mathematics

MDPI

Article

Robust Estimation for Semi-Functional Linear Model with Autoregressive Errors

Bin Yang [1,2], Min Chen [3,4], Tong Su [1] and Jianjun Zhou [1,*]

[1] Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China
[2] City College, Kunming University of Science and Technology, Kunming 650500, China
[3] School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China
[4] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
* Correspondence: jjzhou@ynu.edu.cn

Abstract: It is well-known that the traditional functional regression model is mainly based on the least square or likelihood method. These methods usually rely on some strong assumptions, such as error independence and normality, that are not always satisfied. For example, the response variable may contain outliers, and the error term is serially correlated. Violation of assumptions can result in unfavorable influences on model estimation. Therefore, a robust estimation procedure of a semi-functional linear model with autoregressive error is developed to solve this problem. We compare the efficiency of our procedure to the least square method through a simulation study and two real data analyses. The conclusion illustrates that the proposed method outperforms the least square method, providing random errors follow the heavy-tail distribution.

Keywords: autoregressive errors; heavy-tail distribution; outlier; robust estimation; semi-functional linear model

MSC: 62R10; 62M10

Citation: Yang, B.; Chen, M.; Su, T.; Zhou, J. Robust Estimation for Semi-Functional Linear Model with Autoregressive Errors. *Mathematics* **2023**, *11*, 277. https://doi.org/10.3390/math11020277

Academic Editor: Christophe Chesneau

Received: 6 November 2022
Revised: 31 December 2022
Accepted: 2 January 2023
Published: 5 January 2023

1. Introduction

As a powerful tool in functional data analysis (FDA), functional regression modeling has been extensively studied in the past several decades. For example, Cardot et al. [1,2] introduced the functional linear model in which a scalar response variable is explained by a functional predictor. Yao et al. [3] proposed a functional linear regression model to analyse sparse longitudinal data. Hall and Horowitz [4] obtained the optimal convergence rates of estimators based on functional principal components analysis. Further, Li et al. [5] recently extended classical clusterwise linear regression to incorporate multiple functional predictors and proposed clusterwise functional linear regression models. For the functional nonparametric model, Ferraty and Vieu [6] extended the kernel regression to functional data. Baíllo and Grané [7] proposed a local linear regression and Burba et al. [8] used the k-Nearest Neighbour method to deal with the functional nonparametric models, respectively. Moreover, to improve the power of prediction or interpretation of the functional regression model, some functional semiparametric models have also been developed. For instance, Aneiros-Pérez and Vieu [9] first introduced a semi-functional partial linear regression model, and Aneiros-Pérez and Vieu [10] used this model to predict time series. Further, Shin [11] proposed new estimators of a partial functional linear model which explore the relationship between a scalar response variable and mixed-type predictors. Zhou and Chen [12] introduced a semi-functional linear model by combining the feature of a functional linear regression model and a nonparametric regression model.

It is well-known that the traditional functional regression models are mainly based on the least square or likelihood method. However, small proportions of outliers or the

heavy-tailed nature of the error distribution could dramatically deteriorate the efficiency of estimators. To tackle this problem, many robust estimations have been introduced into the functional regression model. For example, Yu et al. [13] proposed a robust estimation method based on exponential square loss for the semi-functional linear regression model. Subsequently, Yu et al. [14] proposed a robust estimation procedure of a partial function linear model based on modal regression. Cai et al. [15] constructed a robust estimation with a modified Huber's loss for partial functional linear models. Cao and Sun [16] studied the robust estimation of partial functional linear regression models based on functional principal component analysis and weighted compound quantile regression. Tang [17] introduced a robust estimation method for the semi-functional linear model, and Boente et al. [18] constructed new robust estimators for semi-functional linear regression models with a bounded loss function and a preliminary residual scale estimator. Recently, Boente and Daniela [19] considered the robust estimation for functional quadratic regression models. Moreover, Pannu and Billor [20] proposed a functional adaptive group LASSO variable selection method based on the weighted least absolute deviation. In the aforementioned literature, the random error is usually assumed to be independent and identically distributed. However, this assumption may be inappropriate in practice, especially when the observations are collected sequentially over time. However, as far as we know, only a few works have focused on robust estimation for regression models with serially correlated error. For example, Sanjoy et al. [21] considered robust estimation of nonlinear regression with autoregressive errors. Riazoshams et al. [22] studied the performance of a robust two-stage estimator in nonlinear regression with an autocorrelated error. Recently, Serenay and Baris [23] proposed a novel hybrid robust tapering approach in nonlinear regression with the presence of autocorrelation and outliers. Motivated by these previous works, this paper is mainly concerned with a robust semi-functional linear model with autoregressive errors (SFLAR). The rest of the paper is organized as follows. Section 2 introduces robust estimation and algorithm. Some simulation studies are presented in Section 3. Section 4 illustrates the good performance of the robust estimation method with two real data. Section 5 is the conclusion.

2. Model and Estimation

Suppose that $\{x_i(t), y_i, z_i\}_{i=1}^n$ satisfies the following semi-functional linear model with autoregressive errors (SFLAR)

$$\begin{aligned} y_i &= \int_{\mathcal{T}} \beta(t) x_i(t) \mathrm{d}t + g(z_i) + \varepsilon_i, \\ \varepsilon_i &= \sum_{l=1}^{q} a_l \varepsilon_{i-l} + e_i, \end{aligned} \tag{1}$$

where y_i is a real-valued random response variable, and z_i is a real-valued covariate defined on a compact interval $z = [\underline{z}, \bar{z}]$. The functional covariate $x_i(t)$ defined on an interval $\mathcal{T} \subset R$ is a zero mean, second-order (i.e., $\mathbb{E}|x(t)|^2 < \infty$ for all $t \in \mathcal{T}$) stochastic process defined on (Ω, B, P) with sample paths in the Hilbert space $L^2(\mathcal{T})$. The inner product of $L^2(\mathcal{T})$ is defined by $\langle u, v \rangle = \int_{\mathcal{T}} u(t)v(t)\mathrm{d}t$ for any $u, v \in L^2(\mathcal{T})$ and norm $\|u\| = \langle u, u \rangle^{1/2}$. The unknown slope function $\beta(t)$ belongs to $L^2(\mathcal{T})$. Without losing generality, throughout this paper, we assume $\mathcal{T} = [0, 1]$. The nonparametric function $g(z)$ is an unknown smooth function; e_i is a random error with zero mean, finite variance, and independent of $(x_i(t), z_i)$.

Since the unknown functions $\beta(t)$ and $g(z)$ are infinite-dimensional parameters, we cannot obtain their estimators directly via an optimization method. Therefore, some dimension reduction methods need to be applied. Specifically, suppose the covariance operator C_x of the functional variable x is denoted by

$$C_x(s, u) = Cov[x(s), x(u)], \tag{2}$$

where $C_x(s,u)$ is continuous on $[0,1]^2$, and the eigenvalues of C_x, denoted by $\{\lambda_j\}_{j=1}^{\infty}$, satisfy $\lambda_1 > \lambda_2 > \cdots > 0$. By Mercer's Theorem, $C_x(s,u)$ can be expanded by

$$C_x(s,u) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(u), \tag{3}$$

where ϕ_j is the continuous orthogonal eigenfunction, and λ_j is the corresponding eigenvalue. Obviously, the eigenfunction sequence $\{\phi_j\}_{j=1}^{\infty}$ forms an orthonormal basis for $L^2[0,1]$. Then, random function $x_i(t)$ and the slope function $\beta(t)$ can be respectively expressed as

$$x_i(t) = \sum_{j=1}^{\infty} \xi_{ij}\phi_j(t), \quad \beta(t) = \sum_{j=1}^{\infty} b_j\phi_j(t), \tag{4}$$

where $\xi_{ij} = \int_{\mathcal{T}} x_i(t)\phi_j(t)\mathrm{d}t$ are called the principle component scores satisfying $\mathbb{E}\xi_{ij} = 0$, and $\mathbb{E}\xi_{ij}\xi_{ik} = \lambda_j I(j=k)$, and $b_j = \int_{\mathcal{T}} \beta(t)\phi_j(t)\mathrm{d}t$.

Moreover, let S_{r,N_n} be the space of polynomial splines defined on interval $[\underline{z}, \overline{z}]$ with degree $r-1$ and knot sequence $\underline{z} < z_1 < \cdots < z_{N_n} < \overline{z}$. The space S_{r,N_n} is a K-dimensional linear space, $K = K(n) = N_n + r$. Following the arguments of de Boor [24], we can conclude that, if it is sufficiently smooth, the nonparametric function $g(z)$ can be approximately expressed as

$$g(z) \approx \sum_{k=1}^{K} c_k B_k(z), \tag{5}$$

where B_k, $k = 1, 2, \cdots, K$ are B-spline basis functions in S_{r,N_n}. For more details on spline function, we refer to de Boor [24]. Substituting Equation (4) and Equation (5) into Equation (1) yields

$$y_i \approx \sum_{j=1}^{J} \xi_{ij} b_j + \sum_{k=1}^{K} c_k B_k(z_i) + \varepsilon_i. \tag{6}$$

However, because the covariance function C_x is unknown, eigenfunctions ϕ_j are also unknown, and variables ξ_{ij} are unobservable. To tackle this problem, we define the empirical version of covariance function C_x by

$$\widehat{C}_x(s,u) = \frac{1}{n}\sum_{i=1}^{n} x_i(s)x_i(u), \tag{7}$$

define the estimated eigenelements of $\widehat{C}_x$ by

$$\int \widehat{C}_x(s,u)\hat{\phi}_j(u)\mathrm{d}u = \hat{\lambda}_j\hat{\phi}_j(s), \quad (1 < j < n). \tag{8}$$

By Mercer's Theorem, we have

$$\widehat{C}_x(s,u) = \sum_{j=1}^{n} \hat{\lambda}_j\hat{\phi}_j(s)\hat{\phi}_j(u), \tag{9}$$

where $(\hat{\lambda}_j, \hat{\phi}_j)$ are (eigenvalue, eigenfunction) pairs for the linear operator with kernel $\widehat{C}_x$, ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 > \cdots \geq 0$. We regard $(\hat{\lambda}_j, \hat{\phi}_j)$ as an estimator of (λ_j, ϕ_j). Moreover, Equation (6) can be written as

$$y_i \approx \widehat{\boldsymbol{U}}_i^{\top}\boldsymbol{b} + \boldsymbol{B}_i^{\top}\boldsymbol{c} + \sum_{l=1}^{q} a_l\left(y_{i-l} - \widehat{\boldsymbol{U}}_{i-l}^{\top}\boldsymbol{b} - \boldsymbol{B}_{i-l}^{\top}\boldsymbol{c}\right) + e_i, \quad q+1 \leq i \leq n. \tag{10}$$

Furthermore, denote

$$L_n(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) = \frac{1}{n-q}\sum_{i=q+1}^{n}\rho\left(y_i - \widehat{\boldsymbol{U}}_i^{\top}\boldsymbol{b} - \boldsymbol{B}_i^{\top}\boldsymbol{c} - \sum_{l=1}^{q}a_l\left(y_{i-l} - \widehat{\boldsymbol{U}}_{i-l}^{\top}\boldsymbol{b} - \boldsymbol{B}_{i-l}^{\top}\boldsymbol{c}\right)\right), \tag{11}$$

where $\rho(\cdot)$ is a proper robust loss function, $\boldsymbol{a} = (a_1,\cdots,a_q)^{\top}$, $\boldsymbol{b} = (b_1,\cdots,b_J)^{\top}$, $\boldsymbol{c} = (c_1,\cdots,c_K)^{\top}$, $\widehat{\boldsymbol{U}}_i = (\hat{\xi}_{i1},\hat{\xi}_{i2},\cdots,\hat{\xi}_{iJ})^{\top}$, and $\boldsymbol{B}_i = (B_1(z_i),\cdots,B_K(z_i))^{\top}$. Therefore, $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{c}}$ can be obtained by solving the following minimization problem:

$$\min_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} L_n(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}). \tag{12}$$

Finally, the estimator of functional coefficient $\beta(t)$ and the spline estimator of the nonparametric function $g(z)$ are given by $\hat{\beta}(t) = \sum_{j=1}^{J}\hat{b}_j\hat{\phi}_j(t)$, $\hat{g}(z_i) = \sum_{k=1}^{K}\hat{c}_kB_k(z_i)$, respectively.

Various loss functions $\rho(\cdot)$ have been proposed for the robust estimation when outliers are present in the data. For example, the least absolute deviations (LAD) loss, that is, $\rho(r) = |r|$, can be regarded as a median regression. Although the LAD is not sensitive to outliers, its disadvantage is that the LAD is not smooth at the zero point, which will lead to some problems in model training. Therefore, considering the requirements of robustness and differentiability of the loss function, the Huber loss function (HB) [25] is considered in this paper, which is convex, smooth, and symmetric about 0. Hence, the optimization of Equation (12) for estimators is well-defined, and the local minimum problem is avoided.

To implement the proposed method, we take B-spline functions with equally spaced knots and the fixed degree three to approximate the unknown function $g(z)$. We only need to choose the numbers of eigenfunctions J and B-spline functions K. Many methods, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), can be used to select the numbers of eigenfunctions J and B-spline functions K. In this paper, we choose the truncation parameters J and K by minimizing the following Bayesian information criterion (BIC) criteria:

$$\mathrm{BIC}(J,K) = \log\left\{\sum_{i=q+1}^{n}\rho\left(y_i - \sum_{j=1}^{J}\hat{b}_j\hat{\xi}_{ij} - \sum_{k=1}^{K}\hat{c}_kB_k(z_i)\right)\right\} + \frac{\log n}{2n}(J+K), \tag{13}$$

which is a function of J and K, the smaller the value of BIC, the better the fitting.

Furthermore, denote $\boldsymbol{\theta} = (\boldsymbol{b}^{\top},\boldsymbol{c}^{\top})^{\top}$, $\widehat{\boldsymbol{H}}_i^{\top} = (\widehat{\boldsymbol{U}}_i^{\top},\boldsymbol{B}_i^{\top})^{\top}$. Then, Equation (11) can be written as the following equation:

$$L_n(\boldsymbol{a},\boldsymbol{\theta}) = \frac{1}{n-q}\sum_{i=q+1}^{n}\rho\left(y_i - \widehat{\boldsymbol{H}}_i^{\top}\boldsymbol{\theta} - \sum_{l=1}^{q}a_l\left(y_{i-l} - \widehat{\boldsymbol{H}}_{i-l}^{\top}\boldsymbol{\theta}\right)\right). \tag{14}$$

Moreover, Equation (11) can be rewritten as:

$$L_n(\boldsymbol{a},\boldsymbol{\theta}) = \frac{1}{n-q}\sum_{i=q+1}^{n}\rho\left(V_i - \boldsymbol{D}_i^{\top}\boldsymbol{\theta}\right), \tag{15}$$

where $V_i = y_i - \sum_{l=1}^{q}a_ly_{i-l}$, $\boldsymbol{D}_i = \widehat{\boldsymbol{H}}_i - \sum_{l=1}^{q}a_l\widehat{\boldsymbol{H}}_{i-l}^{\top}$. Then, the estimators of parameter vector $\boldsymbol{a} = (a_1,a_2,\cdots,a_q)^{\top}$ and $\boldsymbol{\theta} = (\boldsymbol{b}^{\top},\boldsymbol{c}^{\top})^{\top}$ can be obtained by minimizing Equation (15). However, the objective function $L_n(\boldsymbol{a},\boldsymbol{\theta})$ has no analytic solution for the unknown parameters. Therefore, this paper uses a two-step iteratively reweighted least-squares (TSIRLS) to obtain the local optimal solution $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{\theta}}$ of Equation (15). The TSIRLS algorithm is summarized as follows:

Initialization: given initial value: $\hat{\boldsymbol{\theta}}^{(0)}$, $\hat{\boldsymbol{a}}^{(0)}$ using the LS estimators and let $k = 0$.

Step 1. Update $\hat{\varepsilon}_i^{(k)}$, $\boldsymbol{W}^{(k)} = diag\{w_1^{(k)}, \cdots, w_n^{(k)}\}$:

$$\hat{\varepsilon}_i^{(k)} = y_i - \boldsymbol{H}_i^{\top}\hat{\boldsymbol{\theta}}^{(k)}, \quad w_i^{(k)} = \frac{\psi(\hat{\varepsilon}_i^{(k)})}{\hat{\varepsilon}_i^{(k)}};$$

Step 2. Update $\tilde{\boldsymbol{V}}^{(k)}$, $\tilde{\boldsymbol{D}}^{(k)}$:

$$\tilde{\boldsymbol{V}}^{(k)} = \begin{pmatrix} \hat{\varepsilon}_{q+1}^{(k)} \\ \vdots \\ \hat{\varepsilon}_n^{(k)} \end{pmatrix}, \quad \tilde{\boldsymbol{D}}^{(k)} = \begin{pmatrix} \hat{\varepsilon}_q^{(k)} & \cdots & \hat{\varepsilon}_1^{(k)} \\ \vdots & & \vdots \\ \hat{\varepsilon}_{n-1}^{(k)} & \cdots & \hat{\varepsilon}_{n-q}^{(k)} \end{pmatrix};$$

Step 3. Update $\hat{e}_i^{(k)}$, $\tilde{\boldsymbol{W}}^{(k)} = diag\{\tilde{w}_1^{(k)}, \cdots, \tilde{w}_n^{(k)}\}$:

$$\hat{e}_i^{(k)} = \hat{\varepsilon}_i^{(k)} - \sum_{l=1}^{q} \hat{a}_l^{(k)}\hat{\varepsilon}_{i-l}^{(k)}, \quad \tilde{w}_i^{(k)} = \frac{\psi(\hat{e}_i^{(k)})}{\hat{e}_i^{(k)}};$$

Step 4. Update $\hat{\boldsymbol{a}}^{(k+1)}$:

$$\hat{\boldsymbol{a}}^{(k+1)} = (\tilde{\boldsymbol{D}}^{(k)\top}\tilde{\boldsymbol{W}}^{(k)}\tilde{\boldsymbol{D}}^{(k)})^{-1}\tilde{\boldsymbol{D}}^{(k)\top}\tilde{\boldsymbol{W}}^{(k)}\tilde{\boldsymbol{V}}^{(k)};$$

Step 5. Update $V_i^{(k)}$, $\boldsymbol{D}_i^{(k)}$:

$$V_i^{(k)} = y_i - \sum_{l=1}^{q} a_l^{(k+1)} y_{i-l}, \quad \boldsymbol{D}_i^{(k)} = \widehat{\boldsymbol{H}}_i - \sum_{l=1}^{q} a_l^{(k+1)}\widehat{\boldsymbol{H}}_{i-l};$$

Step 6. Update $\hat{\boldsymbol{\theta}}^{(k+1)}$:

$$\hat{\boldsymbol{\theta}}^{(k+1)} = (\boldsymbol{D}^{(k)\top}\boldsymbol{W}^{(k)}\boldsymbol{D}^{(\boldsymbol{k})})^{-1}\boldsymbol{D}^{(k)\top}\boldsymbol{W}^{(k)}\boldsymbol{V}^{(k)},$$

where

$$\boldsymbol{V}^{(k)} = \begin{pmatrix} V_{q+1}^{(k)} \\ \vdots \\ V_n^{(k)} \end{pmatrix}, \quad \boldsymbol{D}^{(k)} = \begin{pmatrix} \boldsymbol{D}_{q+1}^{(k)} \\ \vdots \\ \boldsymbol{D}_n^{(k)} \end{pmatrix};$$

Step 7. Update $k = k + 1$. Repeat the above steps until $\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\|_1 + \|\hat{\boldsymbol{a}}^{(k+1)} - \hat{\boldsymbol{a}}^{(k)}\|_1 \leq 10^{-8}$. Then, the estimation of the parameter vectors $\hat{\boldsymbol{a}} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_q)^{\top}$, $\hat{\boldsymbol{b}} = (\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_J)^{\top}$, and $\hat{\boldsymbol{c}} = (\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_K)^{\top}$ are obtained.

3. Simulation

In this section, we conduct some simulation studies to investigate the finite sample performance of our proposed method. The simulation data $\{(x_i(t), y_i, z_i), 1 \leq i \leq n\}$ are generated from the following SFLAR model:

$$y_i = \int_{\mathcal{T}} \beta(t)x_i(t)dt + g(z_i) + \varepsilon_i, \quad \varepsilon_i = \sum_{l=1}^{q} a_l\varepsilon_{i-l} + e_i, \tag{16}$$

where z_i is distributed uniformly on $[-1, 1]$, and the functional predictor $x_i(t) = \sum_{j=1}^{50} \xi_{ij}\phi_j(t)$, ξ_{ij} is distributed as independent normal with mean 0 and variance $\lambda_j = ((j - 0.5)\pi)^{-1}$ and $\phi_j(t) = \sqrt{2}\sin((j - 0.5)\pi t)$. For the unknown functions $\beta(\cdot)$ and $g(\cdot)$ and autoregressive co-

efficients $(a_1, \cdots, a_q)$, given $\beta(t) = \sqrt{2}\sin(0.5\pi t) + 3\sqrt{2}\sin(1.5\pi t)$, $g(z) = z + 3\cos(2\pi z)$, the first-order autoregressive error $AR(1) : \varepsilon_t = 0.8\varepsilon_{t-1} + e_t$. We evaluate the performance of various loss functions with different sample sizes $n = 100, 300$ and 500. Each experiment has been repeated 1000 times. Similar to Cai et al. [15], the following four scenarios error distributions are considered:
(**1**) $e_i \sim N(0, 1)$;
(**2**) $e_i \sim \pi N(0, 1) + (1 - \pi)N(0, 100)$, where π = 0.8;
(**3**) $e_i \sim t(2)$;
(**4**) $e_i \sim Cauchy(0, 1)$.

Moreover, according to one of the reviewers' suggestions, we also consider different values of π in Scenario 2. Specifically, the values of π are taken as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9.

To evaluate the performances of our estimators, we define the integrated mean square errors (IMSE) of the estimators of function $\beta(t)$, $g(z)$,

$$\text{IMSE}_1 = \int_0^1 [\hat{\beta}(t) - \beta(t)]^2 \mathrm{d}t, \quad \text{IMSE}_2 = \int_0^1 [\hat{g}(z) - g(z)]^2 \mathrm{d}z, \tag{17}$$

and the mean square prediction error (MSPE):

$$\text{MSPE} = \frac{1}{n} \sum_{i=q+1}^{n} \left(\widehat{y}_i - g(z_i) - \int_0^1 \beta(t) x_i(t) \mathrm{d}t \right)^2, \tag{18}$$

where $\widehat{y}_i = \widehat{g}(z_i) + \int_{\mathcal{T}} \widehat{\beta}(t) x_i(t) \mathrm{d}t$. Further, in order to obtain $\widehat{\beta}(t)$ and $\widehat{g}(z_i)$, the two tuning parameters J and K are selected by minimizing the BIC value. Specifically, the range of J is selected within 1 to 6, and the range of K is selected within 3 to 6.

We compare the finite sample performance of the SFLAR model with ordinary least squares (OLS), least absolute deviations (LAD) loss, and Huber (HB) loss function. Further, according to one of the reviewers' suggestions, we also consider the exponential square (EXP) loss function (Yu et al. [13]). To implement the robust method considered in Yu et al. [13], the tuning parameters h, J and K need to be selected. In this paper, the tuning parameters are chosen by cross-validation and BIC. Specifically, for some fixed h, the tuning parameters J and K are selected by BIC. Then, under the selected J and K, like Jiang [26], the tuning parameter h is chosen by 5−fold cross-validation (CV) within the possible grids points $h = 0.5 \times 1.02^l$ $(l = 0, 1, \cdots, 100)$. We repeat the process until the tuning parameters are invariant.

Table 1 reports the bias and mean square error (MSE) of estimator $\hat{a}$ under different error distributions. Tables 2 and 3 present the sample mean and standard deviation (SD) of IMSE_1 and IMSE_2 under different error distributions. Table 4 lists the MSPE results of model prediction under different error distributions for different estimation methods. The symbol "-" indicates that these values are much higher than others, so they are not listed in the table. Figure 1 displays the sample means of $\text{IMSE} = \text{IMSE}_1 + \text{IMSE}_2$ for functional parameters $\beta(t)$, $g(z)$ and MSE $(\times 10^3)$ for autoregressive coefficient a when $n = 500$, and π takes values in the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ for Scenario (2).

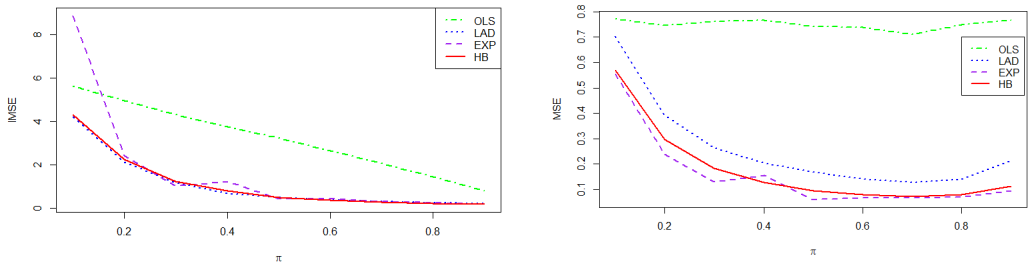
From the simulation results of different estimation methods under different error distributions in Tables 1–4 and Figure 1, we can conclude that:

(**1**) For the autoregressive coefficient vector a, the estimation results of all methods based on OLS, LAD, EXP and HB seem equally good, whenever the error distribution is normal or heavy-tailed, which indicates the estimator $\hat{a}$ is not sensitive to the distribution of random error. At the same time, we can see that the values of Bias and MSE of HB and EXP are less than OLS. Hence, it can be concluded the robust estimators based on HB and EXP outperform the OLS method.

(**2**) For the slope function $\beta(t)$ and nonparametric function $g(z)$, the OLS estimator behaves badly when the error distribution is non-normal, particularly when the random error is heavy-tailed. In contrast, the estimators of LAD, EXP and HB are much better than

the OLS. Even if the expectation and variance of the error do not exist, the robust estimation method is still feasible.

(**3**) Whenever the distribution of random error is normal or heavy-tailed, the bias and MSE of $\hat{a}$, mean and SD of IMSE_1, IMSE_2 and MSPE based on LAD, EXP and HB methods decrease as the sample size increases. Hence, the proposed robust procedure is effective.



(**a**) (**b**)

Figure 1. (**a**) The sample mean of IMSE for functional parameters $\beta(t)$ and $g(z)$ when $n = 500$ and π takes values in the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$; (**b**) the sample mean of MSE ($\times 10^3$) for autoregressive coefficient a under the same setting.

Table 1. The Bias and MSE of $\hat{a}$ under different error distributions.

n	Methods	$N(0,1)$		MN		$t(2)$		*Cauchy* $(0,1)$	
		Bias	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
100	OLS	−0.0106	0.0043	−0.0084	0.0040	−0.0081	0.0037	−0.0225	0.0038
	LAD	−0.1424	0.0264	−0.0613	0.0068	−0.1036	0.0164	−0.0232	0.0017
	EXP	−0.0240	0.0059	−0.0025	0.0008	−0.0084	0.0021	−0.0007	0.0300
	HB	−0.0106	0.0044	−0.0019	0.0006	−0.0049	0.0019	−0.0011	0.0002
300	OLS	−0.0051	0.0014	−0.0067	0.0012	−0.0051	0.0012	−0.0097	0.0011
	LAD	−0.0400	0.0033	−0.0140	0.0004	−0.0226	0.0012	−0.0026	0.0000
	EXP	−0.0143	0.0019	−0.0016	0.0001	−0.0041	0.0005	0.0000	0.0094
	HB	−0.0053	0.0014	−0.0021	0.0002	−0.0031	0.0005	−0.0002	0.0000
500	OLS	−0.0045	0.0008	−0.0042	0.0007	−0.0017	0.0007	−0.0053	0.0005
	LAD	−0.0234	0.0015	−0.0002	0.0001	0.0099	0.0004	−0.0011	0.0000
	EXP	−0.0119	0.0010	−0.0006	0.0001	−0.0018	0.0002	−0.0001	0.0058
	HB	−0.0044	0.0008	−0.0009	0.0001	−0.0008	0.0002	−0.0002	0.0000

Table 2. The mean (SD) of IMSE_1 for $\hat{\beta}(t)$ under different error distributions and sample sizes.

n	Methods	$N(0,1)$	MN	$t(2)$	*Cauchy* $(0,1)$
		Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
100	OLS	0.2664 (0.1266)	1.4068 (1.2305)	0.8923 (4.2805)	- (-)
	LAD	0.3298 (0.1663)	0.4360 (0.2541)	0.4730 (0.3153)	0.4996 (0.3083)
	EXP	0.3243 (0.1714)	0.4247 (0.7672)	0.3786 (0.2029)	0.4602 (0.2534)
	HB	0.2711 (0.1305)	0.3963 (0.2124)	0.3800 (0.2211)	0.4734 (0.2721)
300	OLS	0.1099 (0.0445)	0.4753 (0.3015)	0.4589 (2.0454)	- (-)
	LAD	0.1272 (0.0526)	0.1870 (0.0788)	0.1438 (0.0638)	0.1964 (0.0830)
	EXP	0.1663 (0.0712)	0.1787 (0.0745)	0.1792 (0.0759)	0.1982(0.0815)
	HB	0.1110 (0.0451)	0.1813 (0.0764)	0.1357 (0.0569)	0.1986 (0.0805)
500	OLS	0.0836 (0.0305)	0.3248 (0.1985)	0.7908 (17.3387)	- (-)
	LAD	0.0936 (0.0352)	0.0928 (0.0346)	0.1026 (0.0417)	0.1557 (0.0573)
	EXP	0.1394 (0.0537)	0.1469 (0.0553)	0.1463 (0.0546)	0.1570(0.0583)
	HB	0.0844 (0.0308)	0.1474 (0.0550)	0.0980 (0.0380)	0.1567 (0.0584)

Table 3. The mean (SD) of $IMSE_2$ for $\hat{g}(\cdot)$ under different error distributions and sample sizes.

n	Methods	$N(0,1)$	MN	$t(2)$	*Cauchy* $(0,1)$
		Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
100	OLS	0.3509 (0.3710)	7.1999 (8.9383)	7.0279 (-)	- (-)
	LAD	0.4821 (0.7918)	1.4814 (2.1051)	1.1504 (4.6860)	1.9271 (2.9633)
	EXP	0.4132 (0.4437)	1.0579 (2.0157)	0.7482 (0.8314)	1.2962 (1.5780)
	HB	0.3713 (0.3992)	0.8917 (1.0224)	0.7717 (0.8362)	1.3913 (1.7224)
300	OLS	0.1140 (0.1166)	2.2063 (2.6667)	2.0490 (12.9143)	- (-)
	LAD	0.1597 (0.1726)	0.3016 (0.3629)	0.2406 (0.2751)	0.3699 (0.4413)
	EXP	0.1385 (0.1356)	0.2434 (0.2624)	0.2229 (0.2740)	0.3681 (0.4213)
	HB	0.1189 (0.1242)	0.2388 (0.2597)	0.2074 (0.2528)	0.3629 (0.4087)
500	OLS	0.0677 (0.0754)	1.3708 (1.5933)	3.2699 (-)	- (-)
	LAD	0.1043 (0.1177)	0.1046 (0.1065)	0.1410 (0.1513)	0.1945 (0.2207)
	EXP	0.0914 (0.0891)	0.1414 (0.1428)	0.1376 (0.1440)	0.2129 (0.2366)
	HB	0.0722 (0.0787)	0.1485 (0.1636)	0.1266 (0.1406)	0.2085 (0.2365)

Table 4. The mean (SD) of MSPE under different error distributions and sample sizes.

n	Methods	$N(0,1)$	MN	$t(2)$	*Cauchy* $(0,1)$
		Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
100	OLS	0.3588 (0.3542)	7.3662 (8.8633)	7.3858 (-)	- (-)
	LAD	0.4677 (0.4338)	1.4368 (1.7128)	0.9803 (0.9559)	1.9208 (2.7565)
	EXP	0.4426 (0.4447)	1.0267 (1.6935)	0.7628 (0.8031)	1.3085 (1.3842)
	HB	0.3792 (0.3800)	0.8862 (0.9175)	0.7665 (0.7987)	1.4059 (1.6041)
300	OLS	0.1266 (0.1173)	2.3559 (2.6711)	2.1143 (12.5808)	- (-)
	LAD	0.1767 (0.1725)	0.3310 (0.3619)	0.2613 (0.2739)	0.4033 (0.4421)
	EXP	0.1607 (0.1349)	0.2703 (0.2626)	0.2502 (0.2712)	0.4017 (0.4206)
	HB	0.1317 (0.1246)	0.2662 (0.2597)	0.2274 (0.2509)	0.3969 (0.4090)
500	OLS	0.0775 (0.0751)	1.4681 (1.5899)	3.6659 (-)	- (-)
	LAD	0.1165 (0.1175)	0.1172 (0.1067)	0.1563 (0.1508)	0.2207 (0.2209)
	EXP	0.1105 (0.0889)	0.1635 (0.1429)	0.1601 (0.1448)	0.2399 (0.2368)
	HB	0.0821 (0.0785)	0.1706 (0.1644)	0.1409 (0.1411)	0.2353 (0.2365)

4. Applications to Real Data

In this section, we illustrate the proposed robust estimation procedure for a semi-functional linear model with autoregressive errors by two real data analyses. The first example is the electricity consumed data, and the second one is the Tecator data.

4.1. Electricity Consumption Data

This subsection aims to compare the proposed robust estimation method with the OLS method, that is, whether the proposed robust estimation method is as good as the OLS estimation method in the absence of outliers in the data set. For this purpose, we consider electricity consumption data, which is available on the website http://www.eia.doe.gov/emeu/aer, (accessed on 31 December 2022). The data set includes the electricity consumption c_i (KWH) from January 1973 to January 2016 (517 months) (see Figure 2) and annual average retail price z_i (per KWH, including tax) (43 years).

As shown in Figure 2, the time series obviously shows some linear trends and some heterogeneity in the variance structure. Similar to Aneiros and Vieu [10] and Yu et al. [27], we eliminate the heteroscedasticity and the linear trend of the electricity retail sales data by differencing the log(electricity data) and obtain the time series (see Figure 3): $\mathcal{X}_i = \log(c_{i+1}) - \log(c_i)$, $i = 1, 2, \cdots, 516$. Let $x_i(s) = \left\{\mathcal{X}_{12(i-1)+s}, i = 1, 2, \cdots, 43, s = 1, 2, \cdots, 12\right\}$ be the monthly log (difference electricity data).

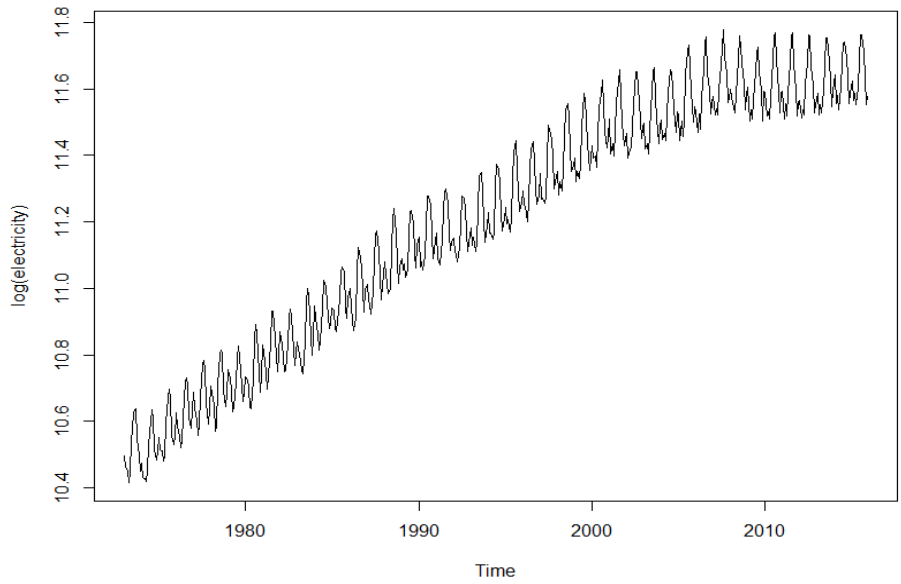


Figure 2. The ln (electricity data) from January 1973 to January 2016.

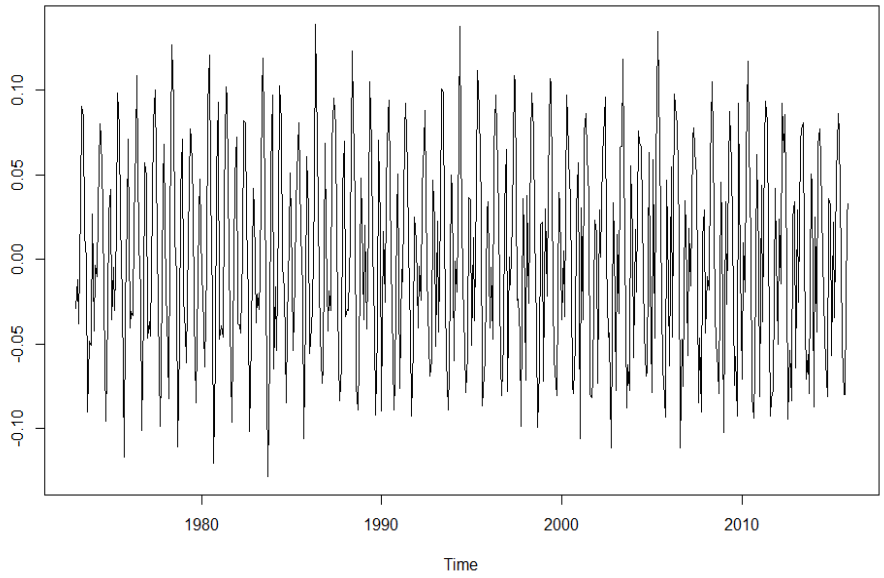


Figure 3. The differenced log (electricity data) from January 1973 to January 2016.

There is only one observation per month, so we use 10 cubic B-spline basis functions to convert the discrete monthly electricity data into functional continuous smooth annual curve predictors $\{x_i(t), t \in [1,12], i = 1,2,\cdots,43\}$ (see Figure 4). The response variable is $y_i(s) = \mathcal{X}_{12i+s}$, which represents the electricity consumption data in the s-month of the i-year, where $i = 1,2,\cdots,42, s = 1,2,\cdots,12$. The electricity data set is divided into the training sample $\{y_i(s), z_i, x_i(t)\}_{i=1}^{41}$ and the test sample $\{y_{42}(s), z_{42}, x_{42}(t)\}$.

To ascertain whether there are outliers in the data set, the boxplot of the electricity consumption data from January to December is displayed in Figure 5. As shown in Figure 5, there are some outliers in March and November. Hence, the robust method may perform better than OLS. Meanwhile, to determine whether the random errors are serially correlated, we first display the autocorrelation function plot (see Figure 6) of the residual sequences $\{\hat{\varepsilon}_i(s)\}_{i=1}^{41}, s = 1,\cdots,12$ of the the following semi-functional linear model (SFLM) based on the HB estimation procedure:

$$y_i(s) = \int_1^{12} \beta(t)x_i(t)\mathrm{d}t + g(z_i) + \varepsilon_i(s), \quad i = 1,2,\cdots,41;\ s = 1,\cdots,12. \tag{19}$$

From Figure 6, it can be seen that there exist serial correlations for random error sequences $\varepsilon_i(8)$ and $\varepsilon_i(9)$. We further use the Ljung–Box (LB) test statistics to test whether first-fourth order serial correlations exist. The corresponding $p-$values are 0.0049 and 0.0013, respectively. Under the given significance level of 0.05, we can also conclude that the random error sequences $\varepsilon_i(8)$ and $\varepsilon_i(9)$ are the fourth-order serial correlation, which is consistent with Figure 6. Hence, we consider using the robust method to estimate the following SFLAR model based on the training set:

$$y_i(s) = \int_1^{12} \beta(t)x_i(t)\mathrm{d}t + g(z_i) + \varepsilon_i, \quad \varepsilon_i = \sum_{l=1}^{q} a_l\varepsilon_{i-l} + e_i, \quad i = 1, 2, \cdots, 41, \tag{20}$$

and then the test set is used to predict the values of $y_{42}(s)$.

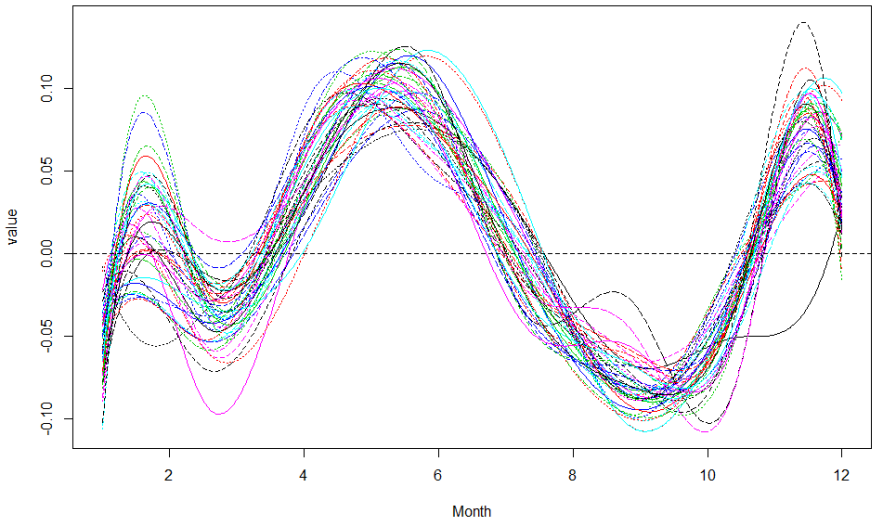


Figure 4. Annual curves of electricity data in the commercial sector from January 1973 to January 2016.

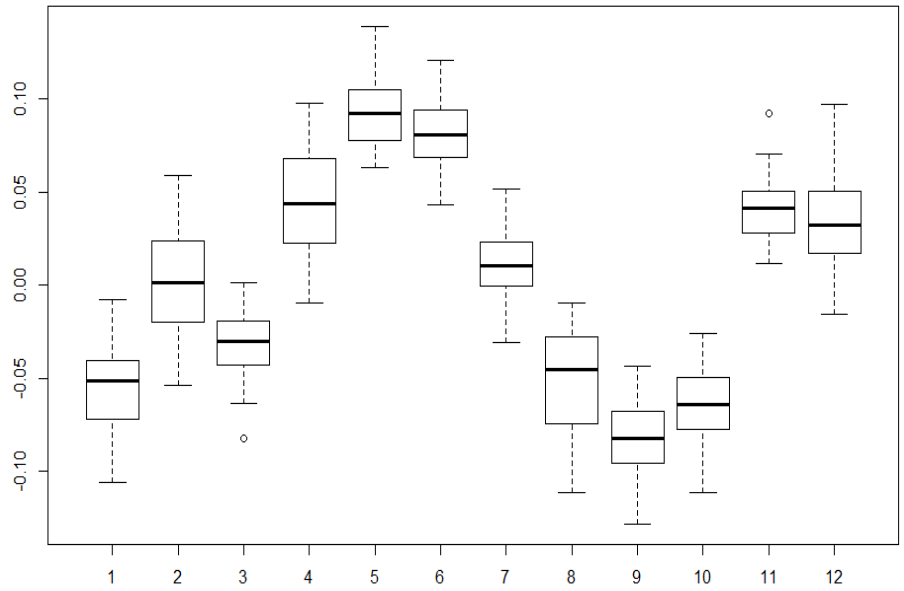


Figure 5. The boxplots from January to December.

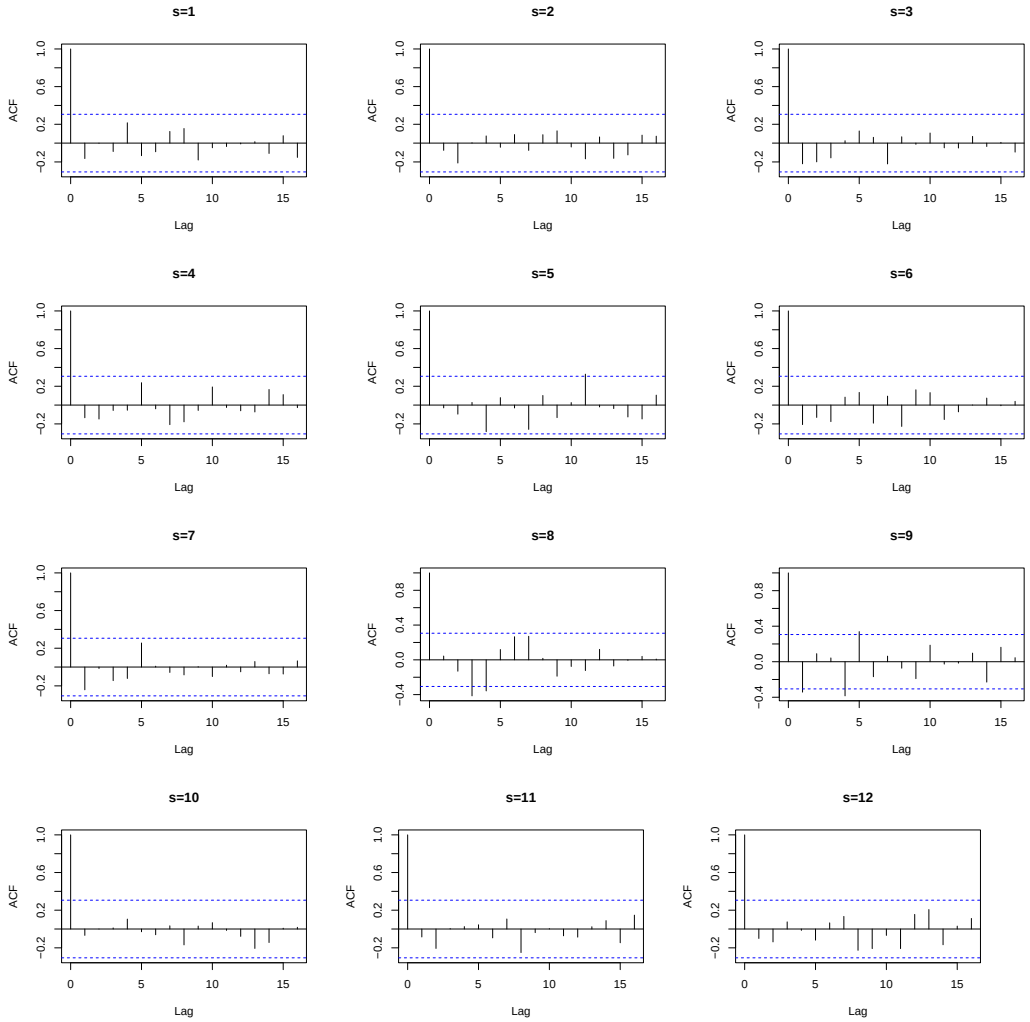


Figure 6. Autocorrelation graph of residual sequences from January ($s = 1$) to December ($s = 12$).

We also compare the robust estimation procedure (LAD, EXP and HB) with OLS to illustrate the proposed mehtod. Further, to implement our approach, the unknown function $g(\cdot)$ is approximated by a cubic B-spline function with equispaced knots. The truncation parameters J and K are selected by the BIC method and calculated according to Equation (13) in Section 2. Similar to Aneiros and Vieu [10] and Yu et al. [27], the following two error criteria, mean quadratic error (MQE):

$$\text{MQE} = \frac{1}{12}\sum_{s=1}^{12}(y_{42}(s) - \hat{y}_{42}(s))^2, \tag{21}$$

and mean relative quadratic error (MRQE):

$$\text{MRQE} = \frac{1}{12}\sum_{s=1}^{12}\frac{(y_{42}(s) - \hat{y}_{42}(s))^2}{\text{Var}(y(s))} \tag{22}$$

are used to evaluate the prediction ability of the model, where $\text{Var}(y(s))$ is the empirical variance of $\{y_i(s)\}_{i=1}^{41}$. Table 5 shows the MQE and MRQE of LAD, HB, EXP and OLS estimation. It can be seen from Table 5 that the robust estimation procedure performs better than OLS when there are some outliers.

Table 5. MQE and MRQE of three estimation methods under the SFLAR model.

Criterions	Methods			
	OLS	**LAD**	**HB**	**EXP**
MQE	0.00017	0.00016	0.00017	0.00017
MRQE	0.34483	0.27929	0.31347	0.34435

4.2. Tecator Data

In this subsection, we further apply the robust procedure to analyze the Tecator spectral data set, which originated from http://lib.stat.cmu.edu/datasets/tecator, (accessed on 31 December 2022). This data set consists of 215 meat samples. Each sample records a spectrometric curve corresponding to a spectrum of absorbance measured at 100 wavelengths between 850 and 1050 nm by the Near-Infrared Transmission (NIT) principle, and the contents of moisture, protein and fat measured in percentages. Chen et al. [28], Aneiros and Vieu [29], Lin et al. [30] and Sang et al. [31] have conducted extensive research on these data, aiming to predict the content of certain components (such as fat) by using spectral absorbance. Recently, Cai et al. [15] predicted fat content (Y) according to water content (Z_1), protein content (Z_2) and spectral curve $X(t)$.

This paper is interested in whether the accuracy of LAD, EXP and HB is better than OLS in predicting fat content (y) when the distribution of the response is not normal, and the random errors are correlated. For this purpose, we first use the Shapiro–Wilk statistic to test whether the fat content distribution is normal. The $p-$value of the test statistic is 0.000, which indicates that the distribution is not normal. Further, the skewness value (0.796) shows that the data have a right-skewed distribution. The least squares estimation may not be effective, so the robust method is adopted. Then, like in Section 4.1, we apply the LB test statistics to test whether there are first-order serial correlations based on the HB method. The corresponding $p-$value is 0.0222, which is less than the given significance level of 0.05. Therefore, we also apply the robust method to estimate the following SFLAR model:

$$y_i = \int_{850}^{1050} \beta(t)x_i(t)\mathrm{d}t + g(z_i) + \varepsilon_i, \quad \varepsilon_i = a_1\varepsilon_{i-1} + e_i, \quad i = 1, 2, \cdots, 215, \tag{23}$$

where the response variable y_i is the fat content of the i-th sample, and the explanatory variables z_i and $x_i(t)$ stand for the protein and the spectrometric curve of the i-th sample, respectively.

Similar to Cai et al. [15], the two criteria, mean absolute prediction error (MAPE):

$$\mathrm{MAPE} = \frac{1}{214}\sum_{i=2}^{215}|y_i - \hat{y}_i|, \tag{24}$$

and the median prediction error (MPE):

$$\mathrm{MPE} = \underset{i\in\{2,3,\cdots,215\}}{\mathrm{median}}|y_i - \hat{y}_i| \tag{25}$$

are used to evaluate the prediction effect of different methods. From the values of MAPE and MPE of LAD, HB, EXP and OLS displayed in Table 6, it can be seen that the robust methods have better performance than the OLS estimation when the distribution of the response is not normal.

Table 6. MAPE and MPE of three estimation methods under SFLAR.

Criterions	Methods			
	OLS	**LAD**	**HB**	**EXP**
MAPE	1.862	1.771	1.783	1.891
MPE	1.528	1.345	1.388	1.138

5. Conclusions

This paper considers the estimation problem of a semi-functional linear model with autoregressive errors when the random error follows a heavy-tailed distribution, or there are some outliers. The traditional estimation methods, such as the least square or likelihood method, are susceptible to outliers or heavily tailed errors, resulting in the corresponding estimation no longer being effective. Therefore, this paper introduces the robust estimation procedure. Simulation research and two real data analyses show that if the random error follows the normal distribution, the robust estimation has the same good statistical properties as the OLS. However, if the random error is heavily tailed, the robust estimation outperforms the OLS method.

Author Contributions: Conceptualization, M.C.; methodology, J.Z. and B.Y.; software, B.Y. and T.S.; formal analysis, J.Z. and B.Y.; resources, J.Z. and B.Y.; data curation, J.Z. and B.Y.; writing—original draft preparation, B.Y.; writing—review and editing, J.Z.; visualization, B.Y.; supervision, M.C. and J.Z.; project administration, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under 210 Grant Nos. 11861074, 11731011, 12261051 and 11731015, and Applied Basic Research Project of Yunnan 211 Province under Grant No. 2019FB138.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data sets generated and analysed are available from the corresponding author on reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike information criterion
BIC	Bayesian information criterion
EXP	exponential square
HB	Huber loss
IMSE	integral of mean square errors
KWH	kilowatt hour
LAD	Least absolute deviations
MN	Mixed normal
MQE	mean quadratic error
MRQE	mean relative quadratic error
MSE	mean square error
MSPE	mean square prediction error
OLS	Ordinary least squares
SD	standard deviation
SFLAR	semi-functional linear model with autoregressive error
TSIRLS	two-step iteratively reweighted least squares

References

1. Cardot, H.; Ferraty, F.; Sarda, P. Functional linear model. *Stat. Probab. Lett.* **1999**, *45*, 11–22. [CrossRef]
2. Cardot, H.; Ferraty, F.; Sarda, P. Spline estimators for the functional linear model. *Stat. Sin.* **2003**, *13*, 571–591. [CrossRef]
3. Yao, F.; Müller, H.G.; Wang, J.L. Functional linear regression analysis for longitudinal data. *Ann. Stat.* **2005**, *33*, 2873–2903. [CrossRef]
4. Hall, P.; Horowitz, J.L. Methodology and convergence rates for functional linear regression. *Ann. Stat.* **2007**, **35**, 70–91. [CrossRef]
5. Li, T.; Song, X.Y.; Zhang, Y.Y.; Zhu, H.T.; Zhu, Z.Y. Clusterwise functional linear regression models. *Comput. Stat. Data Anal.* **2021**, *158*, 0167–9473. [CrossRef]
6. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer: New York, NY, USA, 2006.
7. Baíllo, A.; Grané, A. Local linear regression for functional predictor and scalar response. *J. Multivar. Anal.* **2009**, *100*, 102–111. [CrossRef]
8. Burba, F.; Ferraty, F.; Vieu, P. k-Nearest Neighbour method in functional nonparametric regression. *J. Nonparametric Stat.* **2009**, *21*, 453–469. [CrossRef]
9. Aneiros-Pérez, G.; Vieu, P. Semi-functional partial linear regression. *Stat. Probab. Lett.* **2006**, *76*, 1102–1110. [CrossRef]
10. Aneiros-Pérez, G.; Vieu, P. Nonparametric time series prediction: A semi-functional partial linear modeling. *J. Multivar. Anal.* **2008**, *99*, 834–857. [CrossRef]
11. Shin, H. Partial functional linear regression. *J. Stat. Plan. Inference* **2009**, **139**, 3405–3418. [CrossRef]
12. Zhou J.J.; Chen M. Spline estimators for semi-functional linear model. *Stat. Probab. Lett.* **2012**, *82*, 505–513. [CrossRef]
13. Yu, P.; Zhu, Z.Y.; Zhang, Z.Z. Robust exponential squared loss-based estimation in semi-functional linear regression models. *Comput. Stat.* **2019**, *34*, 503-525. [CrossRef]
14. Yu, P.; Zhu, Z.Y.; Shi, J.H.; Ai, X.K. Robust estimation for partial functional linear regression model based on modal regression. *J. Syst. Sci. Complex.* **2020**, **33**, 527-544. [CrossRef]
15. Cai, X.; Xue, L.G.; Lu, F. Robust estimation with a modified Huber's loss for partial functional linear models based on splines. *J. Korean Stat. Soc.* **2020**, *49*, 1214–1237. [CrossRef]
16. Cao, P.; Sun, J. Robust estimation for partial functional linear regression models based on FPCA and weighted composite quantile regression. *Open Math.* **2021**, *19*, 1493–1509. [CrossRef]
17. Tang Q.G. Estimation for semi-functional linear regression. *Statistics* **2015**, *49*, 1262–1278. [CrossRef]
18. Boente, G.; Salibian-Barrera, M.; Vena, P. Robust estimation for semi-functional linear regression models. *Comput. Stat. Data Anal.* **2020**, *152*, 107041. [CrossRef]
19. Boente, G.; Daniela, P. Robust estimation for functional quadratic regression models. *arXiv* **2022**, arXiv:2209.02742. https://doi.org/10.48550/arXiv.2209.02742.
20. Pannu, J.; Billor, N. Robust sparse functional regression model. *Commun. Stat. - Simul. Comput.* **2022**, *51*, 4883-4903. [CrossRef]
21. Sinha, S.K.; Field, C.A.; Smith, B. Robust estimation of nonlinear regression with autoregressive errors. *Stat. Probab. Lett.* **2003**, *63*, 49–59. [CrossRef]
22. Riazoshams, H.; Midi, H.B; Sharipov, O.S. The performance of robust two-stage estimator in nonlinear regression with autocorrelated error. *Commun. Stat. - Simul. Comput.* **2010**, *39*, 1251–1268. [CrossRef]
23. Serenay, K.; Baris, A. A novel hybrid robust tapering approach for nonlinear regression in the presence of autocorrelation and outliers. *Commun. Stat. - Simul. Comput.* **2021**, 1–17. [CrossRef]
24. de Boor, C. *A Practical Guide to Spline*; Springer: New York, NY. USA, 1978. [CrossRef]
25. Maronna, R.A.; Martin, R.D.; Yohai, V.J.; Salibián-Barrera, M. *Robust Statistics: Theory and Methods (with R)*; Wiley: New York, NY, USA, 2018.
26. Jiang, Y.L. An exponential-squared estimator in the autoregressive model with heavy-tailed errors. *Stat. Its Interface* **2016**, *9*, 233–238. [CrossRef]
27. Yu, P.; Li, T.; Zhu, Z.Y.; Zhang, Z.Z. Composite quantile estimation in partial functional linear regression model with dependent errors. *Metrika* **2019**, *82*, 633–656. [CrossRef]
28. Chen, D.; Hall, P.; Müller, H.G. Single and multiple index functional regression models with nonparametric link. *Ann. Stat.* **2011**, *39*, 1720–1747. [CrossRef]
29. Aneiros, P.; Vieu, P. Partial linear modelling with multi-functional covariates. *Comput. Stat.* **2015**, *30*, 647–671. [CrossRef]
30. Lin, Z.H.; Cao, J.G.; Wang, L.L.; Wang, H.N. Locally sparse estimator for functional linear regression models. *J. Comput. Graph. Stat.* **2017**, *26*, 306–318. [CrossRef]
31. Sang, P.J.; Richard, A.L.; Cao, J.G. Sparse estimation for functional semiparametric additive models. *J. Multivar. Anal.* **2018**, *168*, 105–118. [CrossRef]