = 10$, $X_3 = 10$, where $O \in \{5, 10\}$. The first O rows of γ are randomly generated from a uniform distribution between 11 and 13 and the remaining $n - O$ rows of γ are all zeros. Therefore, the first O observations are high leverage outliers. In order to compare the performance of outlier detection of different methods, we use three benchmark proportions: M, S and JD (She and Owen, 2011). M is the mean masking probability (fraction of undetected true outliers), S denotes the mean swamping probability (fraction of good points labeled as outliers), and JD means the joint outlier detection rate (fraction of simulations with 0 masking). Ideally, M ≈ 0, S ≈ 0 and JD $\approx 100\%$.

As can be seen from Table 7, S and MM have relatively small probabilities of both masking and swamping, and LTS has smaller probability of masking but higher probability of swamping in the presence of 5% outliers. When the proportion of outliers increases to 10%, LMS, LTS, S, MM and REWLSE still have high joint identification rates which are larger than 60%. LMS has lowest masking probability but also has highest swamping probability. As expected, M_H and M_T have very high masking probabilities due to their sensitivity to high leverage outliers.

Example 4: We next use the modified data on wood specific gravity (Rousseeuw, 1984; Rousseeuw and Leroy, 1987; Olive and Hawkins, 2011) to compare OLS with LTS, LMS, S, and MM. The data set is shown in Table 9, which contains 20 points and four of them ($i = 4, 6, 8, 19$) are outliers (Rousseeuw, 1984). A linear regression model is used to investigate the influence of anatomical factors on wood specific gravity. The estimates of the six parameters by OLS, LTS, S, MM, and LMS (LMS estimates are provided by Rousseeuw (1984)) are shown in Table 8. LTS, S, and MM produce similar coefficient estimates and are close to LMS estimates. However, the OLS estimates are quite different from those robust estimates of LTS, S, MM, and LMS. Therefore, the OLS estimates are greatly affected by the outliers.

Table 9 lists the standardized residuals (residuals divided by the estimated scale) for OLS, LTS, S, MM, and LMS (the standardized residuals for LMS are provided by

Rousseeuw (1984)). The corresponding estimated scales are $\hat{\sigma}_{OLS} = 0.02412$, $\hat{\sigma}_{LTS} = 0.0065$, $\hat{\sigma}_S = 0.01351$, $\hat{\sigma}_{MM} = 0.01351$, and $\hat{\sigma}_{LMS} = 0.0195$, respectively. It is not easy to identify the outliers by looking at standardized residuals of OLS, but standardized residuals of LTS, S, MM and LMS could correctly identify the four outliers and the identified four outliers are exactly the same as which were spotted by LMS in Rousseeuw (1984). Therefore, the naive method by looking at the standardized residuals of OLS might miss the outliers due to the masking effect.

Figure 3 shows plots of the residuals versus the fitted values for OLS, LTS, S, MM and LMS, respectively. There are obvious four outliers by looking at residual plots of LTS, S, MM and LMS. However, the four outliers can not be easily detected by naively looking at the residual plot of OLS due to the masking effect.

Example 5: Finally, we apply OLS, LTS, LMS, S, and MM estimators to an artificial three-predictor data set which was created by Hawkins *et al.* (1984) to test the outlier detection of these regression parameter estimates. The dataset contains outliers at cases 1-10. The standardized residuals for OLS, LTS, LMS, S, and MM estimations are given in Table 10. The standardized residuals show that LTS, S, LMS, and MM all correctly flag the outliers and obtain similar coefficient estimation. Nonetheless, Hadi and Simonoff (1993) indicated that MM estimator with high efficiency level masked true outliers and swamped in the cases 11-14, but less-efficient versions of MM estimator (with efficiencies up to about 80%) give results similar to LMS and LTS. Similar to example 4, OLS estimator fails to identify the outliers.

4 Discussion

In this article, we describe and compare different available robust methods. Table 11 summarizes the robustness attributes and asymptotic efficiency of most of the estimators we have discussed. Based on Table 11, it can be seen that MM-estimates and REWLSE have both high breakdown point and high efficiency. Our simulation study also demonstrated that MM-estimates and REWLSE have overall best performance among all compared robust methods. However, Park *et al.* (2012) pointed out that MM-estimates cannot detect any outliers when the contamination percentage is equal to and above 30%. In terms of breakdown point and efficiency, GM-estimates (Mallows, Schweppe), Bounded R-estimates, M-estimates, and LAD estimates are less attractive due to their low breakdown points. Although LMS, LTS, S-estimates, and GS-estimates are strongly resistent to outliers, their efficiencies are low. However, these high breakdown point robust estimates such as S-estimates and LTS are traditionally used as the initial estimates for some other high breakdown point and high efficiency robust estimates.

As one referee pointed out, although MM, S and LTS have high breakdown points, the practical computation of these estimators is very challenging, especially for large datasets (Stromberg *et al.*, 2000; Hawkins and Olive, 2002). The commonly adopted method is to use the elemental resampling algorithm to obtain a number of subsets of data and calculate the initial regression estimate for each element set, and then compute the robust regression estimate from a number of initial estimates. However, based on Hawkins and Olive (2002), the theoretical properties of the above computed estimates depend on the number of elemental sets and their high breakdown properties usually require the number of elementary sets to go to infinity. For example, Hawkins and Olive (2002) proved that LTS estimator computed from the elemental resampling techniques, such as FAST-LTS algorithm, has zero breakdown point. In order to compute MM, S and LTS estimators with high breakdown point, one should consider all possible elemental sets. In addition, Olive and Hawkins (2011) and Park *et al.* (2012) pointed out that if a practical initial estimator which has not been proved to be high breakdown is used in the implementation of the two-stage estimator such as S and MM estimator, the resulting two-stage estimator may be neither consistent nor high breakdown.

We would also like to mention some other directions to provide regression estimates which are robust to outliers. Lee (1989, 1993), Kemp and Santos Silva (2012), and Yao and Li (2014) proposed *modal regression* to robustly estimate the regression function. Modal regression focuses on "most likely" conditional values rather than the conditional average or median. However, when the error distribution is homogeneous, modal regression line is the same as traditional mean regression line, except for intercepts. In addition, Linton and Xiao (2007), Yuan and De Gooijer (2007), Wang and Yao (2012), Yao and Zhao (2013), and Chen et al. (2015) proposed to adaptively estimate the regression functions by estimating the error density using kernel density estimation. Those adaptive estimates are also demonstrated to be robust to outliers and heavy-tail error distributions based on their simulation studies.

References

Bai, X., Yao, W., and Boyer, J. E. (2012). Robust fitting of mixture regression models. *Computational Statistics and Data Analysis*, 56, 2347-2359.

Carroll, R. J. and Welsch, A. H. (1988). A Note on Asymmetry and Robustness in Linear Regression. *Journal of American Statisitcal Association*, 4, 285-287.

Chen, Y., Wang, Q., and Yao, W. (2015). Adaptive Estimation for Varying Coefficient Models. *Journal of Multivariate Analysis*, 137, 17-31.

Coakley, C. W. and Hettmansperger, T. P. (1993). A Bounded Influence, High Breakdown, Efficient Regression Estimator. *Journal of American Statistical Association*, 88, 872-880.

Croux, C., Rousseeuw, P. J., and Hössjer O. (1994). Generalized S-estimators. *Journal of American Statistical Association*, 89, 1271-1281.

Donoho, D. L. and Huber, P. J. (1983). The Notation of Break-down Point, *in A Festschrift for E. L. Lehmann, Wadsworth*.

Freedman, W. L., Wilson, C. D., and Madore, B. F. (1991). New Cepheid Distances to Nearby Galaxies Based on BVRI CCD Photometry. *Astrophysical Journal*, 372, 455-470.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360.

Gervini, D. and Yohai, V. J. (2002). A Class of Robust and Fully Efficient Regression Estimators. *The Annals of Statistics*, 30, 583-616.

Hadi, A. S., Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*,88,1264-1272.

Handschin, E., Kohlas, J., Fiechter, A., and Schweppe, F. (1975). Bad Data Analysis for Power System State Estimation. *IEEE Transactions on Power Apparatus and Systems*, 2, 329-337.

Hawkins, D. M., Bradu, D., Kass, G. V. (1984). Location of several outliers in multiple regression data using elemental sets. *Technometrics*, 26,197-208.

Hawkins, D. M. and Olive, D. J. (2002). Inconsistency of resampling algorithms for high breakdown regression estimators and a new algorithm. *Journal of the American Statistical Association*, 97, 136-148.

Huber, P.J. (1964). Robust Version of a Location Parameter. *Annals of Mathematical Statistics*, 36, 1753-1758.

Huber, P.J. (1981). Robust Statistics. *New York: John Wiley and Sons.*

Jackel, L.A. (1972). Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals. *Annals of Mathematical Statistics*, 5, 1449-1458.

Kemp, G. C. R. and Santos Silva, J. M. C. (2012). Regression towards the mode. *Journal of Economics*, 170, 92-101.

Lee, M. J. (1989). Mode Regression. *Journal of Econometrics*, 42, 337-349.

Lee, M. J. (1993). Quadratic Mode Regression. *Journal of Econometrics*, 57, 1-19.

Lee, Y., MacEachern, S. N., and Jung, Y. (2012). Regularization of Case-Specific Parameters for Robustness and Efficiency. *Statistical Science*, 27, 350-372.

Linton, O. B. and Xiao, Z. (2007). A nonparametric regression estimator that adapts to error distribution of unknown form. *Econometric Theory*, 23, 371-413.

Mallows, C.L. (1975). On Some Topics in Robustness. unpublished memorandum, Bell Tel. Laboratories, Murray Hill.

Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). Robust Statistics. *John Wiley.*

Naranjo, J.D., Hettmansperger, T. P. (1994). Bounded Influence Rank Regression. *Journal of the Royal Statistical Society B*, 56, 209-220.

Olive, D. J. and Hawkins, D. M. (2011). Practical high breakdown regression. Available at: http://lagrange.math.siu.edu/Olive/pphbreg.pdf.

Park, Y., Kim, D. and Kim, S. (2012). Robust regression using data partitioning and M-estimation. *Communications in Statistics - Simulation and Computation*, 41, 1282-1300.

Rousseeuw, P. J.(1983). Multivariate Estimation with High Breakdown Point. Research Report No. 192, Center for Statistics and Operations research, VUB Brussels.

Rousseeuw, P. J.(1984). Least Median of Squares regression. *Journal of American Statistical Association*, 79, 871-880.

Rousseeuw, P. J. and Leroy, A. M.. (1987). Robust Regression and Outlier Detection. *New York: Wiley.*

Rousseeuw, P. J. and Yohai, V. J. (1984). Robust Regression by Means of S-estimators. *Robust and Nonlinear Time series*, J. Franke, W. Härdle and R. D. Martin (eds.), Lectures Notes in Statistics 26, 256-272, New York: Springer.

She, Y. (2009). Thresholding-Based Iterative Selection Procedures for Model Selection and Shrinkage. *Electronic Journal of Statistics*, 3, 384-415.

She, Y. and Owen, A. B. (2011). Outlier Detection Using Nonconvex Penalized regression. *Journal of American Statistical Association*, 106, 626-639.

Siegel, A.F. (1982). Robust Regression Using Repeated Medians. *Biometrika*, 69, 242-244.

Stromberg, A. J., Hawkins, D. M., and Hössjer, O. (2000). The Least Trimmed Differences Regression Estimator and Alternatives. *Journal of American Statistical Association*, 95, 853-864.

Wang, Q. and Yao, W. (2012). An adaptive estimation of MAVE. *Journal of Multivariate Analysis*, 104, 88-100.

Wilcox (1996). A review of some recent developments in robust regression. *British Journal of Mathematical and Statistical Psychology*, 49, 253-274.

Yao, W. and Li, L. (2014). A new regression model: modal linear regression. *Scandinavian Journal of Statistics*, 1-16.

Yao, W., Lindsay, B. G., and Li, R. (2012). Local modal regression. *Journal of Nonparametric Statistics,* 24, 647-663.

Yao, W. and Zhao, Z. (2013). Kernel density based linear regression estimates. *Communications in Statistics-Theory and Methods*, 42, 4499-4512.

Yohai, V. J. (1987). High Breakdown-point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, 15, 642-656.

You, J. (1999). A Monte Carlo Comparison of Several High Breakdown and Efficient Estimators. *Computational Statistics and Data Analysis*, 30, 205-219.

Yu, C., Chen, K., and Yao, W. (2015). Outlier Detection and Robust Mixture Modeling Using Nonconvex Penalized Likelihood. *Journal of Statistical Planning and Inference*, 164, 27-38.

Yuan, A. and De Gooijer, J. G. (2007). Semiparametric regression with kernel error model. *Scandinavian Journal of Statistics*, 34, 841-869.

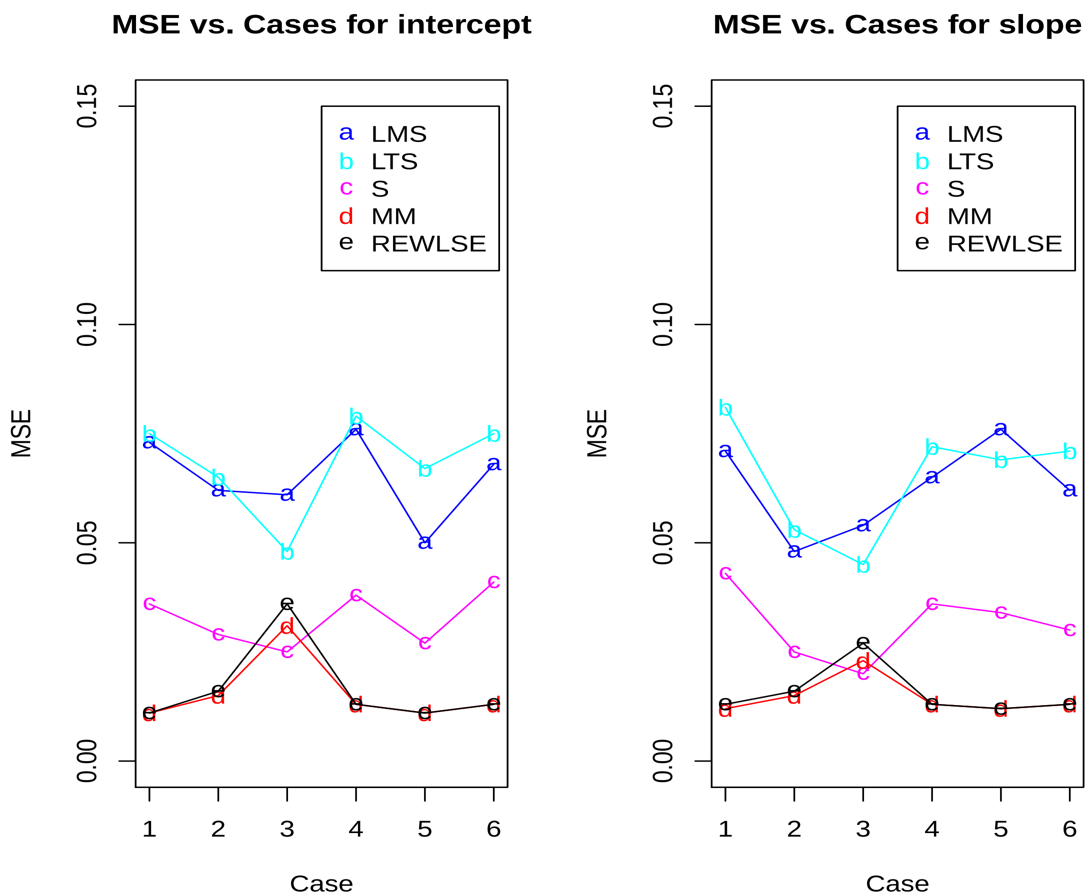

Figure 1: Plot of MSE of intercept (left) and slope (right) estimates vs. different cases for LMS, LTS, S, MM, and REWLSE, for Example 1 when $n = 100$.

Table 1: Comparison of Different Estimates for Example 1 with $n = 20$

	OLS	M_H	M_T	LMS	LTS	S	MM	REWLSE
				Case I: $\varepsilon \sim N(0,1)$				
MSE($\hat{\beta}_0$)	0.0644	0.0679	0.0685	0.2536	0.2332	0.1548	0.0679	0.0654
RB($\hat{\beta}_0$)	-0.0121	-0.0073	0.0011	-0.0093	-0.0482	-0.0045	0.0032	0.0105
MAD($\hat{\beta}_0$)	0.1663	0.1570	0.1473	0.3669	0.3780	0.2519	0.1533	0.1593
MSE($\hat{\beta}_1$)	0.0544	0.0578	0.0584	0.3117	0.2455	0.1843	0.0572	0.0563
RB($\hat{\beta}_1$)	-0.0021	-0.0060	-0.0145	-0.0002	0.0513	0.0247	0.0001	0.0041
MAD($\hat{\beta}_1$)	0.1527	0.1574	0.1543	0.3159	0.3129	0.2667	0.1517	0.1527
				Case II: $\varepsilon \sim t_3$				
MSE($\hat{\beta}_0$)	0.1337	0.0804	0.0829	0.2374	0.2259	0.1293	0.0825	0.0867
RB($\hat{\beta}_0$)	-0.0399	-0.0151	0.0064	-0.0052	0.0128	0.0122	-0.0055	0.0082
MAD($\hat{\beta}_0$)	0.2353	0.1672	0.1746	0.3030	0.2925	0.2357	0.1754	0.1881
MSE($\hat{\beta}_1$)	0.1867	0.0947	0.0952	0.2924	0.2836	0.1739	0.0902	0.1007
RB($\hat{\beta}_1$)	0.0251	-0.0013	-0.0160	0.0223	0.0416	0.0055	-0.0097	-0.0037
MAD($\hat{\beta}_1$)	0.1994	0.1744	0.1818	0.3598	0.3353	0.2496	0.1783	0.1965
				Case III: $\varepsilon \sim t_1$				
MSE($\hat{\beta}_0$)	2201.5	0.2245	0.1655	0.2347	0.2388	0.1752	0.1764	0.1727
RB($\hat{\beta}_0$)	0.0979	0.0227	0.0224	0.0253	0.0095	0.0053	0.0064	-0.0053
MAD($\hat{\beta}_0$)	0.8252	0.2991	0.2553	0.2801	0.2876	0.2471	0.2646	0.3095
MSE($\hat{\beta}_1$)	810.20	0.3723	0.2192	0.3962	0.4527	0.2209	0.2313	0.2272
RB($\hat{\beta}_1$)	-0.0127	0.0030	0.0010	-0.0924	-0.1014	-0.0541	-0.0075	0.0033
MAD($\hat{\beta}_1$)	0.8094	0.3142	0.2659	0.3550	0.3046	0.2586	0.2808	0.2853
				Case IV: $\varepsilon \sim 0.95N(0,1) + 0.05N(0,10^2)$				
MSE($\hat{\beta}_0$)	0.3494	0.0651	0.0641	0.2555	0.2431	0.1697	0.0621	0.0648
RB($\hat{\beta}_0$)	-0.0899	-0.0299	-0.0171	-0.0145	-0.0459	-0.0468	-0.0162	-0.0061
MAD($\hat{\beta}_0$)	0.2961	0.1652	0.1763	0.3273	0.3424	0.2846	0.1696	0.1676
MSE($\hat{\beta}_1$)	0.3261	0.0677	0.0601	0.3433	0.3164	0.1558	0.0585	0.0527
RB($\hat{\beta}_1$)	-0.0364	-0.0200	-0.0178	0.0161	-0.0069	-0.0055	-0.0169	-0.0214
MAD($\hat{\beta}_1$)	0.2344	0.1536	0.1477	0.2996	0.2996	0.2492	0.1476	0.1556
				Case V: $\varepsilon \sim N(0,1)$ with outliers in y direction				
MSE($\hat{\beta}_0$)	9.4097	0.0966	0.0483	0.2238	0.1943	0.1306	0.0473	0.0423
RB($\hat{\beta}_0$)	3.0130	0.2288	0.0267	-0.0356	-0.0230	0.0010	0.0267	0.0250
MAD($\hat{\beta}_0$)	3.0130	0.2385	0.1330	0.3024	0.3062	0.2362	0.1290	0.1425
MSE($\hat{\beta}_1$)	4.9846	0.0912	0.0608	0.3241	0.2599	0.1847	0.0603	0.0603
RB($\hat{\beta}_1$)	0.0443	0.0126	-0.0045	0.0036	-0.0433	-0.0020	0.0013	-0.0004
MAD($\hat{\beta}_1$)	1.3387	0.1793	0.1458	0.3305	0.3065	0.2313	0.1502	0.1500
				Case VI: $\varepsilon \sim N(0,1)$ with high leverage outliers				
MSE($\hat{\beta}_0$)	0.7718	0.8123	0.8293	0.2228	0.2021	0.1370	0.0718	0.0717
RB($\hat{\beta}_0$)	0.2579	0.2347	0.2790	0.0491	0.0303	0.0084	0.0030	0.0009
MAD($\hat{\beta}_0$)	0.5265	0.5485	0.5590	0.2981	0.2892	0.2422	0.1540	0.1559
MSE($\hat{\beta}_1$)	13.397	13.738	13.878	0.3435	0.2287	0.1661	0.0772	0.0731
RB($\hat{\beta}_1$)	3.6644	3.7123	3.7294	0.0129	0.0500	-0.0280	0.0126	0.0129
MAD($\hat{\beta}_1$)	3.6644	3.7123	3.7294	0.2944	0.2882	0.2708	0.1603	0.1588

Table 2: Comparison of Different Estimates for Example 1 with $n = 100$

TRUE	OLS	M_H	M_T	LMS	LTS	S	MM	REWLSE
			Case I: $\varepsilon \sim N(0,1)$					
MSE($\hat{\beta}_0$)	0.0102	0.0117	0.0118	0.0675	0.0728	0.0315	0.0118	0.0109
RB($\hat{\beta}_0$)	0.0049	-0.0037	-0.0031	0.0084	0.0357	0.0272	-0.0015	-0.0039
MAD($\hat{\beta}_0$)	0.0724	0.0737	0.0730	0.1819	0.1829	0.1225	0.0734	0.0727
MSE($\hat{\beta}_1$)	0.0105	0.0112	0.0114	0.0610	0.0762	0.0367	0.0114	0.0115
RB($\hat{\beta}_1$)	-0.0108	0.0002	-0.0032	-0.0167	-0.0028	-0.0007	-0.0047	-0.0035
MAD($\hat{\beta}_1$)	0.0717	0.0700	0.0759	0.1670	0.1785	0.1188	0.0762	0.0761
			Case II: $\varepsilon \sim t_3$					
MSE($\hat{\beta}_0$)	0.0355	0.0168	0.0164	0.0600	0.0612	0.0284	0.0165	0.0166
RB($\hat{\beta}_0$)	-0.0255	-0.0018	0.0037	0.0220	0.0051	0.0167	0.0052	0.0037
MAD($\hat{\beta}_0$)	0.1133	0.0868	0.0736	0.1703	0.1849	0.1010	0.0767	0.0819
MSE($\hat{\beta}_1$)	0.0394	0.0196	0.0189	0.0589	0.0661	0.0324	0.0187	0.0190
RB($\hat{\beta}_1$)	-0.0363	-0.0293	-0.0335	-0.0229	-0.0087	-0.0265	-0.0309	-0.0278
MAD($\hat{\beta}_1$)	0.1235	0.0896	0.0927	0.1501	0.1625	0.1167	0.0892	0.1027
			Case III: $\varepsilon \sim t_1$					
MSE($\hat{\beta}_0$)	8810.4	0.0405	0.0327	0.0713	0.0486	0.0304	0.0342	0.0339
RB($\hat{\beta}_0$)	0.0687	-0.0213	0.0023	-0.0230	-0.0107	-0.0165	0.0020	-0.0250
MAD($\hat{\beta}_0$)	1.1700	0.1385	0.1216	0.1720	0.1374	0.1040	0.1245	0.1235
MSE($\hat{\beta}_1$)	27954	0.0396	0.0298	0.0667	0.0466	0.0337	0.0300	0.0302
RB($\hat{\beta}_1$)	-0.0347	0.0117	0.0011	0.0225	-0.0044	0.0188	0.0003	0.0006
MAD($\hat{\beta}_1$)	1.1136	0.1300	0.1078	0.1556	0.1317	0.1083	0.1078	0.1060
			Case IV: $\varepsilon \sim 0.95N(0,1) + 0.05N(0,10^2)$					
MSE($\hat{\beta}_0$)	49.49	0.0139	0.0131	0.0757	0.0752	0.0344	0.0130	0.0136
RB($\hat{\beta}_0$)	-0.0365	0.0036	-0.0012	0.0607	0.0293	0.0257	-0.0011	0.0002
MAD($\hat{\beta}_0$)	0.1742	0.0723	0.0702	0.1990	0.1848	0.1289	0.0691	0.0716
MSE($\hat{\beta}_1$)	1.4060	0.0125	0.0108	0.0576	0.0698	0.0291	0.0112	0.0112
RB($\hat{\beta}_1$)	-0.0209	-0.0072	-0.0073	0.0150	-0.0009	0.0012	-0.0092	-0.0140
MAD($\hat{\beta}_1$)	0.1448	0.0725	0.0690	0.1675	0.1552	0.1070	0.0704	0.0734
			Case V: $\varepsilon \sim N(0,1)$ with outliers in y direction					
MSE($\hat{\beta}_0$)	8.9967	0.0500	0.0123	0.0701	0.0724	0.0313	0.0122	0.0125
RB($\hat{\beta}_0$)	2.9916	0.1842	-0.0101	-0.0023	-0.0174	-0.0230	-0.0114	-0.0127
MAD($\hat{\beta}_0$)	2.9916	0.1842	0.0838	0.1827	0.1889	0.1228	0.0782	0.0843
MSE($\hat{\beta}_1$)	0.9034	0.0169	0.0115	0.0641	0.0708	0.0305	0.0113	0.0117
RB($\hat{\beta}_1$)	-0.0817	0.0101	0.0090	0.0105	0.0003	0.0149	0.0068	0.0001
MAD($\hat{\beta}_1$)	0.6839	0.0819	0.0734	0.1725	0.1926	0.1060	0.0734	0.0759
			Case VI: $\varepsilon \sim N(0,1)$ with high leverage outliers					
MSE($\hat{\beta}_0$)	0.2942	0.3116	0.3198	0.0678	0.0615	0.0296	0.0123	0.0126
RB($\hat{\beta}_0$)	0.3942	0.3645	0.3709	-0.0020	0.0094	-0.0134	-0.0028	-0.0048
MAD($\hat{\beta}_0$)	0.4118	0.4124	0.4210	0.1694	0.1664	0.1165	0.0768	0.0752
MSE($\hat{\beta}_1$)	13.269	13.621	13.740	0.0643	0.0616	0.0346	0.0119	0.0117
RB($\hat{\beta}_1$)	3.6445	3.6860	3.7033	0.0010	-0.0321	0.0054	0.0101	0.0115
MAD($\hat{\beta}_1$)	3.6445	3.6860	3.7033	0.1786	0.1647	0.1117	0.0634	0.0660

Table 3: Comparison of Different Estimates for Example 2 with $n = 20$

TRUE	OLS	M_H	M_T	LMS	LTS	S	MM	REWLSE
			Case I: $\varepsilon \sim N(0,1)$					
MSE($\hat{\beta}_0$)	0.0574	0.0612	0.0660	0.3700	0.2544	0.1870	0.0622	0.0613
RB($\hat{\beta}_0$)	0.0136	0.0081	0.0053	-0.0068	-0.0673	-0.0435	-0.0033	-0.0016
MAD($\hat{\beta}_0$)	0.1707	0.1879	0.1955	0.4187	0.3356	0.2912	0.1889	0.1871
MSE($\hat{\beta}_1$)	0.0670	0.0732	0.0866	0.4648	0.3008	0.2550	0.0776	0.0730
RB($\hat{\beta}_1$)	-0.0259	-0.0253	-0.0256	-0.0359	0.0227	-0.0100	-0.0228	-0.0252
MAD($\hat{\beta}_1$)	0.1723	0.1824	0.1955	0.4119	0.3317	0.2857	0.1795	0.1728
MSE($\hat{\beta}_2$)	0.0642	0.0624	0.0667	0.4648	0.3275	0.2544	0.0647	0.0648
RB($\hat{\beta}_2$)	0.0086	0.0026	-0.0010	0.0173	0.0365	-0.0386	0.0070	0.0014
MAD($\hat{\beta}_2$)	0.1760	0.1664	0.1643	0.4029	0.3535	0.2994	0.1704	0.1748
MSE($\hat{\beta}_3$)	0.0706	0.0751	0.0829	0.4202	0.3052	0.2283	0.0789	0.0748
RB($\hat{\beta}_3$)	-0.0099	0.0084	0.0026	-0.1113	-0.0516	-0.0244	0.0001	-0.0057
MAD($\hat{\beta}_3$)	0.1680	0.1761	0.1859	0.3983	0.3461	0.2765	0.1802	0.1718
			Case II: $\varepsilon \sim t_3$					
MSE($\hat{\beta}_0$)	0.1736	0.0927	0.0919	0.3861	0.2972	0.1970	0.0904	0.0949
RB($\hat{\beta}_0$)	0.0416	0.0414	0.0292	0.1139	0.1070	0.0601	0.0300	0.0157
MAD($\hat{\beta}_0$)	0.2544	0.2274	0.2110	0.4091	0.3620	0.2878	0.2024	0.1970
MSE($\hat{\beta}_1$)	0.1959	0.1545	0.1560	0.5253	0.4219	0.2548	0.1557	0.1535
RB($\hat{\beta}_1$)	0.0248	0.0135	0.0214	-0.0811	-0.0361	0.0100	0.0239	0.0150
MAD($\hat{\beta}_1$)	0.2470	0.2329	0.2376	0.4159	0.3234	0.3009	0.2300	0.2317
MSE($\hat{\beta}_2$)	0.2878	0.1690	0.1638	0.6129	0.3644	0.2437	0.1580	0.1614
RB($\hat{\beta}_2$)	-0.0509	-0.0188	-0.0179	-0.0465	-0.0039	0.0166	-0.0130	0.0082
MAD($\hat{\beta}_2$)	0.2743	0.2432	0.2428	0.4551	0.3399	0.2901	0.2351	0.2399
MSE($\hat{\beta}_3$)	0.2566	0.1383	0.1359	0.5529	0.4124	0.3023	0.1404	0.1380
RB($\hat{\beta}_3$)	0.0433	0.0427	0.0071	0.0568	0.0804	-0.0006	0.0010	-0.0013
MAD($\hat{\beta}_3$)	0.2517	0.2140	0.2364	0.4065	0.3745	0.2883	0.2346	0.2251
			Case III: $\varepsilon \sim t_1$					
MSE($\hat{\beta}_0$)	33406	0.4019	0.3217	0.8503	0.4983	0.3847	0.3298	0.3258
RB($\hat{\beta}_0$)	-0.2515	-0.1037	-0.0377	-0.0176	-0.0483	-0.0501	-0.0545	-0.0510
MAD($\hat{\beta}_0$)	0.9030	0.3429	0.3327	0.4471	0.3553	0.3459	0.2961	0.3172
MSE($\hat{\beta}_1$)	2069.0	0.4238	0.3401	1.1884	0.6947	0.3958	0.3393	0.3424
RB($\hat{\beta}_1$)	-0.1555	-0.0913	-0.0909	0.0056	-0.0019	0.0112	-0.0561	-0.0483
MAD($\hat{\beta}_1$)	1.0759	0.3655	0.3286	0.4701	0.3801	0.3219	0.3148	0.3274
MSE($\hat{\beta}_2$)	3188.0	0.6564	0.4883	0.9200	0.6419	0.4245	0.4616	0.4526
RB($\hat{\beta}_2$)	0.1034	0.0541	-0.0091	0.0396	0.0210	-0.0583	-0.0606	0.0026
MAD($\hat{\beta}_2$)	0.9409	0.4642	0.3538	0.4468	0.4039	0.3362	0.3965	0.3491
MSE($\hat{\beta}_3$)	1774.0	0.5386	0.4867	1.0639	0.6637	0.4577	0.4923	0.4951
RB($\hat{\beta}_3$)	0.0863	0.0229	-0.0303	-0.1065	-0.0568	-0.0215	-0.0155	-0.0107
MAD($\hat{\beta}_3$)	0.8467	0.3781	0.3606	0.4403	0.3913	0.3599	0.3615	0.3667

Table 4: Comparison of Different Estimates for Example 2 with $n = 20$

TRUE	OLS	M_H	M_T	LMS	LTS	S	MM	REWLSE
	Case IV: $\varepsilon \sim 0.95N(0,1) + 0.05N(0,10^2)$							
MSE($\hat{\beta}_0$)	0.4870	0.0919	0.0754	0.3972	0.2812	0.1929	0.0755	0.0721
RB($\hat{\beta}_0$)	0.0447	0.0215	0.0108	-0.0390	0.0195	-0.0209	0.0184	0.0216
MAD($\hat{\beta}_0$)	0.2980	0.2172	0.1838	0.4019	0.3822	0.3052	0.1875	0.1854
MSE($\hat{\beta}_1$)	0.4734	0.1240	0.0983	0.5312	0.3103	0.2023	0.0976	0.1009
RB($\hat{\beta}_1$)	-0.0162	-0.0055	0.0090	-0.0115	-0.0055	0.0060	0.0003	0.0046
MAD($\hat{\beta}_1$)	0.3040	0.2209	0.1990	0.3813	0.3508	0.2743	0.2090	0.2072
MSE($\hat{\beta}_2$)	0.6811	0.1069	0.0861	0.4851	0.3101	0.2033	0.0939	0.1072
RB($\hat{\beta}_2$)	-0.0642	-0.0265	-0.0146	-0.0774	-0.0313	-0.0361	-0.0177	-0.0092
MAD($\hat{\beta}_2$)	0.2557	0.1722	0.1594	0.4231	0.3303	0.2592	0.1612	0.1805
MSE($\hat{\beta}_3$)	0.5937	0.1055	0.0798	0.5802	0.3336	0.2359	0.0773	0.0736
RB($\hat{\beta}_3$)	-0.0083	-0.0033	-0.0126	0.0150	-0.0341	-0.0323	-0.0065	-0.0055
MAD($\hat{\beta}_3$)	0.3409	0.2131	0.1928	0.3499	0.3382	0.2876	0.1891	0.2133
	Case V: $\varepsilon \sim N(0,1)$ with outliers in y direction							
MSE($\hat{\beta}_0$)	9.8831	0.1470	0.0784	0.3489	0.2487	0.1643	0.0735	0.0739
RB($\hat{\beta}_0$)	2.9410	0.2160	0.0030	-0.0628	-0.0392	-0.0053	0.0025	0.0110
MAD($\hat{\beta}_0$)	2.9410	0.2655	0.1807	0.4015	0.3288	0.2530	0.1818	0.1760
MSE($\hat{\beta}_1$)	4.6973	0.0918	0.0721	0.4046	0.2875	0.2069	0.0668	0.0697
RB($\hat{\beta}_1$)	-0.1366	-0.0105	-0.0385	0.0102	0.0221	-0.0005	-0.0243	-0.0134
MAD($\hat{\beta}_1$)	1.3713	0.1846	0.1844	0.3920	0.3655	0.3075	0.1775	0.1846
MSE($\hat{\beta}_2$)	6.1743	0.1736	0.1001	0.5268	0.2976	0.2143	0.0936	0.0955
RB($\hat{\beta}_2$)	-0.2108	-0.0370	-0.0079	0.0435	0.0053	-0.0081	-0.0117	-0.0304
MAD($\hat{\beta}_2$)	1.667	0.2181	0.1775	0.3829	0.3649	0.2845	0.1833	0.1800
MSE($\hat{\beta}_3$)	5.5139	0.1331	0.0781	0.3906	0.2896	0.2088	0.0746	0.0855
RB($\hat{\beta}_3$)	-0.2013	-0.0291	-0.0322	-0.0773	-0.0335	-0.0129	-0.0322	-0.0337
MAD($\hat{\beta}_3$)	1.5581	0.2392	0.1874	0.3885	0.3493	0.2791	0.1889	0.1868
	Case VI: $\varepsilon \sim N(0,1)$ with high leverage outliers							
MSE($\hat{\beta}_0$)	0.9099	0.9893	1.0550	0.3347	0.29325	0.1537	0.0685	0.0679
RB($\hat{\beta}_0$)	0.1594	0.1545	0.1863	0.0552	0.0736	0.0472	-0.0014	0.0098
MAD($\hat{\beta}_0$)	0.6569	0.7016	0.7226	0.3627	0.3188	0.2750	0.1642	0.1783
MSE($\hat{\beta}_1$)	13.504	13.770	13.8048	0.4387	0.3476	0.2132	0.0980	0.1090
RB($\hat{\beta}_1$)	3.6694	3.7134	3.7218	-0.0265	0.0247	-0.0354	-0.0493	-0.0466
MAD($\hat{\beta}_1$)	3.6694	3.7134	3.7218	0.3999	0.3257	0.2841	0.1791	0.1938
MSE($\hat{\beta}_2$)	0.8393	0.9009	0.9702	0.3716	0.2452	0.1608	0.0797	0.0767
RB($\hat{\beta}_2$)	-0.2032	-0.1788	-0.1737	-0.0156	-0.0173	0.0307	-0.0401	-0.0091
MAD($\hat{\beta}_2$)	0.6166	0.6425	0.6462	0.3405	0.2932	0.2346	0.1693	0.1795
MSE($\hat{\beta}_3$)	0.7862	0.8487	0.9068	0.3964	0.3133	0.1900	0.0839	0.0919
RB($\hat{\beta}_3$)	-0.1069	-0.1278	-0.1327	-0.1022	-0.0969	-0.0599	-0.0260	-0.0374
MAD($\hat{\beta}_3$)	0.6878	0.6706	0.7013	0.3196	0.3171	0.2757	0.1871	0.1830

Table 5: Comparison of Different Estimates for Example 2 with $n = 100$

TRUE	OLS	M_H	M_T	LMS	LTS	S	MM	REWLSE
				Case I: $\varepsilon \sim N(0,1)$				
MSE($\hat{\beta}_0$)	0.0086	0.0090	0.0090	0.0750	0.0778	0.0398	0.0089	0.0089
RB($\hat{\beta}_0$)	0.0119	0.0089	0.0087	0.0127	0.0245	0.0100	0.0076	0.0080
MAD($\hat{\beta}_0$)	0.0669	0.0684	0.0677	0.1862	0.1995	0.1329	0.0682	0.0649
MSE($\hat{\beta}_1$)	0.0102	0.0111	0.0111	0.0660	0.0753	0.0459	0.0111	0.0111
RB($\hat{\beta}_1$)	-0.0025	-0.0062	-0.0053	-0.0016	-0.0077	-0.0222	-0.0067	-0.0065
MAD($\hat{\beta}_1$)	0.0688	0.0704	0.0751	0.1559	0.1875	0.1653	0.0761	0.0761
MSE($\hat{\beta}_2$)	0.0118	0.0121	0.0124	0.0682	0.0786	0.0488	0.0123	0.0120
RB($\hat{\beta}_2$)	0.0084	0.0072	0.0106	0.0395	0.0055	-0.0150	0.0121	0.0045
MAD($\hat{\beta}_2$)	0.0727	0.0710	0.0694	0.1742	0.1718	0.1570	0.0710	0.0713
MSE($\hat{\beta}_3$)	0.0098	0.0102	0.0101	0.0621	0.0564	0.0372	0.0101	0.0101
RB($\hat{\beta}_3$)	0.0073	0.0041	0.0065	-0.0005	-0.0019	-0.0026	0.0054	0.0048
MAD($\hat{\beta}_3$)	0.0705	0.0702	0.0699	0.1714	0.1508	0.1172	0.0708	0.0703
				Case II: $\varepsilon \sim t_3$				
MSE($\hat{\beta}_0$)	0.0214	0.0134	0.0137	0.0546	0.0554	0.0326	0.0138	0.0153
RB($\hat{\beta}_0$)	0.0066	0.0019	-0.0013	-0.0362	-0.0202	-0.0162	-0.0002	0.0010
MAD($\hat{\beta}_0$)	0.1103	0.0815	0.0787	0.1319	0.1640	0.1324	0.0790	0.0811
MSE($\hat{\beta}_1$)	0.0271	0.0139	0.0139	0.0767	0.0703	0.0455	0.0137	0.0140
RB($\hat{\beta}_1$)	-0.0257	-0.0116	-0.0149	-0.034	0.0110	-0.0023	-0.0144	-0.0097
MAD($\hat{\beta}_1$)	0.1135	0.0852	0.0818	0.1658	0.1618	0.1272	0.0781	0.0852
MSE($\hat{\beta}_2$)	0.0320	0.0185	0.0177	0.0644	0.0601	0.0416	0.0181	0.0184
RB($\hat{\beta}_2$)	0.0044	0.0170	0.0311	0.0078	0.0437	0.0196	0.0293	0.0250
MAD($\hat{\beta}_2$)	0.1245	0.0901	0.0858	0.1618	0.1688	0.1369	0.0865	0.0846
MSE($\hat{\beta}_3$)	0.0311	0.0190	0.0190	0.0709	0.0658	0.0451	0.0188	0.0203
RB($\hat{\beta}_3$)	0.0022	-0.0037	0.0055	-0.0014	-0.0057	0.0235	0.0025	0.0038
MAD($\hat{\beta}_3$)	0.1231	0.0968	0.0955	0.1731	0.1600	0.1442	0.0952	0.1002
				Case III: $\varepsilon \sim t_1$				
MSE($\hat{\beta}_0$)	178.66	0.0378	0.0301	0.0624	0.0455	0.0305	0.0311	0.0330
RB($\hat{\beta}_0$)	-0.0912	-0.0267	-0.0017	-0.0239	-0.0108	-0.0169	-0.0087	-0.0124
MAD($\hat{\beta}_0$)	0.9190	0.1154	0.1205	0.1607	0.1351	0.1170	0.1139	0.1185
MSE($\hat{\beta}_1$)	72.519	0.0445	0.0333	0.0693	0.0634	0.0528	0.0359	0.0359
RB($\hat{\beta}_1$)	-0.1049	-0.0095	-0.0007	-0.0002	-0.0089	-0.0068	0.0021	-0.0002
MAD($\hat{\beta}_1$)	0.7958	0.1288	0.1218	0.1686	0.1418	0.1494	0.1292	0.1349
MSE($\hat{\beta}_2$)	198.19	0.0370	0.0342	0.0708	0.0658	0.0535	0.0352	0.0377
RB($\hat{\beta}_2$)	0.0636	0.0200	0.0035	-0.0147	0.0183	-0.0044	0.0049	0.0068
MAD($\hat{\beta}_2$)	0.6610	0.1184	0.1118	0.1645	0.1449	0.1573	0.1143	0.1268
MSE($\hat{\beta}_3$)	68.1196	0.0389	0.0323	0.0579	0.0481	0.0446	0.0338	0.0325
RB($\hat{\beta}_3$)	-0.1051	-0.0111	-0.0265	-0.0210	-0.0284	-0.0221	-0.0283	-0.0293
MAD($\hat{\beta}_3$)	0.812	0.1204	0.1232	0.1598	0.1339	0.1389	0.1294	0.1194

Table 6: Comparison of Different Estimates for Example 2 with $n = 100$

TRUE	OLS	M_H	M_T	LMS	LTS	S	MM	REWLSE
	Case IV: $\varepsilon \sim 0.95N(0,1) + 0.05N(0,10^2)$							
MSE($\hat{\beta}_0$)	0.0600	0.0136	0.0120	0.0666	0.0697	0.0365	0.0119	0.0119
RB($\hat{\beta}_0$)	-0.0074	0.0002	-0.0004	-0.0182	-0.0302	-0.0182	-0.0026	-0.0037
MAD($\hat{\beta}_0$)	0.1605	0.0811	0.0756	0.1874	0.1751	0.1370	0.0768	0.0771
MSE($\hat{\beta}_1$)	0.0638	0.0162	0.01499	0.0661	0.0786	0.0499	0.0150	0.0154
RB($\hat{\beta}_1$)	-0.0306	0.0139	0.0106	0.0049	0.0037	0.0100	0.0118	0.0190
MAD($\hat{\beta}_1$)	0.1640	0.0819	0.0832	0.1667	0.1835	0.1488	0.0829	0.0857
MSE($\hat{\beta}_2$)	0.0522	0.0128	0.0120	0.0621	0.0679	0.0368	0.0120	0.0130
RB($\hat{\beta}_2$)	0.0278	0.0125	0.0012	-0.0112	0.0070	-0.0040	0.0017	0.0040
MAD($\hat{\beta}_2$)	0.1359	0.07363	0.0709	0.1876	0.1647	0.1328	0.0706	0.0749
MSE($\hat{\beta}_3$)	0.0782	0.0170	0.0154	0.0706	0.0746	0.0422	0.0153	0.0156
RB($\hat{\beta}_3$)	-0.0063	-0.0031	0.0038	0.0036	0.0241	-0.0191	0.0048	0.0074
MAD($\hat{\beta}_3$)	0.1804	0.0759	0.0823	0.1651	0.1791	0.1373	0.0824	0.0796
	Case V: $\varepsilon \sim N(0,1)$ with outliers in y direction							
MSE($\hat{\beta}_0$)	9.0524	0.0571	0.0116	0.0669	0.0776	0.0457	0.0115	0.0115
RB($\hat{\beta}_0$)	3.0111	0.2139	0.0153	0.0299	0.0341	0.0315	0.0124	0.0088
MAD($\hat{\beta}_0$)	3.0111	0.2139	0.0698	0.1735	0.1980	0.1384	0.0693	0.0708
MSE($\hat{\beta}_1$)	0.9323	0.0137	0.0110	0.0574	0.0595	0.0378	0.0108	0.0116
RB($\hat{\beta}_1$)	-0.1124	0.0084	0.0098	0.0078	0.0304	0.0118	0.0104	0.0089
MAD($\hat{\beta}_1$)	0.6464	0.0808	0.0712	0.1487	0.1520	0.1182	0.0685	0.0749
MSE($\hat{\beta}_2$)	0.8554	0.01487	0.0121	0.0596	0.0675	0.0387	0.0120	0.0123
RB($\hat{\beta}_2$)	-0.1011	-0.0224	-0.0161	-0.0108	0.0105	-0.0219	-0.0133	-0.0162
MAD($\hat{\beta}_2$)	0.6016	0.0822	0.0724	0.1605	0.1849	0.1479	0.0715	0.0695
MSE($\hat{\beta}_3$)	0.8373	0.0161	0.0125	0.0706	0.0618	0.0378	0.0124	0.0130
RB($\hat{\beta}_3$)	-0.0809	0.0031	-0.0035	-0.0020	-0.0401	-0.0086	-0.0045	-0.0021
MAD($\hat{\beta}_3$)	0.5779	0.0823	0.0774	0.1688	0.1731	0.1381	0.0774	0.0756
	Case VI: $\varepsilon \sim N(0,1)$ with high leverage outliers							
MSE($\hat{\beta}_0$)	0.2092	0.2077	0.2095	0.0594	0.0707	0.0435	0.0111	0.0114
RB($\hat{\beta}_0$)	0.3088	0.2964	0.2991	-0.0003	-0.0203	-0.0215	-0.0107	-0.0131
MAD($\hat{\beta}_0$)	0.3358	0.3399	0.3386	0.1531	0.1871	0.1335	0.0694	0.0720
MSE($\hat{\beta}_1$)	13.303	13.683	13.726	0.0674	0.0759	0.0491	0.0114	0.0118
RB($\hat{\beta}_1$)	3.6473	3.6979	3.7026	0.0183	-0.0120	0.0132	-0.0051	-0.0047
MAD($\hat{\beta}_1$)	3.6473	3.6979	3.7026	0.1645	0.1842	0.1546	0.0758	0.0713
MSE($\hat{\beta}_2$)	0.1728	0.1812	0.1858	0.0640	0.0683	0.0373	0.0120	0.0124
RB($\hat{\beta}_2$)	-0.1141	-0.1184	-0.1281	0.0054	-0.0184	0.0069	-0.0147	-0.0132
MAD($\hat{\beta}_2$)	0.2689	0.2771	0.2796	0.1491	0.1676	0.1106	0.0712	0.0747
MSE($\hat{\beta}_3$)	0.1557	0.1592	0.1600	0.0641	0.0603	0.0342	0.0132	0.0137
RB($\hat{\beta}_3$)	-0.1083	-0.1104	-0.1088	-0.0239	-0.0222	0.0047	-0.0045	0.0034
MAD($\hat{\beta}_3$)	0.2561	0.2614	0.2637	0.1664	0.1564	0.1299	0.0724	0.0783

Table 7: Outlier detection results for Example 3

	5% outliers			10% outliers		
	M	S	JD	M	S	JD
M_H	0.976	0.008	0.000	0.983	0.009	0.000
M_T	0.976	0.008	0.000	0.983	0.009	0.000
LMS	0.122	0.030	0.865	0.285	0.029	0.690
LTS	0.091	0.0249	0.900	0.332	0.025	0.645
S	0.088	0.008	0.905	0.365	0.008	0.625
MM	0.088	0.006	0.910	0.379	0.006	0.615
REWLSE	0.118	0.006	0.875	0.348	0.005	0.645

Table 8: Regression Estimates for Modified Data on Wood Specific Gravity

Estimators	β_1	β_2	β_3	β_4	β_5	Intercept
OLS	0.4407	-1.4750	-0.2612	0.0208	0.1708	0.4218
LTS	0.2384	-0.0699	-0.5695	-0.3839	0.6821	0.2880
S	0.2051	-0.1765	-0.5276	-0.4438	0.6163	0.3931
MM	0.2165	-0.0808	-0.5639	-0.3982	0.6046	0.3784
LMS	0.2687	-0.2381	-0.5357	-0.2937	0.4510	0.4347

Table 9: Modified Data on Wood Specific Gravity With Standardized Residuals From OLS, LTS, S, MM and LMS

								Residual/Scale				
i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	y_i	OLS	LTS	S	MM	LMS
1	0.5730	0.1059	0.4650	0.5380	0.8410	1.000	0.5340	-0.7250	2.4231	0.5728	0.5937	-0.0827
2	0.6510	0.1356	0.5270	0.5450	0.8870	1.000	0.5350	0.0472	0.9400	0.3391	0.3306	0.0013
3	0.6060	0.1273	0.4940	0.5210	0.9200	1.000	0.5700	1.2425	0.0457	0.2364	0.0313	0.2836
4	0.4370	0.1591	0.4460	0.4230	0.9920	1.000	0.4500	0.3546	-31.878	-13.599	-14.067	-7.6137
5	0.5470	0.1135	0.5310	0.5190	0.9150	1.000	0.5480	1.0023	2.5153	1.2121	0.9750	0.9020
6	0.4440	0.1628	0.4290	0.4110	0.9840	1.000	0.4310	-0.4518	-36.752	-15.813	-16.268	-9.1023
7	0.4890	0.1231	0.5620	0.4550	0.8240	1.000	0.4810	0.9066	2.9574	0.5036	0.4830	0.3746
8	0.4130	0.1673	0.4180	0.430	0.9780	1.000	0.4230	-0.0349	-35.947	-15.623	-15.967	-8.9077
9	0.5360	0.1182	0.5920	0.4640	0.8540	1.000	0.4750	-0.3959	0.0457	-0.4185	-0.5688	-0.3746
10	0.6850	0.1564	0.6310	0.5640	0.9140	1.000	0.4860	-0.4150	-0.3174	-0.0580	-0.0245	-0.2071
11	0.6640	0.1588	0.5060	0.4810	0.8670	1.000	0.5540	1.9856	0.0457	-0.2014	-0.2003	0.0013
12	0.7030	0.1335	0.5190	0.4840	0.8120	1.000	0.5190	-1.1975	0.0457	-0.5225	-0.4747	-0.9656
13	0.6530	0.1395	0.6250	0.5190	0.8920	1.000	0.4920	-0.4854	0.8094	0.3226	0.2394	0.0013
14	0.5860	0.1114	0.5050	0.5650	0.8890	1.000	0.5170	-1.2610	-0.7948	-0.4711	-0.5230	-0.6709
15	0.5340	0.1143	0.5210	0.5700	0.8890	1.000	0.5020	-0.5865	0.6440	0.0307	0.0326	-0.1733
16	0.5230	0.1320	0.5050	0.6120	0.9190	1.000	0.5080	0.5237	0.0457	-0.1238	-0.0139	0.0013
17	0.5800	0.1249	0.5460	0.6080	0.9540	1.000	0.5200	-0.2548	-0.6450	0.0344	-0.0546	0.0013
18	0.4480	0.1028	0.5220	0.5340	0.9180	1.000	0.5060	0.2838	-0.9106	-0.4673	-0.6790	-0.1090
19	0.4170	0.1687	0.4050	0.4150	0.9810	1.000	0.4010	-1.0836	-42.291	-18.381	-18.770	-10.726
20	0.5280	0.1057	0.4240	0.5660	0.9090	1.000	0.5680	0.5450	0.0457	0.0141	-0.0990	0.0013

Table 10: Hawkins, Bradu, and Kass Data With Standardized Residuals From OLS, LTS, LMS, S and MM

i	OLS	LTS	LMS	S	MM	i	OLS	LTS	LMS	S	MM
1	1.4999	14.040	16.535	12.371	12.293	39	-0.5745	-0.6933	-0.6690	-0.6965	-0.7829
2	1.7759	14.847	17.406	13.032	12.854	40	-0.0099	-0.0116	-0.0556	-0.1934	-0.4685
3	1.3295	15.055	17.782	13.262	13.140	41	-0.4982	-0.1584	-0.0101	-0.1632	-0.1335
4	1.1369	14.155	16.687	12.433	12.195	42	-0.4855	0.0740	0.1245	-0.1518	-0.5418
5	1.3583	14.693	17.316	12.928	12.763	43	0.7704	1.5833	1.6742	1.1729	0.8701
6	1.5240	14.297	16.857	12.631	12.616	44	-0.8093	-0.3356	-0.2317	-0.4228	-0.6398
7	2.0050	15.516	18.205	13.672	13.623	45	-0.3459	-0.4275	-0.4002	-0.4718	-0.5562
8	1.7026	15.063	17.706	13.239	13.106	46	-0.6416	0.0743	0.2280	-0.0659	-0.2753
9	1.2022	14.338	16.897	12.582	12.337	47	-0.6773	-1.3703	-1.4135	-1.2232	-1.1638
10	1.3477	14.901	17.538	13.048	12.763	48	-0.2427	0.1680	0.3074	0.1052	0.1034
11	-3.4830	-0.1701	0.4452	0.0769	-0.0610	49	0.2863	1.1861	1.4335	0.9974	1.0416
12	-4.1743	-0.4320	0.3040	-0.1144	-0.2274	50	-0.3093	-0.2452	-0.2225	-0.3114	-0.4384
13	-2.7174	0.4759	1.1097	0.7177	0.7903	51	0.3901	1.0390	1.2209	0.8315	0.8249
14	-1.6666	-0.7451	-0.6690	-0.3591	-0.2921	52	-0.5249	-0.8739	-0.8425	-0.7766	-0.6907
15	-0.2979	-0.8288	-0.8548	-0.7288	-0.6373	53	-0.0263	2.2484	2.6351	1.7225	1.2866
16	0.3819	0.5360	0.6539	0.4493	0.5690	54	0.7640	1.3158	1.4109	1.0068	0.8646
17	0.2885	0.4903	0.4754	0.2189	-0.0845	55	0.3242	0.6418	0.6607	0.3775	0.1196
18	-0.1802	0.2414	0.3621	0.1251	0.0473	56	0.3487	0.2622	0.2572	0.1322	0.0642
19	0.2917	0.7430	0.7921	0.4788	0.2351	57	0.2719	1.3283	1.5702	1.0443	0.9420
20	0.1486	0.4199	0.5334	0.3356	0.3832	58	0.1222	-0.1701	-0.1946	-0.2107	-0.2211
21	0.2952	1.3877	1.6517	1.1399	1.1110	59	-0.3336	0.1523	0.2389	-0.0022	-0.2144
22	0.4179	1.1949	1.2841	0.8442	0.5344	60	-0.6017	-0.1735	-0.1587	-0.3697	-0.7796
23	-0.1919	-1.0632	-1.2060	-1.0068	-1.0502	61	-0.0071	0.4612	0.4772	0.1987	-0.1570
24	0.6022	1.3358	1.4797	1.0306	0.8950	62	0.3028	1.4909	1.7243	1.1312	0.9165
25	-0.1403	0.1607	0.2234	-0.0158	-0.2167	63	0.2911	-0.1379	-0.2334	-0.2523	-0.3870
26	-0.2130	-0.5121	-0.6183	-0.6089	-0.8698	64	-0.3994	-0.6549	-0.6636	-0.6225	-0.6498
27	-0.6181	-1.080	-1.1022	-0.9628	-0.9281	65	-0.1370	1.4073	1.7357	1.1044	0.9328
28	-0.1136	0.9259	1.1510	0.6831	0.5220	66	-0.1169	-0.5340	-0.6690	-0.6609	-0.9578
29	0.1722	0.9506	1.1064	0.6807	0.4976	67	-0.2180	-0.6910	-0.7758	-0.7146	-0.8377
30	-0.5705	0.4733	0.6579	0.2399	-0.0686	68	0.2373	1.8989	2.1808	1.4271	1.0622
31	-0.1257	-0.1701	-0.0921	-0.1732	-0.0960	69	0.0814	0.4751	0.6002	0.3214	0.2680
32	0.2492	-0.0328	-0.1266	-0.2005	-0.4262	70	0.2082	1.8047	2.0823	1.3933	1.0981
33	-0.0479	-0.5604	-0.6690	-0.6050	-0.7397	71	0.0041	0.5359	0.6638	0.3755	0.2822
34	-0.3046	-0.1701	-0.1718	-0.3418	-0.6366	72	0.0611	0.4370	0.4695	0.1981	-0.0801
35	-0.1840	0.4784	0.6690	0.3729	0.3664	73	0.1951	1.5509	1.7583	1.1355	0.7729
36	-0.5241	-0.7504	-0.8263	-0.8108	-1.0700	74	-0.1719	-0.3801	-0.4942	-0.5459	-0.9029
37	-0.0999	0.1027	0.0587	-0.1279	-0.5094	75	-0.1588	1.4297	1.6891	1.0198	0.6185
38	0.5546	1.6095	1.8378	1.2772	1.1646						

Table 11: Breakdown Points and Asymptotic Efficiencies of Various Regression Estimators

	Estimator	Breakdown Point	Asymptotic Efficiency
High BP	LMS	0.5	0
	LTS	0.5	0.08
	S-estimates	0.5	0.29
	GS-estimates	0.5	0.67
	MM-estimates	0.5	0.95
	GM-estimates(S1S)	0.5	0.95
	REWLSE	0.5	1.00
Low BP	GM-estimates(Mallows,Schweppe)	$1/(p+1)$	0.95
	Bounded R-estimates	< 0.2	0.90-0.95
	Monotone M-estimates	$1/n$	0.95
	LAD	$1/n$	0.64
	OLS	$1/n$	1.00

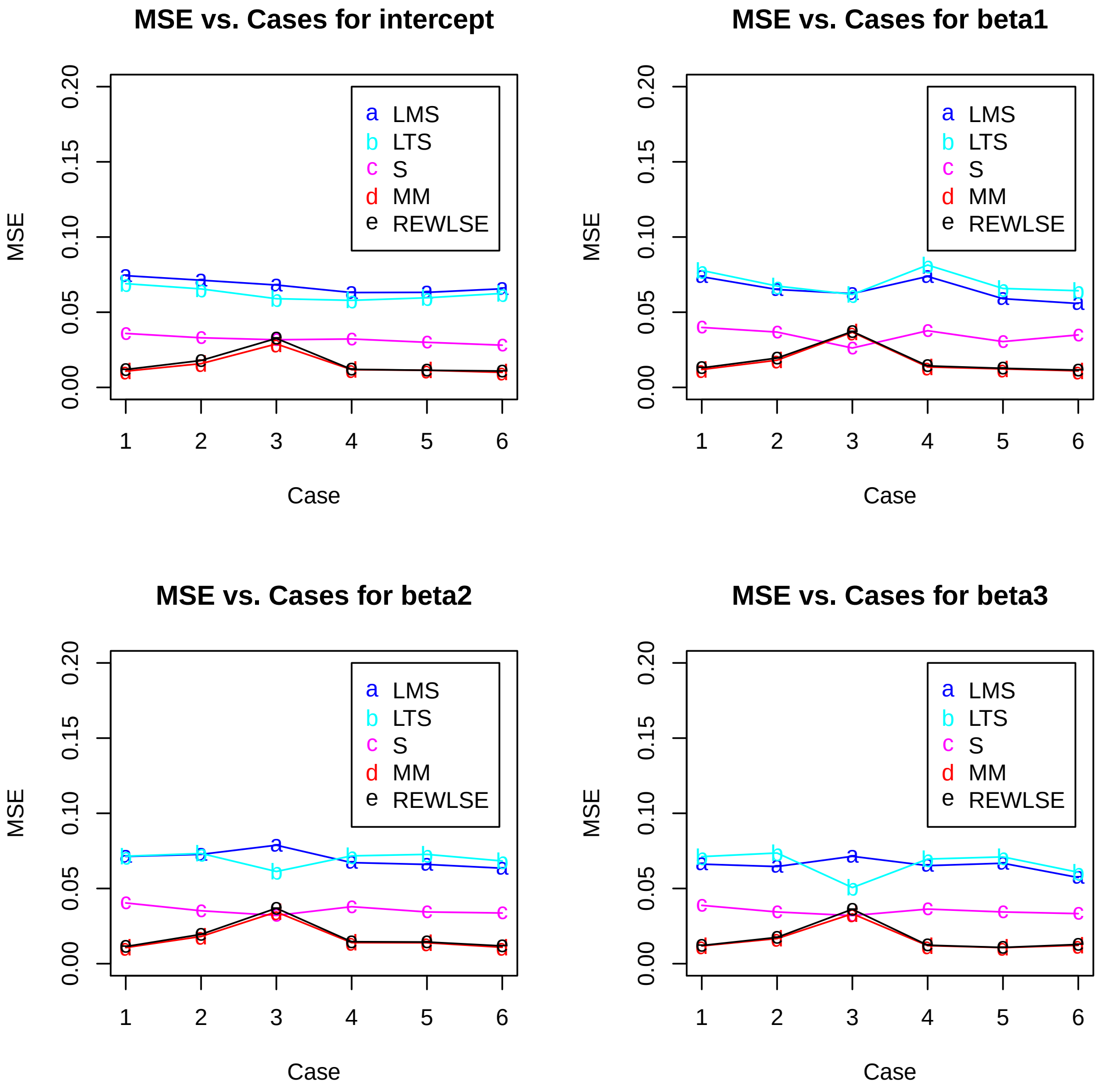

Figure 2: Plot of MSE of different regression parameter estimates vs. different cases for LMS, LTS, S, MM, and REWLSE, for Example 2 when $n = 100$.

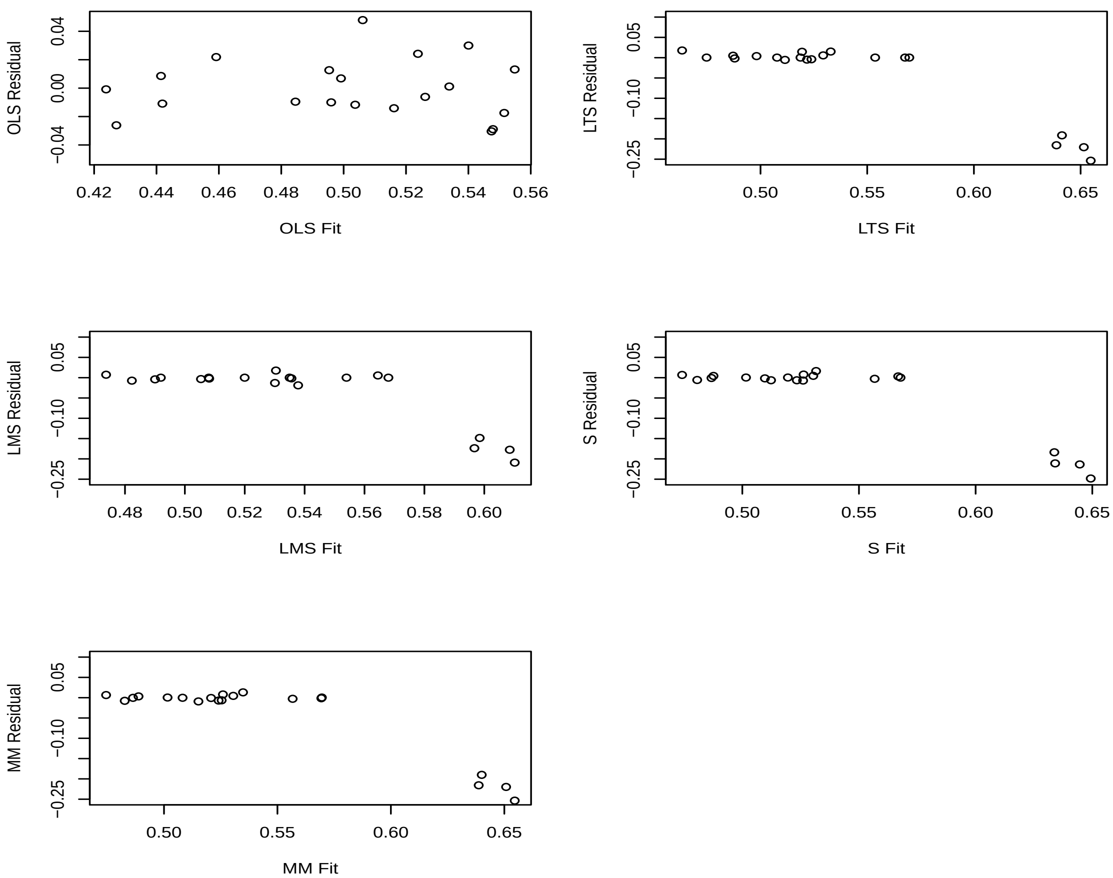

Figure 3: Plots of residual versus fitted values for OLS, LTS, S, MM and LMS for modified wood data