

Article

# Robust Variable Selection with Exponential Squared Loss for the Spatial Single-Index Varying-Coefficient Model

Yezi Wang , Zhijian Wang and Yunquan Song \*

College of Science, China University of Petroleum, Qingdao 266580, China

\* Correspondence: statistics99@163.com; Tel.: +86-0532-8698-2921

**Abstract:** As spatial correlation and heterogeneity often coincide in the data, we propose a spatial single-index varying-coefficient model. For the model, in this paper, a robust variable selection method based on spline estimation and exponential squared loss is offered to estimate parameters and identify significant variables. We establish the theoretical properties under some regularity conditions. A block coordinate descent (BCD) algorithm with the concave–convex process (CCCP) is composed uniquely for solving algorithms. Simulations show that our methods perform well even though observations are noisy or the estimated spatial mass matrix is inaccurate.

**Keywords:** spatial single-index varying-coefficient model; exponential squared loss; variable selection

## 1. Introduction

Spatial econometrics is one of the essential branches of econometrics. Its basic content is to consider the spatial effects of variables in regional scientific models. The most widely used spatial econometric model is the spatial autoregressive (SAR) model, first proposed by [1], which has been extensively studied and applied in the fields of economy, finance, and environment.

The SAR model is mainly a parameter model. However, in the practical application, only the parametric model cannot fully explain the complex economic problems and phenomena. Therefore, in order to improve the flexibility and applicability of the spatial econometric model, the non-parametric spatial econometric model has received more attention. Ref. [2] studied the SAR model in the non-parametric frame, obtained the parameter estimators by using the generalized moment estimation, and proved the consistency and asymptotic property of the estimator. The instrumental variable method was used by [3] to study semi-parametric varying-coefficient spatial panel data models with endogenous explanatory variables.

However, for all practical purposes, data may have spatial correlation and spatial heterogeneity simultaneously, which leads to spatial heterogeneity that cannot be fully considered and reflected by the SAR model in the parametric form and the non-parametric SAR model.

The single-index varying-coefficient model is a generalization of the single-index and varying-coefficient models, which can effectively avoid the “curse of dimension” in multidimensional non-parametric regression. Many domestic and foreign researchers have learned this. Refs. [4,5] studied the evaluation of the single-index varying-coefficient model. Ref. [6] constructed the empirical likelihood confidence region of the single-index varying-coefficient model by using the empirical likelihood method; Ref. [7] proposed a new estimated empirical likelihood ratio statistic, obtained maximum likelihood estimators of the model parameters, and proposed a new Profile empirical likelihood ratio, which was shown to be asymptotically close to the standard chi-square distribution.

In addition, selecting significant explanatory variables is one of the most important problems of statistical learning. Some robust regression methods have been proposed,



**Citation:** Wang, Y.; Wang, Z.; Song, Y. Robust Variable Selection with Exponential Squared Loss for the Spatial Single-Index Varying-Coefficient Model. *Entropy* **2023**, *25*, 230. <https://doi.org/10.3390/e25020230>

Academic Editors: Ernestina Menasalvas, Tomoko Matsui, Pavel Shevchenko, Gareth W. Peters and Francois Septier

Received: 6 December 2022

Revised: 21 January 2023

Accepted: 23 January 2023

Published: 26 January 2023



**Copyright:** © 2023 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/>).

such as quantile regression, composite quantile regression, and modal regression. Ref. [8] presented a new class of robust regression estimators methods for linear models based on exponential square loss. The specific method is as follows: for the linear regression model  $y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$ , minimize  $\sum_{i=1}^n \left\{ 1 - \exp\left(-\frac{(y_i - \mathbf{X}_i^T \boldsymbol{\beta})^2}{h}\right) \right\}$  this objective function to estimate the regression parameters  $\boldsymbol{\beta}$ , in which  $h > 0$  controls the robustness of the estimation. For a large  $h$ ,  $1 - \exp(-r^2/h) \approx r^2/h$ . Therefore, the proposed estimation is similar to the least squares estimation in the extreme case. For a small  $h$ , the value of  $|r|$  is large, and the impact on the estimated value is small. Hence, a small value of  $h$  will limit the influence of outliers on the estimation, thus improving the robustness of estimators. Ref. [8] also pointed out that their method is more robust than the other general robust estimators methods. Ref. [9] made a robust estimation based on exponential square loss for some linear regression models and proposed a data driver to select adjustment parameters. The exponential square loss square is used in data simulation, and positive results were obtained by the method. Ref. [10] suggested a robust variable selection for the high-dimensional single-index varying-coefficient model based on exponential square loss, established and proved the theoretical properties of estimators, and demonstrated the robustness of this method through numerical simulation. Ref. [11] applied exponential square loss to conduct robust structure analysis and variable selection for some linear variable coefficient models and obtained good results.

Inspired by the above article, we introduce the spatial position of the observed objects into a single-index variable coefficient model, and a spatial single-index variable coefficient model is proposed. We also presented a variable selection method for the spatial single-index varying-coefficient model based on spline estimation and the exponential loss function. This method was capable of selecting significant predictors while estimating regression coefficients. The following are the main contributions of this work.

1. We propose a novel model: the spatial single-index varying-coefficient model, which can deal with the spatial correlation and spatial heterogeneity of data at the same time.
2. We construct a robust variable selection method for the spatial single-index varying-coefficient model, which uses exponential square loss function to resist the influence of strong noise and inaccurate spatial weight matrix. Furthermore, we present the BCD (block coordinate descent) algorithm to solve the optimization problem of the objective function.
3. Under reasonable assumptions, we give theoretical properties of this method. In addition, we verify the robustness and effectiveness of the variable selection method through numerical simulation studies. The numerical study shows that the method is more robust than other comparative methods in variable selection and parameter estimation when outliers or noise are presented in the observations.

The rest of this paper is organized as follows. In Section 2, we develop the methodology for variable selection with exponential squared loss and give the theoretical properties of the proposed method in Section 3. In Section 4, we present the related algorithms. The experimental results are carried out in Section 5, and we conclude the paper in Section 6. All of the details of the proofs of the main theorems are collected in the Appendix A.

## 2. Methodology

### 2.1. Model Setup

Consider the following spatial single-index varying-coefficient model:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + g_1(\mathbf{U}_i^T \boldsymbol{\alpha}) z_{i1} + g_2(\mathbf{U}_i^T \boldsymbol{\alpha}) z_{i2} + \cdots + g_q(\mathbf{U}_i^T \boldsymbol{\alpha}) z_{iq} + \varepsilon_i \quad (1)$$

where  $y_i$  is the response variable,  $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iq})^T$  is the  $q$ -dimensional of the observed variable, and  $\mathbf{U}_i = (U_{i1}, U_{i2}, \dots, U_{im})^T$  is the  $m$ -dimensional spatial location parameter. The  $n \times n$  matrix of the spatial weights matrix  $W$  in dimensional space is  $w_{ij}$ .

$\rho$  and  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)^T$  are the parameters to be estimated. It is natural to suppose that  $\varepsilon_i$  is independent and subject to a mean value of zero and a variance of  $\sigma^2$ .  $g(\cdot)$  is an unknown function. For the identifiability of the model, it is assumed that  $\|\alpha\| = 1$  and the first nonzero element of  $\alpha$  is positive.

It can be seen from the model (1) that the spatial single-index varying-coefficient model is a semi-parametric varying-coefficient model, and the unknown function  $g$  changes with the transformation of geographical location. When  $\rho = 0$ , the model becomes the partial linear single-index varying-coefficient model. When  $z_{i1} = 1$  and  $g_1(U_i^T \alpha) = U_i^T \alpha$  while the other  $g(\cdot) = 0$ , the model becomes the SAR model.

### 2.2. Basis Function Expansion

Since  $g(\cdot)$  is unknown, we replace  $g(\cdot)$  with its basis function approximations. The specific estimation steps are as follows:

**Step 1.** The initial value  $\alpha_0$  should be given. This paper uses the method proposed by [12]. We roughly calculate the estimated value of  $\alpha$  by the linear regression model:

$$y_i = U_i^T \alpha z_{i1} + U_i^T \alpha z_{i2} + \dots + U_i^T \alpha z_{iq} + \varepsilon_i$$

set the estimated value of  $\alpha$  as  $\alpha_0$ , in which  $\|\alpha_0\| = 1$  and the first nonzero element in  $\alpha_0$  is positive.

**Step 2.** Set  $a \leq k_1 < k_2 < \dots < k_l \leq b$  as  $l$  nodes on the interval  $[a, b]$ . By the initial value  $\alpha_0$ , let  $t_i = U_i^T \alpha_0$ , then the radial basis function of degree  $p$  is

$$\delta(t_i) = \left(1, t_i, t_i^2, \dots, t_i^{p-1}, |t_i - k_1|^{2p-1}, |t_i - k_2|^{2p-1}, \dots, |t_i - k_l|^{2p-1}\right)^T$$

Suppose that the coefficient of the radial basis function is

$$\gamma_{1-i} = (\gamma_{1-i0}, \gamma_{1-i1}, \dots, \gamma_{1-i(p-1)}, \gamma_{1-ip}, \dots, \gamma_{1-i(p+l-1)})^T,$$

then, the  $s$ th unknown function  $g_{is}(t_i) \approx \delta(t_i)^T \gamma_{1-s}$ , where  $i = 1, 2, \dots, n, s = 1, 2, \dots, q$ . Substituting the radial basis function into model (1), we can obtain the following:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + z_{i1} \delta(t_i)^T \gamma_{1-1} + z_{i2} \delta(t_i)^T \gamma_{1-2} + \dots + z_{iq} \delta(t_i)^T \gamma_{1-q} + \varepsilon_i \quad (2)$$

Let  $Y = (y_1, y_2, \dots, y_n)^T, \gamma_1 = (\gamma_{1-1}^T, \gamma_{1-2}^T, \dots, \gamma_{1-n}^T)^T, D = (D_1, D_2, \dots, D_n)^T$ , where  $D_i = (z_{i1} \delta(t_i)^T, z_{i2} \delta(t_i)^T, \dots, z_{iq} \delta(t_i)^T)^T, \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ , then the matrix form of the model (2) is

$$Y = \rho W Y + D \gamma_1 + \varepsilon \quad (3)$$

As can be seen from model (3), model (1) is transformed from the spatial single-index varying-coefficient model to the classical SAR model under the fitting of the radial basis function. The theory of the SAR model is relatively well-equipped, and the exponential squared loss-based variable selection method for the SAR model is used to estimate the unknown parameters.

### 2.3. The Penalized Robust Regression Estimator

Now, we consider the variable selection for the model (3). To guarantee the model identifiability and to improve the model fitting accuracy and interpretability, we normally assume that the true regression coefficient vector  $\alpha^*$  is sparse with only a small proportion of nonzeros [13,14]. It is natural to employ the penalized method that simultaneously selects important variables and estimates the values of parameters. The constructed model is recast as follows:

$$\min L(\gamma_1, \rho) = \frac{1}{n} \sum_{i=1}^n \phi_{\gamma_2}(Y_i - \rho \tilde{Y}_i - D_i \gamma_1) + \lambda \sum_{j=1}^q P(|\gamma_{1-j}|) \tag{4}$$

where  $\lambda > 0$ ,  $\tilde{Y} = WY$ ,  $\sum_{j=1}^q P(|\gamma_{1-j}|)$  is a penalty term,  $\phi_{\gamma_2(\cdot)}$  is the exponential squared loss function:  $\phi_{\gamma_2(t)=1-\exp(-t^2/\gamma_2)}$ , in which  $\gamma_2$  is the tuning parameter controlling the degree of robustness.

Concerning the choice of the penalty term. The lasso or adaptive lasso penalty could be considered if there is no extra structured information. Assume that  $\hat{\gamma}_1$  is a root- $n$ -consistent estimator for  $\gamma_1$ , for instance, the naive least square estimator  $\hat{\gamma}_{1(ols)}$ . Define the weight vector  $\eta \in \mathbb{R}^q$  with  $\eta_j = 1/|\hat{\gamma}_{1-j}|^r$  ( $j = 1, \dots, q$ ),  $r > 0$ , and then we set  $r = 1$  in this paper as suggested by [15]. An adaptive lasso penalty is described as

$$\sum_{j=1}^q P(|\gamma_{1-j}|) = \sum_{j=1}^q \eta_j |\gamma_{1-j}|. \tag{5}$$

The objective function of penalized robust regression that consists of exponential squared loss and an adaptive lasso penalty is formulated as

$$\min L(\gamma_1, \rho) = \frac{1}{n} \sum_{i=1}^n \phi_{\gamma_2}(Y_i - \rho \tilde{Y}_i - D_i \gamma_1) + \lambda \sum_{j=1}^q \eta_j |\gamma_{1-j}| \tag{6}$$

The selection of tuning parameter  $\gamma_2$  and regularization parameter  $\lambda$  is discussed in Section 4.

#### 2.4. Estimation of the Variance of the Noise

Set  $H = (I_n - \rho W)^{-1}$ , then the variance of the noise is estimated as

$$\hat{\sigma}^2 = \frac{1}{n} (Y - HD\gamma_1)^T (HH^T)^{-1} (y - HD\gamma_1), \tag{7}$$

where  $\rho$  and  $\gamma_1$  could be estimated by the solutions of (6). It is pointed out that  $H$  is a non-singular matrix, then  $(HH^T)^{-1} = (H^T)^{-1}H^{-1} = (I_n - \rho W)^T(I_n - \rho W)$ . Let  $u = HD\gamma_1$ , then  $u = HD\gamma_1 = (I_n - \rho W)^{-1}(D\gamma_1)$  and then  $\hat{\sigma}^2$  defined by (7) can be computed by

$$\hat{\sigma}^2 = \frac{1}{n} \|(I_n - \rho W)(Y - u)\|_2^2 \tag{8}$$

### 3. Theoretical Properties

To discuss the theoretical properties, let the parameters  $\theta = (\rho, \gamma_1^T)^T$ , with  $\theta_0, \alpha_0$  and  $g_0(\cdot)$ , be the true values of  $\theta, \alpha$  and  $g(\cdot)$ . It is generally assumed that  $\alpha_{l0} = 0, l = s + 1, \dots, p$ , and  $\alpha_{l0}, l = 1, \dots, s$ , are all nonzero parts of  $\alpha_0$ . Moreover, we assume that  $g_{j0} = 0, j = d + 1, \dots, q$ , and  $g_{j0}, j = 1, \dots, d$  are all nonzero parts of  $g(\cdot)$ . Set  $\phi = (\alpha_2, \alpha_3, \dots, \alpha_m)^T, \alpha(\phi) = (\sqrt{1 - \|\phi\|^2}, \phi^T)^T$ ; the real parameters of  $\phi_0$  satisfy  $\|\phi_0\| < 1$ . Hence,  $\alpha_\phi$  is differentiable within the neighborhood of  $\phi_0$ , and the Jacobian matrix is

$$J_\phi = \begin{pmatrix} -(1 - \|\phi\|^2)^{-1/2} \phi^T \\ I_{m-1} \end{pmatrix} \tag{9}$$

Assumption:

- (C1) The density function  $f(t)$  of  $U\alpha$  is uniformly bounded on  $T = \{t = U\alpha\}$  and far from 0. Furthermore,  $f(t)$  is assumed to satisfy the Lipschitz condition of order 1 on  $T$ .
- (C2) The function  $g_j(t), j = 1, \dots, q$ , has bounded and continuous derivatives up to order  $r(\geq 2)$  on  $T$ , where  $g_j(t)$  is the  $j$ th components of  $g(t)$ .

- (C3)  $E(\|U^6\|) < \infty, E(\|Z^6\|) < \infty$  and  $E(|\varepsilon|^6) < \infty$ .
- (C4)  $\{(y_i, U_i, z_i), 1 \leq i \leq n\}$  is a strictly stationary and strongly mixing sequence with coefficient  $\gamma(n) = O(\xi^n)$ , where  $0 < \xi < 1$ .
- (C5) Let  $c_1, \dots, c_K$  be the interior knots of  $[a, b]$ , where  $a = \inf\{t : t \in T\}, b = \sup\{t : t \in T\}$ . Moreover, we set  $c_0 = a, c_{K+1} = b, h_i = c_i - c_{i-1}, h = \max\{h_i\}$ . Then, a positive constant  $C_0$  exists such that

$$\frac{h}{\min\{h_i\}} < C_0, \quad \max\{h_{i+1} - h_i\} = o(K^{-1}).$$

- (C6) Let  $b_n = \max_j\{|\dot{p}_j(|\gamma_{1-j0}|) : \gamma_{1-j0} \neq 0\}$  and then  $b_n \rightarrow 0$  as  $n \rightarrow \infty$ . Further, let  $\lim_{n \rightarrow \infty} \inf_0 \lim_{|\gamma_{1-j}| \rightarrow 0} \lambda_j^{-1} |\dot{p}_j(|\gamma_{1-j}|)| > 0$ , where  $j = d + 1, \dots, q$ .
- (C7)  $H(\rho) = (I_n - \rho W)^{-1}$  is a nonsingular matrix, invertible for any  $\rho \in \Theta$ ,  $\Theta$  is a compact parameter space, and the absolute row and column sums of  $H(\rho), H(\rho)^{-1}$  are uniformly bounded on  $\rho \in \Theta$ ;
- (C8) Let

$$I(\phi, \gamma_1; \gamma_2) = \frac{2}{\gamma_2} \int G(\phi) G^T(\phi) e^{-r^2/\gamma_2} \left(\frac{2r^2}{\gamma_2} - 1\right) dF(G, y)$$

where  $r = Y - (I_n - \rho W)^{-1} D(\phi) \gamma_1 = Y - G(\phi) \gamma_1, G(\phi) = (I_n - \rho W)^{-1} D(\phi)$ . Suppose that  $I(\phi, \gamma_1; \gamma_2)$  is negative definite.

- (C9)  $\Sigma = E(GG^T)$  is positive definite.

Under the above preparations, we give the following sampling properties for our proposed estimators. The following theorem presents the consistency of the penalized exponential squared loss estimators.

**Theorem 1.** Assume that conditions C1 ~ C9 hold and the number of knots  $K = O(n^{1/(2r+1)})$ . Further, we suppose that  $\gamma_{2-n} - \gamma_{2-0} = o_p(1)$  for some  $\gamma_{2-0} > 0$  and  $I(\phi_0, \gamma_{1-0}; \gamma_{2-0})$  is negative definite. Then,

- (i)  $\|\alpha - \alpha_0\| = O_p(n^{-1/(2r+1)} + a_n)$ ;
- (ii)  $\|\hat{g}_j(\cdot) - g_{j0}(\cdot)\| = O_p(n^{-r/(2r+1)} + a_n)$ , for  $j = 1, \dots, q$ ,

where  $a_n = \max_j\{|\dot{p}_j(|\gamma_{1-j0}|) : \gamma_{1-j0} \neq 0\}$ ,  $r$  is defined in condition (C2), and  $\dot{p}_\lambda(\cdot)$  represents the first order derivative of  $p_\lambda(\cdot)$ .

In addition, we have proved that when some suitable conditions hold, the consistent estimation must be sparse, as described below.

**Theorem 2.** Suppose that conditions C1 ~ C9 hold, and the number of knots  $K = O(n^{1/(2r+1)})$ . We assume that  $\sqrt{n}a_n = O_p(1)$  and  $\sqrt{n}(\gamma_{2-n} - \gamma_{2-0}) = o_p(1)$ . Let

$$\lambda_{j(\max)} \rightarrow 0, \quad n^{r/(2r+1)} \lambda_{j(\min)} \rightarrow \infty (n \rightarrow \infty).$$

Then, with probability approaching 1,  $\hat{\alpha}$  and  $\hat{g}(\cdot)$  satisfy

- (i)  $\hat{\alpha}_l = 0, l = s + 1, \dots, p$ ;
- (ii)  $\hat{g}_j(\cdot) = 0, j = d + 1, \dots, q$ .

We then show that the estimators of nonzero coefficients for the parameter components have the same asymptotic distribution as the estimators based on the correct submodel. Set

$$\alpha^* = (\alpha_1, \dots, \alpha_s)^T, \quad g^*(t) = (g_1^T(t), \dots, g_d^T(t))^T,$$

and let  $\alpha_0^*$  and  $g_0^*(t)$  be true values of  $\alpha^*$  and  $g^*$ , respectively. Corresponding covariates are denoted by  $U_i^*$  and  $Z_i^*$ ,  $i = 1, \dots, n$ . Furthermore, let  $\Sigma_2 = \text{cov}\left(\exp(-r^2/\gamma_{2-0}) \frac{2r}{\gamma_{2-0}} G_{i1}\right)$ ,  $\Sigma_1 = \text{diag}\{\dot{p}_{\lambda_1}(|\gamma_{1-01}^*|), \dots, \dot{p}_{\lambda_d}(|\gamma_{1-0d}^*|)\}$ ,

$$\Delta = (\dot{p}_{\lambda_1}(|\gamma_{1-01}^*|) \text{sign}(\gamma_{1-01}^*), \dots, \dot{p}_{\lambda_d}(|\gamma_{1-0d}^*|) \text{sign}(\gamma_{1-0d}^*))^T,$$

$$I_1(\phi_{01}^*, \gamma_{1-01}^*; \gamma_{2-0}) = \frac{2}{\gamma_{2-0}} E\left[\exp(-r^2/\gamma_{2-0}) \left(\frac{2r^2}{\gamma_{2-0}} - 1\right)\right] (EG_{i1}G_{i1}^T).$$

The following result presents the asymptotic properties of  $\hat{\alpha}^*$ .

**Theorem 3.** *If the assumptions of Theorem 2 hold, we have*

$$\sqrt{n}(I_1(\phi_{01}^*, \gamma_{1-01}^*; \gamma_{2-0}) + \Sigma_1)\{\hat{\alpha}^* - \alpha_0^* + (I_1(\phi_{01}^*, \gamma_{1-01}^*; \gamma_{2-0})^{-1}\Delta)\} \xrightarrow{L} N\left(\mathbf{0}, J_{\phi_0^*}\Sigma_2J_{\phi_0^*}^T\right)$$

where ' $\xrightarrow{L}$ ' represents the convergence in distribution.

Theorems 1 and 2 show that the proposed variable selection procedure is consistent, and Theorems 1 and 3 show that the penalized estimators have the oracle property. This demonstrates that if the subset of true zero coefficients are known, the penalty estimators perform well.

#### 4. Algorithm

In this section, we talk about a feasible algorithm for the solution of (6). A data-driven procedure for  $\gamma_2$  and a simple selection method for  $\lambda$  are considered. Moreover, effective optimization algorithms have been composed to solve non-convex and non-differentiable objective functions.

##### 4.1. Choice of the Tuning Parameter $\gamma_2$

The tuning parameter  $\gamma_2$  controls the level of robustness and performance of the proposed robust regression estimators. Ref. [16] propose a data-driven procedure to choose  $\gamma_2$  for ordinary regression. We follow its steps and apply it to the spatial single-index varying-coefficient model. Firstly, a set of tuning parameters is determined to ensure that the proposed penalized robust estimators have an asymptotic breakdown point at 1/2. Then, the tuning parameter is selected with the maximum efficiency.

The whole procedures are presented as follows:

**Step 1.** Initialize  $\hat{\rho} = \rho^{(0)}$  and  $\hat{\gamma}_1 = \gamma_1^{(0)}$ . Set  $\rho^{(0)} = \frac{1}{2}$ ,  $\gamma_1^{(0)}$  a robust estimator. The model  $Y = \rho WY + D\gamma_1 + \epsilon$  can be recasted as  $Y^* = D\gamma_1 + \epsilon$ , where  $Y^* = Y - \rho WY$ .

**Step 2.** Find the pseudo outlier set of the sample:

Let  $A_n = \{(D_1, Y_1^*), \dots, (D_n, Y_n^*)\}$ . Calculate  $r_i(\hat{\gamma}_1) = Y_i^* - D_i\hat{\gamma}_1$ ,  $i = 1, \dots, n$  and  $S_n = 1.4826 \times \text{median}_i |r_i(\hat{\gamma}_1) - \text{median}_j(r_j(\hat{\gamma}_1))|$ . Then, take the pseudo outlier set  $A_m = \{(D_i, Y_i) : |r_i(\hat{\gamma}_1)| \geq 2.5S_n\}$ , set  $m = \#\{1 \leq i \leq n : |r_i(\hat{\gamma}_1)| \geq 2.5S_n\}$ , and  $A_{n-m} = A_n/A_m$ .

**Step 3.** Select the tuning parameter  $\gamma_{2-n}$ : construct  $\hat{V}(\gamma_2) = \{\hat{I}(\hat{\gamma}_1)\}^{-1}\tilde{\Sigma}_2\{\hat{I}(\hat{\gamma}_1)\}^{-1}$ , in which

$$\hat{I}(\hat{\gamma}_1) = \frac{2}{\gamma_2} \left\{ \frac{1}{n} \sum_{i=1}^n \exp(-r_i^2(\hat{\gamma}_1)/\gamma_2) \left(\frac{2r_i(\hat{\gamma}_1)}{\gamma_2} - 1\right) \right\} \cdot \left(\frac{1}{n} \sum_{i=1}^n D_i D_i^T\right)$$

$$\tilde{\Sigma}_2 = \text{Cov}\left\{\exp(-r_1^2(\hat{\gamma}_1)/\gamma_2) \frac{2r_1(\hat{\gamma}_1)}{\gamma_2} D_1, \dots, \exp(-r_n^2(\hat{\gamma}_1)/\gamma_2) \frac{2r_n(\hat{\gamma}_1)}{\gamma_2} D_n\right\}.$$

Let  $\gamma_{2-n}$  be the minimizer of  $\det(\hat{V}(\gamma_2))$  in the set  $G = \{\gamma_2 : \zeta(\gamma_2) \in (0, 1]\}$ , where  $\zeta(\cdot)$  has the same definition in [8] and  $\det(\cdot)$  means the determinant operator.

**Step 4.** Update  $\hat{\rho}$  and  $\hat{\gamma}_1$  as the optimal solution of  $\min \sum_{i=1}^n \phi_{\gamma_2}(Y_i - \rho \tilde{Y}_i - D_i \gamma_1)$ , where  $\tilde{Y} = WY$ . Repeat step 2 to step 4 until convergence.

It is noted that an initial robust estimator  $\gamma_1^{(0)}$  is needed in the initial step above. In practice, we make the estimator of the LAD loss as the initial estimator. In this sense, the selection of  $\gamma_2$  does not depend on  $\lambda$  basically. Meanwhile, one could also select the two parameters  $\gamma_2$  and  $\lambda$  jointly by cross-validation as discussed in [8]. Nevertheless, this approach needs huge computation. Moreover, the candidate interval of  $\gamma_2$  is  $\{\gamma_2 : \zeta(\gamma_2) \in (0, 1]\}$ . Practically, we find the threshold of  $\gamma_{2-1}$  subject to  $\zeta(\gamma_{2-1}) = 1$ . The choice of  $\gamma_2$  is usually located in the interval of  $[5\gamma_{2-1}, 30\gamma_{2-1}]$ .

4.2. Choice of the Regularization Parameter  $\lambda$  and  $\eta_j$

With regard to the choice of the regularization parameter  $\lambda$  and  $\eta_j$  in (6), as the parameter  $\lambda$  can be unified with  $\eta_j$ , we set  $\lambda_j = \lambda \cdot \eta_j$ . Generally, many methods can be applied to select  $\lambda_j$ , such as AIC, BIC, and cross-validation. To ensure that variable selection is consistent and that the intensive computation can be reduced, we propose the regularization parameter by minimizing a BIC-type objective function as [16]:

$$\sum_{i=1}^n \left[ 1 - \exp\left\{ -\frac{(Y_i^* - D_i \gamma_1)^2}{\gamma_{2-n}} \right\} \right] + n \sum_{j=1}^q \lambda_j |\gamma_{1-j}| - \sum_{j=1}^q \log(0.5n\lambda_j) \log(n) \quad (10)$$

where  $Y_i^* = Y_i - \rho W_n Y_i$ . This results in  $\lambda_j = \frac{\log(n)}{n^{|\gamma_{1-j}|}}$ .  $\gamma_{1-j}$  can be easily estimated by the unpenalized exponential squares loss estimator  $\widehat{\gamma}_{1-j}$ , where the parameter value of  $\gamma_2$  has been estimated as described in Section 4.1. Note that this simple choice satisfies the conditions  $\sqrt{n}\widehat{\lambda}_j \rightarrow 0$  for  $j \leq d$  and  $\sqrt{n}\widehat{\lambda}_j \rightarrow \infty$  for  $j > d$ , with  $d$  the number of nonzeros in the true value of  $\gamma_1$ . Thus, the consistent variable selection is ensured by the final estimator.

4.3. Block Coordinate Descent (BCD) Algorithm

We seek to compose an effective algorithm to solve the objective function (6). Finding an effective algorithm is difficult because the optimization problem is non-convex and non-differentiable. We embark on using the BCD algorithm proposed by [17] and then overcome the above challenges. The BCD algorithm framework is shown in Algorithm 1 specifically.

---

**Algorithm 1** The block coordinate descent (BCD) algorithm

---

1. Set initial value for  $\gamma_1^0 \in \mathbb{R}^p$  and  $\rho^0 \in (0, 1)$ ;
2. **repeat** for  $k = 0, 1, 2, \dots$
3. Solve the subproblem about  $\rho$  with initial point  $\rho^k$ :

$$\rho^{k+1} \leftarrow \min_{\rho \in [0,1]} L(\gamma_1^k, \rho) \quad (11)$$

4. Solve the subproblem with initial value  $\gamma_1^k$ ,

$$\min_{\gamma_1 \in \mathbb{R}^q} L(\gamma_1, \rho^{k+1}) \quad (12)$$

to get a solution  $\gamma_1^{k+1}$ , ensuring that  $L(\gamma_1^k, \rho^{k+1}) - L(\gamma_1^{k+1}, \rho^{k+1}) \leq 0$ , and  $\gamma_1^{k+1}$  is a stationary point of  $L(\gamma_1, \rho^{k+1})$ .

5. **until** convergence.
-

4.4. DC Decomposition and CCCP Algorithm

An elemental observation for problem (12) is that the exponential squared loss function is a DC function, and the lasso or the adaptive lasso penalty function is convex. As a result, problem (12) is a DC programming. It can be solved by the following algorithms.

We first analyze whether the exponential squared loss function  $\phi_{\gamma_2}(t)$  can be denoted as the difference of two convex functions:

$$\phi_{\gamma_2}(t) := [\phi_{\gamma_2}(t) + v(t)] - v(t) := u(t) - v(t) \tag{13}$$

where  $\phi_{\gamma_2}(t) = 1 - e^{-\frac{t^2}{\gamma_2}}$ ,  $v(t) = \frac{1}{3\gamma_2^2}t^4$ ,  $u(t) = \phi_{\gamma_2}(t) + v(t)$ .

Set

$$\begin{aligned} J_{\text{vex}}(\gamma_1) &= \frac{1}{n} \sum_{i=1}^n u\left(Y_i - \rho^k \langle w_i, Y \rangle - D_i \gamma_1\right) + \lambda \sum_{j=1}^q P(|\gamma_{1-j}|) \\ J_{\text{cav}}(\gamma_1) &= \frac{1}{n} \sum_{i=1}^n v\left(Y_i - \rho^k \langle w_i, Y \rangle - D_i \gamma_1\right) \end{aligned} \tag{14}$$

in which  $u(\cdot)$ ,  $v(\cdot)$  is defined in (13),  $w_i$  is in the  $i$ th row of the weight matrix  $W$ , and  $\sum_{j=1}^q P(|\gamma_{1-j}|)$  a convex penalty with regard to  $\gamma_1$ . Then,  $J_{\text{vex}}(\cdot)$  and  $J_{\text{cav}}(\cdot)$  are convex and concave functions, respectively. Subproblem (12) is recast as follows:

$$\min_{\gamma_1 \in \mathbb{R}^n} L(\gamma_1, \rho^k) = J_{\text{vex}}(\gamma_1) + J_{\text{cav}}(\gamma_1), \tag{15}$$

Furthermore, it can be solved by the concave–convex procedure algorithm structure proposed by [18] as shown in Algorithm 2.

---

**Algorithm 2** The Concave–Convex Procedure

---

1. Initialize  $\gamma_1^0$ . Set  $k = 0$ .
2. **repeat** for  $k = 0, 1, 2, \dots$
- 3.

$$\gamma_1^{k+1} = \operatorname{argmin}_{\gamma_1} J_{\text{vex}}(\gamma_1) + J'_{\text{cav}}(\gamma_1^k) \cdot \gamma_1 \tag{16}$$

4. **until** convergence of  $\gamma_1^k$ .
- 

We focus on the lasso and the adaptive lasso penalty. Since  $J'_{\text{cav}}(\gamma_1^k) \cdot \gamma_1$  is linear to  $\gamma_1$ , according to the definition in (15), the objective function of (16) can be expressed as

$$\min_{\gamma_1 \in \mathbb{R}^q} \psi(\gamma_1) + \lambda \sum_{i=1}^q P(|\gamma_{1-i}|), \tag{17}$$

where  $\psi(\gamma_1)$  is a convex and continuously differentiable function,  $\sum_{i=1}^q P(|\gamma_{1-i}|)$  is the lasso penalty,  $\sum_{i=1}^q |\gamma_{1-i}|$ , or the more general adaptive lasso penalty:  $\sum_{i=1}^q \eta_i |\gamma_{1-i}|$ ,  $\eta_i \geq 0, i = 1, \dots, q$ . Therefore, we can refer to an efficient algorithm ISTA and FISTA proposed by [19] to solve the model with a framework (17) for the lasso penalty. The iterative steps of ISTA is simply  $\gamma_1^k = \Theta_L(\gamma_1^{k-1})$ , where  $L$  is the unknown Lipschitz constant. FISTA is an accelerated version of ISTA that has been shown to have a better convergence rate in theory and practice, proven by [19]. Ref. [17] extended it to solve the model by adaptive lasso penalty and can ensure numerical efficiency.

Now consider solving subproblem (11) to update  $\rho_k$ . Since problem (11) minimizes a function of univariate variable, we employ the classical golden section search algorithm based on parabolic interpolation (see [20] for details).

In accordance with Beck and Teboulle, the value of the iterative function generated by FISTA for solving the subproblem (16) of CCCP converges to the optimal function value at



the speed of  $O(1/k^2)$ , with an iteration step of  $k$ . The ordinary termination criterion of ISTA and FISTA is  $\frac{\|\gamma_1^k - \gamma_1^{k-1}\|}{\max\{\|\gamma_1^k\|, 1\}} \leq \text{tol}_{\gamma_1}$ , where  $\text{tol}_{\gamma_1}$  is a tolerance approaching zero and greater than zero. Under the criterion of either  $\|\gamma_1^k - \gamma_1^{k-1}\| \leq \epsilon_1$ , or  $\|L(\gamma_1^k) - L(\gamma_1^{k+1})\| \leq \epsilon_2$ , Algorithm 1 terminates. Therefore, to obtain an optimal solution of  $\epsilon$ , the required iterations of the FISTA algorithm are  $O(1/\sqrt{\epsilon})$  and the gradient  $\nabla\psi(\gamma_1)$  of (17) is computed for each iteration. Suppose that the BCD algorithm converges with a specified number of iterations and the CCCP algorithm terminates at most  $m$  times in each iteration. Since  $O(np)$  computation is needed to calculate the gradient  $\nabla\psi(\gamma_1)$ , the total computational complexity is  $O(mnp/\sqrt{\epsilon})$ .

### 5. Simulation Studied

In this section, we conduct numerical studies to illustrate the performance of the proposed method, including the cases of normal data and noisy data.

#### 5.1. Simulation Sampling

The data is generated from model (1). We set  $\alpha = (\alpha_1, \alpha_2, 0_q)^T$ , where  $(\alpha_1, \alpha_2)$  generates from a 2-dimensional normal distribution of the mean vector  $(0.6, 0.8)$  and covariance matrix  $0.01 \cdot I_2$ , with  $I_2$  the unit matrix  $\in \mathbb{R}^{2 \times 2}$ ,  $0_q$  is the zero vector of  $q$  dimension. Set the sample size  $n \in \{25, 144, 324\}$ , and spatial coefficient  $\rho$  is generated by uniform distribution on interval  $[\rho_1 - 0.1, \rho_1 + 0.1]$ , where  $\rho_1 \in \{0.8, 0.5, 0.2\}$ . For comparison's sake, we also consider  $\rho = 0$ , which means that there is no spatial dependency, and model (1) changes into the normal single-index varying-coefficient model.

The variable  $Y_n$  follows  $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ , in which  $\sigma^2$  is generated from a uniform distribution by  $\sigma_1 \in \{1, 2\}$  on interval  $[\sigma_1 - 0.1, \sigma_1 + 0.1]$ . We also consider the case when there are outliers in the response. The error term follows a mixed normal distribution  $(1 - \delta_1) \cdot \mathcal{N}(0, 1) + \delta_1 \cdot \mathcal{N}(10, 6^2)$ , where  $\delta_1 \in \{0.01, 0.05\}$ .  $z_{ij}$  is independent and randomly taken from the normal distribution  $N(0, 1)$ , and the space weight matrix  $W_n = I_R \otimes B_m$ , where  $B_m = (1/(m - 1))(\mathbf{1}_m \cdot \mathbf{1}_m^T - I_m)$ ,  $\otimes$  is a Kronecker product, and  $\mathbf{1}_m$  is the  $m$ -dimensional column vector of ones. We take different values of  $m = 2$  and  $R$ , where  $R = 20, 100$ .

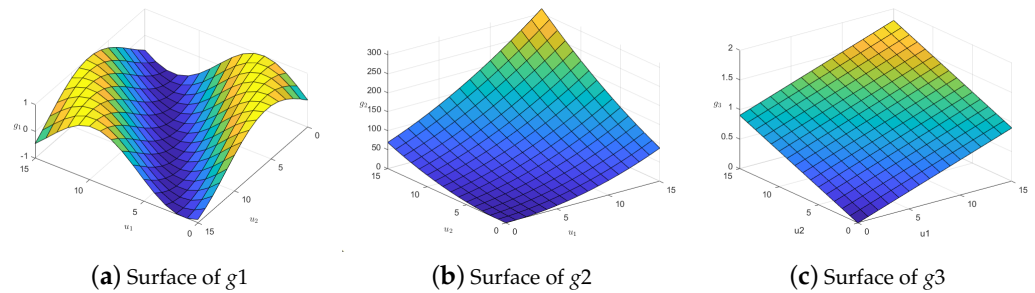
Moreover, we construct the spatial location information, where two-dimensional plane coordinates are used in this paper. Take a square to simulate the geographical area object, set the end point of the lower left corner of the square as the origin, and establish a rectangular coordinate system along the horizontal and vertical directions. Each side is divided into  $h - 1$  equal points, and corresponding equal points are connected along the horizontal and vertical axes to form  $h * h$  crossing points (including the equal points of each square side). Each crossing point is the geographical location point. Set sampling capacity  $n = h^2$ ; then, the geographical location coordinate  $U_i = (u_{i1}, u_{i2}, \dots, u_{iq})^T$  is expressed as:

$$U_i = (0.5\text{mod}(i - 1, h), 0.5\text{floor}(i - 1, h), 0, \dots, 0)^T$$

where  $\text{mod}$  and  $\text{floor}$  are the representations of built-in functions in MATLAB,  $\text{mod}(i - 1, h)$  represents the remainder of  $i - 1$  divided by  $h$ , and  $\text{floor}(i - 1, h)$  represents the integer part of the quotient of  $i - 1$  divided by  $h$ . We set

$$g(t) = (g_1(t), \dots, g_q(t))^T = (\sin(t), 3t^2, \frac{1}{6}t, 0, \dots, 0)^T$$

The true surface of the three coefficient functions is shown in Figure 1.



**Figure 1.** Real surfaces of coefficient functions.

Another important problem of the spatial single-index varying-coefficient model is the estimation of weight matrix  $W$ . Since  $W \in \mathbb{R}^{n \times n}$  is composed of the correlation of every two observations, it is usually difficult to obtain an accurate estimation of the weight matrix  $W$  in practical applications. In order to confirm the effect of inaccurate estimation of the matrix  $W$ , we randomly remove 30%, 50%, and 80% non-zero weights from each row of the true weight matrix  $W$ , respectively.

For each case of the simulation experiment, all of the results shown below are averaged over 100 replications to avoid unintended effects. We adopt the node selection method proposed by [12], with  $step = 10$  and the number of radial basis functions  $p = 3$ .

5.2. Simulation Results

The evaluation of simulation results is shown as follows. We use the median of squared error (MedSE) proposed by [21]. It is defined as  $\|\alpha - \hat{\alpha}\|^2$  in this paper, where  $\|\alpha\| = \sqrt{\sum_{i=1}^n \alpha_i^2}$ ,  $\alpha = (\alpha_1, \dots, \alpha_n)$ ,  $\hat{\alpha}$  is the estimator of  $\alpha$ . The square root of mean deviation (MAISE) is used as the evaluation index for the unknown function. Specifically,  $MAISE = mcn^{-1} \sum_{i=1}^{mcn} \sqrt{n^{-1} \sum_{j=1}^n (g_t(U\alpha) - \delta(U\hat{\alpha})^T \gamma_{1-t})^2}$ ,  $t = 1, 2, \dots, q$ , where  $mcn$  represents the total simulation times of the model, and  $t$  represents the  $t$ th unknown function of the model,  $g_t(\cdot)$ . The smaller the value of each index, the higher the accuracy of parameter estimation and the better fitting effect of the unknown function.

Table 1 illustrates the results of the estimated coefficient by the spatial single-index varying-coefficient model with  $q = 5$ , the null penalty term, and Gaussian noise in  $y$ , where "E", "S", and "L" indicate the exponential squared loss, the square loss, and the LAD loss, respectively. It is shown that both of the three loss functions bring nonzero estimates of  $\alpha_1$  and  $\alpha_2$ , which are close to the true values (the mean of the true values of  $\alpha_1$  and  $\alpha_2$  are 0.6, 0.8 resp.). Comparatively, the model with the square loss produces the most accurate estimation. As the sample size increases, all three loss functions bring an accurate estimate of  $\alpha$  and  $\sigma^2$ .

Table 2 presents the results of the estimated coefficient by the spatial single-index varying-coefficient model when the dimension is comparatively close to the sample size. Similar results in Table 1 have been observed, except for  $q = 5$ . As the sample size is not enough compared with the dimension, these results are as expected.

**Table 1.** Estimation with no regularizer on normal data ( $q = 5$ ).

|                              | $n = 25, q = 5$ |        |        | $n = 144, q = 5$ |        |        | $n = 324, q = 5$ |        |        |
|------------------------------|-----------------|--------|--------|------------------|--------|--------|------------------|--------|--------|
|                              | E+null          | S+null | L+null | E+null           | S+null | L+null | E+null           | S+null | L+null |
| $\rho_1 = 0.8, \sigma_1 = 1$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.80            | 0.61   | 0.83   | 0.50             | 0.66   | 0.28   | 0.53             | 0.62   | 0.67   |
| $\alpha_2$                   | 0.88            | 0.61   | 0.75   | 0.78             | 0.77   | 0.66   | 0.74             | 0.79   | 0.96   |
| $\hat{\rho}$                 | 0.80            | 0.94   | 0.75   | 0.90             | 0.80   | 0.91   | 0.88             | 0.89   | 0.84   |
| $\hat{\sigma}^2$             | 0.45            | 0.78   | 0.82   | 0.79             | 0.72   | 0.70   | 0.86             | 0.71   | 0.80   |
| MedSE                        | 0.28            | 0.44   | 0.70   | 0.22             | 0.23   | 0.42   | 0.17             | 0.16   | 0.19   |
| $\rho_1 = 0.5, \sigma_1 = 1$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.69            | 0.65   | 0.64   | 0.49             | 0.66   | 0.28   | 0.53             | 0.63   | 0.67   |
| $\alpha_2$                   | 0.82            | 0.61   | 0.88   | 0.78             | 0.76   | 0.72   | 0.74             | 0.80   | 0.96   |
| $\hat{\rho}$                 | 0.52            | 0.61   | 0.50   | 0.68             | 0.40   | 0.74   | 0.62             | 0.65   | 0.55   |
| $\hat{\sigma}^2$             | 0.44            | 0.82   | 0.71   | 0.83             | 0.71   | 0.74   | 0.89             | 0.73   | 0.82   |
| MedSE                        | 0.22            | 0.45   | 0.77   | 0.22             | 0.23   | 0.41   | 0.17             | 0.16   | 0.19   |
| $\rho_1 = 0.2, \sigma_1 = 1$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.67            | 0.66   | 0.57   | 0.51             | 0.65   | 0.32   | 0.53             | 0.65   | 0.65   |
| $\alpha_2$                   | 0.81            | 0.59   | 0.96   | 0.80             | 0.75   | 0.72   | 0.76             | 0.81   | 0.98   |
| $\hat{\rho}$                 | 0.14            | 0.00   | 0.27   | 0.33             | 0.00   | 0.50   | 0.19             | 0.26   | 0.16   |
| $\hat{\sigma}^2$             | 0.44            | 0.82   | 0.68   | 0.87             | 0.70   | 0.78   | 0.91             | 0.76   | 0.84   |
| MedSE                        | 0.23            | 0.49   | 0.72   | 0.21             | 0.23   | 0.41   | 0.17             | 0.16   | 0.22   |
| $\rho_1 = 0, \sigma_1 = 1$   |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.68            | 0.66   | 0.61   | 0.51             | 0.66   | 0.32   | 0.53             | 0.65   | 0.65   |
| $\alpha_2$                   | 0.81            | 0.60   | 0.95   | 0.81             | 0.73   | 0.70   | 0.76             | 0.81   | 0.98   |
| $\hat{\rho}$                 | 0.00            | 0.00   | 0.22   | 0.19             | 0.00   | 0.36   | 0.03             | 0.11   | 0.04   |
| $\hat{\sigma}^2$             | 0.44            | 0.83   | 0.69   | 0.88             | 0.72   | 0.79   | 0.91             | 0.76   | 0.84   |
| MedSE                        | 0.22            | 0.47   | 0.70   | 0.21             | 0.25   | 0.39   | 0.16             | 0.16   | 0.22   |
| $\rho_1 = 0.8, \sigma_1 = 2$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.73            | 0.77   | 1.37   | 0.41             | 0.79   | 0.34   | 0.58             | 0.63   | 0.42   |
| $\alpha_2$                   | 0.74            | 0.27   | 0.64   | 0.66             | 0.65   | 0.55   | 0.69             | 0.68   | 0.96   |
| $\hat{\rho}$                 | 0.86            | 0.98   | 0.62   | 0.95             | 0.83   | 0.97   | 0.93             | 0.94   | 0.90   |
| $\hat{\sigma}^2$             | 1.83            | 3.26   | 5.78   | 3.20             | 3.08   | 2.68   | 3.53             | 2.89   | 3.25   |
| MedSE                        | 0.46            | 1.00   | 1.95   | 0.49             | 0.52   | 0.90   | 0.38             | 0.33   | 0.48   |
| $\rho_1 = 0.5, \sigma_1 = 2$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.72            | 0.79   | 0.66   | 0.43             | 0.78   | 0.23   | 0.59             | 0.65   | 0.42   |
| $\alpha_2$                   | 0.74            | 0.32   | 0.90   | 0.68             | 0.68   | 0.54   | 0.72             | 0.70   | 1.01   |
| $\hat{\rho}$                 | 0.58            | 0.68   | 0.50   | 0.77             | 0.29   | 0.86   | 0.70             | 0.74   | 0.62   |
| $\hat{\sigma}^2$             | 1.87            | 3.49   | 3.05   | 3.44             | 3.03   | 3.00   | 3.74             | 3.08   | 3.45   |
| MedSE                        | 0.46            | 0.98   | 1.59   | 0.48             | 0.48   | 0.95   | 0.38             | 0.32   | 0.48   |
| $\rho_1 = 0.2, \sigma_1 = 2$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.76            | 0.77   | 0.53   | 0.46             | 0.78   | 0.19   | 0.60             | 0.68   | 0.46   |
| $\alpha_2$                   | 0.76            | 0.35   | 1.06   | 0.72             | 0.65   | 0.65   | 0.75             | 0.72   | 1.01   |
| $\hat{\rho}$                 | 0.14            | 0.00   | 0.39   | 0.39             | 0.00   | 0.61   | 0.21             | 0.32   | 0.23   |
| $\hat{\sigma}^2$             | 1.88            | 3.52   | 3.00   | 3.68             | 2.99   | 3.19   | 3.88             | 3.24   | 3.57   |
| MedSE                        | 0.45            | 0.97   | 1.51   | 0.46             | 0.50   | 0.84   | 0.36             | 0.32   | 0.48   |
| $\rho_1 = 0, \sigma_1 = 2$   |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                   | 0.77            | 0.78   | 0.57   | 0.47             | 0.79   | 0.23   | 0.59             | 0.68   | 0.47   |
| $\alpha_2$                   | 0.77            | 0.34   | 1.07   | 0.72             | 0.64   | 0.55   | 0.76             | 0.72   | 1.04   |
| $\hat{\rho}$                 | 0.00            | 0.00   | 0.31   | 0.23             | 0.00   | 0.52   | 0.01             | 0.14   | 0.07   |
| $\hat{\sigma}^2$             | 1.88            | 3.54   | 3.04   | 3.74             | 3.09   | 3.32   | 3.90             | 3.26   | 3.60   |
| MedSE                        | 0.45            | 0.97   | 1.50   | 0.46             | 0.52   | 0.89   | 0.35             | 0.33   | 0.48   |

**Table 2.** Estimation with no regularizer on normal data when the dimension is close to the sample size.

|                              | $n = 25, q = 20$ |        |        | $n = 144, q = 80$ |        |        | $n = 324, q = 200$ |        |        |
|------------------------------|------------------|--------|--------|-------------------|--------|--------|--------------------|--------|--------|
|                              | E+null           | S+null | L+null | E+null            | S+null | L+null | E+null             | S+null | L+null |
| $\rho_1 = 0.8, \sigma_1 = 1$ |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.74             | 0.39   | 0.24   | 0.04              | 0.49   | -0.21  | 0.78               | 0.72   | 0.62   |
| $\alpha_2$                   | 0.67             | 0.18   | 2.81   | 1.05              | 0.88   | 2.15   | 0.86               | 0.86   | 0.71   |
| $\hat{\rho}$                 | 0.84             | 0.93   | 0.50   | 0.54              | 0.80   | 0.50   | 0.80               | 0.80   | 0.50   |
| $\hat{\sigma}^2$             | 0.18             | 0.53   | 3.54   | 0.30              | 0.26   | 1.00   | 0.37               | 0.41   | 1.68   |
| MedSE                        | 2.79             | 2.20   | 7.92   | 2.78              | 1.72   | 4.56   | 1.44               | 1.66   | 2.21   |
| $\rho_1 = 0.5, \sigma_1 = 1$ |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.72             | 0.39   | 0.10   | 0.19              | 0.49   | 0.06   | 0.75               | 0.68   | 0.56   |
| $\alpha_2$                   | 0.65             | 0.27   | 1.76   | 0.94              | 0.83   | 1.68   | 0.84               | 0.84   | 0.65   |
| $\hat{\rho}$                 | 0.61             | 0.66   | 0.50   | 0.46              | 0.54   | 0.50   | 0.52               | 0.53   | 0.50   |
| $\hat{\sigma}^2$             | 0.17             | 0.56   | 0.44   | 0.23              | 0.24   | 0.47   | 0.36               | 0.40   | 0.72   |
| MedSE                        | 2.87             | 2.04   | 3.30   | 1.57              | 1.71   | 2.43   | 1.40               | 1.60   | 1.62   |
| $\rho_1 = 0.2, \sigma_1 = 1$ |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.70             | 0.37   | 0.08   | 0.17              | 0.50   | 0.24   | 0.75               | 0.67   | 0.55   |
| $\alpha_2$                   | 0.64             | 0.40   | 1.45   | 0.93              | 0.79   | 1.60   | 0.84               | 0.85   | 0.60   |
| $\hat{\rho}$                 | 0.45             | 0.04   | 0.50   | 0.00              | 0.31   | 0.50   | 0.13               | 0.19   | 0.50   |
| $\hat{\sigma}^2$             | 0.18             | 0.57   | 0.22   | 0.22              | 0.24   | 0.50   | 0.36               | 0.40   | 0.73   |
| MedSE                        | 3.16             | 1.71   | 2.40   | 1.59              | 1.82   | 2.50   | 1.41               | 1.59   | 1.64   |
| $\rho_1 = 0, \sigma_1 = 1$   |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.71             | 0.37   | 0.06   | 0.20              | 0.50   | 0.26   | 0.75               | 0.68   | 0.61   |
| $\alpha_2$                   | 0.64             | 0.38   | 1.47   | 0.94              | 0.80   | 1.60   | 0.84               | 0.85   | 0.62   |
| $\hat{\rho}$                 | 0.35             | 0.00   | 0.50   | 0.00              | 0.16   | 0.50   | 0.00               | 0.04   | 0.50   |
| $\hat{\sigma}^2$             | 0.18             | 0.57   | 0.21   | 0.23              | 0.24   | 0.52   | 0.36               | 0.40   | 0.77   |
| MedSE                        | 3.19             | 1.76   | 2.42   | 1.57              | 1.81   | 2.62   | 1.41               | 1.60   | 1.84   |
| $\rho_1 = 0.8, \sigma_1 = 2$ |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.58             | 0.08   | -0.47  | -1.05             | 0.29   | -1.22  | 0.78               | 0.68   | 0.54   |
| $\alpha_2$                   | 0.45             | -0.57  | 2.66   | 1.31              | 0.87   | 3.60   | 0.91               | 0.71   | 0.72   |
| $\hat{\rho}$                 | 0.77             | 0.97   | 0.50   | 0.61              | 0.81   | 0.50   | 0.84               | 0.87   | 0.50   |
| $\hat{\sigma}^2$             | 4.18             | 2.27   | 12.43  | 2.20              | 1.06   | 4.13   | 1.63               | 1.63   | 6.24   |
| MedSE                        | 8.37             | 4.41   | 9.85   | 5.68              | 3.59   | 9.27   | 2.97               | 3.24   | 4.38   |
| $\rho_1 = 0.5, \sigma_1 = 2$ |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.63             | 0.11   | -0.38  | -0.43             | 0.33   | -0.45  | 0.75               | 0.68   | 0.57   |
| $\alpha_2$                   | 1.23             | -0.23  | 2.89   | 1.28              | 0.78   | 2.71   | 0.93               | 0.71   | 0.62   |
| $\hat{\rho}$                 | 0.64             | 0.60   | 0.50   | 0.47              | 0.59   | 0.50   | 0.60               | 0.61   | 0.50   |
| $\hat{\sigma}^2$             | 2.33             | 2.44   | 1.88   | 1.76              | 1.00   | 2.03   | 1.73               | 1.68   | 3.08   |
| MedSE                        | 5.36             | 3.80   | 6.83   | 4.34              | 3.55   | 5.02   | 2.97               | 3.26   | 3.35   |
| $\rho_1 = 0.2, \sigma_1 = 2$ |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.65             | 0.13   | -0.81  | -0.15             | 0.36   | -0.17  | 0.75               | 0.69   | 0.65   |
| $\alpha_2$                   | 1.14             | -0.04  | 2.37   | 0.88              | 0.74   | 2.48   | 0.96               | 0.72   | 0.51   |
| $\hat{\rho}$                 | 0.37             | 0.00   | 0.50   | 0.00              | 0.31   | 0.50   | 0.26               | 0.22   | 0.50   |
| $\hat{\sigma}^2$             | 1.89             | 2.43   | 0.95   | 1.95              | 1.03   | 2.01   | 1.91               | 1.72   | 3.17   |
| MedSE                        | 5.12             | 3.43   | 5.01   | 3.42              | 3.66   | 4.81   | 3.07               | 3.30   | 3.32   |
| $\rho_1 = 0, \sigma_1 = 2$   |                  |        |        |                   |        |        |                    |        |        |
| $\alpha_1$                   | 0.64             | 0.12   | -0.80  | -0.10             | 0.35   | -0.14  | 0.76               | 0.70   | 0.64   |
| $\alpha_2$                   | 0.58             | -0.18  | 2.23   | 0.88              | 0.76   | 2.47   | 0.93               | 0.72   | 0.53   |
| $\hat{\rho}$                 | 0.21             | 0.00   | 0.50   | 0.00              | 0.15   | 0.50   | 0.01               | 0.01   | 0.50   |
| $\hat{\sigma}^2$             | 0.78             | 2.46   | 0.97   | 2.09              | 1.03   | 2.07   | 1.73               | 1.71   | 3.24   |
| MedSE                        | 5.91             | 3.72   | 4.90   | 3.42              | 3.64   | 4.90   | 2.98               | 3.30   | 3.40   |

Table 3 illustrates the results of the model when the observations of  $y$  have outliers. Compared with the square loss model and LAD loss model, the model with exponential square loss shows advantages in parameter estimation in terms of MedSE, especially when the sample size is large.

We list the results of the estimated coefficients with inaccurate weight matrix  $W$  in Table 4. Compared with the results with normal data (Table 1), the MedSE values increase,

and the estimations of  $\hat{\rho}$  and  $\hat{\sigma}^2$  become worse for each loss functions in total. Particularly, for removing a certain part (30%, 50%, and 80%) of nonzero weights of the matrix  $W$ , MedSE increases as the moving nonzeros increase and decreases as the sample size  $n$  increases for each of the three loss functions. The exponential squared loss has the lowest MedSE among the three loss functions.

**Table 3.** Estimation with no regularizer when the observations of  $y$  have outliers.

|   | $n = 25, q = 5$ |        |        | $n = 144, q = 5$ |        |        | $n = 324, q = 5$ |        |        |
|---|-----------------|--------|--------|------------------|--------|--------|------------------|--------|--------|
|   | E+null          | S+null | L+null | E+null           | S+null | L+null | E+null           | S+null | L+null |
| $\rho_1 = 0.5, \sigma_1 = 1, \delta_1 = 0.01$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                                    | 0.71            | 0.70   | 0.59   | 0.52             | 0.53   | 0.54   | 0.52             | 0.60   | 0.42   |
| $\alpha_2$                                    | 0.49            | 0.76   | 1.20   | 0.73             | 0.94   | 0.86   | 0.72             | 0.66   | 1.14   |
| $\hat{\rho}$                                  | 0.45            | 0.64   | 0.50   | 0.67             | 0.54   | 0.51   | 0.59             | 0.49   | 0.55   |
| $\hat{\sigma}^2$                              | 0.80            | 0.82   | 0.73   | 0.87             | 0.86   | 0.98   | 1.02             | 1.09   | 0.91   |
| MedSE   | 0.46            | 0.34   | 0.51   | 0.30             | 0.25   | 0.21   | 0.18             | 0.22   | 0.35   |
| $\rho_1 = 0.5, \sigma_1 = 2, \delta_1 = 0.01$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                                    | 1.06            | 0.69   | 0.64   | 0.45             | 0.40   | 0.45   | 0.59             | 0.61   | 0.35   |
| $\alpha_2$                                    | 0.14            | 0.86   | 1.58   | 0.62             | 1.28   | 0.80   | 0.71             | 0.54   | 1.38   |
| $\hat{\rho}$                                  | 0.48            | 0.69   | 0.66   | 0.76             | 0.59   | 0.70   | 0.67             | 0.52   | 0.63   |
| $\hat{\sigma}^2$                              | 3.84            | 3.24   | 2.82   | 3.72             | 3.48   | 3.98   | 4.21             | 4.46   | 3.79   |
| MedSE   | 1.52            | 0.70   | 1.07   | 0.55             | 0.63   | 0.56   | 0.39             | 0.40   | 0.67   |
| $\rho_1 = 0.5, \sigma_1 = 1, \delta_1 = 0.05$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                                    | 0.88            | 1.01   | 0.67   | 0.57             | 0.74   | 0.39   | 0.42             | 0.58   | 0.58   |
| $\alpha_2$                                    | 0.43            | 0.72   | 1.01   | 0.56             | 0.78   | 1.00   | 0.65             | 0.51   | 1.05   |
| $\hat{\rho}$                                  | 0.60            | 0.75   | 0.50   | 0.75             | 0.85   | 0.60   | 0.67             | 0.74   | 0.65   |
| $\hat{\sigma}^2$                              | 2.64            | 3.72   | 4.32   | 3.75             | 3.02   | 4.44   | 4.04             | 4.64   | 3.83   |
| MedSE   | 0.67            | 0.59   | 1.07   | 0.73             | 0.45   | 0.47   | 0.35             | 0.48   | 0.28   |
| $\rho_1 = 0.5, \sigma_1 = 2, \delta_1 = 0.05$ |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                                    | 1.09            | 1.00   | 0.74   | 0.56             | 0.61   | 0.61   | 0.48             | 0.61   | 0.41   |
| $\alpha_2$                                    | 0.23            | 0.83   | 1.58   | 0.43             | 1.13   | 0.73   | 0.64             | 0.40   | 1.42   |
| $\hat{\rho}$                                  | 0.39            | 0.76   | 0.50   | 0.77             | 0.82   | 0.50   | 0.69             | 0.69   | 0.64   |
| $\hat{\sigma}^2$                              | 5.02            | 6.17   | 6.23   | 6.13             | 5.67   | 7.62   | 7.08             | 7.94   | 6.33   |
| MedSE   | 0.98            | 0.83   | 1.61   | 0.87             | 0.59   | 0.72   | 0.46             | 0.64   | 0.67   |

**Table 4.** Estimation with no regularizer with noisy weighting matrix  $w$ .

|                            | $n = 25, q = 5$ |        |        | $n = 144, q = 5$ |        |        | $n = 324, q = 5$ |        |        |
|----------------------------|-----------------|--------|--------|------------------|--------|--------|------------------|--------|--------|
|                            | E+null          | S+null | L+null | E+null           | S+null | L+null | E+null           | S+null | L+null |
| Remove 30% nonzero weights |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                 | 0.59            | 0.58   | 0.17   | 0.43             | 0.32   | 0.47   | 0.55             | 0.65   | 0.46   |
| $\alpha_2$                 | 0.59            | 0.97   | 1.63   | 0.82             | 0.80   | 0.97   | 0.73             | 0.78   | 1.12   |
| $\hat{\rho}$               | 0.61            | 0.55   | 0.52   | 0.70             | 0.57   | 0.50   | 0.65             | 0.54   | 0.53   |
| $\hat{\sigma}^2$           | 1.05            | 1.13   | 0.99   | 1.08             | 1.08   | 1.03   | 1.14             | 1.09   | 1.22   |
| MedSE                      | 0.48            | 0.45   | 1.10   | 0.35             | 0.33   | 0.49   | 0.20             | 0.25   | 0.25   |
| Remove 50% nonzero weights |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                 | 0.67            | 0.57   | 0.16   | 0.44             | 0.30   | 0.31   | 0.52             | 0.66   | 0.54   |
| $\alpha_2$                 | 0.61            | 0.90   | 1.48   | 0.73             | 0.81   | 1.15   | 0.74             | 0.79   | 1.10   |
| $\hat{\rho}$               | 0.54            | 0.48   | 0.50   | 0.64             | 0.49   | 0.41   | 0.57             | 0.49   | 0.49   |
| $\hat{\sigma}^2$           | 1.07            | 1.17   | 0.97   | 1.11             | 1.11   | 1.09   | 1.18             | 1.12   | 1.25   |
| MedSE                      | 0.41            | 0.26   | 1.12   | 0.34             | 0.37   | 0.64   | 0.19             | 0.27   | 0.31   |
| Remove 80% nonzero weights |                 |        |        |                  |        |        |                  |        |        |
| $\alpha_1$                 | 0.68            | 0.63   | 0.26   | 0.46             | 0.33   | 0.24   | 0.58             | 0.70   | 0.48   |
| $\alpha_2$                 | 0.59            | 0.94   | 1.29   | 0.71             | 0.82   | 1.31   | 0.77             | 0.78   | 1.19   |
| $\hat{\rho}$               | 0.40            | 0.34   | 0.51   | 0.52             | 0.34   | 0.37   | 0.42             | 0.33   | 0.36   |
| $\hat{\sigma}^2$           | 1.15            | 1.30   | 0.96   | 1.26             | 1.24   | 1.14   | 1.34             | 1.26   | 1.40   |
| MedSE                      | 0.50            | 0.40   | 0.89   | 0.40             | 0.40   | 0.78   | 0.20             | 0.29   | 0.33   |

Correspondingly, Tables 5–8 show the variable selection results compared with other loss functions. The average number of zero coefficients that are correctly chosen is labeled as “Correct”. The label “Incorrect” depicts the average number of nonzero coefficients incorrectly identified as zero. “ $\tilde{l}_1$ ”, “ $l_1$ ”, and “null” express the adaptive lasso penalty, the lasso penalty, and without penalty term, respectively.

Table 5 shows the variable section results of the lasso and the adaptive lasso regularizer on normal data with  $q = 5$ . In almost all of the tested cases, the model with the exponential squared loss and the lasso penalty or the adaptive lasso penalty (i.e.,  $E + l_1$ ,  $E + \tilde{l}_1$ ) identifies more numbers of true zero coefficients (“Correct”) and much lower MedSE.

**Table 5.** Variable section with regularizer on normal data ( $q = 5$ ), E: the exponential loss; S: the square loss; L: the LAD loss;  $l_1$ : the lasso penalty; and  $\tilde{l}_1$ : the adaptive lasso penalty.

|                              | $n = 25, q = 5$ |                   |           |                   |           |                   | $n = 324, q = 5$ |                   |           |                   |           |                   |
|------------------------------|-----------------|-------------------|-----------|-------------------|-----------|-------------------|------------------|-------------------|-----------|-------------------|-----------|-------------------|
|                              | E + $l_1$       | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ | E + $l_1$        | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ |
| $\rho_1 = 0.8, \sigma_1 = 1$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 5.00              | 4.00      | 5.00              | 0.00      | 3.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.99            | 0.86              | 0.86      | 0.97              | 0.73      | 0.82              | 0.89             | 0.88              | 0.88      | 0.89              | 0.89      | 0.92              |
| MedSE                        | 0.42            | 0.37              | 0.44      | 0.36              | 1.43      | 0.54              | 0.14             | 0.14              | 0.20      | 0.16              | 0.21      | 0.24              |
| $\rho_1 = 0.5, \sigma_1 = 1$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 4.00              | 3.00      | 5.00              | 5.00      | 4.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.52            | 0.57              | 0.58      | 0.81              | 0.51      | 0.46              | 0.50             | 0.58              | 0.58      | 0.57              | 0.56      | 0.68              |
| MedSE                        | 0.24            | 0.31              | 0.45      | 0.40              | 0.49      | 0.43              | 0.17             | 0.11              | 0.15      | 0.16              | 0.23      | 0.22              |
| $\rho_1 = 0.2, \sigma_1 = 1$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 4.00              | 3.00      | 5.00              | 5.00      | 3.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 1.00              | 0.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.26            | 0.38              | 0.40      | 0.61              | 0.37      | 0.20              | 0.33             | 0.35              | 0.35      | 0.31              | 0.35      | 0.49              |
| MedSE                        | 0.24            | 0.32              | 0.47      | 0.42              | 0.50      | 0.52              | 0.14             | 0.12              | 0.16      | 0.16              | 0.22      | 0.21              |
| $\rho_1 = 0, \sigma_1 = 1$   |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 4.00              | 3.00      | 5.00              | 5.00      | 3.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 1.00              | 0.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.00            | 0.06              | 0.09      | 0.30              | 0.01      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 0.11              |
| MedSE                        | 0.24            | 0.31              | 0.46      | 0.44              | 0.56      | 0.55              | 0.14             | 0.13              | 0.17      | 0.16              | 0.17      | 0.21              |
| $\rho_1 = 0.8, \sigma_1 = 2$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 2.00              | 1.00      | 3.00              | 0.00      | 1.00              | 5.00             | 5.00              | 5.00      | 4.00              | 2.00      | 4.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 1.00              | 0.00      | 0.00              | 1.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.88            | 0.90              | 0.92      | 0.99              | 0.86      | 0.88              | 0.94             | 0.92              | 0.92      | 0.94              | 0.92      | 0.96              |
| MedSE                        | 0.53            | 0.67              | 0.94      | 0.65              | 2.09      | 1.05              | 0.32             | 0.29              | 0.32      | 0.36              | 0.45      | 0.44              |
| $\rho_1 = 0.5, \sigma_1 = 2$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 2.00              | 1.00      | 2.00              | 3.00      | 3.00              | 5.00             | 5.00              | 5.00      | 4.00              | 2.00      | 4.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 1.00              | 1.00      | 0.00              | 1.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.45            | 0.64              | 0.69      | 0.89              | 0.51      | 0.52              | 0.66             | 0.63              | 0.65      | 0.64              | 0.62      | 0.81              |
| MedSE                        | 0.52            | 0.67              | 0.96      | 0.70              | 0.99      | 0.95              | 0.31             | 0.29              | 0.31      | 0.35              | 0.46      | 0.48              |
| $\rho_1 = 0.2, \sigma_1 = 2$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 2.00              | 1.00      | 1.00              | 2.00      | 3.00              | 5.00             | 5.00              | 5.00      | 4.00              | 2.00      | 3.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 1.00              | 1.00      | 0.00              | 1.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.03            | 0.45              | 0.51      | 0.76              | 0.50      | 0.50              | 0.38             | 0.37              | 0.39      | 0.34              | 0.40      | 0.57              |
| MedSE                        | 0.55            | 0.69              | 0.97      | 0.76              | 1.13      | 1.02              | 0.29             | 0.30              | 0.33      | 0.34              | 0.47      | 0.48              |
| $\rho_1 = 0, \sigma_1 = 2$   |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                      | 4.00            | 2.00              | 1.00      | 1.00              | 2.00      | 2.00              | 5.00             | 5.00              | 5.00      | 4.00              | 3.00      | 3.00              |
| Incorrect                    | 0.00            | 0.00              | 0.00      | 1.00              | 1.00      | 0.00              | 1.00             | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.00            | 0.10              | 0.18      | 0.49              | 0.03      | 0.41              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 0.21              |
| MedSE                        | 0.51            | 0.68              | 0.96      | 0.82              | 1.09      | 1.21              | 0.28             | 0.31              | 0.36      | 0.34              | 0.38      | 0.50              |

Similar results have been observed when the sample dimension is close to the sample size, which is presented in Table 6. In the tested cases of  $n = 360, q = 200$ , the model with  $l_1 \tilde{l}_1$  almost correctly identifies all the zero coefficients. The above performance

of the proposed exponential squared loss and lasso or adaptive lasso penalty is beyond our expectations.

**Table 6.** Variable selection with regularizer on normal data when the dimension is close to the sample size, E: the exponential loss; S: the square loss; L: the LAD loss;  $l_1$ : the lasso penalty; and  $\tilde{l}_1$ : the adaptive lasso penalty.

|                              | $n = 25, q = 20$ |                   |           |                   |           |                   | $n = 324, q = 200$ |                   |           |                   |           |                   |
|------------------------------|------------------|-------------------|-----------|-------------------|-----------|-------------------|--------------------|-------------------|-----------|-------------------|-----------|-------------------|
|                              | E + $l_1$        | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ | E + $l_1$          | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ |
| $\rho_1 = 0.8, \sigma_1 = 1$ |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 7.00             | 9.00              | 5.00      | 6.00              | 8.00      | 16.00             | 195.00             | 200.00            | 187.00    | 195.00            | 180.00    | 187.00            |
| Incorrect                    | 1.00             | 1.00              | 0.00      | 0.00              | 1.00      | 0.00              | 1.00               | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.81             | 0.82              | 0.89      | 0.88              | 0.58      | 0.69              | 0.84               | 0.87              | 0.85      | 0.88              | 0.65      | 0.73              |
| MedSE                        | 2.89             | 1.38              | 3.38      | 2.06              | 1.39      | 0.56              | 1.23               | 0.53              | 1.52      | 1.36              | 1.69      | 1.53              |
| $\rho_1 = 0.5, \sigma_1 = 1$ |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 6.00             | 10.00             | 5.00      | 4.00              | 17.00     | 13.00             | 197.00             | 200.00            | 192.00    | 196.00            | 199.00    | 197.00            |
| Incorrect                    | 0.00             | 1.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00               | 0.00              | 0.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.50             | 0.25              | 0.55      | 0.61              | 0.51      | 0.42              | 0.62               | 0.55              | 0.54      | 0.61              | 0.50      | 0.50              |
| MedSE                        | 2.02             | 1.41              | 3.38      | 2.20              | 0.64      | 0.76              | 1.05               | 0.52              | 1.43      | 1.33              | 0.88      | 1.00              |
| $\rho_1 = 0.2, \sigma_1 = 1$ |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 5.00             | 10.00             | 5.00      | 5.00              | 13.00     | 14.00             | 197.00             | 200.00            | 191.00    | 193.00            | 199.00    | 198.00            |
| Incorrect                    | 1.00             | 1.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00               | 0.00              | 1.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.55             | 0.00              | 0.40      | 0.47              | 0.50      | 0.23              | 0.48               | 0.34              | 0.40      | 0.48              | 0.50      | 0.50              |
| MedSE                        | 2.98             | 1.39              | 3.47      | 2.45              | 0.90      | 0.74              | 1.11               | 0.52              | 1.41      | 1.36              | 0.91      | 1.02              |
| $\rho_1 = 0, \sigma_1 = 1$   |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 9.00             | 11.00             | 5.00      | 4.00              | 13.00     | 13.00             | 197.00             | 200.00            | 192.00    | 193.00            | 200.00    | 196.00            |
| Incorrect                    | 1.00             | 1.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00               | 0.00              | 1.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.55             | 0.00              | 0.00      | 0.12              | 0.38      | 0.00              | 0.13               | 0.00              | 0.03      | 0.16              | 0.50      | 0.49              |
| MedSE                        | 1.78             | 1.24              | 3.41      | 2.28              | 1.04      | 0.82              | 1.07               | 0.52              | 1.42      | 1.36              | 0.92      | 1.13              |
| $\rho_1 = 0.8, \sigma_1 = 2$ |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 6.00             | 6.00              | 5.00      | 1.00              | 5.00      | 13.00             | 162.00             | 172.00            | 133.00    | 137.00            | 156.00    | 138.00            |
| Incorrect                    | 1.00             | 0.00              | 0.00      | 0.00              | 1.00      | 0.00              | 0.00               | 1.00              | 1.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.81             | 0.92              | 1.00      | 0.98              | 0.76      | 0.73              | 0.94               | 0.89              | 0.89      | 0.94              | 0.73      | 0.73              |
| MedSE                        | 3.33             | 3.65              | 7.30      | 5.30              | 2.30      | 0.92              | 2.28               | 1.84              | 2.99      | 2.71              | 2.30      | 2.78              |
| $\rho_1 = 0.5, \sigma_1 = 2$ |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 8.00             | 6.00              | 4.00      | 1.00              | 9.00      | 9.00              | 160.00             | 173.00            | 134.00    | 135.00            | 185.00    | 173.00            |
| Incorrect                    | 1.00             | 1.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00               | 1.00              | 1.00      | 0.00              | 0.00      | 1.00              |
| $\hat{\rho}$                 | 0.77             | 0.53              | 0.91      | 0.90              | 0.60      | 0.47              | 0.74               | 0.63              | 0.61      | 0.75              | 0.50      | 0.50              |
| MedSE                        | 2.52             | 3.39              | 7.63      | 5.95              | 1.87      | 1.44              | 2.34               | 1.81              | 2.96      | 2.77              | 1.64      | 1.94              |
| $\rho_1 = 0.2, \sigma_1 = 2$ |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 7.00             | 7.00              | 4.00      | 1.00              | 8.00      | 9.00              | 162.00             | 167.00            | 129.00    | 130.00            | 181.00    | 173.00            |
| Incorrect                    | 1.00             | 1.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00               | 1.00              | 1.00      | 0.00              | 1.00      | 1.00              |
| $\hat{\rho}$                 | 0.24             | 0.10              | 0.68      | 0.79              | 0.63      | 0.15              | 0.53               | 0.46              | 0.46      | 0.55              | 0.50      | 0.50              |
| MedSE                        | 2.87             | 3.42              | 7.62      | 5.98              | 1.84      | 1.47              | 2.33               | 1.83              | 2.97      | 2.81              | 1.72      | 1.99              |
| $\rho_1 = 0, \sigma_1 = 2$   |                  |                   |           |                   |           |                   |                    |                   |           |                   |           |                   |
| Correct                      | 8.00             | 6.00              | 4.00      | 1.00              | 9.00      | 9.00              | 144.00             | 172.00            | 131.00    | 131.00            | 178.00    | 171.00            |
| Incorrect                    | 1.00             | 0.00              | 0.00      | 0.00              | 0.00      | 0.00              | 0.00               | 1.00              | 1.00      | 0.00              | 1.00      | 1.00              |
| $\hat{\rho}$                 | 0.00             | 0.00              | 0.32      | 0.54              | 0.50      | 0.00              | 0.35               | 0.00              | 0.00      | 0.26              | 0.50      | 0.50              |
| MedSE                        | 2.98             | 3.60              | 7.65      | 5.70              | 1.95      | 1.58              | 2.73               | 1.83              | 2.99      | 2.86              | 1.87      | 2.11              |

Tables 7 and 8 list the variable selections results with noise in the observations and the inaccurate weight matrix. The model with the exponential squared loss and the lasso penalty or the adaptive lasso penalty (i.e., E +  $l_1$ , E +  $\tilde{l}_1$ ) identifies many more numbers of true coefficients (“Correct”) and has much lower MedSE. Compared with the results in the normal cases (Table 5), the superiority of E +  $l_1$  and E +  $\tilde{l}_1$  is more evident.

**Table 7.** Variable selection with regularizer when the observations  $y$  have outliers, E: the exponential loss; S: the square loss; L: the LAD loss;  $l_1$ : the lasso penalty; and  $\tilde{l}_1$ : the adaptive lasso penalty.

|   | $n = 25, q = 5$ |                   |           |                   |           |                   | $n = 324, q = 5$ |                   |           |                   |           |                   |
|---|-----------------|-------------------|-----------|-------------------|-----------|-------------------|------------------|-------------------|-----------|-------------------|-----------|-------------------|
|   | E + $l_1$       | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ | E + $l_1$        | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ |
| $\rho_1 = 0.5, \sigma_1 = 1, \delta_1 = 0.01$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                                       | 4.00            | 4.00              | 4.00      | 5.00              | 4.00      | 3.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                                     | 0.00            | 0.00              | 0.00      | 1.00              | 0.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 1.00      | 0.00              |
| $\hat{\rho}$                                  | 0.77            | 0.64              | 0.63      | 0.77              | 0.70      | 0.79              | 0.74             | 0.66              | 0.66      | 0.70              | 0.64      | 0.61              |
| MedSE   | 0.48            | 0.32              | 0.48      | 0.36              | 0.62      | 0.53              | 0.14             | 0.14              | 0.17      | 0.18              | 0.31      | 0.30              |
| $\hat{\sigma}^2$                              |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| $\rho_1 = 0.5, \sigma_1 = 2, \delta_1 = 0.01$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                                       | 3.00            | 1.00              | 2.00      | 3.00              | 1.00      | 1.00              | 5.00             | 5.00              | 5.00      | 5.00              | 3.00      | 3.00              |
| Incorrect                                     | 0.00            | 0.00              | 0.00      | 0.00              | 2.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 1.00      | 0.00              |
| $\hat{\rho}$                                  | 0.58            | 0.47              | 0.52      | 0.71              | 0.51      | 0.76              | 0.57             | 0.56              | 0.58      | 0.64              | 0.50      | 0.50              |
| MedSE   | 0.61            | 0.97              | 0.93      | 0.79              | 1.19      | 1.17              | 0.18             | 0.35              | 0.35      | 0.34              | 0.66      | 0.63              |
| $\rho_1 = 0.5, \sigma_1 = 1, \delta_1 = 0.05$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                                       | 1.00            | 4.00              | 3.00      | 3.00              | 0.00      | 3.00              | 3.00             | 4.00              | 4.00      | 4.00              | 4.00      | 5.00              |
| Incorrect                                     | 0.00            | 1.00              | 1.00      | 1.00              | 1.00      | 1.00              | 0.00             | 1.00              | 0.00      | 0.00              | 0.00      | 0.00              |
| $\hat{\rho}$                                  | 0.78            | 0.73              | 0.73      | 0.90              | 0.77      | 0.88              | 0.83             | 0.75              | 0.79      | 0.79              | 0.83      | 0.81              |
| MedSE   | 0.69            | 0.87              | 0.92      | 0.69              | 1.75      | 0.66              | 0.38             | 0.30              | 0.35      | 0.32              | 0.52      | 0.22              |
| $\rho_1 = 0.5, \sigma_1 = 2, \delta_1 = 0.05$ |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                                       | 1.00            | 3.00              | 1.00      | 3.00              | 0.00      | 3.00              | 5.00             | 4.00              | 4.00      | 4.00              | 4.00      | 3.00              |
| Incorrect                                     | 0.00            | 0.00              | 1.00      | 1.00              | 1.00      | 0.00              | 0.00             | 1.00              | 0.00      | 0.00              | 1.00      | 0.00              |
| $\hat{\rho}$                                  | 0.65            | 0.42              | 0.41      | 0.83              | 0.51      | 0.79              | 0.75             | 0.59              | 0.64      | 0.67              | 0.63      | 0.57              |
| MedSE   | 0.96            | 1.28              | 1.26      | 0.74              | 1.93      | 0.54              | 0.36             | 0.46              | 0.46      | 0.46              | 0.75      | 0.54              |

**Table 8.** Variable selection with regularizer and noisy weighting matrix  $w$ , E: the exponential loss; S: the square loss; L: the LAD loss;  $l_1$ : the lasso penalty; and  $\tilde{l}_1$ : the adaptive lasso penalty.

|                            | $n = 25, q = 5$ |                   |           |                   |           |                   | $n = 324, q = 5$ |                   |           |                   |           |                   |
|----------------------------|-----------------|-------------------|-----------|-------------------|-----------|-------------------|------------------|-------------------|-----------|-------------------|-----------|-------------------|
|                            | E + $l_1$       | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ | E + $l_1$        | E + $\tilde{l}_1$ | S + $l_1$ | S + $\tilde{l}_1$ | L + $l_1$ | L + $\tilde{l}_1$ |
| Remove 30% nonzero weights |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                    | 4.00            | 5.00              | 5.00      | 5.00              | 0.00      | 5.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                  | 0.00            | 0.00              | 0.00      | 0.00              | 1.00      | 0.00              | 1.00             | 0.00              | 0.00      | 0.00              | 0.00      | 0.00              |
| $\hat{\rho}$               | 0.54            | 0.53              | 0.50      | 0.64              | 0.28      | 0.48              | 0.57             | 0.55              | 0.55      | 0.50              | 0.55      | 0.56              |
| MedSE                      | 0.50            | 0.22              | 0.41      | 0.32              | 0.90      | 0.28              | 0.16             | 0.18              | 0.26      | 0.16              | 0.17      | 0.28              |
| Remove 50% nonzero weights |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                    | 4.00            | 5.00              | 2.00      | 4.00              | 2.00      | 4.00              | 5.00             | 5.00              | 5.00      | 5.00              | 5.00      | 5.00              |
| Incorrect                  | 0.00            | 0.00              | 0.00      | 0.00              | 1.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 0.00              |
| $\hat{\rho}$               | 0.55            | 0.36              | 0.37      | 0.54              | 0.16      | 0.35              | 0.38             | 0.41              | 0.40      | 0.35              | 0.42      | 0.46              |
| MedSE                      | 0.57            | 0.43              | 0.63      | 0.36              | 0.86      | 0.40              | 0.09             | 0.22              | 0.29      | 0.17              | 0.12      | 0.28              |
| Remove 80% nonzero weights |                 |                   |           |                   |           |                   |                  |                   |           |                   |           |                   |
| Correct                    | 4.00            | 5.00              | 5.00      | 3.00              | 0.00      | 5.00              | 5.00             | 4.00              | 4.00      | 5.00              | 4.00      | 4.00              |
| Incorrect                  | 0.00            | 0.00              | 0.00      | 0.00              | 1.00      | 0.00              | 0.00             | 0.00              | 0.00      | 0.00              | 0.00      | 0.00              |
| $\hat{\rho}$               | 0.37            | 0.47              | 0.43      | 0.89              | 0.50      | 0.72              | 0.52             | 0.48              | 0.46      | 0.37              | 0.58      | 0.49              |
| MedSE                      | 0.63            | 0.35              | 0.50      | 0.77              | 0.91      | 0.42              | 0.20             | 0.48              | 0.50      | 0.29              | 0.36      | 0.41              |

For the fitting effect diagram of the coefficient function surface, we select the case at the median position in 100 repeated experiments as the standard. Note, we present the situation when  $h = 16$  on the normal data. The fitting surfaces of  $\hat{g}_1, \hat{g}_2,$  and  $\hat{g}_3$  are shown in Figure 2.

From the fitting effect of each coefficient function, it can be seen that the model has an excellent fitting effect for unknown coefficient functions, which shows that in the case of limited samples, the fitting effect of the spatial single-index varying-coefficient model based on radial basis function and exponential squared loss is excellent. In other cases, the fitting effect of each coefficient function also performs well.

We also present the fitting evaluation index MAISE when  $h = 16, 18$  on normal data, which is shown in Table 9. It can be seen that with the increase in the total number of spatial objects, the value of the unknown function fitting evaluation index MAISE shows a downward trend. That is, the fitting effect is getting better and better. Similarly,



the MAISE value of  $y$  also shows a downward trend, indicating that for the model as a whole, the relevant data is getting closer to the real data.

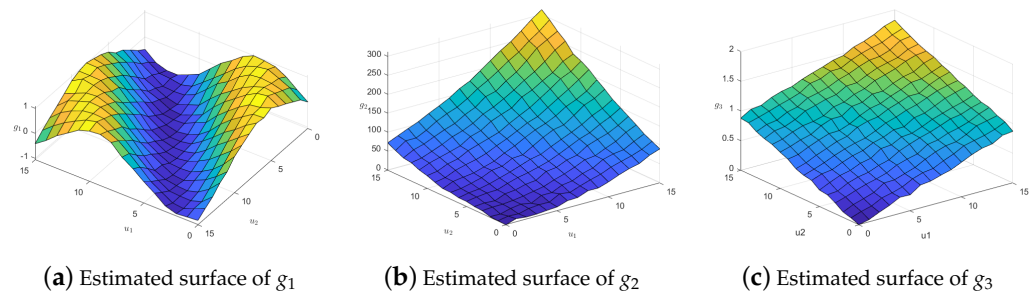


Figure 2. Estimated surfaces of coefficient functions with exponential squared loss.

Table 9. Results of MAISE for the total number of different spatial objects.

|             | $h = 16$ | $h = 18$ |
|-------------|----------|----------|
| $\hat{g}_1$ | 0.0437   | 0.0388   |
| $\hat{g}_2$ | 0.0542   | 0.0539   |
| $\hat{g}_3$ | 0.0515   | 0.0496   |
| $\hat{y}$   | 0.0566   | 0.0548   |

When the observations of  $y$  have outliers, the coefficient function surface fitting effect is compared. We still select the one in the median of 100 repetitions and take the fitted surface  $g_3$  as an example. When  $\rho_1 = 0.5$ ,  $\sigma_1 = 1$ ,  $\delta_1 = 0.05$ , the fitting effect of loss functions with adaptive lasso is shown in Figure 3. This shows that our method performs better. The same conclusion can be conducted in the case of noisy weighting matrix  $W$ . Figure 4 illustrates the results when we remove 50% nonzero weights.

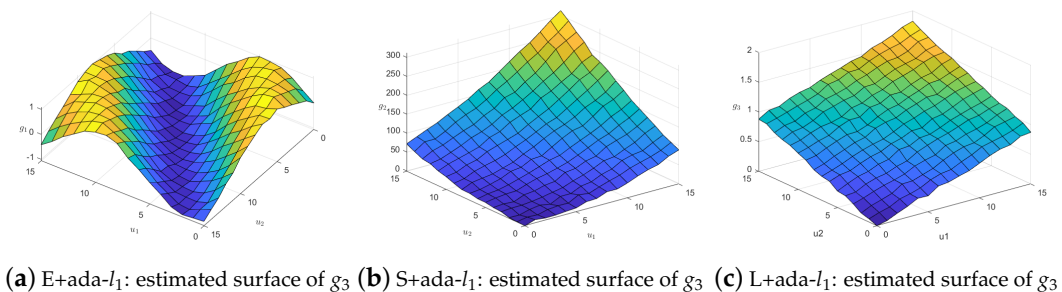


Figure 3. Comparison of  $g_3$  when  $y$  have outliers.

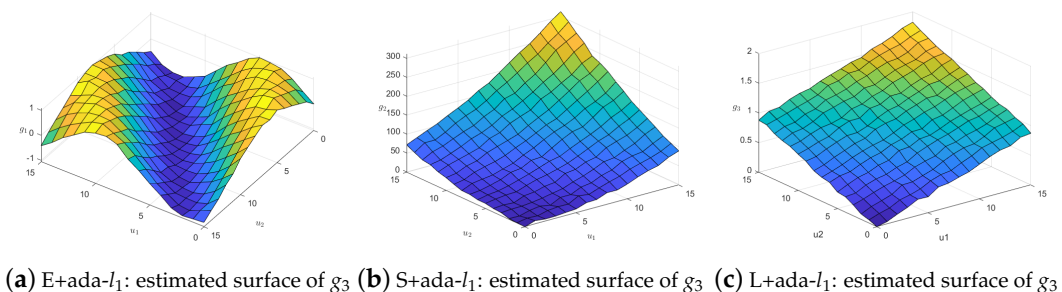


Figure 4. Comparison of  $g_3$  in the case of noisy weighting matrix  $w$ .

### 6. Summary

In this paper, we propose a novel model (the spatial single-index varying-coefficient model) and introduce a robust variable selection based on spline estimation and exponential squared loss for the model. The theoretical properties of the proposed estimators are

established under reasonable assumptions. We especially design a BCD algorithm equipped with a CCCP procedure for efficiently solving the non-convex and non-differentiable mathematical optimization problem about the variable selection process. Numerical studies show that our proposed method is particularly robust and applicable when observations and the weight matrix are noisy.

**Author Contributions:** Conceptualization, Y.S. and Y.W.; methodology, Y.W.; software, Y.W.; validation, Z.W. and Y.S.; formal analysis, Y.W.; investigation, Z.W.; resources, Y.S.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W., Z.W. and Y.S.; visualization, Y.W.; supervision, Y.S.; project administration, Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** The researches are supported by the National Key Research and Development Program of China (2021YFA1000102).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## Abbreviations

The main abbreviations used in this work are as follows:

|                |  |
|----------------|--|
| SAR model:     | Spatial autoregressive model;                    |
| BCD algorithm: | Block-coordinate descent algorithm;              |
| DC function:   | Difference between two convex functions;         |
| CCCP:          | concave–convex procedure;                        |
| ISTA:          | Iterative shrinkage-thresholding algorithm;      |
| FISTA:         | Fast iterative shrinkage-thresholding algorithm; |
| MedSE:         | Median of squared error;                         |
| MAISE:         | Square root of mean deviation.                   |

## Appendix A. Proofs

### Appendix A.1. The Related Lemmas

**Lemma A1** (Convexity Lemma). *Let  $\{h_n(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$  be a sequence of random convex functions defined on a convex, open subset  $\mathbf{T}$  of  $\mathbb{R}^d$ . Assume  $h(\mathbf{t})$  is a real-valued function on  $\mathbf{T}$  for which  $h_n(\mathbf{t}) \rightarrow h(\mathbf{t})$  in probability, for each  $\mathbf{t} \in \mathbf{T}$ . Then, for each compact subset  $\mathbf{J}$  of  $\mathbf{T}$*

$$\sup_{\mathbf{t} \in \mathbf{J}} |h_n(\mathbf{t}) - h(\mathbf{t})| \rightarrow 0 \text{ in probability.}$$

*The function  $h(\cdot)$  is necessary convex on  $\mathbf{T}$ .*

**Proof of Lemma A1.** For this well-known convexity lemma, there are many versions of proof, one of which can be referred to [22].  $\square$

**Lemma A2.** *If  $g_j(t), j = 1, \dots, q$ , satisfy condition (C2), then there a constant  $C > 0$  exists relying only on  $M$  such that*

$$\sup_{t \in T} |g_j(t) - \eta^T(t) \gamma_{1-j}| \leq CK^{-r}.$$

**Proof of Lemma A2.** The proof of Lemma A2 is similar to the proof of inference 6.21 in [23].  $\square$

Appendix A.2. Poof of Main Theorems

**Proof of Theorem 1.** Let

$$\eta = n^{-r/(2r+1)} + a_n, \quad \boldsymbol{\phi} = \boldsymbol{\phi}_0 + \eta \mathbf{t}_1, \quad \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_{1-0} + \eta \mathbf{t}_2, \quad \mathbf{t} = \left( \mathbf{t}_1^T, \mathbf{t}_2^T \right)^T.$$

(i) Let

$$Q(\boldsymbol{\phi}, \boldsymbol{\gamma}_1) = \sum_{i=1}^n \exp \left\{ - \left( Y_i - G_i^T(\boldsymbol{\phi}) \boldsymbol{\gamma}_1 \right)^2 / \gamma_1 \right\}$$

We will present that, for any given  $\epsilon > 0$ , a large constant exists  $C$  such that

$$P \left\{ \inf_{\|\mathbf{t}\|=C} Q(\boldsymbol{\phi}_0 + \eta \mathbf{t}_1, \boldsymbol{\gamma}_{1-0} + \eta \mathbf{t}_2) > Q(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_{1-0}) \right\} \geq 1 - \epsilon, \tag{A1}$$

where the true value of  $\boldsymbol{\phi}$ . and  $\boldsymbol{\gamma}_1$  are  $\boldsymbol{\phi}_0$  and  $\boldsymbol{\gamma}_{1-0}$ . Let

$$W_n(\boldsymbol{\phi}, \boldsymbol{\gamma}_1) = \sum_{i=1}^n \frac{2}{\gamma_2} \exp \left\{ - \left( Y_i - G_i^T(\boldsymbol{\phi}) \boldsymbol{\gamma}_1 \right)^2 / \gamma_2 \right\} \left\{ Y_i - G_i(\boldsymbol{\phi})^T \boldsymbol{\gamma}_1 \right\} \dot{G}_i(\boldsymbol{\phi})^T \boldsymbol{\gamma}_1 J_{\boldsymbol{\phi}}^T U_i$$

and

$$V_n(\boldsymbol{\phi}, \boldsymbol{\gamma}_1) = \sum_{i=1}^n \frac{2}{\gamma_2} \exp \left\{ - \left( Y_i - G_i^T(\boldsymbol{\phi}) \boldsymbol{\gamma}_1 \right)^2 / \gamma_2 \right\} \left\{ Y_i - G_i(\boldsymbol{\phi})^T \boldsymbol{\gamma}_1 \right\} G_i(\boldsymbol{\phi}),$$

where  $\dot{G}_i^T(\boldsymbol{\phi})(\boldsymbol{\phi}_0) = (I_n - \rho W)^{-1} \dot{D}(\boldsymbol{\phi})$  Let  $L_n(\boldsymbol{\tau}) = K^{-1} \{ \ell_n(\boldsymbol{\phi}, \boldsymbol{\gamma}_1) - \ell_n(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_{1-0}) \}$ . Then, through the Taylor expansion and a simple calculation, we obtain

$$\begin{aligned} L_n(\boldsymbol{\tau}) &= \frac{1}{K} \{ \ell_n(\boldsymbol{\phi}_0 + \eta \boldsymbol{\tau}_1, \boldsymbol{\gamma}_{1-0} + \eta \boldsymbol{\tau}_2) - \ell_n(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_{1-0}) \} \\ &\leq \frac{1}{K} \eta \left( W_n(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_0)^T, V_n(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_{1-0})^T \right) (\mathbf{t}_1, \mathbf{t}_2)^T - \frac{1}{2K} (\mathbf{t}_1, \mathbf{t}_2)^T [-I(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_{1-0})] (\mathbf{t}_1, \mathbf{t}_2) n \eta^2 \{ 1 + o_p(1) \} \\ &\quad - \frac{n}{K} \sum_{j=1}^d \left[ p_{\lambda_j}(|\gamma_{1-j0}|) - p_{\lambda_j}(|\gamma_{1-j0}|) \right] \\ &= S_1 + S_2 + S_3 + o_p(1). \end{aligned}$$

Notice that  $S_1 = O_p \left( 1 + n^{r/(2r+1)a_n} \right) \|\mathbf{t}\|$  and  $S_2 = O_p \left( \sqrt{n} K^{-1} \eta^2 \right) \|\mathbf{t}\|^2 = O_p \left( 1 + 2n^{r/(2r+1)a_n} \right) \|\mathbf{t}\|^2$ . Hence, though selecting a sufficiently large  $C$ ,  $S_2$  dominates  $S_1$  uniformly in  $\|\mathbf{t}\| = C$ . Moreover, invoking  $p_{\lambda}(0) = 0$ , and by the standard argument of the Taylor expansion, we obtain

$$\begin{aligned} S_3 &\leq n \sum_{j=1}^d \left[ \eta \dot{p}_{\lambda_j}(|\gamma_{1-j0}|) \operatorname{sgn}(\gamma_{1-j0}) |t_{1j}| + \eta^2 \ddot{p}_{\lambda_j}(|\gamma_{1-j0}|) |t_{1j}|^2 \{ 1 + o(1) \} \right] \\ &\leq \sqrt{s-1} n K^{-1} \eta a_n \|\mathbf{t}\| + n K^{-1} \eta^2 b_n \|\mathbf{t}\|^2 \end{aligned}$$

Then, it is clear to present that  $S_3$  is dominated by  $S_2$  uniformly in  $\|\mathbf{u}\| = C$ . Therefore, selecting a sufficiently large  $C$ , (A.1) holds. Hence, there exist local minimizers  $\hat{\boldsymbol{\phi}}$  and  $\hat{\boldsymbol{\gamma}}_1$  such that

$$\|\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_{1-0}\| = O_p(\eta), \quad \|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\| = O_p(\eta).$$

By calculating, we obtain  $\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\| = O_p(\eta)$ , which finishes the proof of (i).

(ii) Note that

$$\begin{aligned} \|\hat{g}_j(t) - g_{j0}(t)\|^2 &= \int_T \{\hat{g}_j(t) - g_{j0}(t)\}^2 dt \\ &= \int_T \left\{ \delta^T(t) \hat{\gamma}_{1-j} - \delta^T(t) \gamma_{1-j0} + R_j(t) \right\}^2 dt \\ &\leq 2 \int_T \left\{ \delta^T(t) \hat{\gamma}_{1-j} - \delta^T(t) \gamma_{1-j0} \right\}^2 dt + 2 \int_T \left\{ R_j^2(t) \right\} \\ &= 2(\hat{\gamma}_{1-j} - \gamma_{1-j0})^T H(\hat{\gamma}_{1-j} - \gamma_{1-j0}) + 2 \int_T \left\{ R_j^2(t) \right\} \end{aligned}$$

Then, invoking  $\|H\| = O(1)$ , a simple calculation shows

$$(\hat{\gamma}_{1-k} - \gamma_{1-k0})^T H(\hat{\gamma}_{1-k} - \gamma_{1-k0}) = O_p\left(n^{-2r/(2r+1)} + a_n^2\right). \tag{A2}$$

In addition, it is easy to show that

$$\int_T R_j^2(t) dt = O_p\left(n^{-2r/(2r+1)}\right). \tag{A3}$$

Invoking (A.2) and (A.3), we finish the proof of (ii).  $\square$

**Proof of Theorem 2.** (i) From  $\lambda_{\max} \rightarrow 0$ , it is easy to show that  $a_n = 0$  for large  $n$ . Then, by Theorem 1, it is sufficient to show that, for any  $\phi_j$  which satisfies

$$\|\phi_j - \phi_{j0}\| = O_p\left(n^{-r/(2r+1)}\right), \quad j = 1, \dots, s - 1,$$

and some given small  $\epsilon = Cn^{-r/(2r+1)}$  and  $j = s, \dots, p - 1$ , when  $n \rightarrow \infty$ , with probability approximating to one, we obtain  $\partial Q(\phi, \gamma) / \partial \phi_j > 0$  for  $0 < \phi_j < \epsilon$ , and  $\partial Q(\phi, \gamma) / \partial \phi_j < 0$  for  $-\epsilon < \phi_j < 0$ . Let

$$Q_n(\phi, \gamma_1) = \sum_{i=1}^n \exp\left\{-\left(Y_i - G_i^T(\phi) \gamma_1\right)^2 / \gamma_2\right\} \tag{A4}$$

a simple calculation shows that

$$\begin{aligned} \frac{\partial \ell(\phi, \gamma_1)}{\partial \phi_j} &= \frac{\partial Q_n(\phi, \gamma_1)}{\partial \phi_j} \\ &= \sum_{i=1}^n \frac{2}{\gamma_2} \exp\left\{-\left(Y_i - G_i^T(\phi) \gamma_1\right)^2 / \gamma_2\right\} \left\{ Y_i - G_i(\phi)^T \gamma_1 \right\} \dot{G}_i(\phi)^T \gamma_1 e_{\phi_j}^T U_i \end{aligned}$$

where  $e_{\phi_j} = \left(-\left(1 - \|\phi\|^2\right)^{-1/2} \phi_j, 0, \dots, 0, 1, 0, \dots, 0\right)^T$  with  $(j + 1)$ th component 1. Under conditions (C1), (C2), (C3), and Theorem 1, it is easy to present that

$$\frac{\partial \ell(\phi, \gamma)}{\partial \phi_j} = O_p\left(n^{-r/(2r+1)}\right).$$

The sign of the derivative is completely determined by that of  $\phi_j$ . Then,  $\partial Q(\phi, \gamma_1) / \partial \phi_j > 0$  for  $0 < \phi_j < \epsilon$ , and  $\partial Q(\phi, \gamma_1) / \partial \phi_j < 0$  for  $-\epsilon < \phi_j < 0$  hold. This finishes the proof of (i).

(ii) Applying the similar arguments as in the proof of (i), we obtain, with probability approximating to one,  $\hat{\gamma}_{1j} = 0, j = d + 1, \dots, q$ . Then, invoking

$$\sup_t \|\delta(t)\| = O(1),$$

the result of this theorem is obtained from  $\hat{g}_j(t) = \delta^T(t)\hat{\gamma}_{1-j}$ .  $\square$

**Proof of Theorem 3.** Though Theorems 1 and 2, we obtain that, as  $n \rightarrow \infty$ , with probability approximating to one,  $\ell(\boldsymbol{\phi}, \boldsymbol{\gamma}_1)$  reaches the local maximizer at  $(\hat{\boldsymbol{\phi}}^{*T}, 0^T)$  and  $(\hat{\boldsymbol{\gamma}}_1^{*T}, 0^T)$ . Let

$$Q_{1n}(\boldsymbol{\phi}, \boldsymbol{\gamma}_1) = \frac{\partial \ell(\boldsymbol{\phi}, \boldsymbol{\gamma}_1)}{\partial \boldsymbol{\phi}^*}, \quad Q_{2n}(\boldsymbol{\phi}, \boldsymbol{\gamma}_1) = \frac{\partial \ell(\boldsymbol{\phi}, \boldsymbol{\gamma}_1)}{\partial \boldsymbol{\gamma}_1^*}.$$

Then,  $(\hat{\boldsymbol{\phi}}^{*T}, 0^T)$  and  $(\hat{\boldsymbol{\gamma}}_1^{*T}, 0^T)$  satisfy

$$\begin{aligned} & \frac{1}{n} Q_{1n} \left( (\hat{\boldsymbol{\phi}}^{*T}, 0^T)^T, (\hat{\boldsymbol{\gamma}}_1^{*T}, 0^T)^T \right) \\ &= \frac{-2}{n} \sum_{i=1}^n \frac{2}{\gamma_2} \exp \left\{ -(Y_i - G_i^T(\boldsymbol{\phi})\boldsymbol{\gamma}_1)^2 / \gamma_2 \right\} (Y_i - G_i^{*T}(\hat{\boldsymbol{\phi}}_0^*)\hat{\boldsymbol{\gamma}}_1^*) \dot{G}_i^{*T}(\hat{\boldsymbol{\phi}}^*)\hat{\boldsymbol{\gamma}}_1^* J_{\hat{\boldsymbol{\phi}}^*}^T U_i^{*T} - V_1 = 0 \end{aligned}$$

where

$$V_1 = \left( \dot{p}_{\lambda_1}(|\hat{\gamma}_{1-1}|) \frac{\hat{\gamma}_{1-1}^T H}{|\hat{\gamma}_{1-1}|}, \dots, \dot{p}_{\lambda_d}(|\hat{\gamma}_{1-d}|) \frac{\hat{\gamma}_{1-d}^T H}{|\hat{\gamma}_{1-d}|} \right)^T.$$

Applying the Taylor expansion to  $\dot{p}_{\lambda_j}(|\hat{\gamma}_{1-j}|)$ , we have

$$\dot{p}_{\lambda_j}(|\hat{\gamma}_{1-j}|) = \dot{p}_{\lambda_j}(|\hat{\gamma}_{1-j0}|) + \left\{ \ddot{p}_{\lambda_j}(|\hat{\gamma}_{1-j0}|) + o_p(1) \right\} (\hat{\gamma}_{1-j} - \gamma_{1-j0}).$$

Furthermore, condition (C6) implies that  $\ddot{p}_{\lambda_j}(|\hat{\gamma}_{1-j0}|) = o_p(1)$ , and note that  $\dot{p}_{\lambda_j}(|\hat{\gamma}_{1-j0}|) = 0$  as  $\lambda_{\max} \rightarrow 0$ . From Theorems 1 and 2, we have

$$\dot{p}_{\lambda_j}(|\hat{\gamma}_{1-j}|) \frac{\hat{\gamma}_{1-j}^T H}{|\hat{\gamma}_{1-j}|} = o_p(\hat{\boldsymbol{\gamma}}_1^* - \boldsymbol{\gamma}_{1-0}^*).$$

Therefore, a simple calculation shows that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (Y_i - G_i^{*T}(\hat{\boldsymbol{\phi}}_0^*)\hat{\boldsymbol{\gamma}}_1^*) G_i^*(\hat{\boldsymbol{\phi}}^*) &= \frac{1}{n} \sum_{i=1}^n \left\{ \varepsilon_i + R^T(\boldsymbol{\alpha}_0^{*T} U_i^*) Z_i^* - G_i^{*T}(\boldsymbol{\phi}_0^*)(\hat{\boldsymbol{\gamma}}_1^* - \boldsymbol{\gamma}_{1-0}^*) \right. \\ &\quad \left. - [G_i^*(\hat{\boldsymbol{\phi}}_0^*) - G_i^*(\boldsymbol{\phi}_0^*)]^T \hat{\boldsymbol{\gamma}}_1^* \right\} \left\{ G_i^*(\boldsymbol{\phi}_0^*) + [G_i^*(\hat{\boldsymbol{\phi}}^*) - G_i^*(\boldsymbol{\phi}_0^*)] \right\} \end{aligned}$$

Let

$$\Lambda_n = \frac{1}{n} \sum_{i=1}^n G_i^*(\boldsymbol{\phi}_0^*) \left( \varepsilon_i + R^T(U_i^{*T} \boldsymbol{\alpha}_0^*) Z_i^* \right).$$

Then, from conditions (C8) and (C9), Theorem 1, and  $\sup_i \|\delta(t)\| = O(1)$ , we obtain

$$\hat{\boldsymbol{\gamma}}_1^* - \boldsymbol{\gamma}_{1-0}^* = [\Phi_n + o_p(1)]^{-1} \{ \Lambda_n - \Psi_n(\hat{\boldsymbol{\phi}}^* - \boldsymbol{\phi}_0^*) \}.$$

Thus, we can have

$$\begin{aligned} \frac{\ell_n \left( (\hat{\boldsymbol{\phi}}^{*T}, 0^T)^T, (\hat{\boldsymbol{\gamma}}_1^{*T}, 0^T)^T \right)}{\partial \boldsymbol{\phi}_j^*} &= \frac{\partial Q_n \left\{ (\boldsymbol{\phi}_0^{*T}, 0^T)^T, (\boldsymbol{\gamma}_1^{*T}, 0^T)^T \right\}}{\partial \boldsymbol{\phi}_j^*} \\ &= \frac{\partial Q_n \left\{ (\boldsymbol{\phi}_0^{*T}, 0^T)^T, (\boldsymbol{\gamma}_{1-0}^{*T}, 0^T)^T \right\}}{\partial \boldsymbol{\phi}_j^*} \\ &\quad + \sum_{l=1}^{s-1} \left\{ \frac{\partial^2 Q_n \left\{ (\boldsymbol{\phi}_0^{*T}, 0^T)^T, (\boldsymbol{\gamma}_{1-0}^{*T}, 0^T)^T \right\}}{\partial \boldsymbol{\phi}_j^* \partial \boldsymbol{\phi}_l^*} + o_p(1) \right\} (\hat{\boldsymbol{\phi}}_l^* - \boldsymbol{\phi}_{01}^*) \end{aligned}$$

where  $Q_n(\boldsymbol{\phi}, \gamma_1)$  is defined in (A.4). As the definition of  $J_{\boldsymbol{\phi}_0^*}$  that  $\boldsymbol{\alpha} - \boldsymbol{\alpha}_0 = J_{\boldsymbol{\phi}_0^*}(\boldsymbol{\phi} - \boldsymbol{\phi}_0^*) + O_p(n^{-1})$ . Since  $\sqrt{n}(\gamma_{2-n} - \gamma_{2-0}) = o_p(1)$ , the proof is proved by Slutsky's lemma and the central limit theorem. This ends with proof of Theorem 3.  $\square$

Thus, the proof is completed.

## References

1. Cliff, A.D. *Spatial Autocorrelation*; Technical Report; Pion: London, UK, 1973.
2. Su, L. Semiparametric GMM estimation of spatial autoregressive models. *J. Econom.* **2012**, *167*, 543–560. [[CrossRef](#)]
3. Zhang, Y.; Shen, D. Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects. *J. Stat. Plan. Inference* **2015**, *159*, 64–80. [[CrossRef](#)]
4. Fan, J.; Yao, Q.; Cai, Z. Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B* **2003**, *65*, 57–80. [[CrossRef](#)]
5. Lu, Z.; Tjøstheim, D.; Yao, Q. Adaptive varying-coefficient linear models for stochastic processes: Asymptotic theory. *Stat. Sin.* **2007**, *17*, 177–S35.
6. Xue, L.; Wang, Q. Empirical likelihood for single-index varying-coefficient models. *Bernoulli* **2012**, *18*, 836–856. [[CrossRef](#)]
7. Huang, Z.; Zhang, R. Profile empirical-likelihood inferences for the single-index-coefficient regression model. *Stat. Comput.* **2013**, *23*, 455–465. [[CrossRef](#)]
8. Wang, X.; Jiang, Y.; Huang, M.; Zhang, H. Robust variable selection with exponential squared loss. *J. Am. Stat. Assoc.* **2013**, *108*, 632–643. [[CrossRef](#)]
9. Jiang, Y. Robust estimation in partially linear regression models. *J. Appl. Stat.* **2015**, *42*, 2497–2508. [[CrossRef](#)]
10. Song, Y.; Jian, L.; Lin, L. Robust exponential squared loss-based variable selection for high-dimensional single-index varying-coefficient model. *J. Comput. Appl. Math.* **2016**, *308*, 330–345. [[CrossRef](#)]
11. Wang, K.; Lin, L. Robust structure identification and variable selection in partial linear varying coefficient models. *J. Stat. Plan. Inference* **2016**, *174*, 153–168. [[CrossRef](#)]
12. Yu, Y.; Ruppert, D. Penalized spline estimation for partially linear single-index models. *J. Am. Stat. Assoc.* **2002**, *97*, 1042–1054. [[CrossRef](#)]
13. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [[CrossRef](#)]
14. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
15. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
16. Wang, H.; Li, G.; Jiang, G. Robust regression shrinkage and consistent variable selection through the LAD-lasso. *J. Bus. Econ. Stat.* **2007**, *25*, 347–355. [[CrossRef](#)]
17. Song, Y.; Liang, X.; Zhu, Y.; Lin, L. Robust variable selection with exponential squared loss for the spatial autoregressive model. *Comput. Stat. Data Anal.* **2021**, *155*, 107094. [[CrossRef](#)]
18. Yuille, A.L.; Rangarajan, A. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–8 December 2001; Volume 14
19. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [[CrossRef](#)]
20. Forsythe, G.E. *Computer Methods for Mathematical Computations*; Prentice-Hall: Hoboken, NJ, USA, 1977; Volume 259.
21. Liang, H.; Li, R. Variable selection for partially linear models with measurement errors. *J. Am. Stat. Assoc.* **2009**, *104*, 234–248. [[CrossRef](#)] [[PubMed](#)]
22. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **1991**, *7*, 186–199. [[CrossRef](#)]
23. Schumaker, L. *Spline Functions: Basic Theory*; Cambridge Mathematical Library; Cambridge, UK, 1981.

**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.