

Article

Separation of the Linear and Nonlinear Covariates in the Sparse Semi-Parametric Regression Model in the Presence of Outliers[†]

Morteza Amini¹, Mahdi Roozbeh^{2,*} and Nur Anisah Mohamed^{3,*}

¹ Department of Statistics, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran P.O. Box 14155-6455, Iran; morteza.amini@ut.ac.ir

² Department of Statistics, Faculty of Mathematics, Statistics and Computer Sciences, Semnan University, Semnan P.O. Box 35195-363, Iran

³ Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, Kuala Lumpur 50603, Malaysia

* Correspondence: mahdi.roozbeh@semnan.ac.ir (M.R.); nuranisah_mohamed@um.edu.my (N.A.M.)

[†] This paper is an extended version of our paper published in the Proceedings of the 16th Iranian Statistics Conference, Tehran, Iran, 24–26 August 2022; University of Mazandaran: Babolsar, Mazandaran, Iran, 2022; pp. 55–61.

Abstract: Determining the predictor variables that have a non-linear effect as well as those that have a linear effect on the response variable is crucial in additive semi-parametric models. This issue has been extensively investigated by many researchers in the area of semi-parametric linear additive models, and various separation methods are proposed by the authors. A popular issue that might affect both estimation and separation results is the existence of outliers among the observations. In order to address this lack of sensitivity towards extreme observations, robust estimating approaches are frequently applied. We propose a robust method for simultaneously identifying the linear and nonlinear components of a semi-parametric linear additive model, even in the presence of outliers in the observations. Additionally, this model is sparse in that it may be used to determine which explanatory variables are ineffective by giving accurate zero estimates for their coefficients. To assess the effectiveness of the proposed method, a comprehensive Monte Carlo simulation study is conducted along with an application to investigate the dataset, which includes Boston property prices dataset.

Keywords: adaptive LASSO; group LASSO; outlier; penalized approaches; robust methods

MSC: 62G05; 62J07; 62J05



Citation: Amini, M.; Roozbeh, M.; Mohamed, N.A. Separation of the Linear and Nonlinear Covariates in the Sparse Semi-Parametric Regression Model in the Presence of Outliers. *Mathematics* **2024**, *12*, 172. <https://doi.org/10.3390/math12020172>

Academic Editors: Babak Shiri and Zahra Alijani

Received: 21 August 2023

Revised: 14 October 2023

Accepted: 24 October 2023

Published: 5 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Semi-parametric linear additive (SLA) models have both the flexibility of non-parametric regression models as well as the simplicity of linear regression models. These applicable models are broadly used as a popular mechanism for data analysis in many fields. In SLA models, an acceptable relationship of the mean response variable is assumed to connect with some explanatory variables linearly, while it relates to other explanatory variables non-linearly in an additive form.

Suppose $\mathbf{y} = (y_1, \dots, y_n)^T$ is the vector of the response variable and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ is the $n \times p$ design matrix with p covariates and n observations $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$. Without loss of generality, assume that \mathbf{x}_i is partitioned into $\mathbf{x}_{i(1)}^T = (x_{i1}, x_{i2}, \dots, x_{iq})$ and $\mathbf{x}_{i(2)}^T = (x_{i(q+1)}, \dots, x_{ip})$ for some $q \in \{1, \dots, p-1\}$. Then, the semi-parametric linear additive model (see, e.g., [1]) is defined as

$$y_i = \mathbf{x}_{i(1)}^T \boldsymbol{\beta} + \sum_{j=q+1}^p f_j(x_{ij}) + \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_q)^T$ is a q -dimensional vector of unknown parameters, f_{q+1}, \dots, f_p are unknown smooth functions, and ϵ_i 's are random error terms, which are presumed to be independent of (x_i) . It is assumed that the response and the covariates are centered, and thus the intercept term is omitted without loss of generality.

There are several approaches for the estimation of non-parametric additive models, including the back-fitting technique (see [2]), simultaneous estimation and optimization [3–6], mixed model approach [1,7,8], and Boosting approach [9,10]. Ref. [5] has presented a review of some of these methods, up to 2006, and [11] has performed several comparisons between these techniques. The problem of variable selection and penalized estimation in additive models has been investigated by many researchers [12–22].

An essential concern in practice is to identify the linear and nonlinear parts of the SLA model, i.e., whether the explanatory variables can be considered as the linear or nonlinear parts of the model. Ref. [23] studied an additive regression model as the standard model by assuming that each of the functions is decomposed into linear and nonlinear parts. Their proposed approach of estimation was a penalized regression scheme based on a group mini-max concave penalty. Ref. [24] surveyed the additive model and tried to isolate the linear and nonlinear predictors by using two group penalty functions, one for enforcing the sparsity and the other one for enforcing linearity to the components. Ref. [25] introduced a similar model to that of [23], while they imposed the LASSO and group LASSO penalty functions to the coefficients of the linear parts and the coefficients of the spline estimator of the nonlinear part, respectively. Ref. [26] introduced a similar additive model and they enforced linearity to the spline approximation of the functions using the group penalty function of the second derivative of the B-splines. There are also more contributions on the problem of structure recognition and separation of nonlinear and linear parts of the SLA model [27,28]. Some details of the literature review for the proposed separation approaches are considered in Section 2.

The presence of outliers, which are unusual observations that fail to follow the scheme of the bulk of the observations, is a frequent problem in the model fitting of datasets. In such situations, robust regression approaches are used to solve the undesirable effects of the outliers. Some of the most popular robust regression approaches are M-estimation, S-estimation, the least median of squares, and the least trimmed squares; see [29] for more details. Robust methods are well-known statistical techniques to overcome the complication of outliers. The least trimmed squares (LTS), suggested by Rousseeuw and Leroy [30], is one of the most popular robust regression techniques, as it minimizes the sum of h smallest squared residuals instead of the whole sum of them, for a specified positive integer trimming parameter $h \leq n$. The LTS estimator is efficient in reaching the maximum possible breakdown point (50%) [31]. There are several works that have studied robust estimations for the semi-parametric and non-parametric linear models (see, e.g., [32–34]).

In this paper, we consider the effect of outliers on simultaneous separation and estimation methods in SLA models, and we survey the LTS version of the methods by introducing the LTS version of the separation and sparse estimation approach suggested by [24]. The paper is organized as follows. Section 2 presents a literature review of some simultaneous separation and estimation approaches. Section 3 contains the general LTS version of the approaches presented in Section 2, and then we try to apply the LTS version of the proposed method by [24] in our implementation. Then, the finite sample breakdown point of the proposed model is established with the introduction of a computational algorithm. The comprehensive simulation studies are conducted in Section 4, in which many different criteria are evaluated in six different competitive models. The proposed approach is then applied in the Boston housing prices dataset, along with the prediction achievement of different methods. At the same time, we try to reveal the effect of the outliers using the different partial residual plots of all competitive schemes.

2. Literature Review of the Separation Methods

In this section, we review the available penalized models that separate the nonlinear and linear parts of the semi-parametric regression model in the literature.

2.1. Group Penalization of the Spline Coefficients

Ref. [23] studied the following additive regression model as the guideline model:

$$y = \sum_{j=1}^p f_j(x_j) + \epsilon, \tag{2}$$

and they assumed that each of the functions f_j has a linear and a nonlinear part as follows:

$$f_j(x) = \beta_0 + \beta_j x + \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x),$$

where B_{j1}, \dots, B_{jK_n} are basis functions. They suggested estimation of the model parameters by minimizing the following penalized objective function:

$$L(\beta, \theta) = \frac{1}{2n} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} - \sum_{j=1}^p \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x_{ij}) \right)^2 + \sum_{j=1}^p \rho_\gamma(\|\theta_{jn}\|_{A_j}; \sqrt{K}\lambda), \tag{3}$$

where $\rho(\cdot)$ is a penalty function depending on the penalty parameter $\lambda \geq 0$ and a regularization parameter γ . In objective function (3), Lian et al. [23] considered the mini-max concave penalty function for their model, as follows:

$$\rho_\gamma(t; \lambda) = \lambda \int_0^t (1 - x/(\gamma\lambda))_+ dx, \quad t \geq 0.$$

Under some conditions, they proved the consistency of the proposed estimators and studied the correctness of the separation performed by their estimators.

2.2. Affine Group Penalization of the Spline Coefficients

Ref. [24] considered model (2), while they relaxed the assumption that each of the functions f_j has a linear and a nonlinear part; instead, they made an effort to separate the linear and nonlinear covariates by the following affine group penalized model:

$$L(\theta) = \frac{1}{2} \sum_{i=1}^n \left(y_i - \theta_0 - \sum_{j=1}^p \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x_{ij}) \right)^2 + n\lambda_1 \sum_{j=1}^p w_{1j} \|\theta_j\|_{A_j} + n\lambda_2 \sum_{j=1}^p w_{2j} \|\theta_j\|_{D_j}, \tag{4}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are penalty parameters, w_{1j} s and w_{2j} s are the proper weights, which are appropriately chosen in order to reach a suitable consistency in model selection, and for any $K_n \times K_n$ matrix B , $\|\theta_j\|_B = (\theta_j^T B \theta_j)^{1/2}$. Ref. [24] suggest the use of $w_{1j} = 1/\|b_j\|_{A_j}$ and $w_{2j} = 1/\|b_j\|_{B_j}$, for an initial estimate b_j of θ_j , $j = 1, \dots, p$. To enforce sparsity and linearity to the functions f_1, \dots, f_p , they assumed that $\|\theta_j\|_{A_j} = 0$, if and only if $\sum_k \theta_{jk} B_{jk}(x) \equiv 0$ and $\|\theta_j\|_{D_j} = 0$, if and only if $\sum_k \theta_{jk} B_{jk}(x)$ is a linear function of x . Letting

$$A_j = \left\{ \int_0^1 B_{jk}(x) B_{jk'}(x) dx \right\}_{k,k'=1}^{K_n} \quad \text{and} \quad D_j = \left\{ \int_0^1 B_{jk}''(x) B_{jk'}''(x) dx \right\}_{k,k'=1}^{K_n}$$

results in $\|\theta_j\|_{A_j} = \|\sum_k \theta_{jk} B_{jk}\|$ and $\|\theta_j\|_{D_j} = \|\sum_k \theta_{jk} B_{jk}''\|$. They proved that the proposed estimators are asymptotically normal.

2.3. LASSO and Group LASSO Penalization of the Linear and Nonlinear Coefficients

Ref. [25] studied the following scheme as a baseline model:

$$y = Z\alpha + \sum_{j=1}^p f_j(x_j) + \epsilon, \tag{5}$$

in which they assumed that there are some covariates Z that we need to consider in the linear part. Similar to [23], they assumed that

$$f_j(x) = \beta_0 + \beta_j x + \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x)$$

Then, they enforced the LASSO and group LASSO penalty functions to the coefficients of the linear parts and the coefficients of the spline estimator of the nonlinear part, respectively, as follows:

$$L(\beta, \theta) = \sum_{i=1}^n \left(y_i - \beta_0 - Z_i \alpha - \sum_{j=1}^p \beta_j x_{ij} - \sum_{j=1}^p \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x_{ij}) \right)^2 + \lambda_1 \|\alpha\|_1 + \lambda_2 \|\beta\|_1 + \lambda_3 \sum_{j=1}^p \|\theta_j\| \tag{6}$$

2.4. Group Penalization of the Second Derivative of the B-Splines

Ref. [26] considered (2) as their baseline model and imposed linearity to the spline approximation of the functions f_1, \dots, f_p by minimization of the following penalized objective function, which uses the group penalty function of the second derivative of the B-splines:

$$L(\theta) = \sum_{i=1}^n \sum_{\ell=1}^q \rho_{\tau_\ell} \left(y_i - \theta_{0\ell} - \sum_{j=1}^p \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x_{ij}) \right) + nq \sum_{j=1}^p P_\lambda(\|B_j''\|), \tag{7}$$

where $\rho_{\tau_\ell}(\cdot)$ is the quantile regression loss function.

3. Robust Penalized Estimation Methods

All of the penalized loss functions (3), (4), (6), and (7) can be written in the following general form:

$$L_m(\eta) = \sum_{i=1}^n L_{im}(y_i, \eta) + n \sum_{j=1}^p P_{jm}(\eta_j), \quad m = 1, \dots, 4, \tag{8}$$

where L_{im} is the loss function of the i th observation, and P_{jm} is the penalty function of the j th parameter, $i = 1, \dots, n, j = 1, \dots, p$ in the m th model, $m = 1, \dots, 4$.

The least trimmed squares (see [35]) penalized loss function associated with the m th model is then as follows:

$$Q_m(\mathbf{u}, \eta) = \sum_{i=1}^n u_i L_{im}(y_i, \eta) + h \sum_{j=1}^p P_{jm}(\eta_j), \quad m = 1, \dots, 4, \tag{9}$$

where u_i is the binary indicator clarifying whether the i th observation is a normal observation or is an outlier point, such that $\sum_{i=1}^n u_i = h, u_i \in \{0, 1\}$, for $i = 1, \dots, n$, and $h \leq n$ is a starting conjecture for the number of normal observations. Let \mathbf{U} be the diagonal matrix with diagonal elements $\mathbf{u} = (u_1, u_2, \dots, u_n)^\top$.

The resulting robust sparse semi-parametric linear estimator is obtained by the following optimization problem:

$$\begin{aligned} & \min_{\eta, \mathbf{u}} Q_m(\mathbf{u}, \eta) \\ & \text{s.t.} \quad \sum_{i=1}^n u_i = h, \\ & \quad \quad u_i \in \{0, 1\}. \end{aligned} \tag{10}$$

In this work, we only consider the robust version of penalized loss function (4):

$$Q(\mathbf{u}, \theta) = \frac{1}{2} \sum_{i=1}^n u_i \left(y_i - \theta_0 - \sum_{j=1}^p \sum_{k=1}^{K_n} \theta_{jk} B_{jk}(x_{ij}) \right)^2 + h\lambda_1 \sum_{j=1}^p w_{1j} \|\theta_j\|_{A_j} + h\lambda_2 \sum_{j=1}^p w_{2j} \|\theta_j\|_{D_j} \tag{11}$$

Hereafter, we name scheme (4) sparse semi-parametric linear additive (SSLA) and scheme (11) robust sparse semi-parametric linear additive (RSSLA). We also name the special case $\lambda_2 = 0$ of schemes (4) and (11) sparse nonlinear additive (SNLA) and robust sparse nonlinear additive (RSNLA), respectively, because by letting $\lambda_2 = 0$, the schemes change into nonlinear forms. As an alternative competitor for these schemes, the simple linear LASSO regression is also considered, which is called sparse linear (SL), and its robust version based on the LTS method is called robust sparse linear (RSL) in this research.

3.1. The Breakdown Point of the RSSLA Model

The RSSLA estimator is obtained as

$$\hat{\theta}_{RSSLA} = \underset{\theta}{\operatorname{argmin}} \underset{\mathbf{u} \in E_h}{\operatorname{argmin}} Q(\mathbf{u}, \theta),$$

where

$$E_h = \{ \mathbf{u}; u_i \in \{0, 1\}, i = 1, 2, \dots, n, \sum_{i=1}^n u_i = h \}.$$

Conventionally, we consider $h = \lceil n(1 - \alpha) \rceil$, where $\lceil [a] \rceil$ denotes the ceiling of a and $\alpha \in (0, 1)$ is the percent of leverage observed points. Indeed $1 - \alpha$ is a starting guess for the percent of outlier points. Some researchers propose considering $\alpha = 0.75$ (see [36] for more details). Others have proposed considering $h = \lfloor n/2 \rfloor + \lfloor (p + 1)/2 \rfloor$. The finite-sample breakdown point (FBP; see, e.g., [29]) is a size or rate for the consistency of a method. For the complete sample \mathbf{Z} , the FBP of an estimator $S = S(\mathbf{Z})$ is given by

$$\text{BP}(S; \mathbf{Z}) = \min_m \left\{ \frac{m}{n} : \sup_{\mathbf{Z}^*} \|S(\mathbf{Z}^*)\|^2 = \infty \right\},$$

where \mathbf{Z}^* is a corrupted sample obtained from \mathbf{Z} by replacing m of the complete n observations by random samples. In the following theorem, the FBP of LTS-SPSRE is established.

Theorem 1. *The FBP of $\hat{\theta}_{RSSLA}$ estimator is*

$$\text{FBP}(\hat{\theta}_{RSSLA}; \mathbf{y}, \mathbf{X}) = \frac{n - h - 1}{n}. \tag{12}$$

Proof. Let $(\mathbf{y}^*, \mathbf{X}^*)$ be the corrupted sample by replacing the last $m \leq n - h$ sample points. Then the number of normal points in $(\mathbf{y}^*, \mathbf{X}^*)$ is $n - m \geq h$. For an arbitrary sample $(\mathbf{y}^*, \mathbf{X}^*)$, we can write

$$\min_{\mathbf{u} \in E_h} Q(\mathbf{u}, 0) = \min_{\mathbf{u} \in E_h} \mathbf{y}^{*\top} \mathbf{U} \mathbf{y}^* \leq \min_{\mathbf{u} \in E_h} \mathbf{y}^\top \mathbf{U} \mathbf{y} \leq h M_y^2,$$

where $M_y = \max_{i=1, \dots, n} |y_i|$.

Let θ be such that $h\lambda_1 \sum_{j=1}^p w_{1j} \|\theta_j\|_{A_j} + h\lambda_2 \sum_{j=1}^p w_{2j} \|\theta_j\|_{D_j} \geq hM_y^2 + 1$; then

$$\begin{aligned} \min_{u \in E_h} Q(u, \theta) &\geq h\lambda_1 \sum_{j=1}^p w_{1j} \|\theta_j\|_{A_j} + h\lambda_2 \sum_{j=1}^p w_{2j} \|\theta_j\|_{D_j} \\ &\geq hM_y^2 + 1 \\ &> \min_{u \in E_h} Q(u, \mathbf{0}). \end{aligned}$$

Since $\min_{u \in E_h} Q(u, \theta) \leq \min_{u \in E_h} Q(u, \mathbf{0})$, we can write

$$h\lambda_1 \sum_{j=1}^p w_{1j} \|\theta_j\|_{A_j} + h\lambda_2 \sum_{j=1}^p w_{2j} \|\theta_j\|_{D_j} \Big|_{\theta = \hat{\theta}_{\text{RSSLA}}} \leq hM_y^2 + 1,$$

and hence $\text{BP}(\hat{\theta}_{\text{RSSLA}}; \mathbf{y}, \mathbf{X}) \geq \frac{n-h-1}{n}$.

Let Φ be the $n \times (pK_n)$ matrix of $B_{jk}(x_{ij})$ s, $i = 1, \dots, n$, $j = 1, \dots, p$, $k = 1, \dots, K_n$. Change the last $m = n - h + 1$ observations of (\mathbf{y}, \mathbf{X}) such that the last m observations of (\mathbf{y}, Φ) are changed to $(aM, a\mathbf{e})$, with $M > 0$ and $a > 0$, $\mathbf{e} = (e_{j_1}, \dots, e_{j_p})^T$, in which e_i is a vector with 1 as its i th elements and zero elsewhere, and

$$\max(h - m, 0) \left(\max_{i=1, \dots, n} |y_i| + M \max_{i=1, \dots, n} \|\Phi_i\| \right)^2 + h\lambda_1 M/p \sum_{i=1}^p w_{1i} \|e_{j_i}\|_{A_i} + hM/p\lambda_2 \sum_{i=1}^p w_{2i} \|e_{j_i}\|_{D_i} \leq a^2.$$

Let $P(\theta) = h\lambda_1 \sum_{j=1}^p w_{1j} \|\theta_j\|_{A_j} + h\lambda_2 \sum_{j=1}^p w_{2j} \|\theta_j\|_{D_j}$ and consider the point $\theta_M = a(M/p)\mathbf{e}$. Now, for the last m sample points, according to $(\mathbf{y} - \Phi\theta_M) = aM - aM = 0$, it can be written that

$$\min_{u \in E_h} Q(u, \theta_M) = \begin{cases} \min_{u \in E_{h-m}} (\mathbf{y} - \Phi\theta_M)^\top \mathbf{U} (\mathbf{y} - \Phi\theta_M) + P(\theta_M), & h > M \\ P(\theta_M), & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} \min_{u \in E_h} Q(u, \theta_M) &\leq \max(h - m, 0) \left(\max_{i=1, \dots, n} |y_i| + M \max_{i=1, \dots, n} \|\Phi_i\| \right)^2 + P(\theta_M) \\ &\leq a^2. \end{aligned} \tag{13}$$

Also, for the corrupted sample, we can write

$$\min_{u \in E_h} Q(u, \theta) \geq (Ma - a\theta^\top \mathbf{e})^2,$$

in which at least one of the last m points of the corrupted sample is in the set of the least possible h residuals. Now, considering θ such that $|\theta^\top \mathbf{1}| \leq M - 2$, it can be seen that

$$\min_{u \in E_h} Q(u, \theta) \geq a^2(M - \theta^\top \mathbf{e})^2 > a^2, \tag{14}$$

since $\theta^\top \mathbf{e} \leq |\theta^\top \mathbf{1}| \leq M - 2$, which is a contradiction. Thus, we deduce that

$$|\hat{\theta}_{\text{RSSLA}}^\top \mathbf{1}| > M - 2,$$

which means that FBP occurs as M tends to infinity, i.e., $\text{BP}(\hat{\theta}_{\text{RSSLA}}; \mathbf{y}, \mathbf{X}) \leq \frac{n-h-1}{n}$, and the proof completed. \square

3.2. Computational Penalized LTS Algorithm

To find \mathbf{u}^* , we have to look for the minimum of the set E_h overall $\binom{n}{h}$ combinations of the complete set $\{1, \dots, n\}$. Thus, for somewhat large values of sample size, achieving the

optimal value may need too much time and space. To extend the procedure of obtaining the RSSLA model, an analog of the FAST-LTS algorithm developed by [35] is proposed.

Let $\mathbf{u}_k \in E_h$ be the indicator vector obtained at iteration k and $\hat{\theta}_{\text{RSSLA}}^{(k)}$ be the obtained argument that minimizes $Q(\mathbf{u}_k, \theta)$ in the k th iteration. Then,

$$u_{ik+1} = \begin{cases} 1, & e_{ik}^2 \in \{(e_k)_{j:n}^2; j = 1, \dots, h\} \\ 0, & \text{otherwise,} \end{cases}$$

where $(e_k)_{1:n}^2 \leq \dots \leq (e_k)_{n:n}^2$ are the sorted sample of the squared residuals.

It is obvious that

$$Q(\mathbf{u}_{k+1}, \hat{\theta}_{\text{RSSLA}}^{(k+1)}) \leq Q(\mathbf{u}_k, \hat{\theta}_{\text{RSSLA}}^{(k)}),$$

and the algorithm continues until convergence.

To guarantee that the updated solution of the algorithm is as close as possible to the optimal solution of $Q(\mathbf{u}, \hat{\theta})$, the steps of the algorithm are replicated s times with s beginning indicator vectors $\mathbf{u}_0^1, \dots, \mathbf{u}_0^s$. To decrease the computational cost of the algorithmic program, the methodology proposed by [35] is applied, in which only two iterations of the algorithm for each iteration are performed, obtaining $\mathbf{u}_2^1, \dots, \mathbf{u}_2^s$, keeping a small number, k , of them with the lowest values of $Q(\mathbf{u}, \hat{\theta})$, and the algorithm is continued until convergence occurs. The latest result is the indicator with the minimum value for the optimization problem.

4. Simulation Study

In this section, we present an extensive simulation study to examine the performance of the proposed estimators in the presence of the outliers. We consider the simulation scenarios proposed by [24] to generate clean data. The clean data are generated from the model

$$y_i^{\text{clean}} = \sum_{j=1}^p f_j(X_{ij}) + \epsilon_i,$$

where $f_1(x) = 5 \sin(2\pi x)$, $f_2(x) = 10x(l - x)$, $f_3(x) = 3x$, $f_4(x) = 2x$, $f_5(x) = -2x$ and $f_j(x) = 0$, for $j = 6, \dots, p$. The errors ϵ_i are generated from a normal distribution with zero mean and variance σ^2 . The covariates X_{ij} are generated from a multivariate normal distribution with zero mean vector and the covariances $\text{Cov}(X_{ij_1}, X_{ij_2}) = 0.5^{|j_1 - j_2|}$ and then the cumulative distribution function of the standard normal distribution is applied to them to transform their range into $[0, 1]$.

The simulation study is performed for $N = 100$ iterations of the data generation and estimation. For each iteration of the simulation study, we generate $n = 100, 200$ clean train datasets, with $p = 50, 100, 200$ covariates, and $\sigma = 0.2, 0.5$. The clean train data points are denoted by

$$(\mathbf{X}_{i1}^{\text{tr}}, \dots, \mathbf{X}_{ip}^{\text{tr}}, y_i^{\text{tr-clean}}), i = 1, \dots, n.$$

We further generate $n_{ts} = n/2$ test datasets, denoted by

$$(\mathbf{X}_{i1}^{\text{ts}}, \dots, \mathbf{X}_{ip}^{\text{ts}}, y_i^{\text{ts-clean}}), i = 1, \dots, n_{ts}.$$

Then, we contaminate the response values $y_i^{\text{tr-clean}}$ and $y_i^{\text{ts-clean}}$ as follows. From n (n_{ts}) samples, we choose 20% randomly, and we denote this subset by \mathcal{O} . Then, for any $i^* \in \mathcal{O}$, we generate U_{1i^*} and U_{2i^*} from a uniform distribution over $[0, 1]$ independently. Next, we let

$$y_{i^*}^{\text{outlier}} = Y_{i^*}^{\text{clean}} + \left[2I(U_{1i^*} > 0.5) - 1 \right] (2 + U_{2i^*}) S_Y^{\text{clean}}, \text{ for } i^* \in \mathcal{O},$$

where S_Y^{clean} is the sample standard deviation of clean responses. For a Core i5 10210U CPU (1.60 GHz) with 8 GB RAM and R version 4.2.1 (64 Bit), the mean computation time for SSLAM is 16.84 min (with optimization of BIC for choosing penalization parameters), for SLM it is 0.26 s, for SGAM it is 2.41 min, for RSSLAM it is 53.29 s (without optimization

of BIC for choosing penalization parameters), for RSLM it is 0.23 s, and for RSGAM it is 4.54 min. Note that the codes of the proposed models are developed in R, while for the SLM and RSLM models, the R package *glmnet* is used, which implements the main procedures in the C programming language.

Several criteria are considered in this simulation study to examine the performance of the estimators. The mean integrated square error (MISE) for f_j is defined as

$$\text{MISE}(f_j) = \int_0^1 (\hat{f}_j(x) - f_j(x))^2 dx, \quad j = 1, \dots, p.$$

For the purpose of testing the prediction efficiency of the proposed methods in the presence of the outliers, we define the clean data mean square error (CMSE) and clean data prediction error (CPE) for the train and test datasets, respectively, as follows:

$$\text{CMSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i^{\text{tr}} - y_i^{\text{tr-clean}})^2,$$

$$\text{CPE} = \frac{1}{n_{\text{ts}}} \sum_{i=1}^{n_{\text{ts}}} (\hat{y}_i^{\text{ts}} - y_i^{\text{ts-clean}})^2.$$

The false negative rate and the false positive rate are also defined as follows:

$$\text{FNR} = \frac{\#\{j; 1 \leq j \leq p, f_j \neq 0, \hat{f}_j = 0\}}{\#\{j; 1 \leq j \leq p, f_j \neq 0\}},$$

$$\text{FPR} = \frac{\#\{j; 1 \leq j \leq q, f_j = 0, \hat{f}_j \neq 0\}}{\#\{j; 1 \leq j \leq q, f_j = 0\}},$$

where $\#A$ stands for the cardinality of the set A .

We also define the false linear rate (FLR) and false non-linear rate (FNLR) criteria as follows, to examine the separation performance of SSLA and RSSLA models:

$$\text{FLR} = \frac{\#\{j; 1 \leq j \leq p, f_j \text{ is not linear, } \hat{f}_j \text{ is linear}\}}{\#\{j; 1 \leq j \leq p, f_j \text{ is not linear}\}},$$

$$\text{FNLR} = \frac{\#\{j; 1 \leq j \leq p, f_j \text{ is linear, } \hat{f}_j \text{ is not linear}\}}{\#\{j; 1 \leq j \leq p, f_j \text{ is linear}\}},$$

as well as the false outlier rate (FOR) and false non-outlier rate (FNOR) criteria as follows, to examine the outlier detection performance of robust models:

$$\text{FOR} = \frac{\#\{i; 1 \leq i \leq n, y_i \text{ is not outlier, } y_i \text{ is detected as outlier}\}}{\#\{i; 1 \leq i \leq n, y_i \text{ is not outlier}\}},$$

$$\text{FNOR} = \frac{\#\{i; 1 \leq i \leq n, y_i \text{ is an outlier, } y_i \text{ is not detected as outlier}\}}{\#\{i; 1 \leq i \leq n, y_i \text{ is an outlier}\}}.$$

The means and standard errors of all the above criteria are tabulated for all different scenarios in Tables 1–8. From Tables 1 and 2, one can see that the robust separative model RSSLA is the most powerful model for estimating the true regression functions, especially for larger values of n ($n = 200$ in Table 2), while for $n = 100$ (Table 1), the RSNLA model is also a successful model for estimation of the nonlinear regression functions. From Table 3, it can be observed that the RSSLA model is almost more efficient than other competitors (except the RSNLA model in a few cases) in the sense of the clean data MSE (CMSE). The clean data prediction performance of the RSSLA model is of course the best among the six models, based on the CPE values tabulated in Table 4. From Table 5, the RSNLA model is the best model based on the FNR criterion, while the best values of the FPR

criterion are obtained for the SL model, based on the values in Table 6. However, one can see that the FNR and FPR values for the RSSLA model are better than those of the SSLA model, which means that the robust modeling improves the FNR and FPR values in the separative semi-parametric linear model. From Table 7, it can be seen that both SSLA and RSSLA models have near-zero values of the FLR, while the RSSLA model has significantly lower values of the FNLR than the SSLA model. This shows that the robust modeling helps the model to separate the linear and nonlinear covariates more accurately. Finally, from the values of the FOR and FNOR criteria in Table 8, we can deduce that the RSSLA model is the most powerful model among the three robust models for the correct detection of the outliers.

Table 1. The means and standard deviations of MISE values for $n = 100$ from the simulation study for 6 models. The standard deviations are shown in subscripts.

p	σ	f	SSLA	RSSLA	SNLA	RSNLA	SL	RSL
50	0.2	\hat{f}_1	0.334 _{0.309}	0.086 _{0.119}	0.398 _{0.269}	0.239 _{0.217}	2.277 _{0.697}	2.541 _{1.014}
		\hat{f}_2	0.203 _{0.234}	0.058 _{0.106}	0.203 _{0.152}	0.135 _{0.087}	0.213 _{0.177}	0.195 _{0.037}
		\hat{f}_3	0.213 _{0.146}	0.046 _{0.081}	0.276 _{0.116}	0.156 _{0.098}	0.466 _{0.384}	0.357 _{0.167}
		\hat{f}_4	0.150 _{0.169}	0.039 _{0.050}	0.143 _{0.158}	0.084 _{0.061}	0.141 _{0.193}	0.112 _{0.104}
		\hat{f}_5	0.127 _{0.129}	0.043 _{0.054}	0.129 _{0.117}	0.088 _{0.045}	0.102 _{0.077}	0.092 _{0.079}
	0.5	\hat{f}_1	0.444 _{0.415}	0.239 _{0.202}	0.461 _{0.251}	0.333 _{0.217}	2.358 _{0.751}	2.458 _{0.978}
		\hat{f}_2	0.224 _{0.209}	0.138 _{0.123}	0.219 _{0.189}	0.171 _{0.110}	0.198 _{0.059}	0.200 _{0.050}
		\hat{f}_3	0.243 _{0.182}	0.146 _{0.174}	0.275 _{0.128}	0.237 _{0.115}	0.485 _{0.525}	0.388 _{0.225}
		\hat{f}_4	0.152 _{0.165}	0.082 _{0.085}	0.134 _{0.119}	0.096 _{0.062}	0.158 _{0.195}	0.123 _{0.121}
		\hat{f}_5	0.143 _{0.140}	0.080 _{0.054}	0.128 _{0.104}	0.111 _{0.069}	0.121 _{0.210}	0.099 _{0.107}
100	0.2	\hat{f}_1	1.887 _{0.000}	0.887 _{0.000}	0.609 _{0.286}	0.323 _{0.255}	2.332 _{0.827}	2.329 _{0.991}
		\hat{f}_2	0.185 _{0.004}	0.168 _{0.000}	0.193 _{0.059}	0.152 _{0.059}	0.199 _{0.089}	0.191 _{0.026}
		\hat{f}_3	0.321 _{0.000}	0.197 _{0.116}	0.283 _{0.085}	0.315 _{0.000}	0.500 _{0.447}	0.387 _{0.221}
		\hat{f}_4	0.334 _{0.491}	0.080 _{0.000}	0.109 _{0.098}	0.084 _{0.024}	0.134 _{0.242}	0.097 _{0.076}
		\hat{f}_5	0.080 _{0.000}	0.063 _{0.000}	0.106 _{0.087}	0.085 _{0.026}	0.085 _{0.054}	0.079 _{0.006}
	0.5	\hat{f}_1	0.887 _{0.001}	0.587 _{0.003}	0.608 _{0.286}	0.347 _{0.258}	2.151 _{0.635}	2.221 _{0.768}
		\hat{f}_2	0.186 _{0.000}	0.169 _{0.001}	0.194 _{0.083}	0.154 _{0.052}	0.193 _{0.041}	0.190 _{0.023}
		\hat{f}_3	0.321 _{0.001}	0.229 _{0.102}	0.297 _{0.058}	0.318 _{0.001}	0.414 _{0.275}	0.350 _{0.174}
		\hat{f}_4	0.083 _{0.000}	0.080 _{0.000}	0.102 _{0.068}	0.082 _{0.024}	0.092 _{0.072}	0.085 _{0.040}
		\hat{f}_5	0.085 _{0.000}	0.080 _{0.000}	0.121 _{0.151}	0.091 _{0.044}	0.088 _{0.005}	0.084 _{0.029}
200	0.2	\hat{f}_1	0.964 _{0.024}	0.383 _{0.172}	0.574 _{0.281}	0.258 _{0.275}	2.045 _{0.525}	2.093 _{0.719}
		\hat{f}_2	0.124 _{0.003}	0.113 _{0.062}	0.170 _{0.046}	0.086 _{0.109}	0.187 _{0.006}	0.189 _{0.026}
		\hat{f}_3	0.307 _{0.000}	0.141 _{0.101}	0.264 _{0.078}	0.213 _{0.102}	0.394 _{0.237}	0.317 _{0.065}
		\hat{f}_4	0.094 _{0.001}	0.068 _{0.032}	0.086 _{0.040}	0.076 _{0.059}	0.110 _{0.090}	0.088 _{0.054}
		\hat{f}_5	0.088 _{0.002}	0.076 _{0.001}	0.095 _{0.055}	0.079 _{0.020}	0.081 _{0.003}	0.080 _{0.003}
	0.5	\hat{f}_1	0.925 _{0.057}	0.606 _{0.018}	0.615 _{0.314}	0.484 _{0.393}	1.973 _{0.458}	2.092 _{0.584}
		\hat{f}_2	0.186 _{0.001}	0.173 _{0.004}	0.179 _{0.057}	0.167 _{0.055}	0.188 _{0.015}	0.191 _{0.040}
		\hat{f}_3	0.281 _{0.002}	0.205 _{0.098}	0.272 _{0.079}	0.234 _{0.100}	0.350 _{0.130}	0.318 _{0.067}
		\hat{f}_4	0.079 _{0.000}	0.062 _{0.007}	0.085 _{0.033}	0.089 _{0.056}	0.107 _{0.097}	0.098 _{0.092}
		\hat{f}_5	0.081 _{0.003}	0.059 _{0.000}	0.089 _{0.041}	0.079 _{0.005}	0.080 _{0.002}	0.080 _{0.003}

Table 2. The means and standard deviations of MISE values for $n = 200$ from the simulation study for 6 models. The standard deviations are shown in subscripts.

p	σ	f	SSLA	RSSLA	SNLA	RSNLA	SL	RSL
50	0.2	\hat{f}_1	0.287 _{0.686}	0.080 _{0.274}	0.363 _{0.218}	0.330 _{0.098}	2.985 _{0.903}	2.675 _{0.676}
		\hat{f}_2	0.167 _{0.076}	0.055 _{0.044}	0.196 _{0.117}	0.082 _{0.075}	0.196 _{0.033}	0.191 _{0.016}
		\hat{f}_3	0.261 _{0.098}	0.065 _{0.091}	0.246 _{0.135}	0.070 _{0.084}	0.579 _{0.320}	0.485 _{0.268}
		\hat{f}_4	0.076 _{0.040}	0.034 _{0.013}	0.175 _{0.154}	0.041 _{0.046}	0.130 _{0.154}	0.099 _{0.065}
		\hat{f}_5	0.080 _{0.033}	0.035 _{0.001}	0.163 _{0.123}	0.046 _{0.047}	0.099 _{0.079}	0.077 _{0.011}
	0.5	\hat{f}_1	0.290 _{0.252}	0.086 _{0.093}	0.345 _{0.179}	0.187 _{0.139}	2.719 _{0.810}	3.008 _{0.850}
		\hat{f}_2	0.168 _{0.040}	0.050 _{0.033}	0.222 _{0.137}	0.116 _{0.085}	0.211 _{0.121}	0.192 _{0.028}
		\hat{f}_3	0.248 _{0.103}	0.058 _{0.049}	0.275 _{0.185}	0.142 _{0.117}	0.520 _{0.289}	0.466 _{0.253}
		\hat{f}_4	0.075 _{0.015}	0.044 _{0.024}	0.222 _{0.176}	0.089 _{0.052}	0.138 _{0.144}	0.093 _{0.065}
		\hat{f}_5	0.080 _{0.002}	0.050 _{0.027}	0.213 _{0.194}	0.098 _{0.069}	0.107 _{0.105}	0.085 _{0.034}
100	0.2	\hat{f}_1	0.295 _{0.417}	0.076 _{0.122}	0.222 _{0.138}	1.115 _{0.887}	2.607 _{0.701}	2.651 _{0.940}
		\hat{f}_2	0.182 _{0.023}	0.063 _{0.065}	0.162 _{0.056}	0.152 _{0.059}	0.196 _{0.051}	0.188 _{0.016}
		\hat{f}_3	0.275 _{0.097}	0.068 _{0.092}	0.233 _{0.110}	0.207 _{0.123}	0.506 _{0.332}	0.402 _{0.205}
		\hat{f}_4	0.079 _{0.011}	0.048 _{0.035}	0.080 _{0.022}	0.066 _{0.026}	0.096 _{0.077}	0.091 _{0.057}
		\hat{f}_5	0.080 _{0.000}	0.049 _{0.034}	0.082 _{0.019}	0.072 _{0.029}	0.090 _{0.053}	0.083 _{0.036}
	0.5	\hat{f}_1	0.317 _{0.294}	0.112 _{0.124}	0.295 _{0.204}	0.151 _{0.154}	2.413 _{0.672}	2.443 _{0.731}
		\hat{f}_2	0.184 _{0.016}	0.074 _{0.057}	0.144 _{0.056}	0.104 _{0.064}	0.189 _{0.017}	0.188 _{0.010}
		\hat{f}_3	0.299 _{0.069}	0.097 _{0.083}	0.209 _{0.094}	0.120 _{0.096}	0.499 _{0.295}	0.381 _{0.186}
		\hat{f}_4	0.079 _{0.008}	0.065 _{0.020}	0.091 _{0.043}	0.068 _{0.041}	0.103 _{0.099}	0.085 _{0.032}
		\hat{f}_5	0.080 _{0.000}	0.063 _{0.024}	0.094 _{0.045}	0.071 _{0.018}	0.083 _{0.017}	0.079 _{0.010}
200	0.2	\hat{f}_1	0.289 _{0.312}	0.022 _{0.084}	0.149 _{0.170}	0.026 _{0.056}	1.380 _{1.280}	1.392 _{1.288}
		\hat{f}_2	0.191 _{0.020}	0.011 _{0.013}	0.076 _{0.078}	0.018 _{0.030}	0.108 _{0.092}	0.108 _{0.093}
		\hat{f}_3	0.284 _{0.088}	0.020 _{0.079}	0.098 _{0.104}	0.025 _{0.052}	0.260 _{0.300}	0.190 _{0.183}
		\hat{f}_4	0.077 _{0.008}	0.015 _{0.027}	0.048 _{0.045}	0.019 _{0.028}	0.057 _{0.064}	0.051 _{0.055}
		\hat{f}_5	0.081 _{0.001}	0.023 _{0.018}	0.051 _{0.052}	0.025 _{0.033}	0.051 _{0.056}	0.046 _{0.040}
	0.5	\hat{f}_1	0.112 _{0.095}	0.043 _{0.055}	0.176 _{0.240}	0.061 _{0.092}	1.233 _{1.201}	1.226 _{1.226}
		\hat{f}_2	0.082 _{0.011}	0.036 _{0.024}	0.078 _{0.080}	0.045 _{0.062}	0.102 _{0.095}	0.101 _{0.093}
		\hat{f}_3	0.108 _{0.057}	0.033 _{0.028}	0.110 _{0.122}	0.046 _{0.065}	0.254 _{0.320}	0.196 _{0.223}
		\hat{f}_4	0.038 _{0.006}	0.021 _{0.008}	0.043 _{0.044}	0.030 _{0.036}	0.062 _{0.093}	0.056 _{0.086}
		\hat{f}_5	0.023 _{0.000}	0.015 _{0.013}	0.048 _{0.049}	0.034 _{0.039}	0.055 _{0.079}	0.042 _{0.040}

Table 3. The means and standard deviations of CMSE values from the simulation study for 6 models. The standard deviations are shown in subscripts.

n	p	σ	SSLA	RSSLA	SNLA	RSNLA	SL	RSL
100	50	0.2	2.409 _{0.875}	0.736 _{1.082}	2.921 _{0.394}	1.357 _{0.843}	1.427 _{0.269}	1.405 _{0.358}
		0.5	1.657 _{0.388}	1.637 _{0.333}	3.231 _{0.389}	2.126 _{0.814}	2.979 _{0.528}	1.955 _{1.170}
	100	0.2	2.438 _{0.290}	1.539 _{0.741}	2.986 _{0.409}	1.277 _{0.635}	1.486 _{0.334}	1.575 _{0.480}
		0.5	2.615 _{0.288}	1.495 _{0.758}	3.249 _{0.395}	1.743 _{0.651}	1.774 _{0.388}	1.774 _{0.402}
200	50	0.2	2.489 _{0.364}	1.613 _{0.381}	2.984 _{0.410}	0.953 _{0.590}	1.617 _{0.843}	1.695 _{0.501}
		0.5	2.724 _{0.286}	1.628 _{0.643}	3.162 _{0.416}	1.788 _{0.719}	1.745 _{0.362}	1.842 _{0.500}
	100	0.2	0.887 _{1.091}	0.800 _{0.272}	2.874 _{0.934}	0.894 _{0.276}	1.307 _{0.220}	1.284 _{0.169}
		0.5	0.988 _{0.280}	0.825 _{0.370}	3.140 _{0.331}	2.156 _{1.029}	1.512 _{0.189}	1.490 _{0.235}
200	100	0.2	0.836 _{0.420}	0.338 _{0.336}	1.548 _{1.050}	1.879 _{1.304}	1.326 _{0.192}	1.369 _{0.262}
		0.5	1.075 _{0.273}	0.884 _{0.401}	2.621 _{0.922}	1.026 _{0.461}	1.563 _{0.210}	1.627 _{0.298}
200	200	0.2	0.718 _{0.346}	0.218 _{0.352}	1.710 _{1.477}	0.253 _{0.082}	0.780 _{0.680}	0.814 _{0.726}
		0.5	1.023 _{0.249}	0.427 _{0.214}	1.765 _{1.657}	0.573 _{0.705}	0.858 _{0.814}	0.913 _{0.886}

Table 4. The means and standard deviations of CPE values from the simulation study for 6 models. The standard deviations are shown in subscripts.

<i>n</i>	<i>p</i>	σ	SSLA	RSSLA	SNLA	RSNLA	SL	RSL
100	50	0.2	4.193 _{2.098}	0.900 _{1.260}	4.606 _{1.392}	1.602 _{1.084}	1.750 _{0.403}	1.734 _{0.464}
		0.5	2.018 _{0.413}	1.993 _{0.507}	5.451 _{1.546}	2.397 _{0.883}	3.026 _{2.100}	2.190 _{1.345}
	100	0.2	4.220 _{2.657}	1.472 _{0.728}	3.664 _{0.964}	1.861 _{0.300}	1.948 _{0.686}	1.871 _{0.567}
		0.5	2.991 _{1.106}	1.083 _{1.051}	4.073 _{0.889}	1.946 _{0.785}	2.195 _{0.591}	2.167 _{0.598}
	200	0.2	2.167 _{0.306}	1.023 _{0.364}	3.441 _{0.969}	1.119 _{0.664}	1.997 _{0.522}	2.006 _{0.589}
		0.5	2.034 _{0.346}	0.976 _{0.564}	3.901 _{0.929}	2.107 _{0.766}	2.262 _{0.645}	2.197 _{0.592}
200	50	0.2	0.923 _{1.069}	0.910 _{0.336}	7.708 _{1.816}	0.997 _{1.065}	1.450 _{0.250}	1.410 _{0.263}
		0.5	1.128 _{0.334}	0.932 _{0.349}	9.292 _{1.983}	2.353 _{0.973}	1.739 _{0.338}	1.654 _{0.290}
	100	0.2	0.967 _{0.434}	0.434 _{0.425}	2.070 _{1.616}	1.935 _{1.333}	1.525 _{0.289}	1.599 _{0.385}
		0.5	1.182 _{0.372}	1.011 _{0.468}	3.925 _{1.777}	1.136 _{0.464}	1.767 _{0.296}	1.837 _{0.406}
	200	0.2	1.030 _{0.209}	0.252 _{0.389}	1.909 _{1.696}	0.837 _{0.603}	0.932 _{0.827}	0.942 _{0.846}
		0.5	1.202 _{0.386}	0.443 _{0.465}	2.087 _{2.017}	0.636 _{0.784}	1.002 _{0.950}	1.026 _{0.987}

Table 5. The means and standard deviations of FNR values from the simulation study for 6 models. The standard deviations are shown in subscripts.

<i>n</i>	<i>p</i>	σ	SSLA	RSSLA	SNLA	RSNLA	SL	RSL
100	50	0.2	0.299 _{0.239}	0.147 _{0.122}	0.292 _{0.165}	0.233 _{0.176}	0.572 _{0.173}	0.527 _{0.193}
		0.5	0.273 _{0.210}	0.200 _{0.144}	0.288 _{0.171}	0.267 _{0.169}	0.572 _{0.193}	0.533 _{0.192}
	100	0.2	0.631 _{0.056}	0.379 _{0.089}	0.562 _{0.165}	0.337 _{0.176}	0.610 _{0.174}	0.628 _{0.192}
		0.5	0.610 _{0.000}	0.517 _{0.096}	0.520 _{0.160}	0.472 _{0.159}	0.685 _{0.171}	0.632 _{0.187}
	200	0.2	0.642 _{0.237}	0.461 _{0.166}	0.601 _{0.148}	0.403 _{0.137}	0.707 _{0.174}	0.687 _{0.176}
		0.5	0.617 _{0.203}	0.542 _{0.202}	0.611 _{0.154}	0.482 _{0.127}	0.692 _{0.149}	0.715 _{0.171}
200	50	0.2	0.729 _{0.118}	0.237 _{0.240}	0.068 _{0.097}	0.043 _{0.073}	0.468 _{0.161}	0.403 _{0.182}
		0.5	0.699 _{0.121}	0.198 _{0.160}	0.062 _{0.091}	0.052 _{0.088}	0.453 _{0.174}	0.455 _{0.160}
	100	0.2	0.791 _{0.081}	0.478 _{0.174}	0.578 _{0.291}	0.320 _{0.255}	0.674 _{0.264}	0.527 _{0.151}
		0.5	0.813 _{0.064}	0.395 _{0.242}	0.382 _{0.282}	0.237 _{0.230}	0.555 _{0.136}	0.537 _{0.153}
	200	0.2	0.799 _{0.101}	0.411 _{0.129}	0.212 _{0.217}	0.097 _{0.148}	0.322 _{0.288}	0.297 _{0.280}
		0.5	0.818 _{0.093}	0.388 _{0.214}	0.210 _{0.235}	0.133 _{0.188}	0.315 _{0.317}	0.317 _{0.311}

Table 6. The means and standard deviations of FPR values from the simulation study for 6 models. The standard deviations are shown in subscripts.

<i>n</i>	<i>p</i>	σ	SSLA	RSSLA	SNLA	RSNLA	SL	RSL
100	50	0.2	0.463 _{0.216}	0.423 _{0.109}	0.493 _{0.049}	0.349 _{0.054}	0.130 _{0.122}	0.180 _{0.135}
		0.5	0.542 _{0.161}	0.488 _{0.087}	0.492 _{0.051}	0.383 _{0.049}	0.142 _{0.121}	0.173 _{0.131}
	100	0.2	0.251 _{0.154}	0.206 _{0.052}	0.206 _{0.033}	0.150 _{0.033}	0.116 _{0.105}	0.198 _{0.100}
		0.5	0.127 _{0.015}	0.106 _{0.031}	0.208 _{0.034}	0.167 _{0.034}	0.079 _{0.085}	0.094 _{0.075}
	200	0.2	0.451 _{0.163}	0.317 _{0.184}	0.128 _{0.021}	0.090 _{0.017}	0.047 _{0.055}	0.062 _{0.068}
		0.5	0.341 _{0.244}	0.206 _{0.086}	0.125 _{0.020}	0.098 _{0.016}	0.057 _{0.061}	0.064 _{0.075}
200	50	0.2	0.445 _{0.219}	0.166 _{0.119}	0.872 _{0.044}	0.676 _{0.049}	0.022 _{0.010}	0.185 _{0.131}
		0.5	0.623 _{0.190}	0.181 _{0.132}	0.885 _{0.038}	0.697 _{0.051}	0.023 _{0.019}	0.192 _{0.136}
	100	0.2	0.117 _{0.121}	0.103 _{0.096}	0.176 _{0.231}	0.179 _{0.172}	0.032 _{0.013}	0.141 _{0.099}
		0.5	0.198 _{0.175}	0.091 _{0.069}	0.360 _{0.216}	0.311 _{0.135}	0.032 _{0.021}	0.109 _{0.084}
	200	0.2	0.105 _{0.107}	0.076 _{0.048}	0.141 _{0.124}	0.081 _{0.090}	0.037 _{0.054}	0.046 _{0.060}
		0.5	0.205 _{0.104}	0.083 _{0.066}	0.134 _{0.127}	0.084 _{0.095}	0.032 _{0.053}	0.033 _{0.055}

Table 7. The means and standard deviations of FLR and FNLR values from the simulation study for SSLA and RSSLA models. The standard deviations are shown in subscripts.

Criterion			FLR		FNLR	
<i>n</i>	<i>p</i>	<i>σ</i>	SSLA	RSSLA	SSLA	RSSLA
100	50	0.2	0.000 _{0.000}	0.006 _{0.055}	0.549 _{0.203}	0.308 _{0.288}
		0.5	0.005 _{0.052}	0.011 _{0.074}	0.422 _{0.246}	0.384 _{0.265}
	100	0.2	0.000 _{0.000}	0.000 _{0.000}	0.195 _{0.163}	0.010 _{0.001}
		0.5	0.003 _{0.000}	0.000 _{0.000}	0.153 _{0.127}	0.028 _{0.010}
	200	0.2	0.000 _{0.000}	0.000 _{0.000}	0.138 _{0.141}	0.014 _{0.005}
		0.5	0.002 _{0.000}	0.000 _{0.000}	0.106 _{0.082}	0.017 _{0.011}
200	50	0.2	0.000 _{0.000}	0.000 _{0.000}	0.542 _{0.414}	0.125 _{0.162}
		0.5	0.000 _{0.000}	0.000 _{0.000}	0.593 _{0.213}	0.163 _{0.182}
	100	0.2	0.000 _{0.000}	0.000 _{0.000}	0.319 _{0.322}	0.065 _{0.134}
		0.5	0.000 _{0.000}	0.000 _{0.000}	0.571 _{0.333}	0.035 _{0.103}
	200	0.2	0.000 _{0.000}	0.000 _{0.000}	0.275 _{0.249}	0.054 _{0.082}
		0.5	0.000 _{0.000}	0.000 _{0.000}	0.443 _{0.218}	0.027 _{0.095}

Table 8. The means and standard deviations of FOR and FNOR values from the simulation study for 3 robust models. The standard deviations are shown in subscripts.

Criterion			FOR			FNOR		
<i>n</i>	<i>p</i>	<i>σ</i>	RSSLA	RSNLA	RSL	RSSLA	RSNLA	RSL
100	50	0.2	0.090 _{0.157}	0.122 _{0.075}	0.246 _{0.152}	0.085 _{0.039}	0.093 _{0.019}	0.124 _{0.038}
		0.5	0.144 _{0.073}	0.228 _{0.163}	0.316 _{0.124}	0.099 _{0.018}	0.119 _{0.041}	0.142 _{0.031}
	100	0.2	0.029 _{0.049}	0.154 _{0.087}	0.220 _{0.128}	0.070 _{0.012}	0.101 _{0.022}	0.118 _{0.032}
		0.5	0.000 _{0.000}	0.156 _{0.081}	0.253 _{0.118}	0.062 _{0.000}	0.102 _{0.020}	0.126 _{0.029}
	200	0.2	0.017 _{0.031}	0.166 _{0.085}	0.148 _{0.130}	0.065 _{0.001}	0.104 _{0.021}	0.099 _{0.033}
		0.5	0.015 _{0.004}	0.170 _{0.083}	0.260 _{0.138}	0.073 _{0.011}	0.105 _{0.021}	0.128 _{0.034}
200	50	0.2	0.035 _{0.047}	0.098 _{0.056}	0.136 _{0.158}	0.071 _{0.011}	0.087 _{0.014}	0.096 _{0.039}
		0.5	0.041 _{0.045}	0.098 _{0.047}	0.268 _{0.120}	0.073 _{0.011}	0.087 _{0.012}	0.129 _{0.030}
	100	0.2	0.026 _{0.047}	0.124 _{0.065}	0.029 _{0.046}	0.069 _{0.012}	0.093 _{0.016}	0.070 _{0.012}
		0.5	0.059 _{0.064}	0.130 _{0.063}	0.099 _{0.091}	0.077 _{0.016}	0.095 _{0.016}	0.087 _{0.023}
	200	0.2	0.018 _{0.022}	0.078 _{0.087}	0.041 _{0.076}	0.033 _{0.013}	0.056 _{0.050}	0.040 _{0.045}
		0.5	0.063 _{0.070}	0.072 _{0.080}	0.085 _{0.114}	0.030 _{0.010}	0.052 _{0.049}	0.049 _{0.057}

5. Case Study

To evaluate the performance of the proposed method for a real dataset, we analyze the Boston housing prices dataset [37,38] with 506 observations and 14 features. The R package MASS [39] contains these data. Here, we consider the median value of the price of the owner-occupied homes in USD 1000 (Median Price) as the response variable, and the following covariates:

- Crime rate: per capita crime rate by town;
- Nitrogen Oxides: nitrogen oxide concentration (parts per 10 million);
- Rooms: average number of rooms per dwelling;
- Age: proportion of owner-occupied units built prior to 1940;
- Distances: weighted mean of distances to five Boston employment centers;
- Lower Status: lower status of the population (percent).

The following model is considered:

$$\begin{aligned}
 \text{Median Price} = & \mu + f_1(\text{Crime rate}) + f_2(\text{Nitrogen Oxides}) + f_3(\text{Rooms}) + f_4(\text{Age}) \\
 & + f_5(\text{Distances}) + f_6(\text{Lower Status}) + \epsilon
 \end{aligned}
 \tag{15}$$

The leave-one-out cross-validation is considered, by only considering the samples with less than the 90% quantile of the train set square residuals (not considered as outliers)

in all models. We call this criterion trimmed leave-one-out cross-validation (TLOOCV), which is as follows:

$$TLOOCV = \frac{\sum_{i=1}^n u_i (y_i - \hat{y}_i^{(-i)})^2}{\sum_{i=1}^n u_i},$$

where $\hat{y}_i^{(-i)}$ is the prediction of y_i by using all observations except (X_i, y_i) , and

$$u_i = \begin{cases} 1 & (y_i - \hat{y}_i^{(-i)})^2 < \text{Quantile}_{0.9} \left\{ (y_j - \hat{y}_j)^2, j \in \{1, \dots, n\} / i \right\} \\ 0 & \end{cases} .$$

Values of TLOOCV are presented in Table 9, along with the percent of test points $(100(1 - \sum_{i=1}^n u_i/n)\%)$, considered as the outliers. As one can see from Table 9, the RSSLA model has achieved the smallest value of the TLOOCV among all models.

To draw the partial residual plot for the j th covariate ($j = 1, \dots, 6$), we compute the residuals of the regression of the response variable against all covariates except the j th covariate, and then we plot it against the j th covariate. These plots are shown in Figures 1–6 for all six models. The outliers are the points where their square residual is greater than the 90% quantile of the square residuals.

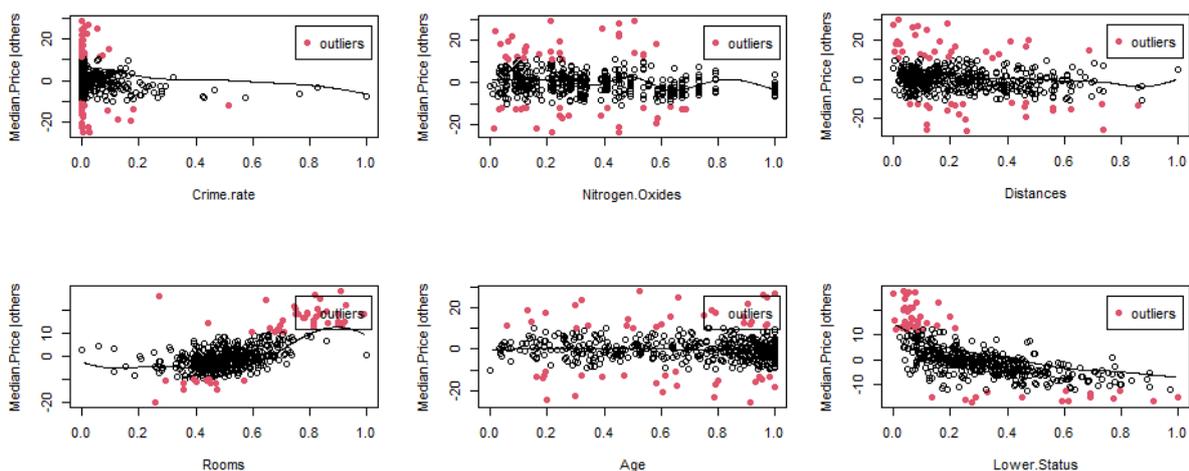


Figure 1. Partial residual plots for SSIA model.

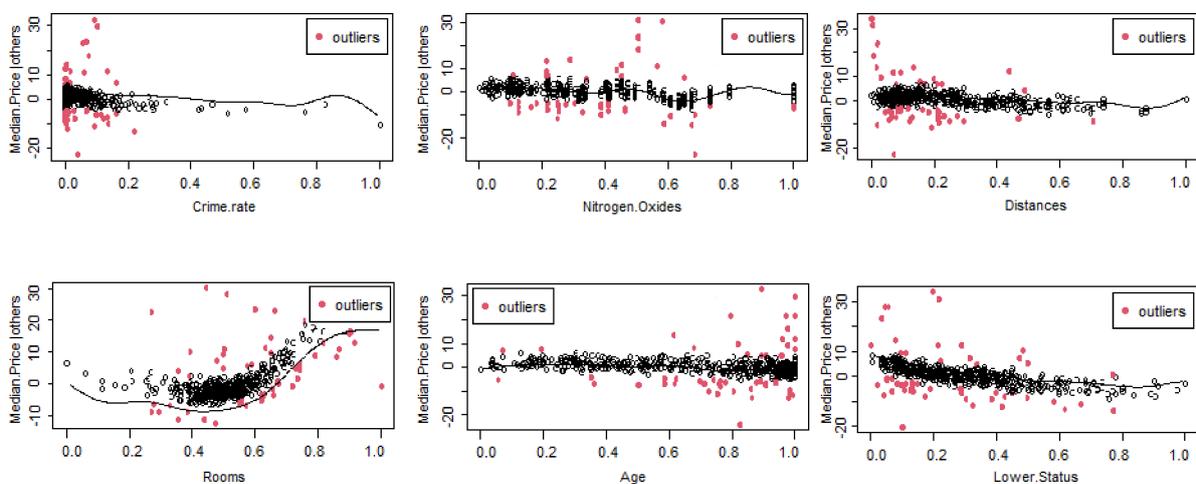


Figure 2. Partial residual plots for RSSLA model.

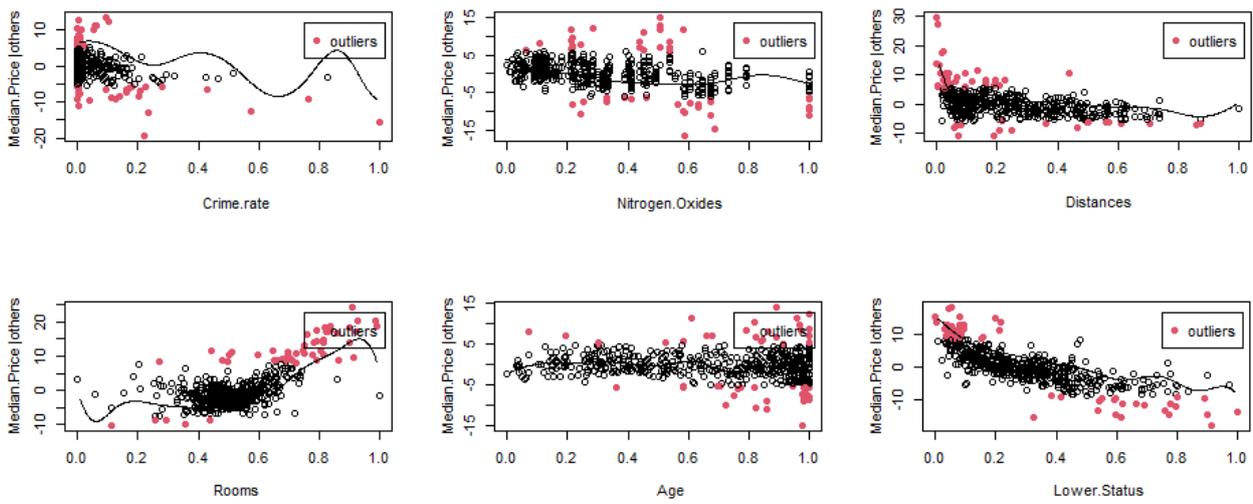


Figure 3. Partial residual plots for SNLA model.

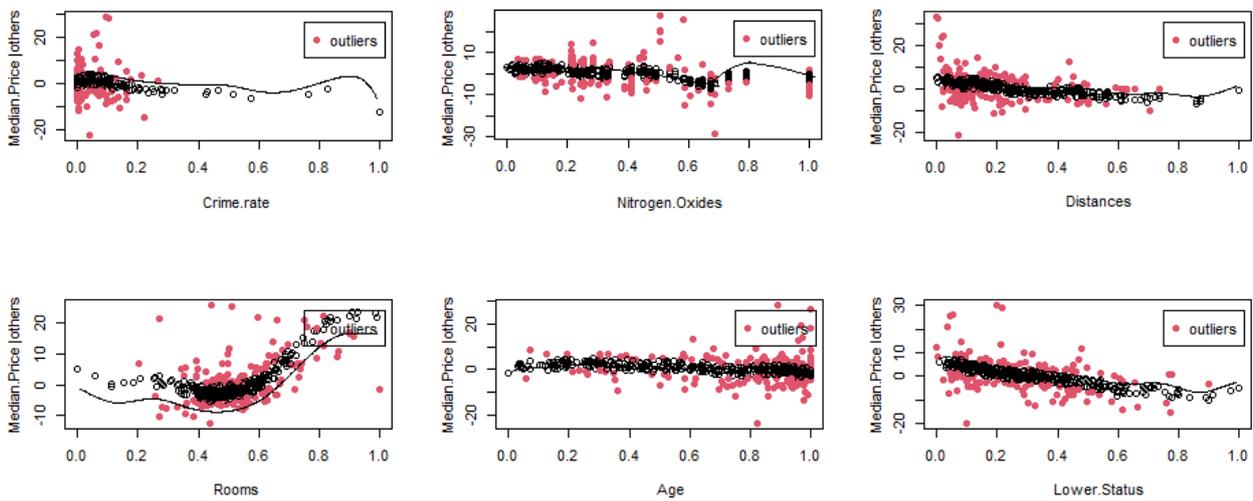


Figure 4. Partial residual plots for RSNLA model.

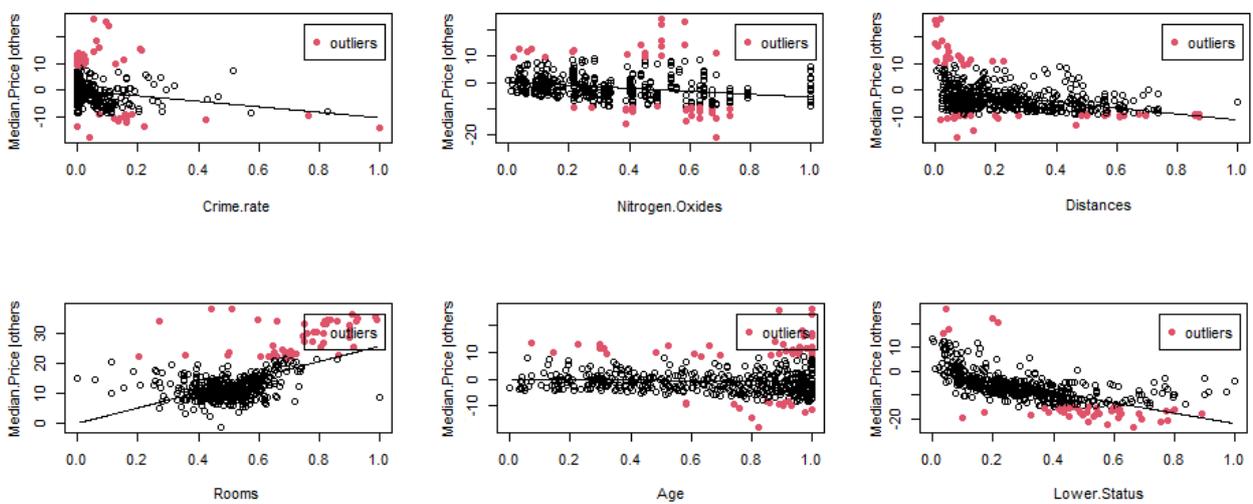


Figure 5. Partial residual plots for SL model.

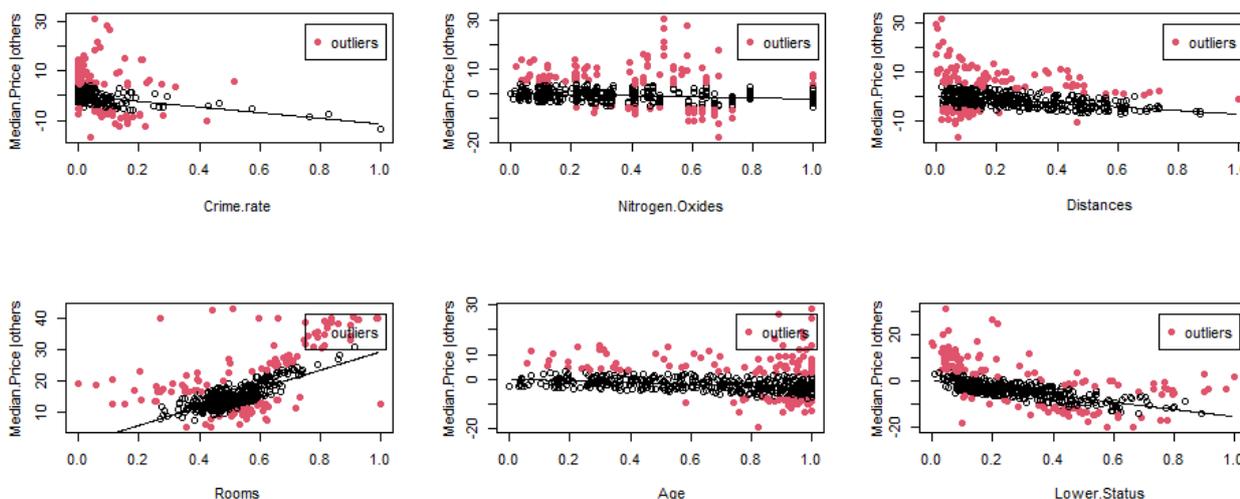


Figure 6. Partial residual plots for RSL model.

Table 9. Trimmed leave-one-out cross-validation and test outlier percent for 6 models.

Model	TLOOCV	Test Outlier Percent
SSLA	5.32	13.8%
RSSLA	4.78	12.6%
SNLA	5.51	14.4%
RSNLA	6.16	12.4%
SL	11.98	10.5%
RSL	10.81	10.5%

6. Conclusions and Summary

This research investigates a robust version of the sparse separative semi-parametric linear regression model introduced by [24] by utilising the LTS regression model [35]. The proposed research might potentially encompass other separative semi-parametric linear regression models put forth by [23,25,26], as well as other upcoming comparable techniques. The simulation analysis demonstrates that the proposed strategy substantially enhances the performance of the sparse separative semi-parametric linear regression model in terms of estimation, prediction, variable selection, and separation. The proposed method outperforms previous robust regression models in outlier detection and has the best performance in terms of a truncated variant of the leave-one-out cross-validation measure [37,38] when applied to the well-known Boston housing prices dataset.

Author Contributions: Conceptualization, M.A.; methodology, M.A. and M.R.; software, M.A. and M.R.; validation, M.A., M.R. and N.A.M.; formal analysis, M.A. and M.R.; investigation, M.A.; resources, M.A. and M.R.; data curation, M.A., M.R. and N.A.M.; writing—original draft preparation, M.A., M.R. and N.A.M.; writing—review and editing, M.A., M.R. and N.A.M.; visualization, M.A., M.R. and N.A.M.; supervision, M.A.; project administration, M.A.; funding acquisition, N.A.M. All authors have read and agreed to the published version of the manuscript.

Funding: We want to thank the Ministry of Higher Education Malaysia for their support in funding this research through the Fundamental Research Grant Scheme (FRGS/1/2023/STG06/UM/02/13) awarded to Nur Anisah Mohamed @ A Rahman.

Data Availability Statement: All used datasets are available in R software (R Foundation for Statistical Computing, Vienna, Austria) at the “MASS” library.

Acknowledgments: The authors would like to thank three anonymous reviewers for their valuable comments and corrections to an earlier version of this paper, which significantly improved the quality of our work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ruppert, D.; Wand, M.P.; Carroll, R.J. *Semiparametric Regression*; Cambridge University Press: Cambridge, UK, 2003.
2. Friedman, J.H.; Stuetzle, W. Projection pursuit regression. *J. Am. Stat. Assoc.* **1981**, *76*, 817–823. [[CrossRef](#)]
3. Marx, B.D.; Eilers, P.H. Direct generalized additive modeling with penalized likelihood. *Comput. Stat. Data Anal.* **1998**, *28*, 193–209. [[CrossRef](#)]
4. Wood, S.N. Modelling and smoothing parameter estimation with multiple quadratic penalties. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2000**, *62*, 413–428. [[CrossRef](#)]
5. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006.
6. Wood, S.N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* **2004**, *99*, 673–686. [[CrossRef](#)]
7. Speed, T. [That BLUP is a good thing: The estimation of random effects]: Comment. *Stat. Sci.* **1991**, *6*, 42–44. [[CrossRef](#)]
8. Wang, Y. Mixed effects smoothing spline analysis of variance. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1998**, *60*, 159–174. [[CrossRef](#)]
9. Breiman, L. Prediction games and arcing algorithms. *Neural Comput.* **1999**, *11*, 1493–1517. [[CrossRef](#)]
10. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
11. Binder, H.; Tutz, G. A comparison of methods for the fitting of generalized additive models. *Stat. Comput.* **2008**, *18*, 87–99. [[CrossRef](#)]
12. Meier, L.; Van de Geer, S.; Bühlmann, P. High-dimensional additive modeling. *Ann. Stat.* **2009**, *37*, 3779–3821. [[CrossRef](#)]
13. Ravikumar, P.; Liu, H.; Lafferty, J.; Wasserman, L. Spam: Sparse additive models. In Proceedings of the 20th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; Curran Associates Inc.: Red Hook, NY, USA, 2007; pp. 1201–1208.
14. Wang, L.; Chen, G.; Li, H. Group SCAD regression analysis for microarray time course gene expression data. *Bioinformatics* **2007**, *23*, 1486–1494. [[CrossRef](#)] [[PubMed](#)]
15. Wang, H.; Xia, Y. Shrinkage estimation of the varying coefficient model. *J. Am. Stat. Assoc.* **2009**, *104*, 747–757. [[CrossRef](#)]
16. Lin, Y.; Zhang, H.H. Component selection and smoothing in multivariate nonparametric regression. *Ann. Stat.* **2006**, *34*, 2272–2297. [[CrossRef](#)]
17. Bach, F.R. Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.* **2008**, *9*, 1179–1225.
18. Huang, J.; Horowitz, J.L.; Wei, F. Variable selection in nonparametric additive models. *Ann. Stat.* **2010**, *38*, 2282. [[CrossRef](#)]
19. Opsomer, J.D.; Ruppert, D. A root-n consistent backfitting estimator for semiparametric additive modeling. *J. Comput. Graph. Stat.* **1999**, *8*, 715–732. [[CrossRef](#)]
20. Wang, L.; Liu, X.; Liang, H.; Carroll, R.J. Estimation and variable selection for generalized additive partial linear models. *Ann. Stat.* **2011**, *39*, 1827. [[CrossRef](#)]
21. Liu, X.; Wang, L.; Liang, H. Estimation and variable selection for semiparametric additive partial linear models (ss-09-140). *Stat. Sin.* **2011**, *21*, 1225. [[CrossRef](#)]
22. Arashi, M.; Asar, Y.; Yüzbaşı, B. SLASSO: A scaled LASSO for multicollinear situations. *J. Stat. Comput. Simul.* **2021**, *91*, 3170–3183. [[CrossRef](#)]
23. Huang, J.; Wei, F.; Ma, S. Semiparametric regression pursuit. *Stat. Sin.* **2012**, *22*, 1403. [[CrossRef](#)]
24. Lian, H.; Liang, H.; Ruppert, D. Separation of covariates into nonparametric and parametric parts in high-dimensional partially linear additive models. *Stat. Sin.* **2015**, *25*, 591–607.
25. Li, X.; Wang, L.; Nettleton, D. Sparse model identification and learning for ultra-high-dimensional additive partially linear models. *J. Multivar. Anal.* **2019**, *173*, 204–228. [[CrossRef](#)]
26. Liu, H.; Ma, J.; Peng, C. Shrinkage estimation for identification of linear components in composite quantile additive models. *Commun. Stat.—Simul. Comput.* **2020**, *49*, 2678–2692. [[CrossRef](#)]
27. Kazemi, M.; Shahsavani, D.; Arashi, M. Variable selection and structure identification for ultrahigh-dimensional partially linear additive models with application to cardiomyopathy microarray data. *Stat. Optim. Inf. Comput.* **2018**, *6*, 373–382. [[CrossRef](#)]
28. Kazemi, M.; Shahsavani, D.; Arashi, M.; Rodrigues, P.C. Identification for partially linear regression model with autoregressive errors. *J. Stat. Comput. Simulation* **2021**, *91*, 1441–1454. [[CrossRef](#)]
29. Maronna, R.A.; Martin, R.D.; Yohai, V.J.; Salibián-Barrera, M. *Robust Statistics: Theory and Methods (with R)*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
30. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*; John Wiley and Sons: New York, NY, USA, 1987.
31. Rousseeuw, P.J. Least median of squares regression. *J. Am. Stat. Assoc.* **1984**, *79*, 871–880. [[CrossRef](#)]
32. Roozbeh, M.; Babaie-Kafaki, S.; Naeimi Sadigh, A. A heuristic approach to combat multicollinearity in least trimmed squares regression analysis. *Appl. Math. Model.* **2018**, *57*, 105–120. [[CrossRef](#)]
33. Amini, M.; Roozbeh, M. Least trimmed squares ridge estimation in partially linear regression models. *J. Stat. Comput. Simul.* **2016**, *86*, 2766–2780. [[CrossRef](#)]
34. Mahmoud, H.F.F.; Kim, B.J.; Kim, I. Robust nonparametric derivative estimator. *Commun. Stat.—Simul. Comput.* **2022**, *51*, 3809–3829. [[CrossRef](#)]

35. Rousseeuw, P.J.; Van Driessen, K. Computing LTS regression for large data sets. *Data Min. Knowl. Discov.* **2006**, *12*, 29–45. [[CrossRef](#)]
36. Alfons, A.; Croux, C.; Gelper, S. Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *Ann. Appl. Stat.* **2013**, *7*, 226–248. [[CrossRef](#)]
37. Belsley D.A.; Kuh, E.; Welsch, R.E. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*; Wiley: New York, NY, USA, 1980.
38. Harrison, D.; Rubinfeld, D.L. Hedonic prices and the demand for clean air. *J. Environ. Econ. Manag.* **1978**, *5*, 81–102. [[CrossRef](#)]
39. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; ISBN 0-387-95457-0. Available online: <https://www.stats.ox.ac.uk/pub/MASS4/> (accessed on 15 August 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.