Computational Statistics (2025) 40:5075–5104
https://doi.org/10.1007/s00180-025-01605-6

ORIGINAL PAPER

Weighted LAD-Liu-LASSO for robust estimation and sparsity

Murat Genç[1] · **Adewale Lukman**[2]

Received: 26 June 2024 / Accepted: 20 January 2025 / Published online: 31 January 2025
© The Author(s) 2025

Abstract
The Least Absolute Shrinkage and Selection Operator (LASSO) is widely used for parameter estimation and variable selection but can encounter challenges with outliers and heavy-tailed error distributions. Integrating variable selection methods such as LASSO with Weighted Least Absolute Deviation (WLAD) has been explored in limited studies to handle these problems. In this study, we proposed the integration of Weighted Least Absolute Deviation with Liu-LASSO to handle variable selection, parameter estimation, and heavy-tailed error distributions due to the advantages of the Liu-LASSO approach over traditional LASSO methods. This approach is demonstrated through a simple simulation study and real-world application. Our findings showcase the superiority of our method over existing techniques while maintaining the asymptotic efficiency comparable to the unpenalized LAD estimator.

Keywords Lasso · Least absolute deviation · Weighted estimator · Liu-Lasso · Outliers · Variable selection

1 Introduction

Regression analysis is a critical statistical tool widely applicable across diverse fields such as science, social sciences, engineering, and many others to investigate and quantify the relationship strength between an outcome variable and one or more covariates. Linear regression is a predominant form of regression analysis, modeling the outcome variable as a linear function of the covariates. The least squares method is renowned for estimating the parameters of a linear regression model. Yet, it is

✉ Adewale Lukman
adewale.lukman@und.edu

Murat Genç
muratgenc@tarsus.edu.tr

[1] Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Tarsus University, 33400 Mersin, Turkey

[2] Department of Mathematics and Statistics, University of North Dakota, Grand Forks, ND, USA

susceptible to outliers in either the outcome variable, covariates, or both due to its dependability on minimizing the sum of squared errors. This vulnerability can lead to biased parameter estimates and unstable predictions in classical methods.

We can improve the prediction accuracy by adopting robust regression methods while tolerating a slight increase in bias. Robust regression effectively analyzes regression models contaminated by outliers (Jiang et al. 2021). Various robust regression methods exist for robust parameter estimation. These include the M-estimator, the MM-estimator, the S-estimator, the least trimmed squared method, the least median of squares method, the Least absolute deviation (LAD), and others (Rousseeuw 1984; Rousseeuw and Leroy 1987; Hawkins 1993; Maronna et al. 2006).

LAD is a form of the M-regression method suitable for models with heavy-tailed error distributions or outliers in the response variable (Wang et al. 2007; Gao and Huang 2010). LAD minimizes the sum of the absolute value of the errors but falls short when outliers occur in the covariates, as noted in the literature (Arslan 2012; Jiang et al. 2021). The LAD method is sensitive to leverage points (outliers in the covariates), which can significantly impact the regression estimates negatively, highlighting its limitations. To address this, researchers have introduced weighted LAD (WLAD) regression, which down-weights leverage points, thereby improving robustness, especially in the presence of leverage (Giloni et al. 2006b, a). This method offers advantages over LAD, particularly under leverage conditions.

Traditional least squares regression and several robust regression methods, such as the LAD and WLAD, do not inherently provide sparse models or perform variable selection. Variable selection is a crucial aspect of model building as it helps identify important features among numerous covariates, reducing modeling biases and improving prediction accuracy. Including unnecessary predictors can lead to inefficient estimation while omitting significant covariates can result in biased parameter estimates and prediction outcomes. Various methods have been proposed, including stepwise regression techniques like forward selection and backward elimination, as well as regularization methods such as the Least Absolute Shrinkage and Selection Operator (LASSO) and elastic net to address these challenges (Hocking 1976; Draper and Smith 1998; Tibshirani 1996; Zou and Hastie 2005; Genç and Özkale 2021; Genç 2022; Lukman et al. 2023b).

LASSO is a powerful regression method that simultaneously performs variable selection and regularization. It accomplishes this by imposing an L_1 penalty on the regression coefficients, effectively shrinking many coefficients to zero and forcing certain coefficients to be excluded from the prediction process, improving the model's accuracy and interpretability. Unlike ridge regression, which also shrinks coefficients but does not enforce zero coefficients, LASSO's ability to set coefficients to zero facilitates variable selection, making it a valuable tool in regression analysis for building more efficient and interpretable models.

It is worth noting that LASSO, being least squares-based, is sensitive to outliers, necessitating the development of a robust variable selection method. A recent approach to address this sensitivity is the integration of LAD regression with the LASSO method, termed the LAD-LASSO regression method. This combination allows for robust parameter estimation and variable selection concurrently (Wang et al. 2007; Wang and

Leng 2007; Xu and Ying 2010). However, it is important to note that the LAD regression estimation method is robust mainly against outliers in the response variable but not against leverage points. Therefore, combining LAD and LASSO methods results in estimators that are resistant to response variable outliers but may not outperform the least squares method when outliers are present in the covariates (Arslan 2012). Arslan (2012) proposed a weighted version of the LAD-LASSO method by integrating the WLAD regression criterion with the adaptive LASSO penalty criterion introduced by Zou (2006). This weighted approach achieves robust parameter estimation even in the presence of outliers and leverage while simultaneously performing variable selection in regression.

Recent research has shed light on the superior performance of the Liu-LASSO method compared to traditional Lasso techniques. Genç (2022) introduced the Liu-LASSO estimator, which integrates the penalization functions of the Liu estimator developed by Liu (1993) with those of LASSO. Studies have shown that Liu-LASSO achieves variable selection and demonstrates significant predictive accuracy with minimal mean squared error across sparse and non-sparse data structures. However, there is a gap in the literature regarding the robust version of Liu-LASSO, which presents an area for further investigation. This study aims to propose the WLAD-Liu-LASSO method by combining the WLAD regression criterion with the Liu-LASSO penalty criterion. The WLAD criterion will mitigate the impact of leverage points and provide resistance to outliers in the response variable. Consequently, the resulting regression estimators will be less sensitive to leverage points and outliers.

Section 2 introduces various methods for parameter estimation, variable selection, and the WLAD-LASSO approach, which serve as the foundation for proposing our new method. Section 3 elaborates on the theoretical properties of the new method, while Sect. 4 presents a simulation study and example to showcase its performance. Finally, we conclude with a discussion section.

2 Methodology

2.1 LAD and WLAD

The linear regression model in matrix form expressed the outcome variable as a linear function of one or more covariates as follows:

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

where $y_i \in \mathbb{R}$ is the outcome variable, $\mathbf{x}_i = \left(x_{i1}, x_{i2}, ..., x_{ip}\right)^T$ are the standardized value of the p-th covariate in the i-th observation, $\boldsymbol{\beta}$ is the vector of the regression coefficient, and ε_i are the independently identically distributed random errors with mean zero and constant variance σ^2,i.e., $\varepsilon_i \sim N\left(0, \sigma^2\right)$. A simplified method for estimating $\boldsymbol{\beta}$ involves minimizing the sum of squared residuals through the least squares method. The least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_n^{LS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{2}$$

where $\mathbf{X}$ is the $n \times p$ matrix whose i-th row is $\mathbf{x}_i^T$ with rank p and $\mathbf{y}$ is the vector of outcome variable. However, the least square method produced unreliable estimates in the presence of outliers/leverage and cannot perform variable selection. However, the least squares method often yields unreliable estimates in the presence of outliers or high-leverage points and lacks the capability for variable selection.

We employ the LAD method as an alternative to the least squares approach, aiming to minimize the sum of absolute residuals (Dielman 2005). While similar to least squares in principle, LAD uses absolute values of the residuals instead of squared. However, a shortcoming of this method is its vulnerability to leverage points (outliers in the covariates) and its inability to perform variable selection (Giloni et al. 2006b). When both outlier and leverage points are present in a model, the breakdown point of LAD matches that of Least Squares, as observed by Rousseeuw and Leroy (1987). The minimum attainable value for the breakdown point is $1/n$.

Ellis and Morgenthaler (1992) addressed the limitation of the LAD regression method by introducing weights, creating the WLAD approach. The WLAD regression estimator is defined to minimize $\sum_{i=1}^{n} w_i \left| y_i - \mathbf{x}_i^T \boldsymbol{\beta} \right|$. The aim was to reduce the influence of outliers in the predictor space. Giloni et al. (2006b) further enhanced WLAD by proposing modifications to increase the breakdown point of LAD regression estimators. Their work also explored the properties of this enhanced WLAD estimator. Through simulations, they demonstrated that the performance of WLAD regression estimators competes favorably with high breakdown point regression estimators, especially in the presence of leverage points. For more details, refer to Hubert and Rousseeuw (1997); Rousseeuw and Leroy (1987).

While WLAD regression addresses robustness to outliers through weighting, it cannot perform variable selection. Variable selection methods, on the other hand, do not directly address outliers but excel in selecting relevant predictors by shrinking coefficients toward zero. Additionally, it is possible to combine robust estimators such as LAD and WLAD with variable selection methods, creating robust variable selection approaches. Sections 2.2 and 2.3 will introduce variable selection and robust variable selection methods.

2.2 LASSO-type estimators

LASSO, the acronym for Least Absolute Shrinkage and Selection Operator, stands out as one of the pioneering regularization techniques for variable selection, garnering widespread popularity in the field. This regularization method introduces a penalty term to the objective function of the LS method, enabling both variable selection and regularization. This dual functionality enhances the model's predictive accuracy and interpretability, making it a crucial tool in regression analysis (Tibshirani 1996). LASSO's strength lies in its ability to perform feature selection and regularization, which is particularly valuable when dealing with high-dimensional datasets that may contain redundant or irrelevant predictors. By shrinking coefficients towards zero, LASSO identifies relevant covariates and reduces overfitting, leading to more stable

and interpretable models. Tibshirani (1996) introduced LASSO by minimizing the following objective function:

$$\hat{\boldsymbol{\beta}}_n^L = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 + \gamma \sum_{j=1}^{p} \left|\beta_j\right| \right\}. \tag{3}$$

where $\gamma > 0$ is a constant tuning parameter that controls the degree of shrinkage applied to the regression coefficients. Constant tuning parameters in LASSO can lead to noticeable bias in the resulting estimates (Fan and Li 2001). Zou (2006) introduced the Adaptive LASSO, which employs different tuning parameters for each regression coefficient to address this limitation. The objective function for Adaptive LASSO is as follows:

$$\hat{\boldsymbol{\beta}}_n^{A-L} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 + n \sum_{j=1}^{p} \gamma_j \left|\beta_j\right| \right\}. \tag{4}$$

where $\gamma_j > 0$ is a weight specific to each regression coefficient, allowing for variable-specific shrinkage. This adaptive weighting helps mitigate the bias introduced by a constant tuning parameter, as used in standard LASSO.

Numerous regularization techniques have emerged as enhancements or possible alternatives addressing the limitations of LASSO. These include the Elastic-net by Zou and Hastie (2005), the Fused LASSO proposed by Tibshirani et al. (2005), Group LASSO by Yuan and Lin (2006), Sparse Group Lasso by Simon et al. (2013), the GO estimator by Genç and Özkale (2021), Liu-LASSO by Genç (2022), Kibria-Lukman estimator by Lukman et al. (2023b), among others. Elastic-net combines the penalties of LASSO and Ridge regression and overcomes Lasso's limitation of selecting at most n variables when the number of predictors exceeds the sample size.

Following the work (Zou and Hastie 2005; Genç 2022) introduced the Liu-LASSO (LL) estimator that merges the penalties of LASSO with those of the Liu estimator, enabling both parameter estimation and variable selection simultaneously. Integrating L_1-norm penalization into the Liu estimator framework enhances its capabilities. The Liu-LASSO estimator not only inherits the automatic variable selection feature of the LASSO estimator but also offers computational advantages. This integration makes it a robust tool for statistical modeling and analysis. In comparative studies, the Liu-LASSO estimator dominates the Elastic Net and Lasso regarding predictive accuracy and model interpretability, making it a preferable choice for this study. Liu-LASSO minimizes the following objective function:

$$\hat{\boldsymbol{\beta}}_n^{LL} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 + \frac{1}{2} \sum_{j=1}^{p} \left(\beta_j - d\hat{\beta}_j^{LS}\right)^2 + \gamma \sum_{j=1}^{p} \left|\beta_j\right| \right\}. \tag{5}$$

where $d > 0$ is a shrinkage parameter controlling the influence of the Liu estimator on the solution. $\hat{\beta}_j^{LS}$ is the least squares estimate of the j-th regression coefficient, providing a baseline for shrinkage. $\gamma > 0$ is the tuning parameter associated with the L_1-norm penalization, controlling variable selection.

2.3 Robust LAD and robust WLAD

The LAD-LASSO method addresses the high susceptibility of the Least Squares (LS) method criterion, used in Lasso, to outliers in the outcome variable/heavy-tailed error distributions (Wang et al. 2007; Wang and Leng 2007). This method substitutes the LS method criterion with the Least Absolute Deviations (LAD) criterion, resulting in the Least Absolute Deviations-Lasso (LAD-L) method. By doing so, LAD-LASSO aims to improve robustness against outliers while still incorporating variable selection through the Lasso regularization. This approach can be beneficial in scenarios where outlier robustness is a priority in regression analysis.

$$\hat{\boldsymbol{\beta}}_n^{LAD-L} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right| + \sum_{j=1}^{p}\gamma_j\left|\beta_j\right| \right\}. \tag{6}$$

The formulation of the LAD-L criterion integrates the robustness of the LAD criterion with the sparsity-inducing property of the LASSO penalty. As a result, the resulting estimator is expected to exhibit resilience against outliers while promoting a sparse representation. To streamline computation, Wang et al. (2007) present the estimator in augmented matrix form by defining $\left\{\left(y_i^*, \mathbf{x}_i^*\right)\right\}$ with $i = 1, \ldots, n+p$, where $\left(y_i^*, \mathbf{x}_i^*\right) = \left(y_i, \mathbf{x}_i\right)$ for $1 \le i \le n$, $\left(y_{n+j}^*, \mathbf{x}_{n+j}^*\right) = \left(0, n\gamma_j\mathbf{r}_j\right)$ for $1 \le j \le p$ and $\mathbf{r}_j$ denotes a p-dimensional vector with its j-th component equal to 1 and all other components equal to 0. Therefore, the estimator can be formulated in LAD form as follows:

$$\hat{\boldsymbol{\beta}}_n^{LAD-L} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n+p}\left|y_i^* - \mathbf{x}_i^{*T}\boldsymbol{\beta}\right|. \tag{7}$$

Despite its advantages, the LAD-L estimator can be sensitive to observations with unusual predictor values (Giloni et al. 2006b; Arslan 2012). In response, they proposed a Weighted LAD-L (WLAD-L) method, which introduces weights to the LAD-L framework. WLAD-L minimizes the following objective function:

$$\hat{\boldsymbol{\beta}}_n^{WLAD-L} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} w_i\left|y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right| + n\sum_{j=1}^{p}\gamma_j\left|\beta_j\right| \right\} \tag{8}$$

where w_i denotes the weights obtained from the robust distances of $\mathbf{x}_i$. Giloni et al. (2006b) defined the weights as follows:

$$w_i = \min\left(1, \frac{\chi^2_{q,0.05}}{(RD(\mathbf{x}_i))^2}\right) \tag{9}$$

where $\chi^2_{q,0.05}$ is the upper 5% critical value of a chi-squared distribution with q degrees of freedom and $RD(\mathbf{x}_i)$ is a robust version of the Mahalanobis distance that can be written by $\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}\right)^T\hat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}\right)$Croux et al. (2003). According to

Rousseeuw and Van Zomeren (1990), $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are a robust estimator of the mean vector and the covariance matrix for the covariates. Arslan (2012) defined the weight as

$$w_i = \min\left(1, \frac{p}{RD(\mathbf{x}_i)}\right). \tag{10}$$

In this study, building on the concepts put forth by Giloni et al. (2006b) and Arslan (2012), we introduce a novel estimator termed Weighted LAD-Liu_LASSO (WLAD-LL), the formulation of which is defined as

$$\hat{\boldsymbol{\beta}}^{WLAD-LL} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} w_i \left| y_i - \mathbf{x}_i^T \boldsymbol{\beta} \right| + \sum_{j=1}^{p} \gamma_j \left| \beta_j \right| + \frac{1}{2}\sum_{j=1}^{p} \left(\beta_j - d\hat{\beta}_j^{LS} \right)^2 \right\}. \tag{11}$$

This new method combines the robustness of the weighted LAD-LASSO (WLAD) with the sparsity-inducing property of the Liu-LASSO (LL), aiming to address outliers and promote a sparse representation simultaneously. The weights are derived using the robust distances of the explanatory variables in Eqs. 9 and 10. This ensures that outliers in the direction of the explanatory variables have large distances and the corresponding weights are small. Therefore, while the Liu-type penalization deals with the multicollinearity in the data, WLAD-LL is expected to produce robust estimates against the outliers in the response and leverage points.

3 ADMM for WLAD-LL

Boyd et al. (2011) introduced the Alternating Direction Method of Multipliers (ADMM), a Lagrangian-based method intended for large-scale optimization problems that can be successfully used for the WLAD-LL problem. The algorithm solves the problem

$$\begin{aligned} &\operatorname{argmin}_{\boldsymbol{\beta},\boldsymbol{\theta}}\ f(\boldsymbol{\beta}) + g(\boldsymbol{\theta}) \\ &\text{subject to } \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\theta} = \mathbf{c} \end{aligned}$$

over the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with sizes $p \times 1$ and $q \times 1$, respectively. Here, $f(\cdot)$ and $g(\cdot)$ are closed convex functions, $\mathbf{A}$ and $\mathbf{B}$ are known constraint matrices with sizes $n \times p$ and $n \times q$, respectively and $\mathbf{c}$ is an $n \times 1$ known constraint vector. The ADMM iterations consist of primal updates over $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and a dual update over $\mathbf{v}$ given by

$$\begin{aligned} \boldsymbol{\beta}^{(k+1)} &= \operatorname{argmin}_{\boldsymbol{\beta}} L_\rho\left(\boldsymbol{\beta}, \boldsymbol{\theta}^{(k)}, \mathbf{v}^{(k)}\right) \\ \boldsymbol{\theta}^{(k+1)} &= \operatorname{argmin}_{\boldsymbol{\theta}} L_\rho\left(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\theta}, \mathbf{v}^{(k)}\right) \\ \mathbf{v}^{(k+1)} &= \mathbf{v}^{(k)} + \rho\left(\mathbf{A}\boldsymbol{\beta}^{(k+1)} + \mathbf{B}\boldsymbol{\theta}^{(k+1)} - \mathbf{c}\right), \end{aligned} \tag{12}$$

where $\rho > 0$ is a fixed penalty parameter and k is the iteration number. The scaled form of the ADMM achieved by transforming the dual variable with $\mathbf{u}^k = (1/\rho)\mathbf{v}^k$,

which we prefer to use due to simpler iterations. To apply the ADMM algorithm to the WLAD-LL problem, we define the weighted-augmented design matrix and outcome, respectively, as

$$\mathbf{X}_W^* = \begin{pmatrix} \mathbf{WX} \\ n\lambda\mathbf{I} \end{pmatrix} \quad \text{and} \quad \mathbf{y}_W^* = \begin{pmatrix} \mathbf{Wy} \\ \mathbf{0} \end{pmatrix}$$

where $\mathbf{W} = \text{diag}\left(w_1, w_2, ..., w_n\right)$, $\mathbf{0}$ is a $p \times 1$ vector of zeroes and $\mathbf{I}$ is the $p \times p$ identity matrix. The WLAD-LL problem in ADMM form corresponding to Equation (11) becomes

$$\begin{aligned} &\text{argmin}_{\boldsymbol{\beta},\boldsymbol{\theta}} \ \frac{1}{n}\|\boldsymbol{\theta}\|_1 + \frac{1}{2}\left\|\boldsymbol{\beta} - d\hat{\boldsymbol{\beta}}_{LS}\right\|_2^2. \\ &\text{subject to } \mathbf{X}_W^*\boldsymbol{\beta} - \boldsymbol{\theta} = \mathbf{y}_W^* \end{aligned}$$

Therefore, the updates of the parameter vectors in the scaled form of the ADMM take the following form:

$$\begin{aligned} \boldsymbol{\beta}^{(k+1)} &= \text{argmin}_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}} \frac{1}{2}\left\|\boldsymbol{\beta} - d\hat{\boldsymbol{\beta}}_{ls}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{X}_W^*\boldsymbol{\beta} - \boldsymbol{\theta}^{(k)} - \mathbf{y}_W^* + \mathbf{u}^{(k)}\right\|_2^2 \\ \boldsymbol{\theta}^{(k+1)} &= \text{argmin}_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}} \frac{1}{n}\|\boldsymbol{\theta}\|_1 + \frac{\rho}{2}\left\|\mathbf{X}_W^*\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\theta} - \mathbf{y}_W^* + \mathbf{u}^{(k)}\right\|_2^2 \\ \mathbf{u}^{(k+1)} &= \mathbf{u}^{(k)} + \mathbf{X}_W^*\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\theta}^{(k+1)} - \mathbf{y}_W^*. \end{aligned} \tag{13}$$

After some algebraic operations, the updates in Equation (13) are obtained as

$$\begin{aligned} \boldsymbol{\beta}^{(k+1)} &= \left(\mathbf{X}_W^{*T}\mathbf{X}_W^* + \frac{1}{\rho}\mathbf{I}_p\right)^{-1}\left(\mathbf{X}_W^{*T}\left(\mathbf{y}_W^* + \boldsymbol{\theta}^{(k)} - \mathbf{u}^{(k)}\right) + \frac{d}{\rho}\hat{\boldsymbol{\beta}}_{LS}\right) \\ \boldsymbol{\theta}^{(k+1)} &= S_{1/(n\rho)}\left(\mathbf{X}_W^*\boldsymbol{\beta}^{(k+1)} - \mathbf{y}_W^* + \mathbf{u}^{(k)}\right) \\ \mathbf{u}^{(k+1)} &= \mathbf{u}^{(k)} + \mathbf{X}_W^*\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\theta}^{(k+1)} - \mathbf{y}_W^*. \end{aligned} \tag{14}$$

where $S(\cdot)$ is the soft-thresholding operator defined by Donoho and Johnstone (1994) as $S(z, \lambda) = \text{sign}(z)\max\left(|z| - \lambda, 0\right)$.

In this study, we implemented the functions in the Eigen library of C++ and called them from R functions for computational efficiency.

4 Theoretical properties

To study the theoretical properties of WLAD-LL, the regression parameter $\boldsymbol{\beta}$ is decomposed as follows: $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_c^T, \boldsymbol{\beta}_d^T\right)$ where $\boldsymbol{\beta}_c = \left(\beta_1, \beta_2, \ldots, \beta_{p_0+1}\right)^T$, and $\boldsymbol{\beta}_d = \left(\beta_{p_0+1}, \beta_{p_0+2}, \ldots, \beta_p\right)^T$. Hence, the corresponding WLAD-LL is given by $\hat{\boldsymbol{\beta}}_W = \left(\hat{\boldsymbol{\beta}}_{Wc}^T, \hat{\boldsymbol{\beta}}_{Wd}^T\right)^T$ and the objective function is denoted by $G(\boldsymbol{\beta}) = G\left(\boldsymbol{\beta}_c, \boldsymbol{\beta}_d\right)$. The predictors are decomposed as follows: $\mathbf{x}_i = \left(\mathbf{x}_{ic}^T, \mathbf{x}_{id}^T\right)$ such that $\mathbf{x}_{ic} = \left(x_{i1}, x_{i2}, \ldots, x_{p_0+1}\right)^T$, and $\mathbf{x}_{id} = \left(x_{ip_0+1}, x_{ip_0+2}, \ldots, x_{ip}\right)^T$. Define

$c_n = \max_j \{\gamma_j : 1 \le j \le p_0\}$ and $d_n = \min\{\gamma_j : p_0{+}1 \le j \le p\}$, where γ_j is a function of n, as described by Arslan (2012). Following (Arslan 2012), we adopted the following assumptions:

A1 The errors ε_i exhibit independence and identical distribution characteristics, with a median value of 0 and a continuous, positively valued density function f that remains unbroken at the origin.
A2 Let $W = \mathrm{diag}(w_1, w_2, ..., w_n)$and assume that $w_i$are known positive values that satisfy $\max(w_i) = O(1)$.
A3 The variance-covariance matrix, $\mathrm{Cov}(\mathbf{x}_i)$ exists and is positive definite.

These assumptions have been widely cited in the literature for establishing the $\sqrt{n}$ -consistency of an estimator. Key works contributing to this understanding include those by Pollard (1991); Bloomfield and Steiger (1983); Knight (1998); Wang et al. (2007); Arslan (2012). We established the following theorem based on these assumptions.

Theorem 1 *Let* $(\mathbf{x}_i, y_i)$ *for* $i = 1, 2, 3, \ldots, n$ *be independent and identically distributed, satisfying assumptions A1-A3 and* $P(\hat{\boldsymbol{\beta}}_d = 0) \to 1$. *If* $\sqrt{n}c_n \to 0$ *and* $\sqrt{n}(1 - d_n) \to \infty$ *as* $n \to \infty$, *then the WLAD-LL estimator is* $\sqrt{n}$*-consistent.*

Proof We follow the arguments of Arslan (2012) for the proof of the theorem. Specifically, considering the objective function

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ w_i \left| y_i - x_i^T \boldsymbol{\beta} \right| \right\} + n \sum_{j=1}^{p} \gamma_j \left| \beta_j \right| + \frac{n}{2} \sum_{j=1}^{p} \left(\beta_j - d_n \hat{\beta}_j^{LS} \right)^2,$$

we define

$$\begin{aligned} D_n(\mathbf{u}) = \sum_{i=1}^{n} \left\{ \left| w_i y_i - x_i^T \left(w_i \boldsymbol{\beta} + \frac{1}{\sqrt{n}} \mathbf{u} \right) \right| - w_i \left| y_i - x_i^T \boldsymbol{\beta} \right| \right\} \\ + n \sum_{j=1}^{p} \gamma_j \left\{ \left| \beta_j + \frac{1}{\sqrt{n}} u_j \right| - \left| \beta_j \right| \right\} \\ + \frac{n}{2} \sum_{j=1}^{p} \left\{ \left(\beta_j + \frac{1}{\sqrt{n}} u_j - d_n \hat{\beta}_j^{LS} \right)^2 - \left(\beta_j - d_n \hat{\beta}_j^{LS} \right)^2 \right\} \end{aligned} \tag{15}$$

Denote the first item in Equation (15) by $L_n(\mathbf{u})$. By using the identity (Knight and Fu 2000)

$$x - y = -y[I(x > 0) - I(x < 0)] + 2 \int_0^y [I(x \le s) - I(x \le 0)] ds, \quad x \ne y,$$

$L_n(\mathbf{u})$ can be written as

$$L_n(\mathbf{u}) = -n^{-1/2}\mathbf{u}^T \sum_{i=1}^{n} \mathbf{x}_i[I(\varepsilon_i > 0) - I(\varepsilon_i < 0)] + 2\sum_{i=1}^{n} \int_0^{n^{-1/2}\mathbf{u}^T\mathbf{x}_i} [I(\varepsilon_i \le s) - I(\varepsilon_i \le 0)]ds. \tag{16}$$

In Equation (16), the first item of $L_n(\mathbf{u})$ converges in distribution to $\mathbf{u}^T\mathbf{W}$, where $\mathbf{W}$ is a $p-$dimensional normal random vector with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{x}_1)$, and the second item converges to $0.5f(0)\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u}$ in probability (Wang et al. 2007). Here, the second item dominates the first item uniformly in $\|\mathbf{u}\|_2 = C$ for sufficiently large C. Also, the second item in Equation (15) converges to 0 in probability since $\sqrt{n}c_n \to 0$. Denote the third item in Equation (15) by $\ell_n(\mathbf{u})$. After algebraic operations, we can write

$$\ell_n(\mathbf{u}) = \sqrt{n}\sum_{j=1}^{p} \left(\beta_j - d_n\hat{\beta}_j\right)u_j + \frac{1}{2}\|\mathbf{u}\|_2^2.$$

Since $\hat{\boldsymbol{\beta}}^{LS} \to \boldsymbol{\beta}$ in probability and $\sqrt{n}\left(1 - d_n\right) \to 0$, $\ell_n(\mathbf{u}) \to \frac{1}{2}\|\mathbf{u}\|_2^2 \le \frac{1}{2\lambda_{\min}}\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u}$ in probability, where $\lambda_{\min}$ is the smallest eigenvalue of $\boldsymbol{\Sigma}$. Hence, the first item in Equation (15) dominates the second item and has the same order as the third item, which completes the proof of the theorem. □

To obtain the asymptotic distribution of WLAD-LL, in addition to assumptions A1, A2 and A3, we will assume that the following assumption is satisfied:

A4 $\frac{1}{n}\mathbf{X}_n^\top\mathbf{X}_n \to \boldsymbol{\Sigma}$ as $n \to \infty$ where $\boldsymbol{\Sigma}$ is a positive definite matrix.

Theorem 2 *Let* $\left(\mathbf{x}_i, y_i\right)$ *for* $i = 1, 2, 3, \ldots, n$ *be independent and identically distributed, satisfying assumptions A1 − A4. If* $\gamma_j = \lambda_j + o(1)$ *as* $n \to \infty$ *with* $\lambda_j > 0$ *for* $j = 1, 2, \ldots, p$, *then*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{WLAD-LL} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{argmin}_{\mathbf{u}} V(\mathbf{u}), \tag{17}$$

where $\boldsymbol{\beta}_0$ *is the true parameter vector*,

$$V(u) = -2\mathbf{u}^T\mathbf{W} + f(0)\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u} + \sum_{j=1}^{p} \lambda_j\left[u_j\mathrm{sign}(\beta_{0j})I(\beta_{0j} \ne 0) + |u_j|I(\beta_{0j} = 0)\right].$$

with $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}/4)$ *is a Gaussian random vector.*

Proof We follow the arguments in Knight and Fu (2000); Gao and Huang (2010) for the proof of the theorem. Without loss of generality, we assume that w_i are normalized such that $\frac{1}{n}\sum_{i=1}^{n} w_i = 1$. Define $V_n(\mathbf{u})$ by

$$V_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} w_i \left|\varepsilon_i - \mathbf{x}_i^T \frac{\mathbf{u}}{\sqrt{n}}\right| + \sum_{j=1}^{p} \gamma_j \left|\beta_{0j} + \frac{u_j}{\sqrt{n}}\right| + \frac{1}{2}\sum_{j=1}^{p}\left(\frac{u_j}{\sqrt{n}} - d\sqrt{n}\hat{u}_j\right)^2 \tag{18}$$

Using the Taylor expansion in Equation (18), we obtain

$$\begin{aligned} V_n(\mathbf{u}) =& \frac{1}{n}\sum_{i=1}^{n} w_i|\varepsilon_i| - \frac{2}{\sqrt{n}}\mathbf{u}^T\mathbf{X}_n^T\mathbf{w}_\varepsilon + \frac{1}{n}\mathbf{u}^T\mathbf{X}_n^T\mathbf{X}_n\mathbf{u} \\ &+ \sum_{j=1}^{p}\gamma_j\left[\left|\beta_{0j}\right| + \frac{u_j}{\sqrt{n}}\text{sign}(\beta_{0j})\right] \\ &+ \frac{1}{2n}\|\mathbf{u}\|_2^2 - d\mathbf{u}^T\hat{\mathbf{u}} + \frac{d^2 n}{2}\|\hat{\mathbf{u}}\|_2^2, \\ \approx& -\frac{2}{\sqrt{n}}\mathbf{u}^T\mathbf{X}_n^T\mathbf{w}_\varepsilon + \frac{1}{n}\mathbf{u}^T\mathbf{X}_n^T\mathbf{X}_n\mathbf{u} + \sum_{j=1}^{p}\gamma_j\left[\frac{u_j}{\sqrt{n}}\text{sign}(\beta_{0j}) + \frac{1}{2n}u_j^2 - du_j\hat{u}_j\right] \end{aligned}$$

where $\mathbf{w}_\varepsilon = (w_1\varepsilon_1, w_2\varepsilon_2, \ldots, w_n\varepsilon_n)^T$ are weighted errors and $\hat{\mathbf{u}} = \sqrt{n}(\hat{\boldsymbol{\beta}}^{LS} - \boldsymbol{\beta}_0)$. By the central limit theorem, the LAD term contributes $-2\mathbf{u}^T\mathbf{W} + f(0)\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u}$ since $\frac{1}{\sqrt{n}}\mathbf{X}_n^T\mathbf{w}_\varepsilon \xrightarrow{d} \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}/4)$ as $n \to \infty$. Also, the lasso penalty converges to $\sum_{j=1}^{p}\lambda_j\left[u_j\text{sign}(\beta_{0j})I(\beta_{0j} \neq 0) + |u_j|I(\beta_{0j} = 0)\right]$. This indicates that $V_n(\mathbf{u}) \xrightarrow{d} V(\mathbf{u})$. Hence, the asymptotic distribution is given by

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{WLAD-LL} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\min_{\mathbf{u}} V(u),$$

which completes proof. □

5 Simulation studies

In this section, we report on comprehensive simulation simulations to assess WLAD-LL's finite sample performance under heavy-tailed distribution errors. The compared methods with WLAD-LL are the OLS, Liu, LAD, LAD-L, LAD-LL, WLAD-L. In simulation studies, the true model is

$$y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \sigma\varepsilon_i \tag{19}$$

where ε_i is generated from the standard double-exponential distribution and the t distribution with 3 degrees of freedom. The values of σ are chosen as 0.5 and 1.5. Also, the samples sizes are determined as $n = 50$ and $n = 100$ to evaluate the effect of sample size in the simulation settings. Moreover, for every simulation run, we divided the datasets into 70% training and 30% test sets in order to assess the estimators' performance. Based on the model (19), we consider two different simulation schemes.

5.1 Simulation setting 1

The true coefficient vector is determined as

$$\boldsymbol{\beta} = \big(\underbrace{0.5, 1, \ldots, s/2}_{s}, \underbrace{0, 0, \ldots, 0}_{8-s}\big)^T$$

where s is the number of non-zero coefficients of the true coefficient vector. In this setting, the values of s are chosen as 2, 4 and 6. The features are generated by following (McDonald and Galarneau 1975) so that

$$x_{ij} = \left(1 - r^2\right)^{1/2} z_{ij} + r z_{i,p+1},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p \tag{20}$$

where z_{ij} are independent standard normal random numbers. With this approach, the correlation coefficient between any two covariates is r^2. In the simulation setting, we fix the correlations between the covariates to 0.5.

We consider the regularization parameter selection methods in Liu (1993) and Wang et al. (2007) to choose the tuning parameters of the WLAD-LL estimator. Specifically, for Liu parameter d, we select

$$\hat{d}_{C_L} = 1 - \hat{\sigma}^2 \left(\frac{\sum_{j=1}^{p} \frac{1}{(1+\lambda_j)}}{\sum_{j=1}^{p} \frac{\lambda_j \hat{\alpha}_j^2}{(1+\lambda_j)^2}} \right) \tag{21}$$

proposed by Liu (1993) based on the modified C_p statistic from Mallows (1973), where $\hat{\boldsymbol{\alpha}} = \mathbf{Q}^T \hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are the estimates $\boldsymbol{\alpha}$ and σ^2, λ_j's are the eigenvalues of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Q}$ is the matrix of eigenvectors corresponding to the eigenvalues of $\mathbf{X}^T\mathbf{X}$. Also, by following (Wang et al. 2007), we select γ_j parameters

$$\hat{\gamma}_j = \frac{\log(n)}{n\left|\tilde{\beta}_j\right|} \tag{22}$$

where $\tilde{\beta}_j$'s are the estimations from the unpenalized LAD estimator.

5.1.1 Simulation setting 2

This simulation setting investigates the prediction performance of the methods in the presence of contamination in the dataset. Specifically, the contamination rate takes the values $\varepsilon_C = 0, 0.1, 0.3, 0.5$, where $\varepsilon_C = 0$ means that there is no contamination in the data. The dataset consists of $m = \lfloor \varepsilon_C n \rfloor$ contaminated observations and $n - m$ uncontaminated observations, where $\lfloor \cdot \rfloor$ denotes the integer part. The data generation process for the first $n - m$ observations is as follows: We set $\boldsymbol{\beta}_1 = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$and generate $\mathbf{x}_{1i} \sim N_p(\mathbf{0}, \mathbf{V})$, $i = 1, 2, \ldots, n - m$ where $\mathbf{V} = \left(v_{ij}\right)$ with $v_{ij} = 0.5^{|i-j|}$. The response y_{1i} is generated with respect to the model

(19). The data generation process for the remaining m contaminated $(\mathbf{x}_2, \mathbf{y}_2)$ observations is as follows: For, $\boldsymbol{\mu} \neq \mathbf{0}$, generate $\mathbf{x}_{2i} \sim N_p(\boldsymbol{\mu}, \mathbf{1})$. Take the vector $\boldsymbol{\beta}_2$ such that $\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$. Obtain the response as $y_{2i} = \mathbf{x}_{2i}^T \boldsymbol{\beta}_2$. Finally, we combine the contaminated and uncontaminated data and fit the models.

We measure the model performance by the median of mean squared errors (TMSE) and median of mean absolute errors (TMAE) where TMSE and TMAE are calculated as TMSE $= \left\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_E\right\|_2^2 / n$ and TMAE $= \left\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_E\right\|_1 / n$ on the test set, respectively, where $\hat{\boldsymbol{\beta}}_E$ is the relevant estimator. Also, we report the accuracy and false positive rate (FPR) measures to compare the ability of estimators to select the true model. It is desirable for a predictor to have and a large accuracy and a small FPR.

The results from the simulation study, presented in Tables 1, 2, 3, 4, provide a comprehensive comparison of various estimators under different scenarios involving correlated data or contaminated data and errors following the standard double-exponential distribution and t-distribution. The estimators evaluated are OLS (ordinary least squares), Liu (Liu estimator), LAD (least absolute deviation), LAD-L (LAD - Lasso estimator), LAD-LL (LAD-Liu-Lasso), WLAD-L (Weighted LAD-Lasso), and WLAD-LL (Weighted LAD-Liu-Lasso). In this study, WLAD-LL is the proposed method. The performance metrics used for comparison are the median of mean squared errors (TMSE) and the median of mean absolute errors (TMAE). From the simulation results, it is evident that the proposed WLAD-LL estimator outperforms the other estimators in terms of TMSE and TMAE across different sample sizes, correlations, and error variances. The WLAD-LL method consistently achieves the lowest TMSE and TMAE values, indicating its robustness and efficiency in minimizing both squared and absolute errors. This superior performance can be attributed to the integration of weighted LAD with the Liu-Lasso, which combines robustness against outliers with the ability to handle multicollinearity effectively.

The OLS estimator, while simple, shows the highest TMSE and TMAE values, particularly in scenarios with higher error variances and correlations, reflecting its sensitivity to outliers and multicollinearity. The Liu estimator provides some improvement but still falls short compared to LAD-based methods. The LAD and LAD-L estimators offer better performance than OLS and Liu, highlighting the benefit of using LAD for robustness. However, the integration with Lasso (LAD-L) and Liu-Lasso (LAD-LL) further enhances performance by addressing sparsity and multicollinearity. The WLAD-L and WLAD-LL methods show the best performance overall, with WLAD-LL consistently yielding the lowest error measures. This demonstrates the effectiveness of weighting in LAD methods, especially when combined with Liu-Lasso, making WLAD-LL the preferred choice in most scenarios involving correlated or contaminated data.

The summary of the performances of these estimators is easily viewed in Figs. 1, 2, 3, 4. These figures illustrate the performance of WLAD-LL compared to other estimators, verifying its superiority in terms of TMSE and TMAE across various conditions. In conclusion, the simulation results strongly support the use of the

Table 1 Measurements of the quality of the estimators from the simulation study for the errors with the standard double-exponential distribution

n	s	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
50	2	0.5	TMSE	1.9304	1.5047	1.8763	1.3818	1.1543	0.7188	0.6768
			TMAE	1.1165	1.0019	1.1042	0.9531	0.8510	0.6604	0.6387
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.33	0.33	0.33	0.33
50	2	1.5	TMSE	7.9331	4.8739	7.3627	4.7943	6.3335	3.2039	4.0238
			TMAE	2.2584	1.7716	2.1953	1.7035	1.9618	1.3334	1.5328
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.50	0.50	0.50	0.50
100	2	0.5	TMSE	1.2522	1.1209	1.207	1.0059	1.0117	0.6695	0.6650
			TMAE	0.8806	0.8211	0.8746	0.7907	0.7912	0.6279	0.6514
			Accuracy				0.88	0.75	0.88	0.75
			FPR				0.17	0.33	0.17	0.33
100	2	1.5	TMSE	6.151	4.9576	5.7485	4.7682	5.4370	3.4098	3.5993
			TMAE	1.8922	1.6757	1.8575	1.661	1.7699	1.3872	1.4072
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.50	0.50	0.33	0.50
50	4	0.5	TMSE	11.9646	11.3054	11.9322	11.175	10.8090	6.6289	6.9795
			TMAE	2.7971	2.725	2.7441	2.7156	2.7028	2.0705	2.0526
			Accuracy				0.88	0.88	0.88	0.88
			FPR				0.25	0.25	0.25	0.25
50	4	1.5	TMSE	16.3025	12.0838	16.6371	14.0765	7.3218	8.1810	5.3220
			TMAE	3.3318	2.7953	3.3165	3.0303	2.1808	2.1866	1.759
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.50	0.50	0.50	0.50
100	4	0.5	TMSE	6.262	6.1383	6.2087	6.0365	5.9965	4.0894	4.2669
			TMAE	2.0012	1.9756	2.0211	1.9726	1.9392	1.6423	1.6515
			Accuracy				0.88	0.75	0.88	0.88
			FPR				0.25	0.50	0.25	0.25
100	4	1.5	TMSE	11.9563	10.4391	11.7959	10.5635	7.7084	6.4768	5.0313
			TMAE	2.7256	2.5109	2.7517	2.5758	2.1881	2.0031	1.7275
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.50	0.50	0.25	0.50
50	6	0.5	TMSE	45.0025	44.181	46.1528	45.0835	43.5538	25.2801	25.6245
			TMAE	5.4809	5.4313	5.4697	5.3948	5.3967	4.0234	4.0294
			Accuracy				1.00	0.88	1.00	0.88
			FPR				0.00	0.50	0.00	0.50
50	6	1.5	TMSE	50.793	44.7277	50.1482	45.4704	30.4401	26.5878	17.8614
			TMAE	5.6323	5.3676	5.6624	5.4759	4.3972	4.0244	3.3867
			Accuracy				0.88	0.88	0.88	0.88
			FPR				0.50	0.50	0.50	0.50
100	6	0.5	TMSE	25.283	24.9707	25.1365	25.1856	24.8943	15.1609	15.0482

Table 1 (continued)

n	s	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
			TMAE	4.0489	4.0274	4.0296	4.0006	4.0068	3.193	3.1565
			Accuracy				1.00	0.88	1.00	0.88
			FPR				0.00	0.50	0.00	0.50
100	6	1.5	TMSE	31.6274	29.3115	31.5614	30.5409	23.9792	18.8556	15.7116
			TMAE	4.5575	4.4053	4.5078	4.4105	4.0072	3.5264	3.1587
			Accuracy				0.88	0.88	0.88	0.88
			FPR				0.50	0.50	0.50	0.50

proposed WLAD-LL estimator for achieving $\sqrt{n}$-consistency and robustness in the presence of outliers and multicollinearity, thus providing a reliable and efficient estimation method for practical applications.

6 Real data analysis

In this section, we analyze the performance of OLS, Liu, LAD, LAD-L LAD-LL, WLAD-L and WLAD-LL estimators on two real datasets, specifically the pollution and crime datasets. Prior to the analysis of the datasets, we center the predictors, scale them using unit length scaling, and then center the responses for each dataset. We divided the data into 70% training and 30% test sets. We get the tuning parameters from the training set and calculate the TMSE and TMAE on the test set. We construct the coefficient estimates using the ADMM algorithm described in Sect. 3. We repeat this procedure 200 times and present the median of TMSE and TMAE.

6.1 Pollution data

The pollution dataset previously examined by Lukman et al. (2023a) and further investigated by Yüzbaşı et al. (2020). This dataset consists of $p = 15$ measurements on mortality rate, along with explanatory variables covering air pollution, socioeconomic factors, and meteorological conditions for $n = 60$ cities across the United States in the year 1960. The dataset is publicly available from the StatLib repository of Carnegie Mellon University (http://lib.stat.cmu.edu/datasets/). Details regarding the variables utilized in the regression analysis are provided in Table 5.

The variance inflation factors for HC and NOX are 98.64 and 104.98, respectively, indicating that the model suffers from multicollinearity. Additionally, a data point has high leverage if its hat value exceeds the cut-off point of $2(p + 1)/n$, calculated as 0.533. Observations 18, 29, 32, 48, and 49 exhibit high leverage since their hat values are greater than 0.533. For standardized residuals, a common rule of thumb is to consider values beyond ± 2 as potential outliers. The standardized residuals reveal that observations 2, 28, 37, and 59 are outliers in the response variable. Therefore, the model suffers from both multicollinearity and the presence of outliers in the response variable and covariates.

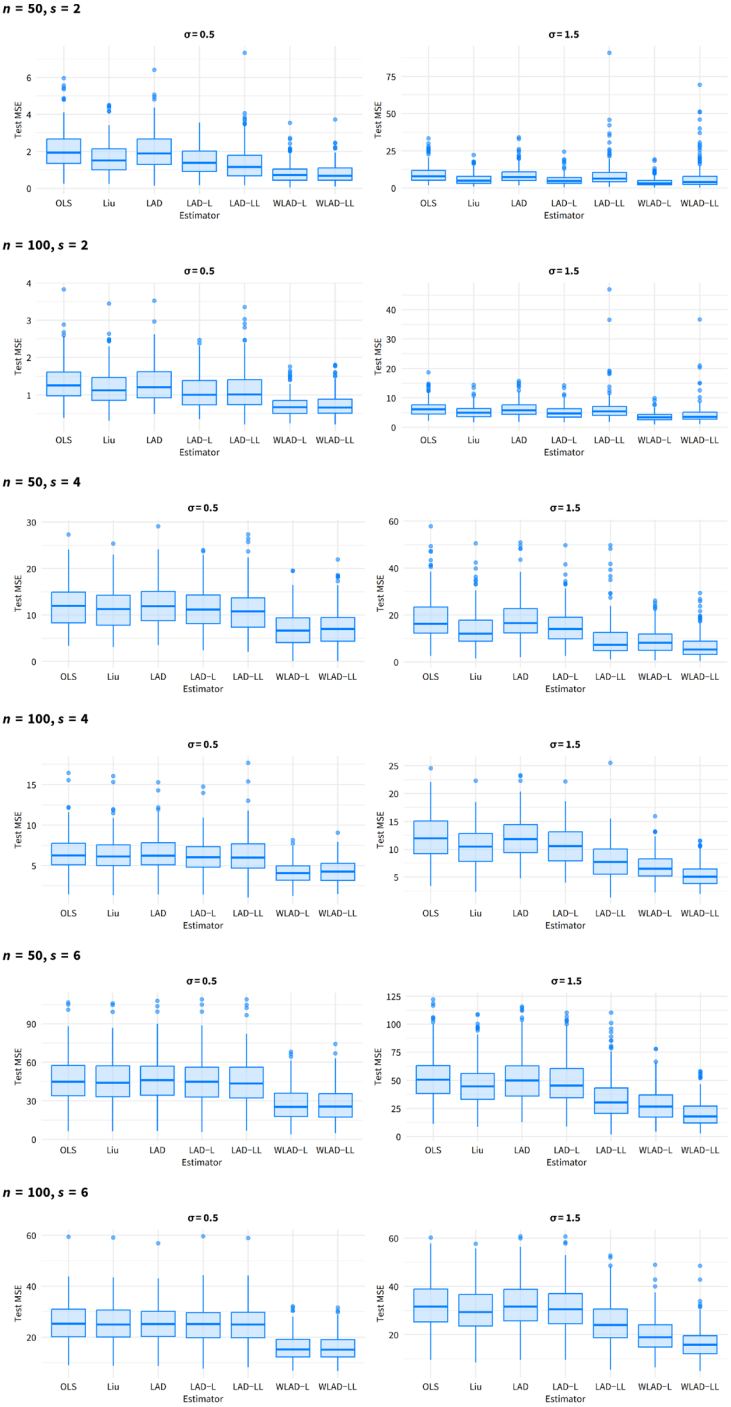


Fig. 1 Box plot of TMSE values based on the simulation setting 1 with laplace errors

Table 2 Measurements of the quality of the estimators from the simulation study for the errors with the $t(3)$ distribution

n	s	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
50	2	0.5	TMSE	1.3971	1.1574	1.267	1.0477	1.0615	0.763	0.7777
			TMAE	0.9448	0.8487	0.9069	0.8082	0.8085	0.6787	0.67
			Accuracy				0.75	0.75	0.875	0.75
			FPR				0.3333	0.3333	0.1667	0.3333
50	2	1.5	TMSE	7.648	6.0779	7.0359	5.4793	6.4461	4.0126	4.6412
			TMAE	2.1456	1.8506	2.0369	1.7918	1.9197	1.4616	1.5998
			Accuracy				0.75	0.625	0.75	0.625
			FPR				0.3333	0.5	0.3333	0.5
100	2	0.5	TMSE	2.0999	1.5355	2.036	1.3196	1.1277	0.8171	0.7928
			TMAE	1.1811	1.0048	1.1744	0.9422	0.8341	0.7111	0.6908
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.3333	0.3333	0.3333	0.3333
100	2	1.5	TMSE	10.2409	6.2	10.173	5.9496	7.7247	3.9309	4.5752
			TMAE	2.542	1.9559	2.5376	1.8998	2.1945	1.4493	1.6069
			Accuracy				0.625	0.625	0.625	0.625
			FPR				0.5	0.5	0.3333	0.5
50	4	0.5	TMSE	12.1529	11.0135	11.8408	11.0789	10.3902	5.743	5.7939
			TMAE	2.8323	2.7103	2.8089	2.7114	2.5610	1.914	1.9228
			Accuracy				0.875	0.75	0.875	0.75
			FPR				0.25	0.5	0.25	0.5
50	4	1.5	TMSE	21.6681	14.9425	20.1126	15.9846	9.4639	9.3543	6.4696
			TMAE	3.7041	3.0631	3.5383	3.2227	2.4455	2.4332	1.8485
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.5	0.5	0.5	0.5
100	4	0.5	TMSE	6.9673	6.6861	6.5753	6.4162	6.4384	4.0249	4.2905
			TMAE	2.0832	2.0496	2.0762	2.0365	2.0019	1.6393	1.6514
			Accuracy				0.875	0.75	0.875	0.75
			FPR				0.25	0.5	0.25	0.5
100	4	1.5	TMSE	13.0921	11.1261	12.3511	11.4079	8.3654	7.942	6.0835
			TMAE	2.8568	2.6108	2.7981	2.6596	2.2410	2.1927	1.8738
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.25	0.5	0.25	0.5
50	6	0.5	TMSE	45.7772	44.648	45.9906	44.4167	41.9775	24.7436	25.0302
			TMAE	5.5736	5.5207	5.5283	5.4467	5.3591	3.9682	3.9792
			Accuracy				0.875	0.875	0.875	0.875
			FPR				0.5	0.5	0	0.5
50	6	1.5	TMSE	52.4899	42.7726	50.1459	46.5844	27.7636	29.7375	19.5878
			TMAE	5.8218	5.2332	5.775	5.5772	4.2688	4.3997	3.5044
			Accuracy				0.875	0.875	0.875	0.875
			FPR				0.5	0.5	0.5	0.5
100	6	0.5	TMSE	23.7012	23.377	23.681	23.4881	22.8883	15.6448	15.1530

Table 2 (continued)

n	s	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
			TMAE	3.9228	3.8918	3.9101	3.9084	3.8722	3.1586	3.1383
			Accuracy				1	0.875	1	0.875
			FPR				0	0.5	0	0.5
100	6	1.5	TMSE	32.5775	29.3215	30.6417	30.0887	22.1661	19.0945	15.2224
			TMAE	4.5208	4.33	4.436	4.3476	3.7109	3.5482	3.066
			Accuracy				0.875	0.875	0.875	0.875
			FPR				0.5	0.5	0.5	0.5

Table 6 shows the medians of the coefficient estimates based on the replications and performance criteria of the compared estimators on the pollution dataset. Table 6 shows that the proposed method WLAD-LL gives the best TMSE and TMAE performance for the data. WLAD-LL is followed by WLAD-L. Although WLAD-L and WLAD-LL choose the same predictors, it is observed that the nonzero coefficients are estimated differently due to the effect of Liu-type regularization.

Table 7 shows the medians of the proportion of the zero-estimated coefficients based on the LAD-L, LAD-LL, WLAD-L and WLAD-LL methods. According to the table, as the case of LAD-L and LAD-LL, WLAD-L and WLAD-LL show a very similar sparsity characteristic.

As a result of the analysis, the WLAD-LL method has a similar sparsity level to WLAD-L and performs better than the other methods with smaller TMSE and TMAE for the pollution dataset.

6.2 Crime data

The crime data are from the Statistical Abstract of the United States, provided in Agresti and Finlay (2009) and the data is available in the R package `smss`Arnold (2015). The dataset consists of $p = 7$ measurements including the violent crime rates for $n = 51$ observations. The detailed expressions for the variables in the dataset is given in Table 8. Observations 9 and 25 of the dataset exhibit high leverage and the observations 1, 11, 25 and 51 are the outliers in the response variable.

Table 9 presents the median coefficient estimates based on the replications and TMSE and TMAE's of the compared estimators on the crime dataset. Table 9 indicates that WLAD-LL outperforms the other estimators in terms of performance criteria. When the TMSE and TMAE values are considered, WLAD-LL improves 47% and 28% over its closest follower WLAD-L for both measures, respectively. In these criteria, the improvement level of WLAD-LL over the other methods is much higher. In addition, the methods excluded the same variables from the model in most cases.

Table 10 shows the medians of the proportion of the zero-estimated coefficients for the underlying estimators. According to Table 10, WLAD-L and WLAD-LL have similar behavior in the sparsity of the coefficients.

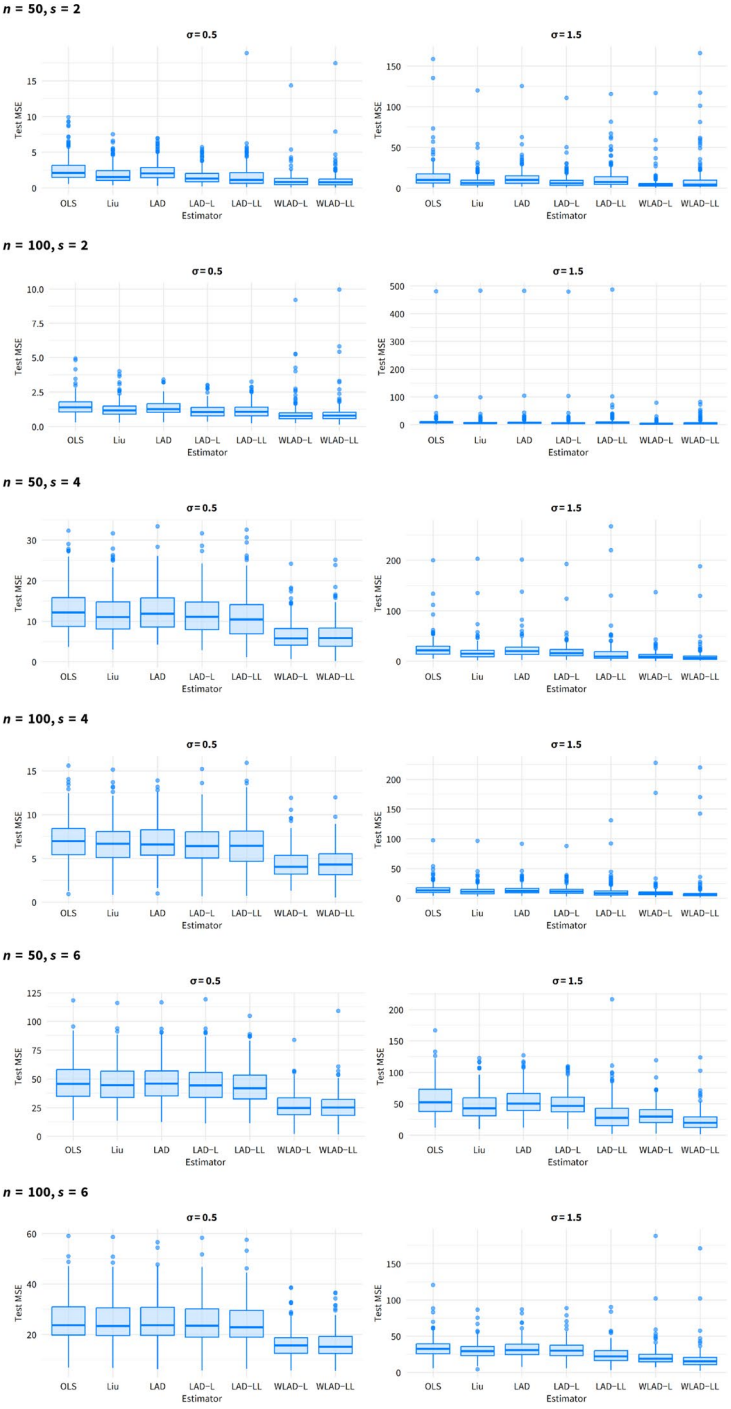


Fig. 2 Box plot of TMSE values based on the simulation setting 1 with $t(3)$ errors

Table 3 Measurements of the quality of the estimators from the simulation study for the errors with the standard double-exponential distribution

n	ε_C	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
50	0	0.5	TMSE	19.2735	17.732	18.4247	15.9841	15.5236	8.9985	8.5080
			TMAE	3.5536	3.3478	3.4552	3.2558	3.1697	2.3951	2.3425
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.40	0.40	0.40	0.40
50	0	1.5	TMSE	16.5726	14.6189	16.2902	14.6689	13.4268	8.2175	8.5662
			TMAE	3.2173	3.0478	3.2939	3.0399	2.9292	2.3318	2.3097
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.40	0.40	0.40	0.40
100	0	0.5	TMSE	10.5502	9.9795	10.3593	9.7838	9.5544	6.4633	6.4675
			TMAE	2.6316	2.5499	2.5848	2.541	2.4842	2.0343	2.0346
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.40	0.40	0.40	0.40
100	0	1.5	TMSE	10.7452	10.2738	10.6413	9.5906	9.6407	6.1724	6.2531
			TMAE	2.5915	2.5153	2.5811	2.4805	2.4445	2.0058	2.0190
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.40	0.40	0.20	0.40
50	0.1	0.5	TMSE	20.9596	17.0586	20.5745	19.0088	12.8043	9.0088	6.9245
			TMAE	3.6267	3.2941	3.6735	3.4008	2.8442	2.3329	2.0667
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.40	0.60	0.40	0.40
50	0.1	1.5	TMSE	20.8576	17.604	21.9736	20.1405	14.1211	8.8235	6.3903
			TMAE	3.7152	3.3206	3.7473	3.5571	2.9900	2.3500	1.9520
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.40	0.60	0.40	0.40
100	0.1	0.5	TMSE	12.4931	11.3043	14.1657	13.3193	10.7987	6.8165	5.6361
			TMAE	2.779	2.6417	2.9365	2.7382	2.4908	2.0209	1.8634
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.40	0.60	0.40	0.40
100	0.1	1.5	TMSE	11.733	10.5525	13.3459	12.8226	9.8816	6.6655	5.5974
			TMAE	2.6923	2.5304	2.8479	2.7555	2.4166	2.0529	1.8624
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.60	0.60	0.40	0.60
50	0.3	0.5	TMSE	21.9345	16.3192	25.7799	20.3864	13.8496	9.9801	6.1387
			TMAE	3.7771	3.2414	3.9773	3.5311	2.8966	2.5564	2.0004
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.00	0.00	0.00	0.00
50	0.3	1.5	TMSE	19.1347	14.2806	22.7153	17.9154	11.5677	10.0566	6.4720
			TMAE	3.5061	2.9814	3.8696	3.3554	2.7293	2.4840	1.9868
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.40	0.60
100	0.3	0.5	TMSE	14.0725	12.1801	16.1976	15.5946	11.0095	7.3656	5.5541

Table 3 (continued)

n	ε_C	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
			TMAE	2.8737	2.6745	3.1569	3.0014	2.5304	2.0777	1.8159
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.60	0.60
100	0.3	1.5	TMSE	13.52	11.8145	16.3312	14.6834	10.8136	8.3583	6.0090
			TMAE	2.8352	2.6305	3.1485	2.9526	2.5656	2.2272	1.8749
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.60	0.60
50	0.5	0.5	TMSE	18.9048	13.4748	21.7614	14.8811	11.6927	9.9131	6.9806
			TMAE	3.5394	3.0611	3.7735	3.2388	2.7773	2.4449	2.1002
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.40	0.60
50	0.5	1.5	TMSE	18.7046	13.9403	23.2192	16.5443	11.4875	10.3101	7.0979
			TMAE	3.599	2.9461	3.933	3.3198	2.7302	2.5372	2.0720
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
100	0.5	0.5	TMSE	12.394	10.5864	13.886	11.5923	9.9090	7.4478	6.4606
			TMAE	2.7447	2.5609	2.9355	2.7095	2.4739	2.1161	2.0150
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
100	0.5	1.5	TMSE	12.2936	10.5148	14.4069	12.3125	10.1507	7.9394	6.7848
			TMAE	2.8056	2.5552	2.9543	2.7845	2.5343	2.2181	2.0642
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60

7 Conclusion

In this study, we recommend the WLAD-LL estimator as an effective solution for parameter estimation and variable selection in linear regression models when there is multicollinearity and outliers. The WLAD-LL estimator combines the strengths of Liu-Lasso regularization and the robustness of the Weighted Least Absolute Deviations (WLAD) criterion to address these challenges effectively. The Liu-Lasso regularization technique incorporated in the WLAD-LL estimator adds an L_1 penalty to the Liu estimator, encouraging sparsity in the model coefficients and resulting in automatic variable selection by excluding irrelevant variables, thus simplifying the model and enhancing interpretability. Additionally, the penalized optimization approach allows for a balance between fitting the data well and maintaining model simplicity, further aiding the selection of the most relevant variables. To accommodate outliers, the WLAD-LL estimator employs the LAD criterion, which minimizes the sum of absolute residuals rather than the sum of squared residuals. This reduces the influence of large residuals, often caused by outliers, making the estimator less sensitive to such anomalies. Furthermore, the weighted approach of the WLAD-LL

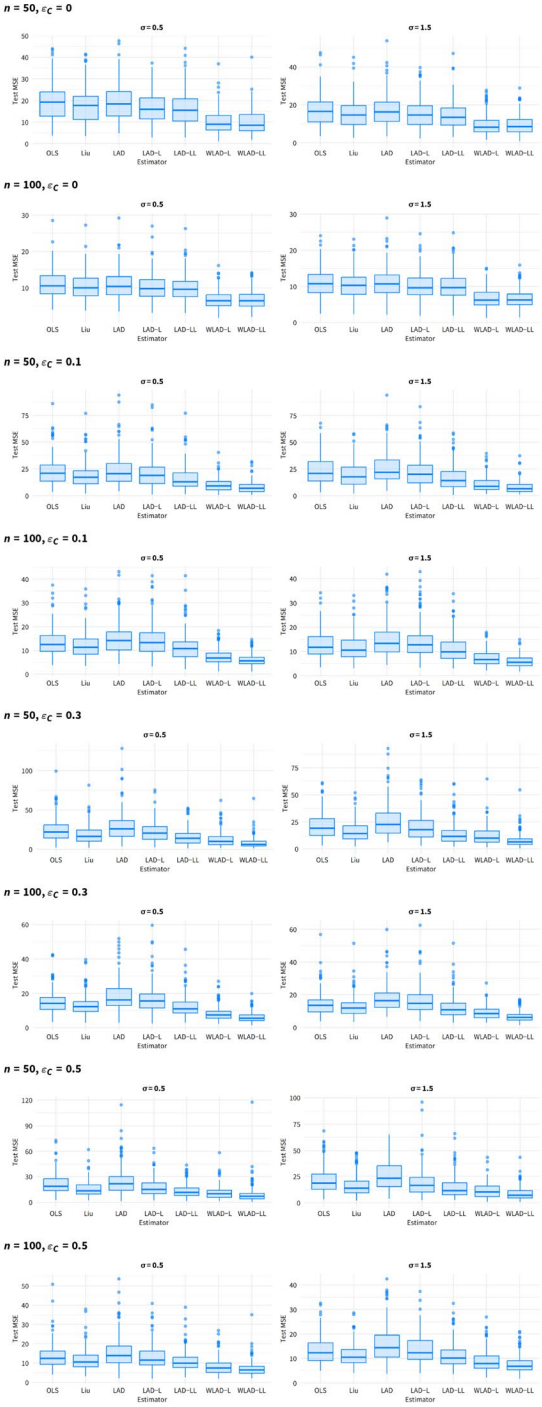


Fig. 3 Box plot of TMSE values based on the simulation setting 2 with laplace errors

Table 4 Measurements of the quality of the estimators from the simulation study for the errors with the $t(3)$ distribution

n	ε_C	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
50	0	0.5	TMSE	12.394	10.5864	13.886	11.5923	9.909	7.4478	6.4606
			TMAE	2.7447	2.5609	2.9355	2.7095	2.4739	2.1161	2.0150
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
50	0	1.5	TMSE	12.2936	10.5148	14.4069	12.3125	10.1507	7.9394	6.7848
			TMAE	2.8056	2.5552	2.9543	2.7845	2.5343	2.2181	2.0642
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
100	0	0.5	TMSE	11.4329	10.6489	11.2194	10.2576	10.3377	6.4714	6.4901
			TMAE	2.7086	2.6069	2.6662	2.5768	2.5592	2.0731	2.0488
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.40	0.40	0.40	0.40
100	0	1.5	TMSE	11.5841	10.6221	11.1359	10.0153	9.8559	6.0232	6.0477
			TMAE	2.6827	2.5629	2.6535	2.541	2.4887	1.9722	1.9745
			Accuracy				0.75	0.75	0.75	0.75
			FPR				0.40	0.40	0.40	0.40
50	0.1	0.5	TMSE	20.8007	16.7681	22.1115	18.5436	14.22	9.6923	7.423
			TMAE	3.6157	3.2317	3.8587	3.5268	2.9461	2.4634	2.1462
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.40	0.60	0.40	0.40
50	0.1	1.5	TMSE	21.6685	16.4937	23.0623	20.1125	14.1872	9.6374	6.4966
			TMAE	3.699	3.2763	3.8656	3.5827	3.0644	2.3746	2.0035
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.40	0.60	0.40	0.40
100	0.1	0.5	TMSE	13.5155	11.7954	15.2647	13.992	10.7704	7.0848	5.7064
			TMAE	2.8195	2.6581	3.0288	2.8505	2.4857	2.0683	1.8611
			Accuracy				0.75	0.63	0.75	0.75
			FPR				0.40	0.60	0.40	0.40
100	0.1	1.5	TMSE	13.2193	11.3702	14.566	14.0678	10.5079	6.8894	5.4729
			TMAE	2.7649	2.583	2.9131	2.762	2.4576	2.0593	1.8193
			Accuracy				0.75	0.69	0.75	0.75
			FPR				0.40	0.50	0.40	0.40
50	0.3	0.5	TMSE	19.6819	15.1769	23.8312	18.8174	12.0167	9.988	6.3174
			TMAE	3.6185	3.115	3.9034	3.5121	2.7006	2.3814	1.9369
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.40	0.60
50	0.3	1.5	TMSE	18.8367	13.8107	22.7981	16.6935	11.8057	11.4101	7.1504
			TMAE	3.4446	2.9534	3.8361	3.2931	2.7552	2.6098	2.0501
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.40	0.60
100	0.3	0.5	TMSE	14.4463	12.5162	16.7234	15.5602	10.8273	8.2342	6.1088

Table 4 (continued)

n	ε_C	σ	Measure	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
			TMAE	2.9016	2.7021	3.1081	3.0535	2.6163	2.2028	1.9028
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.60	0.60
100	0.3	1.5	TMSE	13.7807	11.9333	16.8006	14.6157	10.7172	7.8475	5.9151
			TMAE	2.8893	2.7074	3.1649	2.9241	2.5957	2.1742	1.9098
			Accuracy				0.63	0.63	0.75	0.63
			FPR				0.60	0.60	0.40	0.60
50	0.5	0.5	TMSE	19.6458	14.2	21.7458	16.3892	11.8768	8.9778	7.5778
			TMAE	3.5138	3.0486	3.6949	3.3031	2.7688	2.3352	2.1620
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
50	0.5	1.5	TMSE	18.6393	14.5159	21.3329	16.6131	13.0105	10.1588	7.7265
			TMAE	3.5633	3.099	3.7712	3.209	2.8574	2.4596	2.1639
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
100	0.5	0.5	TMSE	12.1788	10.8109	13.9186	11.6978	10.2281	7.9241	6.9272
			TMAE	2.81	2.5614	2.9744	2.687	2.5579	2.2141	2.0550
			Accuracy				0.63	0.50	0.63	0.63
			FPR				0.60	0.60	0.60	0.60
100	0.5	1.5	TMSE	12.6444	10.9399	14.3161	11.1053	10.2601	7.8267	6.7702
			TMAE	2.8186	2.6006	2.9385	2.6227	2.4925	2.1483	1.9782
			Accuracy				0.63	0.63	0.63	0.63
			FPR				0.60	0.60	0.60	0.60

estimator allows for the down-weighting of observations with large residuals, focusing more on the majority of the data and leading to more robust parameter estimates. Both simulation studies and real-life applications demonstrate that the WLAD-LL estimator outperforms other well-known methods, exhibiting superior TMSE (Total Mean Squared Error) and TMAE (Total Mean Absolute Error) performances. For practitioners, the WLAD-LL estimator offers a robust and efficient alternative, effectively handling variable selection and accommodating outliers. This makes it a valuable tool for achieving reliable and accurate model predictions in practical applications. The performance summary of these estimators is illustrated in Figs. 1, 2, 3, clearly demonstrating the superiority of the WLAD-LL method over others. In the broader context of theoretical properties, we note that previous works by Wang et al. (2007) and Arslan (2012) focused primarily on the consistency property of LAD-LASSO and WLAD methods. Other theoretical properties of LAD-LASSO, such as sparsity and asymptotic normality, were addressed by Gao and Huang (2010). In this study, we further advanced the literature by incorporating the asymptotic normality properties of WLAD-LL. While a detailed examination of additional properties is beyond the scope of this paper, these will be the focus of future research.

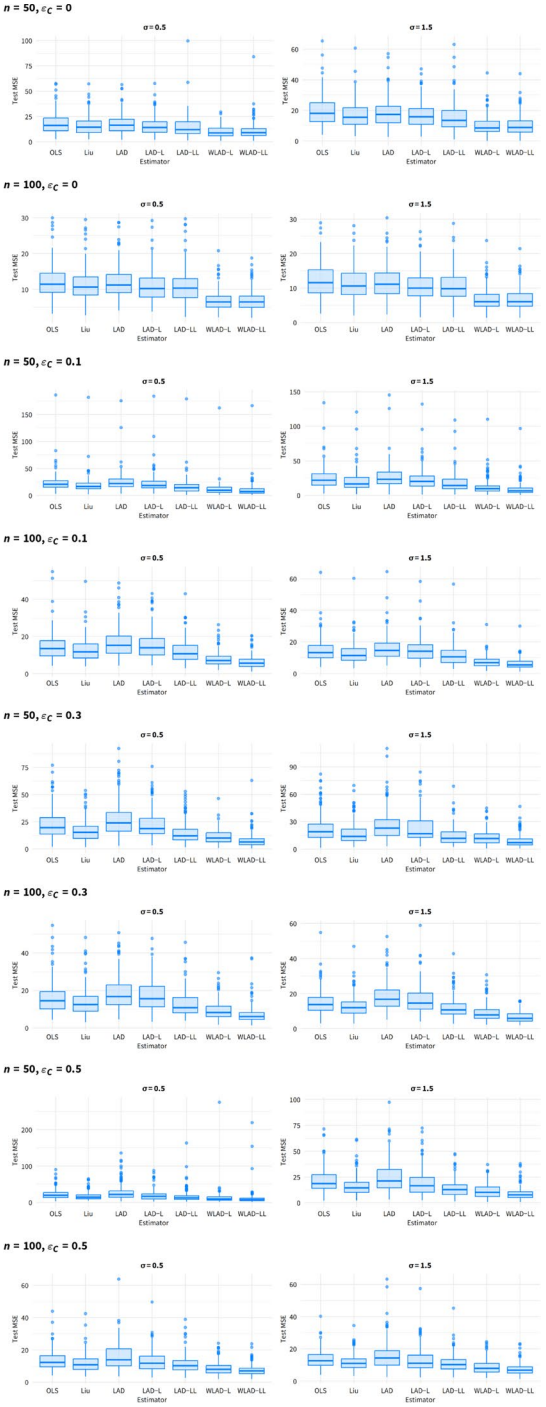


Fig. 4 Box plot of TMSE values based on the simulation setting 2 with $t(3)$ errors

Table 5 Detailed descriptions of the features and response for the pollution dataset

Variable	Description
PREC	Average annual precipitation in inches
JANT	Average January temperature in degrees F
JULT	Average July temperature in degrees F
OVR65	% of 1960 SMSA population aged 65 or older
POPN	Average household size
EDUC	Median school years completed by those over 22
HOUS	% of housing units which are sound and with all facilities
DENS	Population per sq. mile in urbanized areas, 1960
NONW	% non-white population in urbanized areas, 1960
WWDRK	% employed in white collar occupations
POOR	% of families with income< $3000
HC	Relative hydrocarbon pollution potential
NOX	Relative hydrocarbon pollution potential of nitric oxides
SO2	Relative hydrocarbon pollution potential of sulphur dioxide
HUMID	Annual average % relative humidity at 1pm
MORT	Total age-adjusted mortality rate per 100,000

Table 6 TMSE and TMAE for education and asphalt datasets

	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
PREC	134.0726	103.8788	104.9367	122.5787	78.7872	61.1956	53.0394
JANT	−126.1217	−81.6119	−58.9592	0.0000	0.0000	0.0000	0.0000
JULT	−95.0800	−54.3022	−73.6888	0.0000	0.0000	45.1127	193.8120
OVR65	−99.6735	−61.7885	−35.6708	0.0000	0.0000	0.0000	0.0000
POPN	−101.9182	−48.4997	−60.8485	0.0000	0.0000	0.0000	0.0000
EDUC	−101.3398	−76.1752	−69.4400	0.0000	0.0000	0.0000	0.0000
HOUS	−23.1841	−28.8420	−48.0963	0.0000	0.0000	51.6061	85.3495
DENS	40.7503	41.0512	80.8801	81.5524	23.2903	8.4532	4.8076
NONW	260.4439	185.8634	262.5852	191.9216	148.5203	11.7628	14.2621
WWDRK	−8.4443	−10.9229	−25.7623	0.0000	0.0000	0.0000	0.0000
POOR	−2.1962	9.0779	−37.5978	0.0000	0.0000	0.0111	15.1297
HC	−425.4288	−247.6216	−306.3231	0.0000	0.0000	0.0000	0.0000
NOX	439.5380	246.8993	319.1853	−32.5980	245.3796	6.6399	8.6916
SO2	17.4799	40.5399	42.9411	123.3520	7.6352	0.0000	0.0000
HUMID	10.9162	10.2290	−2.5172	0.0000	0.0000	158.2289	61.7143
TMSE	4911.4810	2910.6643	5078.5671	2993.3045	5518.0438	2106.7988	2092.3341
TMAE	55.2175	41.4852	55.8692	43.1131	54.8624	41.7376	41.2775

In summary, the WLAD-LL estimator provides a robust, efficient, and reliable approach for linear regression modeling, making it an indispensable tool for practitioners dealing with multicollinearity and outliers.

Table 7 Proportions of zero-estimated coefficients

	LAD-L	LAD-LL	WLAD-L	WLAD-LL
PREC	0.085	0.085	0.045	0.045
JANT	0.855	0.850	0.945	0.940
JULT	0.925	0.925	0.005	0.005
OVR65	0.730	0.720	0.660	0.655
POPN	0.880	0.880	0.570	0.560
EDUC	0.885	0.885	0.820	0.810
HOUS	0.770	0.765	0.180	0.170
DENS	0.105	0.105	0.155	0.140
NONW	0.035	0.030	0.115	0.110
WWDRK	0.730	0.730	0.605	0.585
POOR	0.655	0.655	0.320	0.315
HC	0.935	0.935	0.495	0.485
NOX	0.060	0.060	0.275	0.260
SO2	0.280	0.275	0.740	0.720
HUMID	0.535	0.530	0.125	0.120

Table 8 Detailed descriptions of the features and response for the pollution dataset

Variable	Description
Murder	Murder rate (per 100,000 people in population)
Pctmetro	The percent of the population living in metropolitan areas
Pctwhite	The percent of the population that is white
pcths	Percent of population with a high school education
Poverty	Percent of population living under poverty line
Single	Percent of population that are single parents
Crime	Violent crime rate (per 100,000 people in population)

Table 9 TMSE and TMAE for education and asphalt datasets

	OLS	Liu	LAD	LAD-L	LAD-LL	WLAD-L	WLAD-LL
Murder	1373.9472	700.9794	1119.1450	1388.7085	0.1911	216.2060	0.1247
Pctmetro	887.2395	515.3148	701.8021	784.5451	0.1522	43.1890	0.0846
Pctwhite	−45.6881	−363.3872	−394.7492	0.0000	0.0000	0.0000	0.0000
Pcths	196.2367	−17.4409	141.8037	0.0000	−0.0016	−199.1978	0.1424
Poverty	439.9220	292.8795	413.2781	372.0037	0.0895	0.0000	0.0000
Single	686.7937	637.5160	684.4423	699.3880	0.1375	294.1430	0.1174
TMSE	274427.6506	128155.0244	270343.7731	285378.4309	88500.2846	4008.1042	2122.4151
TMAE	402.2642	266.8042	405.4937	422.6960	253.8702	54.1792	39.0144

Table 10 Proportions of zero-estimated coefficients

	LAD-L	LAD-LL	WLAD-L	WLAD-LL
Murder	0.000	0.000	0.000	0.000
Pctmetro	0.000	0.000	0.365	0.360
Pctwhite	0.715	0.710	0.990	0.990
Pcths	0.305	0.305	0.035	0.035
Poverty	0.085	0.085	0.870	0.870
Single	0.020	0.020	0.000	0.000

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Agresti A, Finlay B (2009) Statistical methods for the social sciences. Pearson Prentice Hall, Upper Saddle River

Arnold Jeffrey B (2015) smss: Datasets for Agresti and Finlay's "statistical methods for the social sciences". URL https://CRAN.R-project.org/package=smss. R package version 1.0-2

Arslan O (2012) Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. Computat Stat Data Anal 56(6):1952–1965

Bloomfield P, Steiger WL (1983) Least absolute deviations: theory, applications, and algorithms, vol 6. Springer, Berlin

Stephen B, Neal P, Eric C, Borja P, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn 3(1):1–122

Croux C, Filzmoser P, Pison G, Rousseeuw PJ (2003) Fitting multiplicative models by robust alternating regressions. Stat Comput 13(1):23–36. https://doi.org/10.1023/A:1021979409012

Dielman TE (2005) Least absolute value regression: recent contributions. J Stat Comput Simul 75(4):263–286. https://doi.org/10.1080/0094965042000223680

Donoho DL, Johnstone JM (1994) Ideal spatial adaptation by wavelet shrinkage. Biometrika 81(3):425–455

Draper Norman R, Harry S (1998) Applied regression analysis. Wiley series in probability and statistics, 1st edn. Wiley, New Jersey. https://doi.org/10.1002/9781118625590

Ellis Steven P, Stephan M (1992) Leverage and Breakdown in L 1 Regression. J Am Stat Assoc 87(417):143. https://doi.org/10.2307/2290462

Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 96(456):1348–1360

Gao X, Huang J (2010) Asymptotic analysis of high-dimensional LAD regression with LASSO. Stat Sin, 1485–1506

Genç M (2022) A new double-regularized regression using Liu and lasso regularization. Comput Stat 37(1):159–227

Genç M, Özkale MR (2021) Usage of the GO estimator in high dimensional linear models. Comput Stat 36(1):217–239

Giloni A, Sengupta B, Simonoff JS (2006) A mathematical programming approach for improving the robustness of least sum of absolute deviations regression. Naval Res Logist NRL 53(4):261–271. https://doi.org/10.1002/nav.20139

Giloni A, Simonoff JS, Sengupta B (2006) Robust weighted LAD regression. Comput Stat Data Anal 50(11):3124–3140. https://doi.org/10.1016/j.csda.2005.06.005

Hawkins DM (1993) The feasible set algorithm for least median of squares regression. Comput Stat Data Anal 16(1):81–101. https://doi.org/10.1016/0167-9473(93)90246-P

Hocking RR (1976) A biometrics invited paper. The analysis and selection of variables in linear regression. Biometrics 32(1):1. https://doi.org/10.2307/2529336

Hubert M, Rousseeuw PJ (1997) Robust regression with both continuous and binary regressors. J Stat Plan Inference 57(1):153–163. https://doi.org/10.1016/S0378-3758(96)00041-9

Jiang Y, Wang Y, Zhang J, Xie B, Liao J, Liao W (2021) Outlier detection and robust variable selection via the penalized weighted LAD-LASSO method. J Appl Stat 48(2):234–246. https://doi.org/10.1080/02664763.2020.1722079

Knight K (1998) Limiting distributions for L1 regression estimators under general conditions. Ann Stat, 755–770

Knight K, Fu W (2000) Asymptotics for lasso-type estimators. Ann Stat 28(5):1356–1378

Liu K (1993) A new class of blased estimate in linear regression. Commun Stat-Theory Methods 22(2):393–402

Lukman AF, Farghali RA, Golam KBM, Oluyemi OA (2023) Robust-stein estimator for overcoming outliers and multicollinearity. Sci Rep 13(1):9066

Lukman AF, Allohibi J, Jegede SL, Adewuyi ET, Oke S, Alharbi AA (2023) Kibria-Lukman-type estimator for regularization and variable selection with application to cancer data. Mathematics 11(23):4795

Mallows CL (1973) Some comments on Cp. Technometrics 15(4):661–675

Maronna RA, Douglas MR, Yohai VJ (2006) Robust statistics: theory and methods. Wiley series in probability and statistics, 1st edn. Wiley, New Jersey. https://doi.org/10.1002/0470010940

McDonald GC, Galarneau DI (1975) A Monte Carlo evaluation of some ridge-type estimators. J Am Stat Assoc 70(350):407–416

Pollard D (1991) Asymptotics for least absolute deviation regression estimators. Econom Theory 7(2):186–199

Rousseeuw PJ (1984) Least Median of Squares Regression. J Am Stat Assoc 79(388):871–880. https://doi.org/10.1080/01621459.1984.10477105

Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley series in probability and statistics, 1st edn. Wiley, New Jersey. https://doi.org/10.1002/0471725382

Rousseeuw PJ, Van Zomeren BC (1990) Unmasking multivariate outliers and leverage points. J Am Stat Assoc 85(411):633–639. https://doi.org/10.1080/01621459.1990.10474920

Simon N, Friedman J, Hastie T, Tibshirani R (2013) A Sparse-Group Lasso. J Comput Graph Stat 22(2):231–245. https://doi.org/10.1080/10618600.2012.681250

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol 58(1):267–288

Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K (2005) Sparsity and smoothness via the fused Lasso. J R Stat Soc Ser B Stat Methodol 67(1):91–108. https://doi.org/10.1111/j.1467-9868.2005.00490.x

Wang H, Leng C (2007) Unified LASSO estimation by least squares approximation. J Am Stat Assoc 102(479):1039–1048

Wang H, Li G, Jiang G (2007) Robust regression shrinkage and consistent variable selection through the LAD-Lasso. J Bus Econ Stat 25(3):347–355

Xu J, Ying Z (2010) Simultaneous estimation and variable selection in median regression using Lasso-type penalty. Ann Inst Stat Math 62(3):487–514. https://doi.org/10.1007/s10463-008-0184-2

Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B Stat Methodol 68(1):49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x

Yüzbaşı B, Arashi M, Ahmed SE (2020) Shrinkage estimation strategies in generalised ridge regression models: Low/high-dimension regime. Int Stat Rev 88(1):229–251

Zou H (2006) The adaptive lasso and its oracle properties. J Am Stat Assoc 101(476):1418–1429. https://doi.org/10.1198/016214506000000735

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 67(2):301–320

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.