Computational Statistics and Data Analysis 56 (2012) 1952–1965

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

# Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression

Olcay Arslan *

*Department of Statistics, Faculty of Science, Ankara University, 06100 Tandogan, Ankara, Turkey*

ARTICLE INFO

*Article history:*
Received 1 March 2011
Received in revised form 28 November 2011
Accepted 29 November 2011
Available online 8 December 2011

*Keywords:*
LAD
LASSO
Median regression
Robustness
Regression
WLAD-LASSO

ABSTRACT

The weighted least absolute deviation (WLAD) regression estimation method and the adaptive least absolute shrinkage and selection operator (LASSO) are combined to achieve robust parameter estimation and variable selection in regression simultaneously. Compared with the LAD-LASSO method, the weighted LAD-LASSO (WLAD-LASSO) method will resist to the heavy-tailed errors and outliers in explanatory variables. Properties of the WLAD-LASSO estimators are investigated. A small simulation study and an example are provided to demonstrate the superiority of the WLAD-LASSO method over the LAD-LASSO method in the presence of outliers in the explanatory variables and the heavy-tailed error distribution.



## 1. Introduction

There are two main problems of the regression analysis. One is the parameter estimation and the other is the variable selection. Estimation of the parameters are usually carried out with the ordinary least squares regression estimation method. But, it is well-known that the ordinary least squares (OLS) regression estimators are highly sensitive to the outliers in the datasets and as a result robust regression methods have been proposed as alternatives. One of these robust estimation methods is the least absolute deviation (LAD) regression method, where the regression parameters are estimated through the minimization of the sum of the absolute value of the errors. The LAD regression method, which is a special case of the $M$-regression method, is particularly well-suited to the heavy-tailed error distributions. However, if the outliers occur in the explanatory variables (leverage points) the performance of the LAD regression estimators are not better than the OLS estimators. They are highly sensitive to the leverage points. Even a single leverage point can badly affect the LAD regression estimator. To deal with the outliers in the explanatory variables the weighted LAD regression estimation has been proposed (Ellis and Morgenthaler, 1992; Hubert and Rousseeuw, 1997; Giloni et al., 2006a,b). In this method weights which are only dependent on the explanatory variables are introduced to downweight the leverage points and hence reduce the effect of the outliers on the estimation procedure.

Variable selection is another important problem in regression analysis. A large number of regressors are usually introduced at the initial stage of the regression model to attenuate possible modeling biases. However, including unnecessary predictors can degrade the efficiency of the resulting estimation procedure and yields less accurate predictions. On the other hand, omitting an important explanatory variable may produce biased parameter estimates and prediction results. Therefore, since selecting the significant explanatory variables is an important task, a number of methods, including

* Tel.: +90 312 2126720x1210; fax: +90 312 2233202.
*E-mail address:* oarslan@ankara.edu.tr.

0167-9473/$ – see front matter 
doi:10.1016/j.csda.2011.11.022

the Akaike Information Criterion (ACI), the Bayesian Information Criteria (BIC) and Mallows $C_p$, have been proposed and well established (Shao, 1997; Hurvich and Tsai, 1989, 1990; Shi and Tsai, 2002, 2004; Miller, 2002; Ronchetti and Staudte, 1994).

The least absolute shrinkage and selection operator (LASSO) was proposed by Tibshirani (1996) for model selection and estimation in regression models. The LASSO method is proposed to carry on the variable (model) selection and the parameter estimation simultaneously. These two procedures have been combined in a single minimization problem. However, since the LASSO is a special case of the penalized least squares regression with $L_1$-*penalty* function it suffers from the heavy-tailed errors and/or outliers in data. To make the LASSO resistant to the outliers and heavy-tailed error distributions the robust versions of the LASSO method have been briefly mentioned by Fan and Li (2001), Owen (2007), Rosset and Zhu (2004) and Hesterberg et al. (2008).

Recently, the LAD regression and the LASSO methods have been combined (the LAD-LASSO regression method) to carry out robust parameter estimation and variable selection simultaneously (see Xu (2005), Wang and Leng (2007), Gao (2008), and Xu and Ying (2010)). However, it is well-known that the LAD regression estimation method is only resistant to the outlier in the response variable, but not resistant to the leverage points. Thus, as Wang and Leng (2007) point out, combining the LAD and the LASSO methods can only produce estimators that are only resistant to the outliers in the response variable. If there are outliers in predictors, the performance of the LAD regression estimators are not better than the OLS regression estimators so that the performance of the LAD-LASSO estimators will not be better than the ordinary LASSO estimators.

In this paper we will propose a weighted version of the LAD-LASSO method to find the robust regression estimators and select the appropriate predictors. In our proposal we will combine the WLAD regression criterion and the adaptive LASSO penalty criterion introduced by Zou (2006). The WLAD criterion will downweight the leverage points and be resistant to the outliers in the response variable so that the resulting regression estimators will be less sensitive to the leverage points and the outliers. On the other hand, adaptive LASSO penalty criterion will assign different weights to different coefficient to penalize different coefficient in $L_1$ penalty criterion.

The rest of the paper is organized as follows. In Section 2 we introduce the WLAD-LASSO regression method to achieve robust parameter estimation and the variable selection simultaneously in a regression analysis. We discuss some of its theoretical properties in Section 3. In Section 4 we provide a small simulation study and an example to demonstrate the performance of the proposed method. The paper is finalized with a discussion section.

## 2. Weighted LAD-LASSO (WLAD-LASSO)

Consider the linear regression model

$$y_i = \beta_0 + \mathbf{x}_i^T \beta + \varepsilon_i, \quad i = 1, 2, 3, \ldots, n, \tag{1}$$

where $y_i \in R$ is the response variable, $\mathbf{x}_i = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)^T$ is the $p$-dimensional covariate vector, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ are the regression parameters, and $\varepsilon_i$'s are the iid random errors. Without loss of generality, we assume that $\beta_0 = 0$. This can be achieved by centering the covariates and response variable. That is, from now on we will consider the model

$$y_i = \mathbf{x}_i^T \beta + \varepsilon_i, \quad i = 1, 2, 3, \ldots, n. \tag{2}$$

The most popular way of estimating $\beta$ is to minimize the ordinary least squares (OLS) criterion

$$\sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \beta\right)^2, \tag{3}$$

which yields the estimator $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$, where $X$ is the $n \times p$ matrix whose $i$th row is $\mathbf{x}_i^T$ with rank $p$, and $Y = (y_1, y_2, \ldots, y_n)^T$ is the response vector. The usual assumption for using the OLS method is that the random errors $\varepsilon_i$ are normally distributed with mean zero and variance $\sigma^2$. However, it is well-known that the OLS method is very sensitive to the heavy-tailed errors and/or the outliers.

One of the simple robust alternative to the OLS is the LAD regression method. The usual assumption for the LAD method is that the random errors $\varepsilon_i$ have median zero. The regression parameters are estimated by minimizing the LAD criterion

$$\sum_{i=1}^{n} \left|y_i - \mathbf{x}_i^T \beta\right|. \tag{4}$$

However, it is also well known that the LAD regression method is resistant to the outliers in response variable, but it is very sensitive to the outlying observations in the explanatory variables. To correct this problem of the LAD regression method the WLAD regression estimation method has been proposed, which is obtained by minimizing the WLAD objective function

$$\sum_{i=1}^{n} w_i \left|y_i - \mathbf{x}_i^T \beta\right|, \tag{5}$$

where $w_i$, for $i = 1, 2, 3, \ldots, n$, is the weights which are determined by a robust measure of $\mathbf{x}_i$ and will be chosen to downweight the leverage points. Ellis and Morgenthaler (1992) appear to be the first to mention weighted LAD regression estimation method to gain robustness against to the outliers in the regressors. They explicitly showed that suitably chosen weights can improve the finite sample breakdown point of the LAD regression estimators. Hubert and Rousseeuw (1997) used the WLAD regression method to estimate the regression parameters with both continuous and binary regressors. Recently, Giloni et al. (2006) proposed a version of the WLAD regression method to increase the breakdown of the LAD regression estimators. They also discussed some properties of the resulting regression estimators. They showed via simulation that the performance of the weighted LAD regression estimators is competitive with that of the high breakdown point regression estimators in the presence of leverage points. Further, since the WLAD regression estimation problem on our original dataset $y_i$ and $\mathbf{x}_i$ can be reduced to the LAD problem on the transform dataset $\tilde{y}_i = w_i y_i$ and $\tilde{\mathbf{x}}_i = w_i \mathbf{x}_i$ for $i = 1, 2, 3, \ldots, n$ the WLAD regression estimators can be easily computed.

Tibshirani (1996) introduced the LASSO method for simultaneous parameter estimation and variable selection by minimizing the following penalized least squares regression criterion

$$\sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \beta\right)^2 + n\lambda \sum_{j=1}^{p} \left|\beta_j\right|, \tag{6}$$

where $\lambda > 0$ is the tuning parameter. Later, Zou (2006) introduced the adaptive (or weighted) LASSO by using the different tuning parameters for different regression coefficients. He suggests to minimize the following objective function

$$\sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \beta\right)^2 + n \sum_{j=1}^{p} \lambda_j \left|\beta_j\right|. \tag{7}$$

Thus, the adaptive LASSO procedure will be more efficient than the traditional LASSO procedure to estimate the parameters and select the significant variables. However, the LASSO regression estimation is based on the least squares criterion so that the resulting estimators will be very sensitive to the outliers and /or heavy-tailed errors.

Recently, the LAD-LASSO method has been introduced as an alternative to the LASSO method to achieve robust regression estimation and variable selection simultaneously (see Xu (2005), Wang and Leng (2007), Gao (2008), and Xu and Ying (2010)). In this case, the regression parameters are estimated by minimizing the following objective function

$$Q(\beta) = \sum_{i=1}^{n} \left|y_i - \mathbf{x}_i^T \beta\right| + n \sum_{j=1}^{p} \lambda_j \left|\beta_j\right|. \tag{8}$$

Xu (2005), Wang and Leng (2007), Gao (2008), and Xu and Ying (2010) study several properties of the resulting estimators and show that the LAD-LASSO regression estimator has oracle property, and compare with the traditional LASSO, it is resistant to the outliers in the response variable. They discuss how one can find the LAD-LASSO estimates, and choose the tuning parameters $\lambda_j$.

The LAD-LASSO method has been introduced to gain robustness against the heavy-tailed error distributions. But, the LAD-LASSO will not be robust against the outliers in the explanatory variables. In this paper, we will further extend the LAD-LASSO methods to get regression estimators that are also robust against the leverage points. We will achieve this by combining the WLAD criterion with the $L_1$-type penalty function. To obtain the WLAD-LASSO regression estimators we will minimize the following objective function

$$Q_w(\beta) = \sum_{i=1}^{n} w_i \left|y_i - \mathbf{x}_i^T \beta\right| + n \sum_{j=1}^{p} \lambda_j \left|\beta_j\right|, \tag{9}$$

where $w_i$ are the weights obtained from the robust distances of $\mathbf{x}_i$, for $i = 1, 2, 3, \ldots, n$, and $\lambda_j, j = 1, 2, 3, \ldots, p$, are the tuning parameters in the adaptive LASSO objective function and will be estimated from the data. The weights $w_i$ will be obtained using the robust distances of the explanatory variables so that the outlying observations in the direction of the explanatory variables will have large distances and the corresponding weights will be small. Therefore, it is expected that the resulting regression estimator will be robust against the outliers in the response variable and leverage points.

The WALD-LASSO estimator can be easily computed by augmenting the data and using the LAD algorithms. This can be done as follows. First, find $\tilde{y}_i = w_i y_i$ and $\tilde{\mathbf{x}}_i = w_i \mathbf{x}_i$ for $i = 1, 2, 3, \ldots, n$, and define $\left(y_i^*, \mathbf{x}_i^{*^T}\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$, where $\left(y_i^*, \mathbf{x}_i^{*T}\right) = \left(\tilde{y}_i, \tilde{\mathbf{x}}_i^T\right)$ for $i = 1, 2, 3, \ldots, n$ and $\left(y_i^*, \mathbf{x}_i^{*T}\right) = \left(0, n\lambda_j \mathbf{e}_j^T\right)$ for $j = 1, 2, 3, \ldots, p$ and $\mathbf{e}_j$ is a $p$-dimensional vector with the jth term equals to 1 and all others equal to zero. Then, it can be easily verify that

$$Q_w(\beta) = \sum_{i=1}^{n} w_i \left|y_i - \mathbf{x}_i^T \beta\right| + n \sum_{j=1}^{p} \lambda_j \left|\beta_j\right| = \sum_{i=1}^{n+p} \left|y_i^* - \mathbf{x}_i^{*^T} \beta\right|. \tag{10}$$

This is just the traditional LAD criterion for the dataset $\left(y_i^*, \mathbf{x}_i^{*^T}\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$. Therefore, any of the LAD regression algorithms can be used to obtain the WLAD-LASSO estimator.

However, before we use any of the standard LAD regression algorithm to compute the WLAD-LASSO estimator we have to find the weights $w_i$ and also estimate the tuning parameters $\lambda_j$. The weights will be computed using the weight definition given in Hubert and Rousseeuw (1997). Consider $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ as a dataset in $R^p$ and find robust location and scatter estimators $\hat{\mu}$ and $\hat{\Sigma}$. One can use high breakdown point location and scatter estimators such as MVE or MCD. Compute the robust distances defined as $RD(\mathbf{x}_i) = (\mathbf{x}_i - \hat{\mu})^T \hat{\Sigma}^{-1}(\mathbf{x}_i - \hat{\mu})$ for $i = 1, 2, 3, \ldots, n$. Observations for which $RD(\mathbf{x}_i)$ large can be identified as leverage points. Then, based on the robust distances $RD(\mathbf{x}_i)$ we can compute strictly positive weights $w_i$ through the relation $w_i = \min\left\{1, \frac{p}{RD(\mathbf{x}_i)}\right\}$, for $i = 1, 2, 3, \ldots, n$. Since the $w_i$ decrease as $RD(\mathbf{x}_i)$ increase the leverage points will receive smaller weights and will be downweighted. Let $\hat{\boldsymbol{\beta}}$ be a root $n$-consistent estimator of $\beta$, then, for any $\gamma > 0$, $\lambda_j$ can be estimated using the relation $\hat{\lambda}_j = 1/(\left|\hat{\beta}_j\right|^{\gamma})$ for $j = 1, 2, 3, \ldots, p$ (Zou, 2006; Wang and Leng, 2007; Wang et al., 2007). For the case $\gamma = 1$, the estimator $\hat{\lambda}_j$ can be obtained by viewing the LASSO estimator as a Bayesian estimator with each regression parameter having the symmetric univariate Laplace distribution with location zero and scale $\lambda_j$. Then, $\lambda_j$ can be estimated using the posterior log-likelihood function. Here, $\hat{\beta}$ can be any robust estimator of $\beta$, for instance, it can be taken as the unpenalized WLAD regression estimator.

Then, the estimation procedure for the WALD-LASSO regression method can be described as follows:

- First, find the robust estimates $\hat{\mu}$ and $\hat{\Sigma}$ for the location vector and the scatter matrix of the data $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in R^p$ and compute the weights $w_i$ for $i = 1, 2, 3, \ldots, n$.
- Estimate $\beta$ using any of the robust regression procedure. For example, one can estimate $\beta$ by minimizing the unpenalized WALD objective function $\sum_{i=1}^n w_i \left|y_i - \mathbf{x}_i^T\beta\right|$. Using $\hat{\beta}$ compute $\hat{\lambda}_j$, for $j = 1, 2, 3, \ldots, p$.
- Form the dataset $\left(y_i^*, \mathbf{x}_i^{*^T}\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$, where $\left(y_i^*, \mathbf{x}_i^{*T}\right) = \left(\tilde{y}_i, \tilde{\mathbf{x}}_i^T\right) = (w_i y_i, w_i \mathbf{x}_i^T)$ for $i = 1, 2, 3, \ldots, n$ and $\left(y_i^*, \mathbf{x}_i^{*T}\right) = \left(0, n\hat{\lambda}_j \mathbf{e}_j^T\right)$ for $j = 1, 2, 3, \ldots, p$.
- Use one of the LAD regression algorithms to find the LAD regression estimators for the dataset $\left(y_i^*, \mathbf{x}_i^{*^T}\right)$ for $i = 1, 2, 3, \ldots, n, n+1, \ldots, n+p$.

## 3. Theoretical properties

In this section we discuss some relevant asymptotic theories of the weighted LAD-LASSO estimators. To study the asymptotic theory we need the following assumptions and the definitions.

A1. The errors $\varepsilon_i$ are independent and identically distributed with median 0 and a density function $f$ which is continuous and positive at the origin.

A2. Let $W = \mathrm{diag}(w_1, w_2, w_3, \ldots, w_n)$ and assume that $w_i$, for $i = 1, 2, 3, \ldots, n$, are known positive values that satisfy $\max\{w_i\} = O(1)$.

A3. $\lim_{n\to\infty} \frac{X^T W X}{n} = Q$, for some positive definite matrix $Q$.

These assumptions are needed to establish asymptotic normality of the WLAD estimator. Under these assumptions it has been proved that $\sqrt{n}\left(\hat{\beta}_{W0} - \beta\right)$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $Q^{-1}(X^T W^2 X)Q^{-1}/(2f(0))^2$, where $f$ is the density function of $\varepsilon_i$ (see (Giloni et al., 2006a,b)). Note that where $(2f(0))^2$ is the asymptotic variance of the sample median of the random samples from the distribution with the density function $f$.

To study the asymptotic properties of the WLAD-LASSO we also need the following definitions and the notations. Decompose the parameter vector $\beta$ as $\beta = (\beta_a^T, \beta_b^T)^T$, where $\beta_a = (\beta_1, \beta_2, \ldots, \beta_{p_0})^T$ is the vector corresponding to all of the nonzero coefficients and $\beta_b = (\beta_{p_0+1}, \beta_{p_0+2}, \ldots, \beta_p)^T$ is the vector corresponding to all the zero coefficients. Let $\hat{\beta}_w = (\hat{\beta}_{wa}^T, \hat{\beta}_{wb}^T)^T$ be the corresponding WLAD-LASSO estimator. We also decompose the regressors $\mathbf{x}_i$ for $i = 1, 2, \ldots, n$ as $\mathbf{x}_i = (\mathbf{x}_{ia}^T, \mathbf{x}_{ib}^T)^T$ with $\mathbf{x}_{ia} = (x_{i1}, x_{i2}, \ldots, x_{ip_0})^T$ and $\mathbf{x}_{ib} = (x_{ip_0+1}, x_{ip_0+2}, \ldots, x_{ip})^T$. Finally, define $a_n = \max\{\lambda_j \colon 1 \le j \le p_0\}$ and $b_n = \min\{\lambda_j : p_0 < j \le p\}$, where $\lambda_j$ is a function of $n$ (see Wang and Leng (2007), and Wang et al. (2007)). Under these assumptions the following results can be established.

**Result 1.** *Suppose that $(y_i, \mathbf{x}_i)$ for $i = 1, 2, 3, \ldots, n$ are independent and identically distributed and the assumptions given in A1–A3 are satisfied. Then,*

(i) *if $\sqrt{n}a_n \to 0$ as $n \to \infty$ the WLAD-LASSO estimator is $\sqrt{n}-$ consistent.*

(ii) *If $\sqrt{n}a_n \to 0$ and $\sqrt{n}b_n \to \infty$ as $n \to \infty$, then, with probability 1, the WLAD − LASSO estimator $\hat{\beta}_w = (\hat{\beta}_{wa}^T, \hat{\beta}_{wb}^T)^T$ must satisfy $\hat{\beta}_{wb} = \mathbf{0}$.*

These results imply that if the tuning parameters are properly chosen the WLAD-LASSO estimator is $\sqrt{n}-$ consistent and possesses the sparsity property. Thus, the WLAD-LASSO estimators perform as well as the WLAD estimators for estimating $\beta_a$ knowing that $\beta_b = \mathbf{0}$. Further, the following result states the asymptotic normality of the WLAD-LASSO estimator.

**Result 2.** *Suppose that $(y_i, \mathbf{x}_i)$ for $i = 1, 2, 3, \ldots, n$ are independent and identically distributed and the assumptions given in A1–A3 are satisfied. Then, if $\sqrt{n}a_n \to 0$ and $\sqrt{n}b_n \to \infty$ as $n \to \infty$,*

$$\sqrt{n}\left(\hat{\beta}_{wa} - \beta_a\right) \to N_{p_0}\left(\mathbf{0}, (2f(0))^{-2}\boldsymbol{\Sigma}_a\right)$$

*in distribution, where $f$ is the density of $\varepsilon_i$, $\boldsymbol{\Sigma}_a = Q_a^{-1}(X_a^T W^2 X_a)Q_a^{-1}$ and $Q_a = \lim_{n\to\infty} \frac{X_a^T W X_a}{n}$.*

Since the proofs of the above results follow from the adaptations of the proofs of Wang and Leng (2007, Lemmas 1 and 2 and Theorem) and Giloni et al. (2006a,b, Lemmas 1 and 2 and Theorem 1) they will be omitted.

## 4. Numerical studies

### 4.1. Simulation

In this section we provide a small simulation study to illustrate the performance of the WLAD-LASSO estimates for $n > p$. An extensive comparison of the LASSO and LAD-LASSO has been reported in the papers by Wang and Leng (2007) and Xu and Ying (2010). Here, we investigate the finite-sample performance of the WLAD-LASSO method over the LAD-LASSO method in terms of the variable selection and the robust estimation of the regression parameters.

Let $\epsilon$ be the contamination rate with the values 0, 0.1, 0.2, 0.3 and 0.4. For a given contamination rate $\epsilon$ let $m = [\epsilon n]$ be the number of contaminated data, where $n$ is the sample size with the values 50, 100 and 200 and $[\cdot]$ denotes the integer part. The first $n - m$ data $(\mathbf{x}_{1i}, y_{1i})$ are generated as follows. We set $\beta_1 = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$ and generate $\mathbf{x}_{1i} \sim N_p(\mathbf{0}, \mathbf{V})$ where $\mathbf{V} = \left(v_{ij}\right)$ with $v_{ij} = 0.5^{|i-j|}$. The response variables are generated according to the linear regression model $y_{1i} = \mathbf{x}_{1i}^T\beta_1 + \sigma\varepsilon_i$, for $i = 1, 2, \ldots, n - m$, where $\varepsilon$ is generated from the standard $t$-distributions with 1 (Cauchy distribution), 3 and 5 degrees of freedoms and two different values of $\sigma$ (0.5 and 1) are taken. These distributions allow us to have heavy-tailed error distribution and some possible outliers in $y$ direction. These $n-m$ data points will form the main bulk of the dataset. Next, the $m$ contaminated data points $(\mathbf{x}_{2i}, y_{2i})$ are produced as follows. We generate $\mathbf{x}_{2i} \sim N_p(\mu, \mathbf{I})$ with $\mu \neq \mathbf{0}$, take a vector $\beta_2 \neq \beta_1$ and let $y_{2i} = \mathbf{x}_{2i}^T\beta_2$, $i = 1, 2, \ldots, m$. These $m$ points will form the contaminated part of the dataset. We combine these two datasets to make one dataset $(\mathbf{x}_i, y_i)$ for $i = 1, 2, \ldots, n$. Finally, we fit the linear regression model $y = \mathbf{x}^T\beta + \varepsilon$ to our final dataset and try to estimate the unknown regression parameter $\beta$. For each case we repeated the simulation 200 times. The above procedures are repeated for different values of $\gamma$. Since the results are very similar we only report the results for $\gamma = 7$. In Tables 1–3 we summarize the simulation results, which include the percentage of correctly estimated regression models, the average correct 0 coefficients and the average incorrect 0 coefficients along with the mean and median of the mean-squared errors (*MeanMSE* and *MedianMSE*) over 200 simulated datasets.

From these results we see that the performance of the LAD-LASSO and WLAD-LASSO are very similar when there are no leverage points. However, we observe that when there are some leverage points the WLAD-LASSO estimation method performs much better than the LAD-LASSO method in terms of model selection and robust estimation of the regression parameters. The WLAD-LASSO method can correctly identify the number of zero coefficients and it has the smallest average and median MSEs among the other methods considered in this simulation study. The worst performance of the WLAD-LASSO method can be observed for the case when the error distribution is Cauchy and the sample size is small. For the other cases considered in this simulation study its performance seems much better that the LAD-LASSO method. We can further observe that with the increasing sample size, the WLAD-LASSO has higher ratios of finding the right model even the ratio of the outliers is very high. For example, when $n = 100$, $\sigma = 0.5$ and $\epsilon = 0.4$ the ratio of the correctly obtained model is 0.93 for the Cauchy distributed errors.

Figs. 1 and 2 are the boxplots of the regression estimates $\hat{\beta}_1 - \hat{\beta}_8$ from 200 simulated datasets for the case $n = 100$, $\epsilon = 0, 0.3$ and $\sigma = 1$, respectively. In both cases the random errors have standard Cauchy distribution. The horizontal dotted lines shows the true parameter values. In Fig. 1 we observe that the variabilities of the estimates obtained from the LAD-LASSO and the WLAD-LASSO are not significantly different. Also, all of the methods are successfully estimating the regression parameters. However, Fig. 2 shows the significant difference between the variabilities of the estimates obtained from the LAD-LASSO and the WLAD-LASSO methods. From Fig. 2 we can see how the leverage points are affecting the LAD and LAD-LASSO estimates. In this case we can see how the regression estimate $\hat{\beta}$ is attracted by the direction of the contaminated points $\beta_2$. Note that the similar results are obtained for the other cases used in the simulation study.

As a result of this limited simulation study we can conclude that the WLAD-LASSO can be used to achieve the model selection and robust parameter estimation simultaneously if the datasets are subject to the heavy-tailed errors and the leverage points.

**Table 1**
Simulation results for $t_1$ (Cauchy distribution) error.

| $\sigma$ | $n$ | $\epsilon$ | Method | Correctly fitted | No. of zeros | | MeanMSE | MedianMSE |
|---|---|---|---|---|---|---|---|---|
| | | | | | Correct | Incorrect | | |
| 0.5 | 50 | 0 | LAD | 0.00 | 0.01 | 0.00 | 0.21 | 0.17 |
| | | | LAD-LASSO | 0.95 | 5.00 | 0.05 | 0.20 | 0.05 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.24 | 0.19 |
| | | | WLAD-LASSO | 0.99 | 4.99 | 0.00 | 0.10 | 0.04 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 189.92 | 191.26 |
| | | | LAD-LASSO | 0.00 | 1.32 | 0.80 | 189.39 | 187.74 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.31 | 0.23 |
| | | | WLAD-LASSO | 0.97 | 4.96 | 0.00 | 0.10 | 0.05 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 227.70 | 222.42 |
| | | | LAD-LASSO | 0.00 | 1.30 | 0.84 | 228.00 | 220.81 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.48 | 0.30 |
| | | | WLAD-LASSO | 0.90 | 4.90 | 0.02 | 0.19 | 0.07 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 248.54 | 240.90 |
| | | | LAD-LASSO | 0.00 | 1.49 | 0.76 | 248.65 | 239.35 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.89 | 0.42 |
| | | | WLAD-LASSO | 0.81 | 4.78 | 0.08 | 0.57 | 0.10 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 274.06 | 271.34 |
| | | | LAD-LASSO | 0.00 | 1.70 | 0.54 | 277.34 | 270.58 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 71.27 | 1.69 |
| | | | WLAD-LASSO | 0.44 | 3.68 | 0.26 | 85.84 | 0.74 |
| | 100 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.07 | 0.06 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.02 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.08 | 0.07 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 161.16 | 163.64 |
| | | | LAD-LASSO | 0.00 | 1.48 | 0.82 | 162.31 | 164.31 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.11 | 0.08 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.02 |
| | | 0.2 | LAD | 0.00 | 0.01 | 0.00 | 205.87 | 203.57 |
| | | | LAD-LASSO | 0.00 | 1.44 | 0.92 | 207.46 | 206.44 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.13 | 0.10 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 229.15 | 227.81 |
| | | | LAD-LASSO | 0.00 | 1.64 | 0.88 | 231.61 | 230.21 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.21 | 0.16 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.04 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 263.76 | 255.16 |
| | | | LAD-LASSO | 0.00 | 2.16 | 0.42 | 272.38 | 254.53 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 2.83 | 0.30 |
| | | | WLAD-LASSO | 0.93 | 4.88 | 0.06 | 2.46 | 0.08 |
| | 200 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.03 | 0.03 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.01 |
| | | | WLAD | 0.00 | 0.03 | 0.00 | 0.04 | 0.03 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.01 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 147.52 | 148.56 |
| | | | LAD-LASSO | 0.00 | 1.18 | 0.98 | 149.41 | 148.74 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.04 | 0.04 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 193.28 | 191.63 |
| | | | LAD-LASSO | 0.00 | 1.30 | 1.08 | 197.13 | 196.76 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.06 | 0.05 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.02 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 216.79 | 215.18 |
| | | | LAD-LASSO | 0.00 | 1.60 | 0.86 | 218.66 | 214.85 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.08 | 0.07 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 262.03 | 257.13 |
| | | | LAD-LASSO | 0.00 | 2.26 | 0.28 | 273.38 | 261.69 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.28 | 0.11 |
| | | | WLAD-LASSO | 0.99 | 4.97 | 0.00 | 0.17 | 0.04 |

(*continued on next page*)

Table 1 (*continued*)

| $\sigma$ | $n$ | $\epsilon$ | Method | Correctly fitted | No. of zeros | | MeanMSE | MedianMSE |
|---|---|---|---|---|---|---|---|---|
| | | | | | Correct | Incorrect | | |
| 1 | 50 | 0 | LAD | 0.00 | 0.00 | 0.00 | 0.82 | 0.59 |
| | | | LAD-LASSO | 0.76 | 4.94 | 0.22 | 0.77 | 0.27 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 1.00 | 0.72 |
| | | | WLAD-LASSO | 0.67 | 4.59 | 0.06 | 0.74 | 0.25 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 196.58 | 194.02 |
| | | | LAD-LASSO | 0.00 | 1.28 | 0.78 | 195.14 | 193.29 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 1.25 | 0.97 |
| | | | WLAD-LASSO | 0.61 | 4.52 | 0.09 | 1.07 | 0.36 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 229.09 | 224.73 |
| | | | LAD-LASSO | 0.00 | 1.30 | 0.80 | 229.78 | 222.36 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 1.89 | 1.12 |
| | | | WLAD-LASSO | 0.52 | 4.30 | 0.16 | 1.94 | 0.46 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 253.89 | 245.84 |
| | | | LAD-LASSO | 0.00 | 1.48 | 0.72 | 253.44 | 247.37 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 4.02 | 1.76 |
| | | | WLAD-LASSO | 0.40 | 4.00 | 0.18 | 3.75 | 1.00 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 279.25 | 271.34 |
| | | | LAD-LASSO | 0.00 | 1.80 | 0.54 | 283.52 | 270.58 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 749.46 | 7.28 |
| | | | WLAD-LASSO | 0.19 | 2.84 | 0.48 | 871.43 | 6.83 |
| | 100 | 0 | LAD | 0.00 | 0.00 | 0.00 | 0.31 | 0.26 |
| | | | LAD-LASSO | 0.91 | 5.00 | 0.09 | 0.37 | 0.12 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.35 | 0.28 |
| | | | WLAD-LASSO | 0.95 | 4.94 | 0.00 | 0.14 | 0.09 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 166.81 | 166.65 |
| | | | LAD-LASSO | 0.00 | 1.42 | 0.82 | 166.96 | 165.70 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.42 | 0.34 |
| | | | WLAD-LASSO | 0.92 | 4.93 | 0.01 | 0.19 | 0.10 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 208.32 | 205.84 |
| | | | LAD-LASSO | 0.00 | 1.36 | 0.98 | 209.14 | 206.32 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.55 | 0.46 |
| | | | WLAD-LASSO | 0.88 | 4.86 | 0.00 | 0.22 | 0.14 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 226.74 | 223.54 |
| | | | LAD-LASSO | 0.00 | 1.74 | 0.91 | 228.47 | 224.42 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.81 | 0.63 |
| | | | WLAD-LASSO | 0.78 | 4.76 | 0.03 | 0.42 | 0.19 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 271.30 | 259.82 |
| | | | LAD-LASSO0 | 0.00 | 2.20 | 0.53 | 280.36 | 260.18 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 11.31 | 1.29 |
| | | | WLAD-LASS | 0.50 | 4.24 | 0.20 | 11.45 | 0.62 |
| | 200 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.13 | 0.11 |
| | | | LAD-LASSO | 0.97 | 5.00 | 0.03 | 0.14 | 0.05 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.15 | 0.13 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.05 | 0.03 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 152.53 | 151.88 |
| | | | LAD-LASSO | 0.00 | 1.32 | 0.98 | 154.38 | 154.83 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.16 | 0.14 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.07 | 0.04 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 195.16 | 195.01 |
| | | | LAD-LASSO | 0.00 | 1.23 | 1.04 | 198.58 | 199.53 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.21 | 0.18 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.08 | 0.06 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 218.08 | 217.38 |
| | | | LAD-LASSO | 0.00 | 1.76 | 0.91 | 221.22 | 219.66 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.33 | 0.30 |
| | | | WLAD-LASSO | 0.98 | 4.98 | 0.00 | 0.13 | 0.08 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 267.66 | 259.71 |
| | | | LAD-LASSO | 0.00 | 2.57 | 0.24 | 287.53 | 267.04 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.60 | 0.51 |
| | | | WLAD-LASSO | 0.91 | 4.92 | 0.02 | 0.30 | 0.16 |

**Table 2**
Simulation results for $t_3$ error.

| $\sigma$ | $n$ | $\epsilon$ | Method | Correctly fitted | No. of zeros | | MeanMSE | MedianMSE |
|---|---|---|---|---|---|---|---|---|
| | | | | | Correct | Incorrect | | |
| 0.5 | 50 | 0 | LAD | 0.00 | 0.00 | 0.00 | 0.11 | 0.09 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.03 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.12 | 0.10 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 177.50 | 178.31 |
| | | | LAD-LASSO | 0.00 | 1.28 | 0.76 | 177.36 | 178.56 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.13 | 0.12 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.05 | 0.04 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 228.37 | 224.82 |
| | | | LAD-LASSO | 0.00 | 1.36 | 0.88 | 228.73 | 224.82 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.19 | 0.15 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.04 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 247.85 | 244.01 |
| | | | LAD-LASSO | 0.00 | 1.45 | 0.68 | 247.53 | 242.06 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.31 | 0.22 |
| | | | WLAD-LASSO | 0.97 | 4.99 | 0.02 | 0.15 | 0.06 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 278.70 | 269.20 |
| | | | LAD-LASSO | 0.00 | 1.94 | 0.54 | 282.80 | 267.94 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 46.28 | 0.48 |
| | | | WLAD-LASSO | 0.72 | 4.30 | 0.20 | 52.03 | 0.17 |
| | 100 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.04 | 0.04 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.05 | 0.05 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.02 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 157.45 | 159.50 |
| | | | LAD-LASSO | 0.00 | 1.43 | 1.03 | 156.99 | 156.96 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.06 | 0.05 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.02 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 205.99 | 204.04 |
| | | | LAD-LASSO | 0.00 | 1.46 | 0.92 | 207.09 | 205.57 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.08 | 0.06 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 225.41 | 222.81 |
| | | | LAD-LASSO | 0.00 | 1.76 | 0.88 | 227.54 | 223.29 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.01 | 0.08 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.03 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 260.63 | 249.70 |
| | | | LAD-LASSO | 0.00 | 2.13 | 0.60 | 268.04 | 253.05 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 1.16 | 0.15 |
| | | | WLAD-LASSO | 0.97 | 4.97 | 0.03 | 0.55 | 0.04 |
| | 200 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.02 | 0.02 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.03 | 0.02 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 143.72 | 145.82 |
| | | | LAD-LASSO | 0.00 | 1.24 | 0.98 | 144.26 | 146.49 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.03 | 0.02 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 195.00 | 194.87 |
| | | | LAD-LASSO | 0.00 | 1.42 | 1.00 | 198.01 | 198.03 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.03 | 0.03 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 216.70 | 215.57 |
| | | | LAD-LASSO | 0.00 | 1.56 | 1.03 | 218.35 | 215.80 |
| | | | WLAD | 0.00 | 0.03 | 0.00 | 0.05 | 0.05 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.02 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 259.23 | 254.18 |
| | | | LAD-LASSO | 0.00 | 2.31 | 0.37 | 273.34 | 258.32 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.08 | 0.07 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |

(continued on next page)

Table 2 (*continued*)

| $\sigma$ | $n$ | $\epsilon$ | Method | Correctly fitted | No. of zeros | | MeanMSE | MedianMSE |
|---|---|---|---|---|---|---|---|---|
| | | | | | Correct | Incorrect | | |
| 1 | 50 | 0 | LAD | 0.00 | 0.03 | 0.00 | 0.43 | 0.36 |
| | | | LAD-LASSO | 0.88 | 5.00 | 0.12 | 0.43 | 0.16 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.48 | 0.41 |
| | | | WLAD-LASSO | 0.92 | 4.92 | 0.00 | 0.22 | 0.14 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 185.72 | 180.98 |
| | | | LAD-LASSO | 0.00 | 1.28 | 0.84 | 183.03 | 179.87 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.58 | 0.52 |
| | | | WLAD-LASSO | 0.88 | 4.86 | 0.01 | 0.29 | 0.17 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 229.35 | 224.90 |
| | | | LAD-LASSO | 0.00 | 1.34 | 0.92 | 228.21 | 224.92 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.76 | 0.58 |
| | | | WLAD-LASSO | 0.82 | 4.77 | 0.03 | 0.45 | 0.22 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 247.51 | 243.16 |
| | | | LAD-LASSO | 0.00 | 1.54 | 0.70 | 248.03 | 243.70 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 1.24 | 0.87 |
| | | | WLAD-LASSO | 0.64 | 4.59 | 0.14 | 0.97 | 0.40 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 278.59 | 269.25 |
| | | | LAD-LASSO | 0.00 | 1.86 | 0.56 | 282.07 | 267.51 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 48.63 | 1.91 |
| | | | WLAD-LASSO | 0.36 | 3.76 | 0.37 | 54.22 | 1.46 |
| | 100 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.17 | 0.15 |
| | | | LAD-LASSO | 0.99 | 5.00 | 0.01 | 0.13 | 0.06 |
| | | | WLAD | 0.00 | 0.03 | 0.00 | 0.19 | 0.17 |
| | | | WLAD-LASSO | 0.99 | 4.99 | 0.00 | 0.07 | 0.05 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 160.46 | 159.88 |
| | | | LAD-LASSO | 0.00 | 1.36 | 1.03 | 159.83 | 160.04 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.02 | 0.22 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.09 | 0.07 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 203.57 | 204.07 |
| | | | LAD-LASSO | 0.00 | 1.36 | 0.92 | 204.59 | 204.23 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.30 | 0.27 |
| | | | WLAD-LASSO | 0.98 | 4.97 | 0.00 | 0.12 | 0.09 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 224.59 | 221.54 |
| | | | LAD-LASSO | 0.00 | 1.64 | 0.95 | 227.07 | 223.89 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.43 | 0.38 |
| | | | WLAD-LASSO | 0.96 | 4.95 | 0.01 | 0.22 | 0.10 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 256.48 | 253.60 |
| | | | LAD-LASSO | 0.00 | 2.17 | 0.50 | 261.69 | 252.64 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 7.11 | 0.57 |
| | | | WLAD-LASSO | 0.84 | 4.74 | 0.10 | 7.82 | 0.19 |
| | 200 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.08 | 0.08 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.05 | 0.03 |
| | | | WLAD | 0.00 | 0.03 | 0.00 | 0.09 | 0.09 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 146.00 | 142.91 |
| | | | LAD-LASSO | 0.00 | 1.16 | 0.98 | 147.28 | 144.60 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.10 | 0.09 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 193.27 | 192.83 |
| | | | LAD-LASSO | 0.00 | 1.44 | 1.06 | 196.74 | 197.08 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.15 | 0.13 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.05 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 217.80 | 215.77 |
| | | | LAD-LASSO | 0.00 | 1.66 | 1.00 | 218.06 | 215.41 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.20 | 0.18 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.08 | 0.06 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 253.03 | 247.46 |
| | | | LAD-LASSO | 0.00 | 2.22 | 0.36 | 262.89 | 251.82 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.37 | 0.31 |
| | | | WLAD-LASSO | 0.90 | 4.98 | 0.00 | 0.13 | 0.10 |

**Table 3**
Simulation results for $t_5$ error.

| $\sigma$ | $n$ | $\epsilon$ | Method | Correctly fitted | No. of zeros | | MeanMSE | MedianMSE |
|---|---|---|---|---|---|---|---|---|
| | | | | | Correct | Incorrect | | |
| 0.5 | 50 | 0 | LAD | 0.00 | 0.03 | 0.00 | 0.09 | 0.08 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.1 | 0.09 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 181.43 | 182.29 |
| | | | LAD-LASSO | 0.01 | 1.40 | 0.83 | 180.22 | 179.91 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.13 | 0.1 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 228.21 | 220.92 |
| | | | LAD-LASSO | 0.00 | 1.36 | 0.90 | 228.91 | 221.88 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.15 | 0.12 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.04 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 245.13 | 242.58 |
| | | | LAD-LASSO | 0.00 | 1.48 | 0.92 | 244.55 | 239.87 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.22 | 0.18 |
| | | | WLAD-LASSO | 0.98 | 4.99 | 0.01 | 0.11 | 0.05 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 265.09 | 258.33 |
| | | | LAD-LASSO | 0.00 | 1.78 | 0.57 | 266.8 | 262.47 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 53.39 | 0.49 |
| | | | WLAD-LASSO | 0.74 | 4.22 | 0.14 | 57.65 | 0.13 |
| | 100 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.04 | 0.04 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.01 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.05 | 0.04 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.01 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 159.05 | 159.95 |
| | | | LAD-LASSO | 0.00 | 1.44 | 0.86 | 158.62 | 162.45 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.06 | 0.05 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.02 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 203.38 | 199.62 |
| | | | LAD-LASSO | 0.00 | 1.40 | 0.98 | 204.21 | 202.89 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.07 | 0.06 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 225.53 | 223.25 |
| | | | LAD-LASSO | 0.00 | 1.66 | 0.91 | 226.53 | 224.00 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.08 | 0.07 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 257.94 | 248.71 |
| | | | LAD-LASSO | 0.00 | 2.06 | 0.52 | 265.94 | 251.21 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 4.14 | 0.13 |
| | | | WLAD-LASSO | 0.97 | 4.91 | 0.02 | 4.42 | 0.04 |
| | 200 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.02 | 0.02 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.02 | 0.02 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 141.79 | 145.10 |
| | | | LAD-LASSO | 0.00 | 1.32 | 0.91 | 142.11 | 140.47 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.03 | 0.02 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 193.67 | 195.06 |
| | | | LAD-LASSO | 0.00 | 1.31 | 1.03 | 197.58 | 197.81 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.03 | 0.03 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.01 | 0.01 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 217.80 | 215.77 |
| | | | LAD-LASSO | 0.00 | 1.70 | 0.96 | 220.72 | 217.82 |
| | | | WLAD | 0.00 | 0.04 | 0.00 | 0.05 | 0.04 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.02 | 0.02 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 253.34 | 248.84 |
| | | | LAD-LASSO | 0.00 | 2.24 | 0.33 | 264.09 | 254.17 |
| | | | WLAD | 0.01 | 0.00 | 0.00 | 0.07 | 0.06 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.03 | 0.02 |

(*continued on next page*)

Table 3 (*continued*)

| $\sigma$ | $n$ | $\epsilon$ | Method | Correctly fitted | No. of zeros | | MeanMSE | MedianMSE |
|---|---|---|---|---|---|---|---|---|
| | | | | | Correct | Incorrect | | |
| 1 | 50 | 0 | LAD | 0.00 | 0.00 | 0.02 | 0.35 | 0.30 |
| | | | LAD-LASSO | 0.90 | 5.00 | 0.10 | 0.38 | 0.14 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.42 | 0.36 |
| | | | WLAD-LASSO | 0.94 | 4.94 | 0.01 | 0.21 | 0.13 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 182.49 | 181.29 |
| | | | LAD-LASSO | 0.00 | 1.40 | 0.85 | 181.07 | 179.51 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.49 | 0.45 |
| | | | WLAD-LASSO | 0.95 | 4.96 | 0.00 | 0.25 | 0.15 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 229.11 | 225.65 |
| | | | LAD-LASSO | 0.00 | 1.33 | 0.77 | 229.30 | 224.48 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.71 | 0.60 |
| | | | WLAD-LASSO | 0.82 | 4.82 | 0.04 | 0.41 | 0.22 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 245.62 | 240.50 |
| | | | LAD-LASSO | 0.00 | 1.46 | 0.80 | 245.94 | 238.57 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 1.02 | 0.71 |
| | | | WLAD-LASSO | 0.77 | 4.74 | 0.07 | 0.62 | 0.23 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 265.14 | 256.92 |
| | | | LAD-LASSO | 0.00 | 1.76 | 0.60 | 265.79 | 259.63 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 54.98 | 1.96 |
| | | | WLAD-LASSO | 0.40 | 3.68 | 0.28 | 59.04 | 1.10 |
| | 100 | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.17 | 0.15 |
| | | | LAD-LASSO | 0.98 | 5.00 | 0.02 | 0.16 | 0.06 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.19 | 0.18 |
| | | | WLAD-LASSO | 0.98 | 4.99 | 0.00 | 0.08 | 0.06 |
| | | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 159.40 | 160.90 |
| | | | LAD-LASSO | 0.00 | 1.38 | 0.84 | 159.26 | 160.97 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.23 | 0.20 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.09 | 0.06 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 203.30 | 200.58 |
| | | | LAD-LASSO | 0.00 | 1.38 | 0.96 | 204.50 | 202.79 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.28 | 0.24 |
| | | | WLAD-LASSO | 0.98 | 4.99 | 0.00 | 0.13 | 0.09 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 225.33 | 221.38 |
| | | | LAD-LASSO | 0.00 | 1.68 | 0.88 | 226.24 | 222.35 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.34 | 0.29 |
| | | | WLAD-LASSO | 0.98 | 4.97 | 0.00 | 0.15 | 0.10 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 258.65 | 250.46 |
| | | | LAD-LASSO | 0.00 | 2.09 | 0.54 | 265.54 | 249.55 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 5.39 | 0.50 |
| | | | WLAD-LASSO | 0.86 | 4.80 | 0.06 | 5.82 | 0.17 |
| | | 0 | LAD | 0.00 | 0.02 | 0.00 | 0.08 | 0.06 |
| | | | LAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.02 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.09 | 0.08 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.02 |
| | 200 | 0.1 | LAD | 0.00 | 0.00 | 0.00 | 142.46 | 146.16 |
| | | | LAD-LASSO | 0.00 | 1.32 | 0.92 | 143.53 | 143.48 |
| | | | WLAD | 0.00 | 0.01 | 0.00 | 0.10 | 0.09 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.04 | 0.03 |
| | | 0.2 | LAD | 0.00 | 0.00 | 0.00 | 194.04 | 195.77 |
| | | | LAD-LASSO | 0.00 | 1.34 | 1.03 | 198.23 | 198.48 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.13 | 0.11 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.06 | 0.04 |
| | | 0.3 | LAD | 0.00 | 0.00 | 0.00 | 217.89 | 215.67 |
| | | | LAD-LASSO | 0.00 | 1.66 | 0.96 | 221.10 | 217.29 |
| | | | WLAD | 0.00 | 0.02 | 0.00 | 0.19 | 0.17 |
| | | | WLAD-LASSO | 1.00 | 5.00 | 0.00 | 0.08 | 0.06 |
| | | 0.4 | LAD | 0.00 | 0.00 | 0.00 | 254.10 | 250.76 |
| | | | LAD-LASSO | 0.00 | 2.24 | 0.32 | 264.95 | 256.17 |
| | | | WLAD | 0.00 | 0.00 | 0.00 | 0.29 | 0.26 |
| | | | WLAD-LASSO | 0.99 | 4.99 | 0.00 | 0.11 | 0.08 |

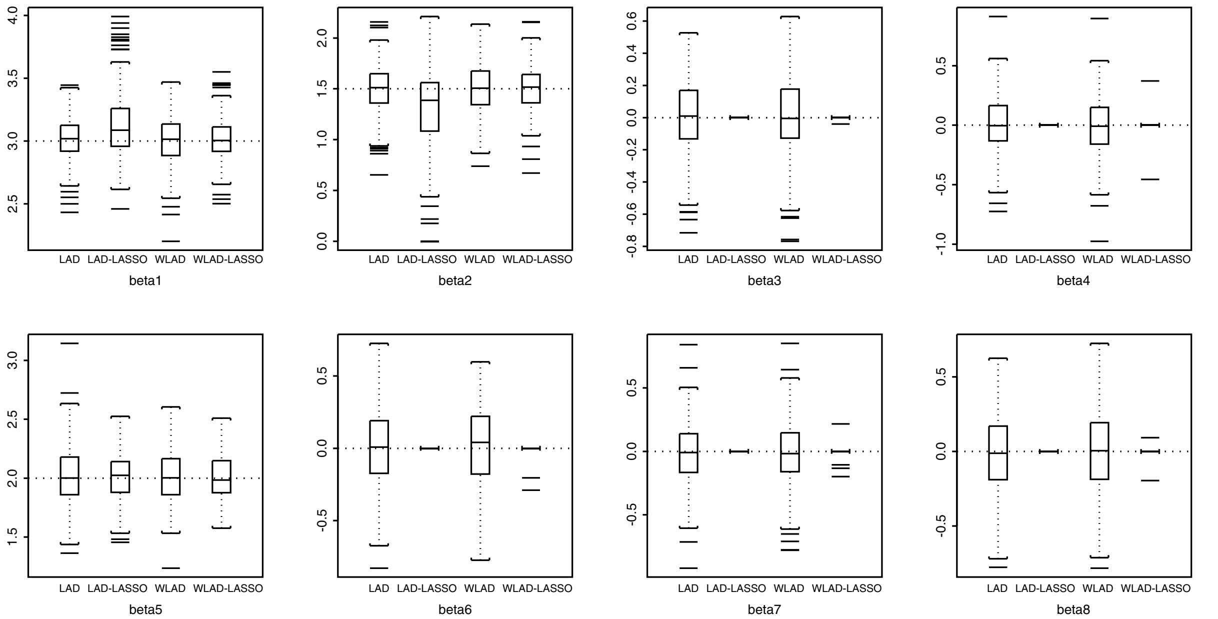


**Fig. 1.** Boxplots of estimates from 200 simulated datasets in simulation study. Random errors have Cauchy distribution $n = 100, \epsilon = 0, \sigma = 1$. Horizontal dotted lines show the true values of the regression coefficients.

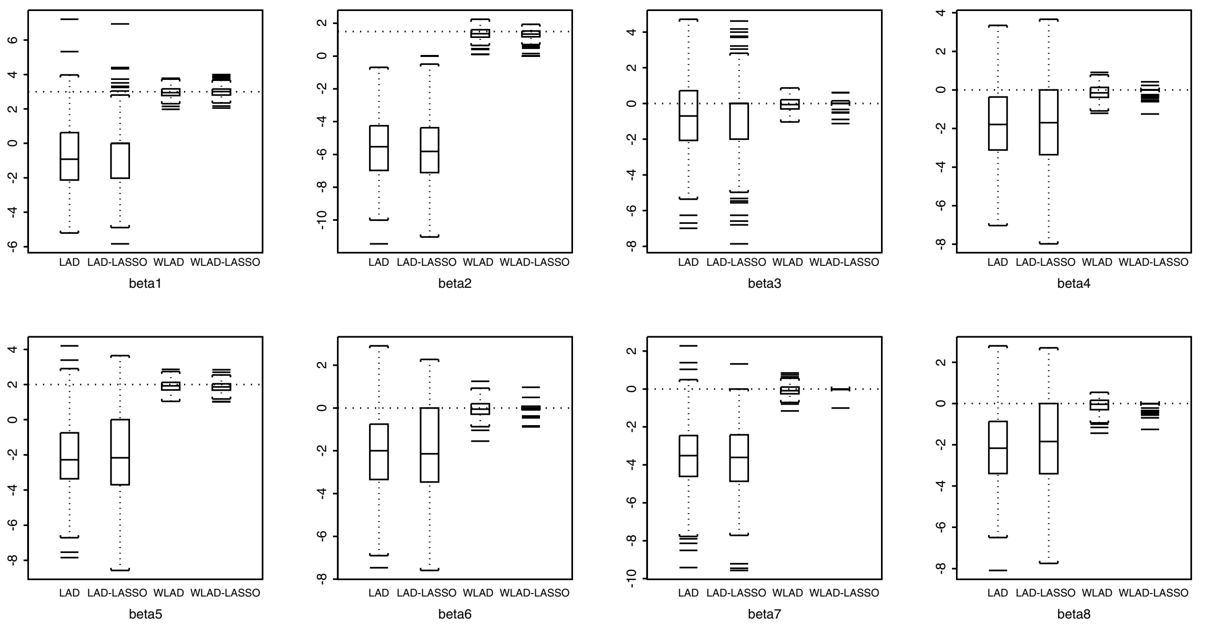


**Fig. 2.** Boxplots of estimates from 200 simulated datasets in simulation study. Random errors have Cauchy distribution $n = 100, \epsilon = 0.3, \sigma = 1$. Horizontal dotted lines show the true values of the regression coefficients.

### *4.2. Example*

In this section we will use the dataset constructed by Hawkins et al. (1984). This is an artificial dataset but it offers the advantage that at least the positions of the outliers are known exactly so that we can measure the effectiveness of the proposed method against the outliers. The dataset has 75 observations one response and three explanatory variables. In the regression sense, the first 10 observations are designed as the bad leverage points and the next four observations (Observations 11–14) are designed as the good leverage points for the illustration of identifying bad/good leverage points. This dataset has been extensively used in the statistical literature to illustrate the strength of the robust regression methods (see e.g. Rousseeuw and Leroy (1987)).

**Table 4**
Results for the Hawkins et al. (1984) dataset.

| | OLS | LAD | LAD-LASSO | WLAD | WLAD-LASSO |
|---|---|---|---|---|---|
| $X1$ | 0.851 (1.79) | 2.811 (0.14) | 2.800 (0.13) | 2.187 (0.17) | 2.001 (0.16) |
| $X2$ | 5.392 (1.35) | 2.941 (0.11) | 2.781 (0.06) | 2.046 (0.17) | 1.903 (0.14) |
| $X3$ | −3.971 (1.01) | −3.824 (0.08) | −3.714 (0.13) | −0.166 (0.16) | 0.000 |

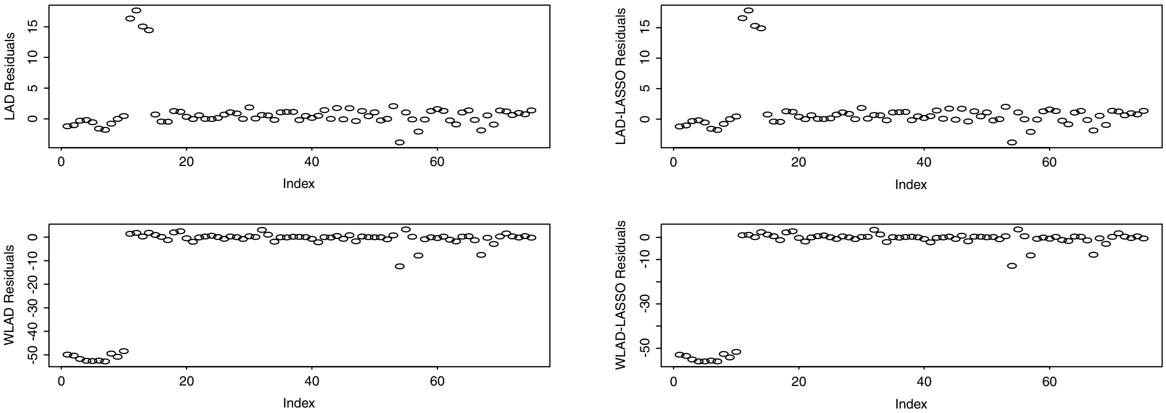


**Fig. 3.** Index plot of the residuals for the Hawkins–Bradu–Kass dataset.

In this paper we first modify the dataset and use the WLAD-LASSO method along with the other methods to estimate the regression parameters. We do the modification in order to perform the robust estimation and the variable selection. We modify the dataset as follows. We take the covariates $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ as our $\mathbf{X}$ matrix and generated the response variable using the following regression models $y_1 = X_1\beta_1 + \varepsilon$, where $\beta_1 = (2, 2, 0)^T$, $X_1$ is the matrix of the observations from 11 to 75, and $\varepsilon$ has the Cauchy distribution. This is our main model. Here we take the main observations and exclude the outliers. For the outliers we use the following equations $y_1 = X_2\beta_2$, where $\beta_2 = (-1, -1, 0)^T$ and $X_2$ is the matrix of first 10 observations which are outliers.

Table 4 summaries the result for this dataset. We take several values for $\gamma$ and observe the similar results so we only report the results for $\gamma = 2$. From these results we observe that the WLAD-LASSO method selects the true model and correctly estimates the non-zero coefficients. Index plots of the residuals given in Fig. 3 also shows that unlike the LAD-LASSO the outliers can be correctly detected by the WLAD-LASSO procedure.

## 5. Conclusion

In this article we have presented the weighted LAD-LASSO method to improve the robustness of the LS and LAD based LASSO methods. The WLAD-LASSO combines the ideas of WLAD regression method and LASSO method for robustly estimating the regression parameters and selecting the right model. Our limited simulation study showed that the WLAD-LASSO method fares comparably well in terms of simultaneous robust estimation and variable selection and retains appealing robustness property of the weighted LAD regression method. Further, since the computation of the WLAD-LASSO method is not very difficult and the calculations can be done using the standard robust estimation and LAD regression algorithms it gains additional benefits in real data analysis. Finally, since WLAD regression methods is a special case of the generalized $M$-regression (GM-regression) method, similarly the GM regression method and the LASSO method can be combined to get simultaneously robust parameter estimation and model selection, or in other words, the LASSO method can be used to carry on model selection in the GM regression.

## Acknowledgments

The author would like to thank the editor, the associate editor and two anonymous referees for their constructive comments that greatly improved the work.

## References

Ellis, S., Morgenthaler, S., 1992. Leverage and breakdown in $L_1$ regression. Journal of the American Statistical Association 87, 143–148.
Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.
Gao, X., 2008, Penalized methods for high-dimensional least absolute deviations regression, unpublished Ph.D. Thesis, University of Iowa.
Giloni, A., Simonoff, J., Sengupta, B., 2006a. Robust weighted LAD regression. Computational Statistics & Data Analysis 50, 3124–3140.

Giloni, A., Sengupta, B., Simonoff, J., 2006b. A Mathematical Programming Approach for Improving the Robustness of Least Sum of Ansolute Deviations Regression. Wiley InterScience, doi:10.1002/nav.20139.
Hawkins, D.M., Bradu, D., Kas, G.V., 1984. Location of several outliers in multiple regression using elemental subsets. Technometrics 26, 197–208.
Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.
Hurvich, C.M., Tsai, C.L., 1990. Model selection for least absolute deviations regressions in small samples. Statistics and Probability Letters 9, 259–265.
Hesterberg, T., Choi, N.H., Meier, L., Fraley, C., 2008. Least angle and $L_1$ penalized regression: a review. Statistics Surveys 2, 61–93.
Hubert, M., Rousseeuw, P., 1997. Robust regression with both continuous and binary regressors. Journal of Statistical Planning and Inference 57, 153–163.
Miller, A., 2002. Subset Selection in Regression, second ed.. Chapman & Hall.
Owen, A.B., 2007. A robust hybrid of lasso and ridge regression. Contemparary Mathematics 443, 59–71.
Ronchetti, E., Staudte, R.G., 1994. A robust version of Mallows $C_p$. Journal of the American Statistical Association 89, 550–559.
Rosset, S., Zhu, J., 2004. Least angle regression: discussion. The Annals of Statistics 32, 469–475.
Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley, New York.
Shao, J., 1997. An asymptotic theory for linear model selection. Statistica Sinica 7, 221–264.
Shi, P., Tsai, C.L., 2002. Regression model selection: a residual likelihood approach. Journal of the Royal Statistical Society, Series B 64, 237–252.
Shi, P., Tsai, C.L., 2004. A joint regression variable and autoregressive order selection criterion. Journal of Time Series 25, 923–941.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.
Wang, H., Leng, C., 2007. Unified lasso estimation by least squares approximation. Journal of the American Statistical Association 102, 1039–1048.
Wang, H., Li, G., Jiang, G., 2007. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. Journal of Business & Economic Statistics 25, 347–355.
Xu, J., 2005, Parameter estimation, model selection and inference in $L_1$-based linear regression, unpublished Ph.D. Thesis, Columbia University.
Xu, J., Ying, Z., 2010. Simultaneous estimation and variable selection in median regression using Lasso-type penalty. Annals of the Institute of Statistical Mathematics 62, 487–514.
Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.