ON THE MEAN-SHIFT OUTLIER MODEL FOR LAD REGRESSION

FELIPE OSORIO

Abstract. The least absolute deviation (LAD) regression technique is well known for providing an estimation mechanism insensitive to outliers in the response variable. However, relatively few studies have highlighted that these models still can be susceptible to influential observations; therefore the development of appropriate diagnostic measures is required. In this paper, we introduce a procedure for detecting outlying observations in LAD regression based on the mean-shift outlier model. This formulation provides a framework for defining standardized residuals, which enables the construction of a QQ-plot with simulated envelope. A noteworthy by-product of this work is the derivation of the gradient statistic for testing linear hypotheses in LAD regression. We illustrate our findings through the analysis of two datasets commonly examined in the literature.

1. Introduction

It is well known that outlying observations can have a disproportionate effect on various aspects of statistical modeling-namely, parameter estimation, goodness of fit, and inferential results. This has made the process of identifying outliers a crucial step in the analysis of regression models. The presence of such anomalous observations can cast doubts on the proposed model and suggest the use of, for example, robust alternatives. Interestingly, even robust estimation procedures can still be vulnerable to outliers (see, for instance [1]). However, relatively few studies have developed diagnostic measures in the context of LAD regression. It should be noted that the least absolute deviation estimation technique (see [2–5]) has been widely used in practical contexts. For example, some interesting developments in LAD estimation in nonparametric regression are presented in [6–8], whereas [9, 10], and [11, 12] study the least absolute deviation estimation for the detection of change-points and ridge estimation in linear regression models as well as seemingly unrelated regression equations, respectively. Applications of LAD regression for estimating electricity demand and characterizing the job preferences of healthcare workers are presented in [13, 14]. An excellent compendium with various applications, theoretical developments and the study of the properties of the LAD estimation method, can be found in [15]. Considering that the LAD regression method is widely used in fields as diverse as engineering, biology, and economics, it is necessary to develop measures for detecting outliers.

Some diagnostic measures have been proposed to assess the effect that unusual observations may exert on the parameter estimates in LAD regression based mainly on analogy with diagnostics developed for ordinary least squares [1, 16–18]. To the best of our knowledge, the first proposal to define measures grounded in case deletion were introduced by Narula and Wellington [1], whereas the leverage and likelihood displacement were studied by Ellis and Morgenthaler [16] and Elian et al. [17], respectively. We should emphasize that Sun and Wei extended Cook's distance and proposed a quasi-likelihood distance that originates from the objective function associated with LAD regression (see [18] for details). Moreover, in the latter work, the equivalence between case deletion diagnostics and the mean-shift outlier model was established, thus extending the results presented in [19]. Several of these diagnostic measures specific to LAD regression have recently been implemented in the R package india [20] which is available from the CRAN repository (see also the diagL1 package [21]).

Key words and phrases. Diagnostics, Gradient statistic, Hypothesis testing, Least absolute deviation, Outliers, Residuals.
Email address: faosorios.stat@gmail.com; *Orcid ID:* 0000-0002-4675-5201.
This preprint has not undergone improvements or corrections. The version of record of this article is published in *Journal of Statistical Computation and Simulation*, and is available online at https://doi.org/10.1080/00949655.2026.2716387.

The main aim of this paper is to develop diagnostic measures specifically oriented towards the detection of outlying observations for LAD regression. Our approach further explores the mean-shift outlier model (see, for instance, [22], Sec. 2.2.2) using different test statistics introduced by [23] in order to assess the role of these atypical observations, as a precursor to the development of these diagnostics. This work complements the diagnostic measures discussed in [17] and [19]. The remainder of the paper unfolds as follows. In Section 2 we review the definition of the least absolute deviation estimation procedure for linear regression, as well as its asymptotic properties and the definition of various statistics for testing linear hypotheses, as an interesting side result, in this section we introduce the gradient statistic to test hypotheses about subvectors in the context of LAD regression. Diagnostic measures by using the mean-shift outlier model are discussed in Section 3, whereas the proposed methodology is illustrated in Section 4 considering two datasets previously analyzed using case-deletion influence measures. Some concluding remarks are given in Section 5. In addition, the Supplementary Material provides the analysis of the Brownlee's Stack loss dataset [24] as well as data related to aircraft manufacturing [25, pp. 154].

2. Background

Consider the linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{Y}$ is an n-dimensional vector of observed responses, $\boldsymbol{X}$ is an $n \times p$ model matrix, whose rows are the p-dimensional vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, $\boldsymbol{\beta} \in \mathbb{R}^p$ denotes the vector of unknown regression coefficients, and $\boldsymbol{\epsilon}$ represents an n-dimensional vector of random disturbances whose elements are assumed to be independent and identically distributed with joint distribution F. The LAD procedure consists of fitting the linear model in (1) by choosing the coefficients estimate $\widehat{\boldsymbol{\beta}}$ through

$$S(\widehat{\boldsymbol{\beta}}) = \min_{\beta \in \mathbb{R}^p} S(\boldsymbol{\beta}), \qquad S(\boldsymbol{\beta}) = \sum_{i=1}^n |y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}|, \tag{2}$$

where $S(\boldsymbol{\beta})$ is the objective function associated with LAD regression. Several authors have proposed alternative approaches for solving the optimization problem in (2). An excellent review of these works is presented in [2]. In this paper, we will use the algorithm proposed by Barrodale and Roberts [26, 27] to obtain the estimate of the regression coefficients $\widehat{\boldsymbol{\beta}}$. This procedure is considered one of the best methods available for estimation in LAD regression and has been implemented in the R package `L1pack` [28].

The asymptotic properties of the LAD estimator arising from optimizing the objective function in (2) have been characterized by [29], who introduced the following assumptions:

- *A1.* The distribution function F has strictly positive density $f(\cdot)$ such that $F^{-1}(1/2) = 0$, and F is continuous with $F'(0) > 0$;
- *A2.* There exists a positive definite matrix $\boldsymbol{Q}$ satisfying $\lim_{n\to\infty} \boldsymbol{X}^\top \boldsymbol{X}/n = \boldsymbol{Q}$.

Under conditions *A1* and *A2*, the LAD estimator is asymptotically normally distributed, that is

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathsf{N}_p(\boldsymbol{0}, \lambda^2 \boldsymbol{Q}^{-1}), \tag{3}$$

with $\lambda^2 = 1/(4f^2(0))$, where λ^2/n is the asymptotic variance of the sample median of the error disturbances. Different sets of conditions for establishing the asymptotic normality of the LAD estimator are discussed in [30] (see also [2]).

In [23], the asymptotic distribution of the Wald, likelihood ratio and score statistics was studied to test hypotheses about subvectors in LAD regression. Specifically, for the linear regression model given in (1) [23] addressed hypotheses of the form $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$, where $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ is partitioned such that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, with $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ being coefficient vectors of dimensions $p - k$ and k, respectively. Based on the partition for $\boldsymbol{X}$ and $\boldsymbol{\beta}$, it leads:

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix}. \tag{4}$$

Let $\boldsymbol{Q}_{22\cdot 1} = \boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$, be the Schur complement of $\boldsymbol{Q}_{11}$. The following assumption is required to establish the asymptotic distribution of the test statistics proposed by [23]

A3. There exists a fixed vector $\boldsymbol{\eta} \in \mathbb{R}^k$ such that $\boldsymbol{\beta}_2 = \boldsymbol{\eta}/\sqrt{n}$.

Given the Assumptions *A1*, *A2* and *A3*, the Wald, likelihood-ratio and score statistics given by

$$W = \frac{n}{\lambda^2}\widehat{\boldsymbol{\beta}}_2^\top \boldsymbol{Q}_{22\cdot 1}\widehat{\boldsymbol{\beta}}_2, \qquad LR = \frac{2}{\lambda}\{S(\widetilde{\boldsymbol{\beta}}) - S(\widehat{\boldsymbol{\beta}})\}, \qquad R = \widetilde{\boldsymbol{g}}_2^\top \boldsymbol{Q}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{g}}_2, \tag{5}$$

which are asymptotically equivalent under $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$ and follow a chi-square distribution with k degrees of freedom, where $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}_1^\top, \boldsymbol{0})^\top$ and $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1^\top, \widehat{\boldsymbol{\beta}}_2^\top)^\top$ represent the LAD estimates for $\boldsymbol{\beta}$ under the null and alternative hypotheses, respectively, and $\widetilde{\boldsymbol{g}}_2$ is a partition of the normalized gradient of the LAD objective function associated with $\boldsymbol{\beta}_2$, which must be evaluated at the restricted estimate $\widetilde{\boldsymbol{\beta}}$. Besides the "holy trinity" [31], a hypothesis test that has gained considerable popularity is the gradient statistic [32]. Next, we define the gradient statistic for hypotheses testing in LAD regression, which is given by

$$T = \frac{\sqrt{n}}{\lambda}\widetilde{\boldsymbol{g}}_2^\top \widehat{\boldsymbol{\beta}}_2. \tag{6}$$

The following proposition establishes that the gradient test defined in (6) is asymptotically equivalent to the other statistics given in Equation (5) and it has a chi-square asymptotic distribution with k degrees of freedom.

Proposition 1. *The test statistic T has asymptotically a chi-squared distribution with k degrees of freedom and noncentral parameter $\mu = \boldsymbol{\eta}^\top \boldsymbol{Q}_{22\cdot 1}\boldsymbol{\eta}/\lambda$.*

Proof. Under the Assumptions *A1*, *A2* and *A3*, yields

$$\sqrt{n}\widehat{\boldsymbol{\beta}}_2 \xrightarrow{D} \mathsf{N}_k(\boldsymbol{\eta}, \lambda^2 \boldsymbol{Q}_{22\cdot 1}), \qquad \widetilde{\boldsymbol{g}}_2 \xrightarrow{D} \mathsf{N}_k(\lambda^{-1}\boldsymbol{Q}_{22\cdot 1}\boldsymbol{\eta}, \boldsymbol{Q}_{22\cdot 1}).$$

It is straightforward to note that,

$$\sqrt{n}\lambda^{-1}\boldsymbol{Q}_{22\cdot 1}^{1/2}\widehat{\boldsymbol{\beta}}_2 \xrightarrow{D} \mathsf{N}_k(\lambda^{-1}\boldsymbol{Q}_{22\cdot 1}^{1/2}\boldsymbol{\eta}, \boldsymbol{I}), \qquad \boldsymbol{Q}_{22\cdot 1}^{-1/2}\widetilde{\boldsymbol{g}}_2 \xrightarrow{D} \mathsf{N}_k(\lambda^{-1}\boldsymbol{Q}_{22\cdot 1}^{1/2}\boldsymbol{\eta}, \boldsymbol{I}).$$

Therefore,

$$T = \left\{\boldsymbol{Q}_{22\cdot 1}^{-1/2}\widetilde{\boldsymbol{g}}_2\right\}^\top \sqrt{n}\lambda^{-1}\boldsymbol{Q}_{22\cdot 1}^{1/2}\widehat{\boldsymbol{\beta}}_2 = \frac{\sqrt{n}}{\lambda}\widetilde{\boldsymbol{g}}_2^\top \widehat{\boldsymbol{\beta}}_2 \xrightarrow{D} \chi^2(k; \mu),$$

with $\mu = \boldsymbol{\eta}^\top \boldsymbol{Q}_{22\cdot 1}\boldsymbol{\eta}/\lambda^2$ and the proof is complete. ☐

We emphasize that, under $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$ it follows that $T \xrightarrow{D} \chi^2(k)$. Furthermore, the Wald, likelihood-ratio, and gradient statistics require the estimation of the nuisance scale parameter λ.

Consider the following curious property of the residuals from the fit of the objective function in (2). For simplicity, we will now adopt the notation introduced by [33]. Let $N = \{1, 2, \ldots, n\}$ be the set denoting all observations, and let M be any subset of N. Additionally, we denote by $|M|$ the cardinality of the set M. For a vector $\boldsymbol{c} \in \mathbb{R}^n$, define $\operatorname{supp}(\boldsymbol{c})$ as the set of all its nonzero components, that is,

$$\operatorname{supp}(\boldsymbol{c}) = \{i \in N : c_i \neq 0\}.$$

Let $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ be the LAD residual vector, whose components are given by $e_i = y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}$, for $i \in N$. We should note that [34] were the first to establish that $M = \operatorname{supp}(\boldsymbol{e})$ satisfies $|M| = n - p$, or equivalently, $B = M^c \subset N$ implies $|B| = p$, that is, in the LAD fit there are exactly p zero residuals. Those observations that satisfy $e_i = 0$ for $i \in B$ are known as basic observations and, interestingly, completely define the LAD estimator obtained from Equation (2). Indeed, the LAD fit can be written as

$$\begin{pmatrix} \boldsymbol{Y}_B \\ \boldsymbol{Y}_M \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_B & \boldsymbol{0} \\ \boldsymbol{X}_M & \boldsymbol{I}_{n-p} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \boldsymbol{e}_M \end{pmatrix},$$

where B and M denote the sets of indices associated with zero and nonzero residuals, respectively, while $\boldsymbol{e}_M$ is the vector of nonzero residuals (see, for instance, [1, 17]). Thus, the procedure proposed by Barrodale and Roberts [26, 27] allows the basic observations to be determined using linear programming techniques.

In this study, we consider the procedure proposed by [35] for the consistent estimation of the nuisance scale parameter λ (see also [2, 36]). Let $e_{(1)} \leq e_{(2)} \leq \cdots \leq e_{(m)}$ be the nonzero residuals ordered in ascending order, where $m = n - p$. Thus, the estimator of λ is calculated as:

$$\widehat{\lambda} = \frac{\sqrt{m}(e_{(m-r-1)} - e_{(r)})}{z_{\alpha/2}}, \tag{7}$$

where $r = (m+1)/2 - z_{\alpha/2}\sqrt{m/4}$ is rounded to the nearest integer. In our numerical experiments, we followed the guidelines provided in [2, 36], so we have used $\alpha = 0.05$.

3. The mean-shift outlier model for LAD regression

The mean-shift outlier model has become a fundamental tool for developing diagnostic measures for various statistical models (see, for instance [37]). The first mention of this model appears in the book by Cook and Weisberg [22, Sec. 2.2.2]. There is a strong connection between the mean-shift outlier model and classical influence diagnostics based on case deletion. To introduce these ideas, consider the mean-shift outlier model, which is given by:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{d}_i\gamma + \boldsymbol{\epsilon}, \tag{8}$$

where $\gamma \in \mathbb{R}$, and $\boldsymbol{d}_i = (\delta_{i1}, \delta_{i2}, \dots, \delta_{in})^\top$ is an $n \times 1$ vector, with

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

that is, $\boldsymbol{d}_i$ is a vector with one at the ith position and zero elsewhere. This leads us to note that

$$\mathsf{E}(Y_j) = \begin{cases} \boldsymbol{x}_i^\top\boldsymbol{\beta} + \gamma, & j = i, \\ \boldsymbol{x}_j^\top\boldsymbol{\beta}, & j \neq i, \end{cases}$$

for $j = 1, \dots, n$. Thus, non-zero values of γ are an indication that the ith case is an outlier. The main feature of the mean-shift outlier model is that it enables the identification of outliers through hypothesis testing

$$H_0 : \gamma = 0, \qquad \text{against} \qquad H_1 : \gamma \neq 0. \tag{9}$$

Interestingly, for the linear regression model under the assumption of normality, the test statistic for the hypothesis in (9) corresponds to the studentized residual. Details on the derivation of this type of residual, as well as some of its basic properties, can be found in [22, Sec. 2.2.2]. It is worth noting that the connection between the mean-shift outlier model and case-deletion techniques has been addressed for more general models. For instance, for nonlinear regression models under the exponential family in [19], whereas for models with correlated data as well as in generalized estimation equations, this has been addressed by [38] and [39], respectively. It is particularly relevant that the mean-shift outlier model provides a mechanism not only for detecting outliers but also for defining standardized residuals in more complex models.

In the context of LAD regression, the equivalence between the mean-shift outlier model and the estimation when the ith observation is removed has been established by [18], which is particularly important because the objective function is not well-behaved, leading to the absence of one-step approximations that allow for a simple evaluation of case deletion measures. Hence, our aim is to identify outliers by testing the hypothesis $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$. Evidently, we can rewrite the model in (8) as

$$\boldsymbol{Y} = \boldsymbol{Z\theta} + \boldsymbol{\epsilon}, \tag{10}$$

with $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{d}_i)$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \gamma)^\top$. Next, we introduce several test statistics to identify outlying observations. Moreover, the mean-shift outlier model allows us to define a standardized type of residual that is particularly useful for building QQ-plots with simulated envelope.

3.1. **Measures for outlier detection.** The LAD estimators for the mean-shift outlier model, denoted by $\widehat{\boldsymbol{\beta}}_*$ and $\widehat{\gamma}_i$, are selected satisfying:

$$S_i^*(\widehat{\boldsymbol{\beta}}_*, \widehat{\gamma}_i) = \min_{\beta \in \mathbb{R}^p, \gamma \in \mathbb{R}} S_i^*(\boldsymbol{\beta}, \gamma), \qquad S_i^*(\boldsymbol{\beta}, \gamma) = \sum_{j=1}^n |y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta} - \delta_{ij}\gamma|. \tag{11}$$

Using Theorem 1 established in [18], it is known that the solution for $\boldsymbol{\beta}$ in problem (11) corresponds to $\widehat{\boldsymbol{\beta}}_{(i)}$, the estimator with the ith case deleted. Indeed,

$$\widehat{\boldsymbol{\beta}}_* = \widehat{\boldsymbol{\beta}}_{(i)}, \qquad \widehat{\gamma}_i = y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_{(i)}, \qquad i = 1, \dots, n.$$

Evaluating the objective function defined in (11) at $\widehat{\boldsymbol{\beta}}_*$ and $\widehat{\gamma}_i$, it follows that

$$S_i^*(\widehat{\boldsymbol{\beta}}_*, \widehat{\gamma}_i) = S_i^*(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\gamma}_i) = \sum_{j=1}^n |y_j - \boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}_{(i)} - \delta_{ij}\widehat{\gamma}_i| = S_{(i)}(\widehat{\boldsymbol{\beta}}_{(i)}),$$

where

$$S_{(i)}(\widehat{\boldsymbol{\beta}}_{(i)}) = \sum_{j \neq i} |y_j - \boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}_{(i)}|,$$

corresponds to the sum of absolute deviations once the ith observation has been removed from the dataset, which must be evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_{(i)}$. As noted in [18], the calculation of $\widehat{\boldsymbol{\beta}}_{(i)}$, for $i = 1, \dots, n$, can be time-consuming. It is important to note that, unlike in other contexts such as nonlinear least squares or generalized linear models in regression, in LAD there are no formulas available to approximate $\widehat{\boldsymbol{\beta}}_{(i)}$, and therefore full iteration of the estimation procedure is required. Inspired by the simplex method, which leads to the Barrodale-Roberts fitting procedure [26, 27], Appendix B of [1] describes the estimation of $\widehat{\boldsymbol{\beta}}_{(i)}$, a computation that may become computationally prohibitive for large n. Despite this, in their experiments, Narula and Wellington [1] consider the computational cost to be reasonable. In our numerical experiments, the computational burden was inexpensive. A brief remark of this aspect is presented in Section 5.

Next, we will denote the restricted estimator under the hypothesis $H_0 : \gamma = 0$ as $\widetilde{\boldsymbol{\beta}}$. We will also use the notation:

$$SAD = S(\widetilde{\boldsymbol{\beta}}) = \sum_{i=1}^n |y_i - \boldsymbol{x}_i^\top \widetilde{\boldsymbol{\beta}}|, \qquad SAD_{(i)} = S_{(i)}(\widehat{\boldsymbol{\beta}}_{(i)}).$$

Thus, the likelihood ratio statistic for testing $H_0 : \gamma = 0$ takes the form,

$$LR_i = \frac{2}{\lambda}(SAD - SAD_{(i)}), \tag{12}$$

and the null hypothesis is rejected, i.e., we identify the ith observation as an outlier if LR_i is greater than a quantile of the chi-square distribution with one degree of freedom.

To obtain the Wald statistic, first note that

$$\boldsymbol{Z}^\top \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{d}_i \\ \boldsymbol{d}_i^\top \boldsymbol{X} & \boldsymbol{d}_i^\top \boldsymbol{d}_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{x}_i \\ \boldsymbol{x}_i^\top & 1 \end{pmatrix}.$$

We know that for the model in (8), we have

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\beta}}_* - \boldsymbol{\beta} \\ \widehat{\gamma}_i - \gamma \end{pmatrix} \xrightarrow{D} \mathsf{N}_{p+1}\left(\begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \lambda^2 \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{r}_{12} \\ \boldsymbol{r}_{21} & r_{22} \end{pmatrix}\right),$$

provided that assumption *A1* is satisfied and analogously to assumption *A2*, we must have that $\boldsymbol{Q} = \lim_{n\to\infty} \boldsymbol{Z}^\top \boldsymbol{Z}/n$. In which case, $\boldsymbol{R} = \boldsymbol{Q}^{-1}$ is partitioned as in (4). Under the null hypothesis, $H_0 : \gamma = 0$, it follows that $\sqrt{n}\widehat{\gamma}_i \xrightarrow{D} \mathsf{N}(0, \lambda^2 r_{22})$, which leads to the Wald test statistic,

$$W_i = \left(\frac{\sqrt{n}\widehat{\gamma}_i}{\lambda\sqrt{r_{22}}}\right)^2 = \frac{n\widehat{\gamma}_i^2}{\lambda^2 r_{22}} \xrightarrow{D} \chi^2(1).$$

Following [23], we can estimate n/r_{22} by its sample counterpart,

$$\boldsymbol{d}_i^\top (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{d}_i = 1 - h_{ii},$$

where h_{ii} is the ith element of the diagonal of the matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Hence, we can decide whether the ith observation is an outlier using the following criterion: reject $H_0 : \gamma = 0$ if $W_i \geq \chi^2_{1-\alpha}(1)$, where

$$W_i = \frac{(1 - h_{ii})\widehat{\gamma}_i^2}{\lambda^2} = \frac{(1 - h_{ii})e_{i(i)}^2}{\lambda^2}, \tag{13}$$

with $e_{i(i)} = y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_{(i)}$, is frequently referred to as the ith deleted residual. To obtain the gradient statistic, note that we can write the jth row of the mean-shift outlier model given in (8) as:

$$y_j = (\boldsymbol{x}_j^\top, \delta_{ij}) \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} + \epsilon_j, \qquad j = 1, \dots, n.$$

Thus, the normalized gradient associated with γ evaluated at the unrestricted estimate $\widetilde{\boldsymbol{\beta}}$ takes the form

$$\widetilde{g}_2 = \frac{1}{\sqrt{n}} \sum_{j=1}^n \delta_{ij} \operatorname{sgn}(y_j - \boldsymbol{x}_j^\top \widetilde{\boldsymbol{\beta}}),$$

where $\operatorname{sgn}(\cdot)$ denotes the sign function. Let $e_j = y_j - \boldsymbol{x}_j^\top \widetilde{\boldsymbol{\beta}}$, $j = 1, \dots, n$, be the LAD residuals [40]. Noticing that $\widetilde{g}_2 = \operatorname{sgn}(e_i)/\sqrt{n}$, and using Proposition 1 follows that the gradient statistic for testing $H_0 : \gamma = 0$ is given by:

$$T_i = \frac{\operatorname{sgn}(e_i)\widehat{\gamma}_i}{\lambda} = \frac{\operatorname{sgn}(e_i)e_{i(i)}}{\lambda},$$

for $i = 1, \dots, n$. This allows us to identify the ith observation as an outlier if $T_i \geq \chi^2_{1-\alpha}(1)$.

It was demonstrated in [41] that the gradient and Rao score statistics are asymptotically equivalent for a fairly general class of estimators. These estimators are defined through the optimization of a general objective function and are particularly popular in econometrics, where they are often referred to as extremum estimators. This class of estimators includes the LAD regression problem as a special case. Unlike the Rao score statistic, in the case of the gradient statistic, it is necessary to estimate the nuisance parameter λ.

3.2. **Standardized residuals and QQ-plot with envelope.** The objective of this section is to introduce two types of residuals. Residuals are crucial for assessing model goodness of fit, evaluating deviations from distributional assumptions, and identifying outliers. The behavior of LAD residuals was characterized by [40], who noted that the zero-residual property can lead to misleading interpretations. To the best of our knowledge, no proposals for standardized residuals in LAD regression have been presented. Hence, a natural option is to consider the randomized quantile residuals defined by [42]. To introduce ideas, note that the LAD estimator of $\boldsymbol{\beta}$ corresponds to the maximum likelihood (ML) estimator under Laplace errors, that is, assuming that the random disturbances are independent, each with a standard Laplace distribution [43] with zero position and scale $\phi > 0$. This leads to writing, $y_i \sim \mathsf{Laplace}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \phi)$, with density function

$$f(y_i; \boldsymbol{\beta}, \phi) = \frac{1}{\sqrt{2}\phi} \exp\Big(-\frac{\sqrt{2}}{\phi}|y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}|\Big), \qquad i = 1, \dots, n. \tag{14}$$

By optimizing the log-likelihood function derived from (14) with respect to ϕ, we obtain

$$\widehat{\phi} = \frac{\sqrt{2}}{n} \sum_{i=1}^n |y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}|.$$

Thus, for LAD regression, the randomized quantile residual is defined as

$$q_i = \Phi^{-1}(F(y_i; \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}, \widehat{\phi})), \qquad i = 1, \dots, n, \tag{15}$$

where $F(\cdot\,; \boldsymbol{x}_i^\top \boldsymbol{\beta}, \phi)$ is the cumulative distribution function of the random variable y_i, which must be evaluated at $\widehat{\boldsymbol{\psi}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$, whereas $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. An interesting property of this type of residual is that for consistent estimators of $\boldsymbol{\beta}$ and ϕ, it follows that these are asymptotically standard normally distributed. Using the Wald test statistic given in Equation (13), we introduce the standardized residual

$$r_i = \frac{\sqrt{1-h_{ii}}\, e_{i(i)}}{\lambda}, \qquad i = 1, \dots, n. \tag{16}$$

By consistently estimating the nuisance parameter λ (see Equation (7)), the r_i's are asymptotically standard normally distributed. We should note that the quantile residual inherits the zero residual property possessed by the raw residuals, i.e., $q_i = 0$ if $i \in B$. However, the standardized residual r_i depends on the eliminated residual $e_{i(i)}$ and does not have, in general, this property.

The suggested residuals can be used to check the adequacy of the proposed model, for example, using the modifications of the Jarque-Bera test proposed in [44, 45]. Additionally, we suggest the use of graphical techniques such as the QQ-plot of residuals with simulated envelope (see, for instance [46] and [47]). The fundamental idea underlying this type of graph is that residuals that lie outside the envelope indicate that they are outliers or may be indicative of the model being inappropriate. Below we present an algorithm for constructing a simulated envelope for either the quantile residuals defined in (15) or the standardized residuals given in (16). In the following algorithm, we will denote any of these selected residuals as $\boldsymbol{z} = (z_1, \dots, z_n)^\top$.

Algorithm 1: Simulated envelope for residuals

Input: Number of replications M, fitted values $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, estimate of scale $\widehat{\phi}$, consistent estimate of nuisance parameter $\widehat{\lambda}$ using Equation (7) and confidence level $1-\alpha$.

Output: A matrix $\boldsymbol{R} \in \mathbb{R}^{n\times 2}$ of simulated envelope for $\boldsymbol{z}$ and selected residual $\boldsymbol{z} = (z_1, \dots, z_n)^\top$.

Compute the residuals $\boldsymbol{z} = (z_1, \dots, z_n)^\top$
for $j = 1$ **to** M **do**
  Generate n observations $\boldsymbol{Y}^* = (y_1^*, \dots, y_n^*)^\top$, where the y_i^*'s are independently drawn from the $\mathsf{Laplace}(\widehat{y}_i, \widehat{\phi})$ distribution.
  Obtain LAD parameter estimates $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\phi}^*$ by regressing $\boldsymbol{Y}^*$ on $\boldsymbol{X}$.
  Compute the residuals $\boldsymbol{z}^* = (z_1^*, \dots, z_n^*)^\top$ (i.e., the quantile residual q_i^* using definition in (15) or the standardized residual r_i^* by using the formula (16)).
  Sort the residuals $\boldsymbol{z}^*$ in non-decreasing order, obtaining $z^*_{(1)} \le z^*_{(2)} \le \cdots \le z^*_{(n)}$.
  Save the ordered residuals for each simulated dataset.
end
for $i = 1$ **to** n **do**
  Compute empirical quantiles $\alpha/2$ and $1-\alpha/2$ of the ordered residuals, and save the lower and upper bounds $R_{i,1} = z^*_{(\alpha/2)}$, $R_{i,2} = z^*_{(1-\alpha/2)}$, respectively.
end
return $\boldsymbol{R}$, $\boldsymbol{z}$.

The procedure for constructing a QQ-plot with a simulated envelope for the residuals from the least absolute deviation fit described in the previous algorithm has been implemented in the `envelope` function available from the india[1] [20] package for the R statistical software.

4. Examples with real-life datasets

In this section we illustrate the proposed methodology through the analysis of two datasets previously studied in the literature. The diagnostic procedures described in the previous section have been implemented in R, and related code is available on github.[2] In our analyses we also used routines available in the india [20] and L1pack [28] packages.

[1] Available on CRAN repository at https://cran.r-project.org/package=india
[2] URL: https://github.com/faosorios/lad_MSOM

4.1. **Martin Marietta data.** This dataset was introduced in [48], who proposed modeling the excess return of Martin Marietta Corporation shares using the Capital Asset Pricing Model (CAPM) [49], has also been studied using a class of asymmetric models with heavier tails than the normal distribution by [50, 51]. The data correspond to monthly returns of excess returns traded on the New York stock exchange between January 1982 and December 1986 and it is available in the `L1pack` R package. The regression model can be written as:

$$r_t - r_{ft} = \beta_0 + \beta_1(r_{mt} - r_{ft}) + \epsilon_t, \qquad t = 1, \dots, 60,$$

where r_t denotes the return of the asset in the period t, r_{ft} is the risk free rate return, for this dataset the 30-day treasury bill rate is used, and r_{mt} is the market returns reported by the Center of Research in Security Prices (CRSP). The scatter plot and the fitted model are displayed in Figure 1.

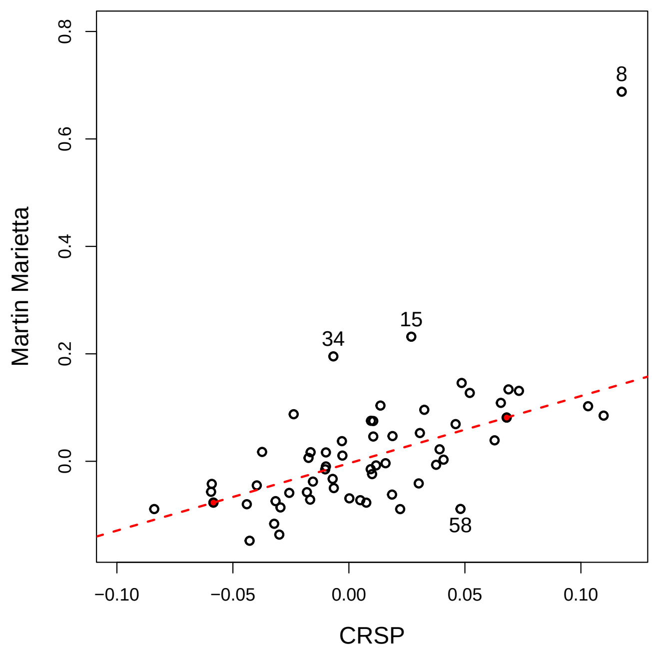
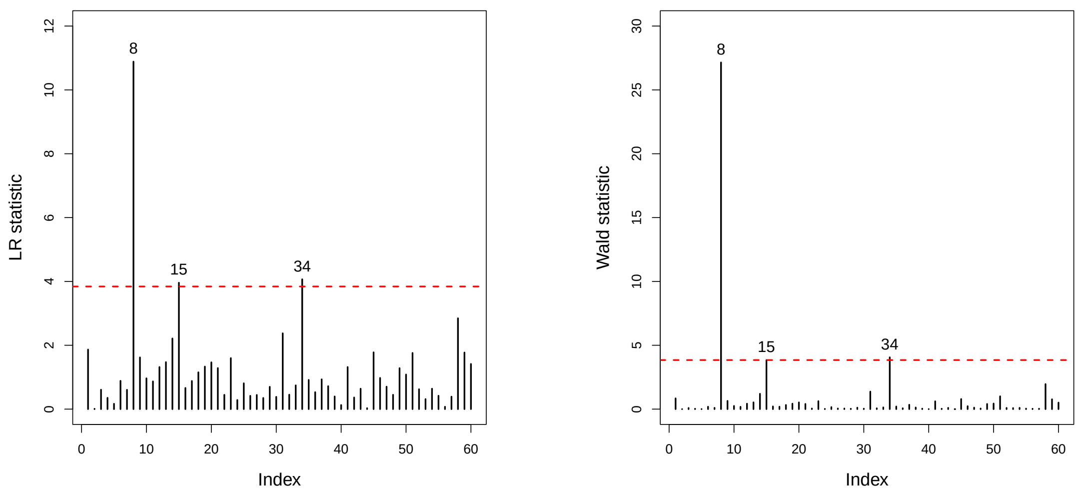


FIGURE 1. Martin Marietta data: Scatter plot of excess of returns for Martin Marietta corporation and the market portfolio with fitted curve (segmented line).

In [51], observations 8, 15, 34, and 58 are identified as outliers, whereas in [50], observations 8 and 58 are assessed as influential under the least squares estimation method. Table 1 presents the parameter estimates using LAD regression as well as the estimates when some observations are removed. Figure 2 displays the statistics proposed for testing the hypothesis $H_0 : \gamma = 0$. Using LR and Wald statistics, observations 8, 15 and 34 can be identified as outliers. In this figure, the cutoff point corresponds to the quantile $\chi^2_{1-\alpha}(1)$ with $\alpha = 0.05$. We should also note that case 8, in addition to being an outlier, is highly influential (which can be detected, for instance, using likelihood displacement). In fact, observation 8 causes a change in the intercept coefficient of 219.75% (see Table 2) and is the only observation that falls outside the envelope shown in Figure 3 (c) and (d). This figure also highlights the zero-residual property of quantile residuals, which leads to unusual behavior in the central part of Figure 3 (c). Note that removing the subset of observations $I = \{8, 15, 34\}$ not only affects the estimation of the coefficients but also affects the estimation of the scale parameters ϕ and λ, by -23.93% and 6.86%, respectively.



(a) (b)

FIGURE 2. Test statistics of the null hypothesis $H_0 : \gamma = 0$ for Martin Marietta data: (a) likelihood ratio LR_i and (b) Wald W_i statistics.

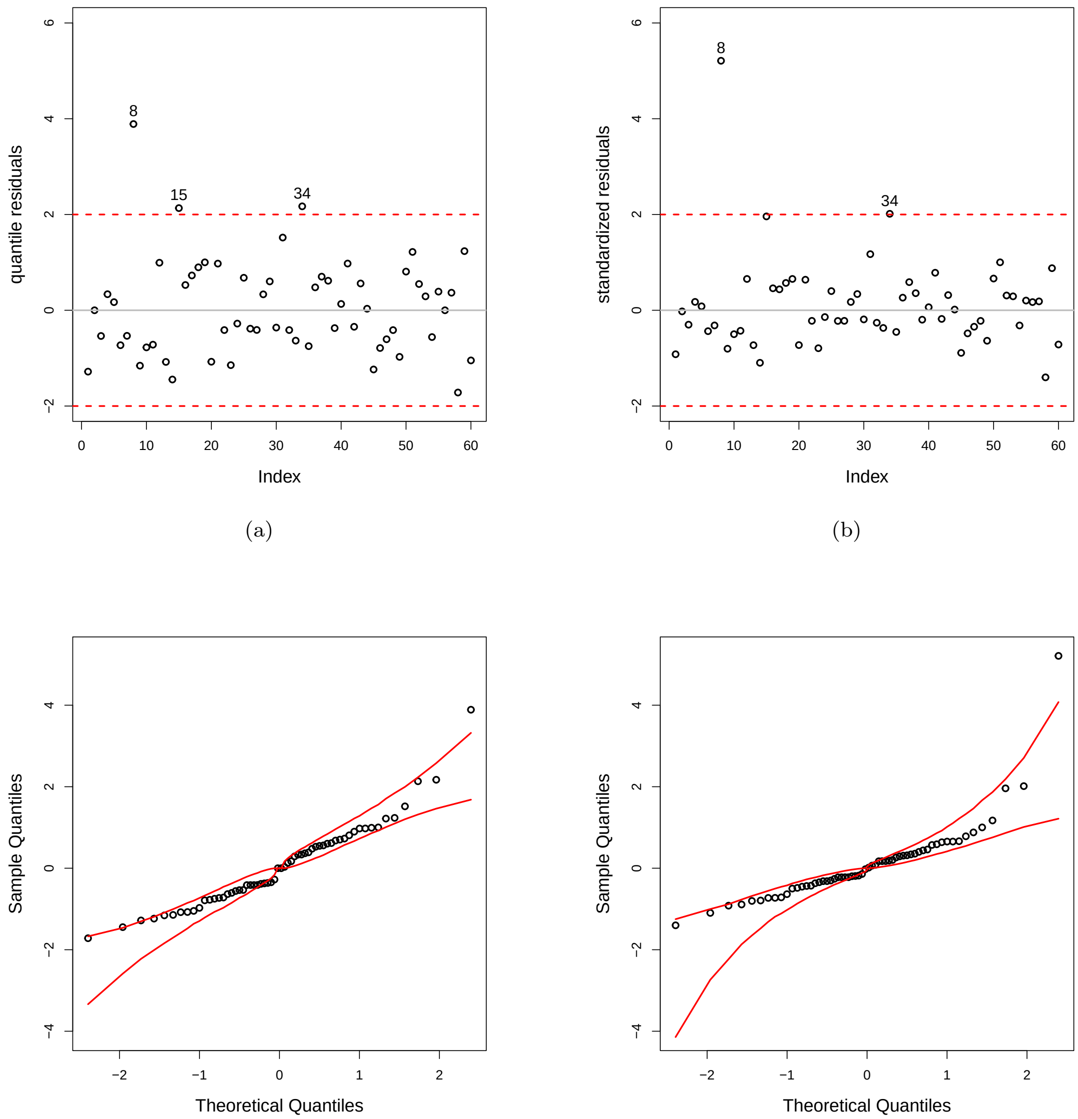
TABLE 1. Parameter estimates and standard errors (in parenthesis) for Martin Marietta data.

Deleted observations	Intercept β_0	CRSP β_1	scale parameter ϕ	scale parameter λ
none	-0.0038 (0.0135)	1.2524 (0.3132)	0.0836	0.1020
8	**-0.0120** (0.0133)	**1.1108** **(0.3275)**	**0.0717**	0.1004
15	-0.0038 (0.0135)	1.2524 (0.3120)	0.0802	0.1014
34	-0.0038 (0.0135)	1.2524 (0.3119)	0.0801	0.1014
58	-0.0038 **(0.0152)**	1.2524 **(0.3538)**	0.0816	**0.1144**
8,15,34	**-0.0120** **(0.0146)**	**1.1108** **(0.3566)**	**0.0636**	0.1090

TABLE 2. Percentage of relative change in parameter estimates for Martin Marietta data.

Deleted observations	Intercept β_1	CRSP β_2	scale parameter ϕ	scale parameter λ
8	**219.7535**	**-11.3015**	**-14.2171**	-1.5700
15	0.0000	0.0000	-4.0935	-0.5470
34	0.0000	0.0000	-4.2506	-0.5470
58	0.0000	0.0000	-2.4658	**12.1030**
8,15,34	**219.7535**	**-11.3015**	**-23.9280**	6.8598

4.2. **Skeena river sockeye salmon data.** This dataset has been reported by [52] and previously analyzed by [53] and [18]. In [54] was proposed characterizing the size of the annual spawning population, or spawners



(a) (b)

(c) (d)

FIGURE 3. Plot of residuals for Martin Marietta data: Index plot of (a) randomized quantile residuals, (b) standardized residuals, simulated envelopes for randomized quantile residuals (c), and standardized residuals (d).

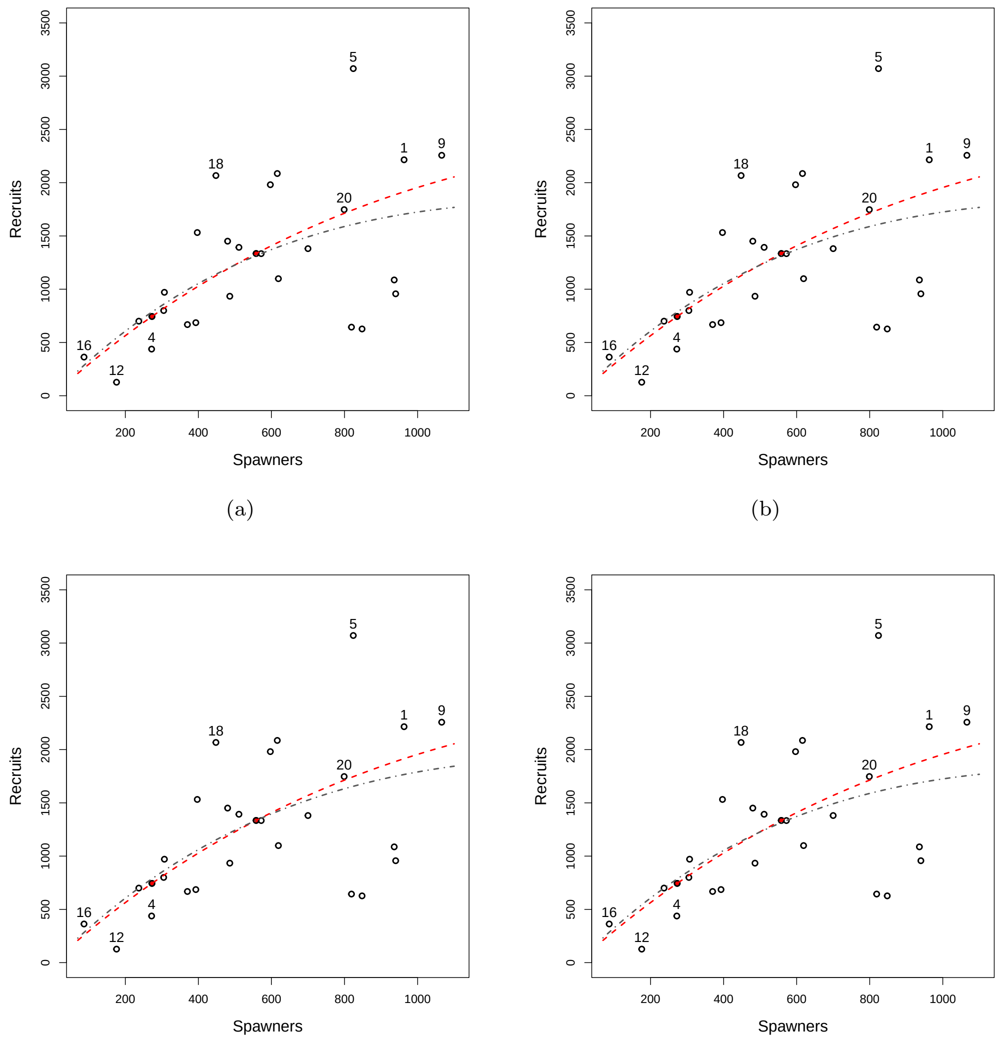
(s), with the number of fish of catchable-size, called recruits (R), using the following model

$$R = \theta_1 \, s \exp(-\theta_2 s), \qquad \theta_2 \geq 0. \tag{17}$$

The dataset is included in the R package india and corresponds to 28 observations of spawners and recruits recorded between 1940 and 1967 for the Skeena River sockeye salmon stock.

The model in (17) can be linearized by using the logarithmic transformation, that is,

$$\log(R/s) = \log\theta_1 - \theta_2 s = \beta_1 + \beta_2 s,$$



(a) (b) (c) (d)

Figure 4. Scatter plot of recruits and spawners with fitted curve (segmented line) and fitted curve with deleted observations (dashed-dotted line): (a) Obs. 5, (b) Obs. 9, (c) Obs. 12, and (d) Obs. 20. Skeena river sockeye salmon data.

with $\beta_1 = \log \theta_1$ and $\beta_2 = -\theta_2$, which allows us to use LAD estimation for a simple linear regression model. Figure 4 shows a graph with the Skeena River sockeye salmon data and the fitted curve. We should note that cases 21 and 23 are determined as basic observations by the estimation algorithm, whereas the estimation results for the linearized model are presented in Table 3.

Observations 1, 5, 9, and 20 were detected as influential by [18] using LAD regression considering the nonlinear model in (17), and notably, case 9 has a highly influential effect. Using various robust estimation procedures [53] additionally identified observations 4, 12, 16, and 18 as influential. In contrast, the diagnostic measures proposed in this paper identify observation 12 as an outlier (see Figure 5). Interestingly, [53, Example 6.1] points out that observation 12, i.e., the year 1951, corresponds to a year in which a rockslide interfered with spawning and thus altering recruitment.

Table 3. Parameter estimates and standard errors (in parenthesis) for Skeena river sockeye salmon data.

Deleted observations	Intercept β_1	spawners β_2	scale parameter ϕ	λ
none	1.1273	-0.0005	0.5083	0.6454
	(0.2882)	(0.0005)		
1	**1.2501**	**-0.0007**	0.5157	**0.7506**
	(0.3418)	(0.0006)		
4	1.2386	**-0.0007**	0.4984	**0.7367**
	(0.3415)	(0.0005)		
5	**1.2501**	**-0.0007**	0.4956	**0.7506**
	(0.3369)	(0.0006)		
9	**1.2501**	**-0.0007**	0.5163	**0.7506**
	(0.3481)	(0.0006)		
12	1.2386	**-0.0007**	**0.4531**	**0.7367**
	(0.3487)	(0.0006)		
16	1.0808	**-0.0004**	0.5088	0.7005
	(0.3396)	(0.0005)		
18	1.1273	-0.0005	0.4953	0.6609
	(0.2987)	(0.0005)		
20	**1.2501**	**-0.0007**	0.5244	**0.7688**
	(0.3445)	(0.0006)		

Table 4. Percentage of relative change in parameter estimates for Skeena river sockeye salmon data.

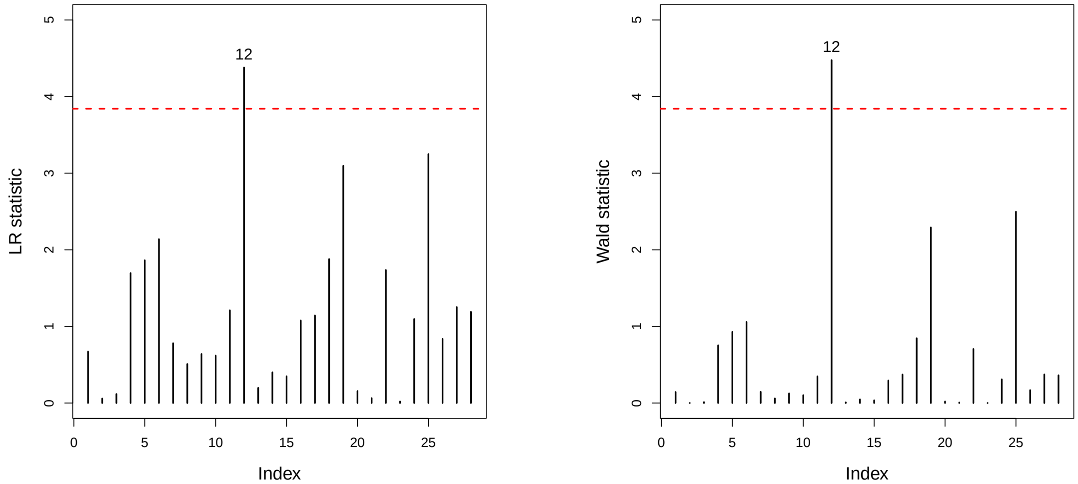
Deleted observations	Intercept β_1	spawners β_2	scale parameter ϕ	λ
1	**10.8940**	**54.3016**	1.4680	**16.3001**
4	9.8686	**43.6281**	-1.9389	**14.1375**
5	**10.8940**	**54.3016**	-2.4957	**16.3001**
9	**10.8940**	**54.3016**	1.5731	**16.3001**
12	9.8686	**43.6281**	**-10.8605**	**14.1375**
16	-4.1284	**-18.2513**	0.1199	8.5278
18	0.0000	0.0000	-2.5461	2.3999
20	**10.8940**	**54.3016**	3.1818	**19.1166**

Figure 4 also shows the adjusted curve (dashed-dotted line) when observation 12 is removed. Remarkably, the case-deletion diagnostics for LAD regression (see [17, 18]) fail to identify case 12, despite its considerable impact on the fitted curve (see Figure 4 (c)). The residual plot reveals that case 12 has a large negative residual and produces the maximum relative change in the estimation of the scale parameters (see Table 4). Thus, the hypothesis testing technique correctly identifies case 12 as an anomalous year.

5. Concluding remarks

In this paper, we have proposed the detection of outliers in LAD regression using different test statistics in the mean-shift outlier model. To the best of our knowledge, the available literature on diagnostics in the context of LAD regression had not addressed this aspect in detail. Thus, the measures developed as part of this work represent a very useful alternative for better understanding the role that outliers can play in the proposed model.

A key contribution of this work is the introduction of two alternatives for defining standardized residuals, the first based on the randomized quantile residual, while the second was derived, by analogy to studentized



(a) (b)

FIGURE 5. Test statistics of the null hypothesis $H_0 : \gamma = 0$ for Skeena river sockeye salmon data: (a) likelihood ratio LR_i and (b) Wald W_i statistics.

residuals in regression under normal errors, i.e, from a test statistic for the detection of outliers. Our findings show that for all the datasets analyzed (see also additional examples in the supplementary material), the proposed methods have been able to identify those observations with large residuals. Furthermore, these types of residuals allowed the introduction of an algorithm for the construction of QQ-plots of residuals with simulated envelope. We should note that this graphical tool has allowed us to appreciate that the distributional assumption that errors follow a standard Laplace distribution is satisfied for all datasets. However, for the construction of the QQ-plot with simulated envelopes described in Algorithm 1, we recommend the use of standardized residuals r_i since they do not have the zero residual property that is inherited by the quantile residuals q_i, which leads to strange behavior of the related simulated envelope (see Figures 3 (c), and 6 (c). Also compare with Figures 2 (c) and 4 (c) from the supplementary material).

TABLE 5. Summaries of the number of iterations until reach convergence of full iterated deletion estimates and elapsed total time considering Barrodale and Roberts (BR) and IRLS algorithms for each dataset.

Dataset	Estimation procedure	Iterations				Total time (sec.)
		minimum	median	average	maximum	
Martin-Marietta	BR	2	2.0	2.6	5	0.013
$(n = 60, p = 2)$	IRLS	17	43.0	72.6	500	0.046
Skeena	BR	2	3.5	3.8	6	0.002
$(n = 28, p = 2)$	IRLS	12	36.5	59.0	267	0.010
Stackloss	BR	5	6.0	6.5	10	0.001
$(n = 21, p = 4)$	IRLS	17	56.0	64.8	190	0.010
Aircraft	BR	6	7.0	7.6	11	0.004
$(n = 23, p = 5)$	IRLS	55	161.0	159.3	419	0.016

The confirmatory analysis presented in Tables 2 and 4 (See also, Tables 2 and 4 of the supplementary material) provides a better understanding of the role of outliers detected using the proposed methodology. Interestingly, for the datasets analyzed in this study, it is common for observations identified as outliers to have a significant effect on the estimation of the scale parameter. We should highlight that for the Skeena

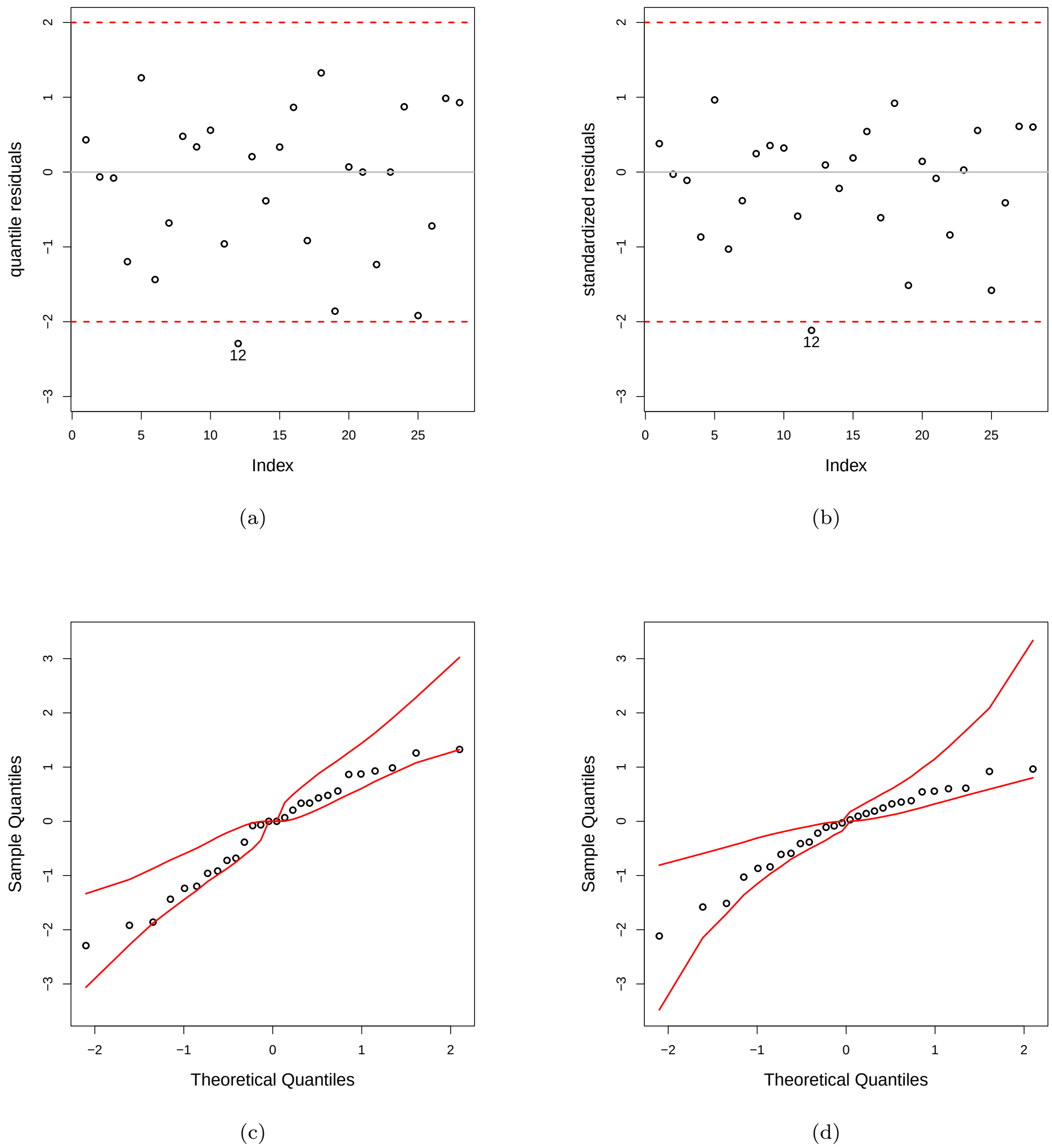
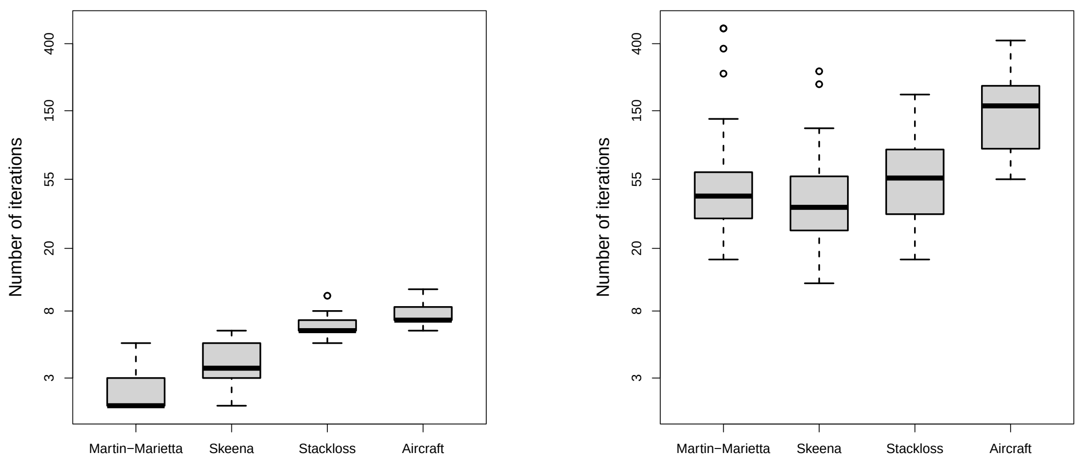


Figure 6. Plot of residuals for Skeena river sockeye salmon data: Index plot of (a) randomized quantile residuals, (b) standardized residuals, simulated envelopes for randomized quantile residuals (c), and standardized residuals (d).

River sockeye salmon data, the detection procedure correctly identified an observation corresponding to a year in which a natural event occurred that could have had a strong impact on the model. Additionally, [53, pp. 192] explored observations that could potentially be influential by examining the estimated weights from several robust procedures. It is important to note that due to the type of algorithm used to obtain the LAD estimators, this type of analysis is not available. Therefore, the proposed methodology helps to remedy this problem.

In order to reduce the computational effort required to evaluate diagnostic measures based on case deletion techniques, the use of one-step approximations has been suggested (see, for instance, [22, Sec. 5.2]). Since such approximations are not available for LAD regression, in this work we have followed the suggestion



(a) (b)

FIGURE 7. Iterations until reach convergence (in a logarithmic scale): Full iterated deletion estimates using (a) Barrodale and Roberts and (b) IRLS algorithms for each dataset.

given by of Sun and Wei [18], who obtained $\widehat{\boldsymbol{\beta}}_{(i)}$, for $i = 1, \dots, n$, using the Barrodale and Roberts (BR) method. Table 5 and Figure 7 present a brief comparison of the computational burden measured in terms of the number of iterations required to reach convergence and the total computation time (in seconds) for estimating the regression coefficients when the ith observation is removed for $i = 1, \dots, n$. An approximate estimation method based on iteratively reweighted least squares (IRLS) [55] is also used. It should be noted that IRLS for LAD regression was first identified as an EM algorithm in [56] and that, despite its simplicity and the fact that it is often fast, it generally does not guarantee convergence to the exact solution. In our comparison, we used the implementation of the IRLS procedure described in [57], which is available in the `L1pack` [28] package. The results were obtained on an Intel i7 desktop computer running at 4.8 GHz with 32 GB of RAM. It is interesting to note that for all the datasets analyzed (see also the supplementary material), the total computation time is less than 50 milliseconds, with the IRLS-based variant being significantly slower than its counterpart using the Barrodale and Roberts procedure. Although differences in computation time depend heavily on the data configuration, it is easy to see from Table 5 that the computation times for the BR method range from 3.5 to 10 times faster than those for IRLS, whereas in terms of the number of iterations required for convergence, the differences in magnitude can be significant (see Figure 7). Following [58], we can note that these results for small-scale data are expected. An interesting finding from this experiment is that one-step approximations for $\widehat{\boldsymbol{\beta}}_{(i)}$ are not reliable; in particular, they are not appropriate when the IRLS method is used.

An important side result presented in this paper corresponds to Proposition 1, where the gradient statistic is developed to test hypotheses about subvectors in LAD regression, thus complementing the results described in [23]. We should note that the gradient statistic is not always able to correctly identify observations as outliers (see Figure 5 in the supplementary material), so further research on the properties of this test statistic is required.

Although the model defined in Equation (17) for the Skeena River sockeye salmon data is intrinsically linear, which allowed the analysis to be conducted with the tools developed in this paper, the development of procedures for detecting outliers in nonlinear regression models using the absolute minimum residual methodology can be challenging and deserves further investigation.

Acknowledgments

The author is grateful to the Associate Editor and two anonymous reviewers for their careful and insightful reviews, which greatly enhanced the clarity and overall quality of this manuscript. The author wishes to express his gratitude to Professor Ronny Vallejos for his encouragement and insightful comments on an earlier version of this paper.

Code and software availability

All analyses were conducted in the R environment for statistical computing. The replication files related to this article are available online at https://github.com/faosorios/lad_MSOM

Disclosure statement

The author declares no potential conflict of interests.

Funding

No funding was received.

References

[1] Narula, S.C., Wellington, J.F. (1985). Interior analysis for the minimum sum of absolute errors regression. *Technometrics* **27**, 181-188.

[2] Dielman, T.E. (2005). Least absolute value regression: recent contributions. *Journal of Statistical Computation and Simulation* **75**, 263-286.

[3] Farebrother, R.W. (2013). *L_1-Norm and L_∞-Norm Estimation: An Introduction to the Least Absolute Residuals, the Minimax Absolute Residual and Related Fitting Procedures*. Springer, New York.

[4] Yüzbaşı, B., Ahmed, S.E., Arashi, M., Norouzirad, M. (2020). LAD, LASSO and related strategies in regression models. In: *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management. ICMSEM 2019. Advances in Intelligent Systems and Computing, vol 1001*. J. Xu, S. Ahmed, F. Cooke, G. Duca (Eds.) Springer, Cham. pp. 429-444.

[5] Radojičić, U., Nordhausen, K. (2023). Least absolute value. In *Encyclopedia of Mathematical Geosciences*, B.S.D. Sagar, Q. Cheng, J. McKinley, F. Agterberg (Eds.) Springer, Cham. pp. 720-724.

[6] Ziegelmann, F.A. (2008). A local linear least-absolute-deviations estimator of volatility. *Communications in Statistics: Simulation and Computation* **37**, 1543-1564.

[7] Honda, T. (2013). Nonparametric LAD cointegrating regression. *Journal of Multivariate Analysis* **117**, 150-162.

[8] Wang, W.W., Yu, P., Lin, L., Tong, T. (2019). Robust estimation of derivatives using locally weighted least absolute deviation regression. *Journal of Machine Learning Research* **20**, 1-49.

[9] Ciuperca, G. (2011). Estimating nonlinear regression with and without change-points by LAD method. *Annals of the Institute of Statistical Mathematics* **63**, 717-743.

[10] Yang, F. (2014). Robust mean change-point detecting through Laplace linear regression using EM algorithm. *Journal of Applied Mathematics*, ID 856350.

[11] Zeebari, Z. (2012). Developing ridge estimation method for median regression. *Journal of Applied Statistics* **39**, 2627-2638.

[12] Zeebari, Z., Shukur, G. (2023). On the least absolute deviations method for ridge estimation of SURE models. *Communications in Statistics: Theory and Methods* **52**, 4773-4791.

[13] VandenHeuvel, D., Wu, J., Wang, Y.G. (2023). Robust regression for electricity demand forecasting against cyberattacks. *International Journal of Forecasting* **39**, 1573-1592.

[14] Guest, L.M., McCabe, J., O'Halloran, C., Rana, M., Sun, W., Rudoler, D. (2024). Perspectives on work in the continuing care sector during and after the COVID-19 pandemic: A mixed-method design. *Journal of Nursing Management*, ID 7187263.

[15] Dodge, Y. (2002). *Statistical Data Analysis Based on the L_1-Norm and Related Methods*. Springer, Basel.

[16] Ellis, S.P., Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression. *Journal of the American Statistical Association* **87**, 143-148.

[17] Elian, S.N., André, C.D.S., Narula, S.C. (2000). Influence measure for the L_1 regression. *Communications in Statistics: Theory and Methods* **29**, 837-849.

[18] Sun, R.B., Wei, B.C. (2004). On influence assessment for LAD regression. *Statistics and Probability Letters* **67**, 97-110.

[19] Wei, B.C., Shih, J.Q. (1994). On statistical models for regression diagnostics. *Annals of the Institute of Statistical Mathematics* **46**, 267-278.

[20] Osorio, F. (2026). india: Influence diagnostics in statistical models. R package version 0.1-2. doi: 10.32614/CRAN.package.india

[21] Rodrigues, K.A.S., Elian, S.N. (2006). Influence measures for L_1 regression: An analysis with the R package diagL1. *Journal of Applied Statistics* **53**, 146-168.

[22] Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, New York.

[23] Koenker, R., Bassett, G. (1982). Test of linear hypotheses and l_1 estimation. *Econometrica* **50**, 1577-1583.

[24] Brownlee, K.A. (1965). *Statistical Theory and Methodology in Science and Engineering, 2nd Edition*. Wiley, New York.

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

[26] Barrodale, I., Roberts, F.D.K. (1973). An improved algorithm for discrete l_1 linear approximation. *SIAM Journal of Numerical Analysis* **10**, 839-848.

[27] Barrodale, I., Roberts, F.D.K. (1974). Algorithm 478: Solution of an overdetermined systems of equations in the l_1 norm. *Communications of the ACM* **17**, 319-320.

[28] Osorio, F., Wołodźko, T. (2026). L1pack: Routines for *L1* estimation. R package version 0.62. doi: 10.32614/CRAN.package.L1pack

[29] Bassett, G., Koenker, R. (1978). Asymptotic theory for least absolute error regression. *Journal of the American Statistical Association* **73**, 618-622.

[30] Rao, C.R (1988). Methodology based on the L_1-norm in statistical inference. *Sankhyā, Series A* **50**, 289-313.

[31] Muggeo, V.M.R, Lovison, G. (2014). The "Three plus one" likelihood-based test statistics: Unified geometrical and graphical interpretations. *The American Statistician* **68**, 302-306.

[32] Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics* **34**, 206-215.

[33] Flores, S. (2015). Sharp non-asymptotic performance bounds for l_1 and Huber robust regression estimators. *TEST* **24**, 796-812.

[34] Appa, G., Smith, C. (1973). On L_1 and Chebyshev estimation. *Mathematical Programming* **5**, 73-87.

[35] McKean, J.W., Schrader, R.M. (1984). A comparison of methods for studentizing the sample mean. *Communications in Statistics - Simulation and Computation* **13**, 751-773.

[36] Dielman, T.E., Rose, E.L. (1996). A note on hypothesis testing in LAV multiple regression: A small sample comparison. *Computational Statistics & Data Analysis* **21**, 463-470.

[37] Beckman, R.J., Cook, R.D. (1983). Outlier..........s. *Technometrics* **25**, 119-149.

[38] Wei, W.H., Fung, W.K. (1999). The mean-shift outlier model in general weighted regression and its applications. *Computational Statistics & Data Analysis* 30, 429-441.

[39] Osorio, F., Gárate, Á., Russo, C.M. (2024). The gradient test statistic for outlier detection in generalized estimating equations. *Statistics and Probability Letters* **209**, 110087.

[40] Morgenthaler, S. (1997). Properties of L_1 residuals. In: Dodge, Y., editor. L_1 *Statistical Procedures and Related Topics. IMS Lecture Notes - Monograph Series* **31**. Hayward, Institute of Mathematical Statistics. p. 79-90.

[41] Crudu, F., Osorio, F. (2020). Bilinear form test statistics for extremum estimation. *Economics Letters* **187**, 108885.

[42] Dunn, P.K., Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**, 236-244.

[43] Kotz, S., Kozubowski, T.J., Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance*. Springer, New York.

[44] Gel, Y.R., Gastwirth, J.L. (2008). A robust modification of the Jarque-Bera test of normality. *Economics Letters* **99**, 30-32.

[45] Gel, Y.R. (2010). Test of fit for a Laplace distribution against heavier tailed alternatives. *Computational Statistics & Data Analysis* **54**, 958-965.

[46] Atkinson, A.C. (1985). *Plots, Transformations and Regression*. Oxford University Press, Oxford.

[47] Paula, G.A. (2025). *Modelos de Regressão e Aplicações*. Instituto de Matemática e Estatística, Universidade de São Paulo.

[48] Butler, R.J., McDonald, J.B., Nelson, R.D., White, S.B. (1990). Robust and partially adaptive estimation of regression models. *The Review of Economics and Statistics* **72**, 321-327.

[49] Sharpe, W.F. (1964). Capital asset prices: A theory of equilibrium under conditions of risk. *Journal of Finance* **19**, 425-442.

[50] Ferreira, C.S., Lachos, V.H., Bolfarine, H. (2015). Inference and diagnostics in skew scale mixtures of normal regression models. *Journal of Statistical Computation and Simulation* **85**, 517-537.

[51] Ferreira, C.S., Mattos, T.B., Balakrishnan, N. (2016). Mean-shift outliers model in skew scale-mixtures of normal distributions. *Journal of Statistical Computation and Simulation* **86**, 2346-2361.

[52] Ricker, W.E., Smith, H.D. (1975). A revised interpretation of the history of the Skeena river sockeye salmon. *Journal of the Fisheries Research Board of Canada* **32**, 1369-1381.

[53] Carroll, R.J., Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, New York.

[54] Ricker, W.E. (1954). Stock and recruitment. *Journal of the Fisheries Research Board of Canada* **11**, 559-623.

[55] Schlossmacher, E.J. (1973). An iterative technique for absolute deviations curve fitting. *Journal of the American Statistical Association* **68**, 857-859.

[56] Lange, K., Sinsheimer, J.S. (1993). Normal/Independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* **2**, 175-198.

[57] Phillips, R.F. (2002). Least absolute deviations estimation via the EM algorithm. *Statistics and Computing* **12**, 281-285.

[58] Armstrong, R.D., Frome, E.L. (1976). A comparison of two algorithms for absolute deviation curve fitting. *Journal of the American Statistical Association* **71**, 328-330.