Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(1): 1602–1627.
DOI: 10.3934/math.2025074
Received: 03 October 2024
Revised: 04 January 2025
Accepted: 14 January 2025
Published: 23 January 2025

Research article

Forecasting the monthly retail sales of electricity based on the semi-functional linear model with autoregressive errors

Bin Yang[1,2], **Min Chen**[3,4] **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 650051, 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:** Email: jjzhou@ynu.edu.cn.

Abstract: In many statistical applications, data are collected sequentially over time and exhibit autocorrelation characteristics. Ignoring this autocorrelation may lead to a decrease in the model's prediction accuracy. To this end, assuming that the error process is an autoregressive process, this paper introduced a semi-functional linear model with autoregressive errors. Based on the functional principal component analysis and the spline method, we obtained the estimators of the slope function, nonparametric function, and autoregressive coefficients. Under some regular conditions, we found the convergence rate of the proposed estimators. A simulation study was conducted to investigate the finite sample performance of the proposed estimators. Finally, we applied our model to forecast the monthly retail sales of electricity, which illustrates the validity of our model from a predictive perspective.

Keywords: semi-functional linear model; autoregressive errors; convergence rate; functional principal component analysis; spline estimation
Mathematics Subject Classification: 62J05, 62P12

1. Introduction

Functional regression modeling has been extensively studied as an important research field in functional data analysis (FDA). For example, to investigate the linear relationship between a scalar response and a functional explanatory variable, comprehensive studies focused on functional linear regression models have been conducted by Manteiga and Calvo [1] and Bande et al. [2]. However, the linear assumption is sometimes too restrictive to describe the data structure comprehensively. To this end, the functional nonparametric regression model, as an extension of the nonparametric model in the functional data setting, has also been proposed (see Ferraty and Vieu [3] and Belarbi et al. [4]). Furthermore, to improve the power of prediction and interpretation of functional regression models, various functional semiparametric models have been proposed by introducing some additional real-valued predictors. For example, Pérez and Vieu [5] proposed a semi-functional partial linear regression model by combining a traditional linear model with a nonparametric treatment of functional data. Shin [6] and Shin and Lee [7] studied the problem of estimation and prediction of the partial functional linear regression model, respectively. Yao et al. [8] extended the partial functional linear regression to quantile regression settings and established a more flexible and robust method. Kong et al. [9] investigated a class of partial function linear regression models for processing multiple functional and non-functional explanatory variables and automatically identifying essential risk factors through appropriate regularization. In addition, Zhou and Chen [10] first introduced the semi-functional linear model (SFL) by combining the features of a functional linear regression model and a traditional nonparametric model and studied its spline estimation. Peng et al. [11] proposed a varying coefficient partially functional linear regression model, which combines a varying coefficient regression model and a functional linear regression model, and investigated the estimation problem based on the polynomial spline method. Novo et al. [12] developed an automatic and location-adaptive estimation procedure for the functional single-index model. For a recent survey on this field, we refer to some literature like Cuevas [13], Goia and Vieu [14], Ling and Vieu [15, 16], and Aneiros et al. [17].

Meanwhile, given that electricity cannot be stored, modeling and forecasting electricity demand and price are of primary interest to electric market agents. Traditional prediction models, like simple time series models, often fail to address the complexities of electricity demand, so functional regression models have been applied to electricity demand forecasting. For example, Vilar et al. [18] utilized functional nonparametric and semi-functional partial linear models to forecast next-day electricity demand and price. Based on the robust functional principal component analysis and nonparametric models with both functional response and covariate, Aneiros et al. [19] proposed two functional regression models to forecast daily curves of electricity demand and price. In addition, using residual-based bootstrap algorithms, Vilar et al. [20] provided two procedures to obtain prediction intervals for electricity demand and price based on the functional nonparametric autoregressive model and the partial linear semiparametric model. Peláez et al. [21] also proposed three model-based procedures to construct bootstrap prediction regions for daily curves of electricity demand and price.

All the models mentioned above are based on a critical assumption that the random errors are independently and identically distributed. However, the independence assumption may be inappropriate in many applications. For example, the electricity consumption dataset that has been extensively studied by Pérez and Vieu [22], Yu et al. [23], and Yang et al. [24] exhibits a specific autocorrelation structure. If this autocorrelation structure is ignored, it may lead to inaccurate prediction. To this end, many works have been developed to deal with the autocorrelation structure in functional regression models. For example, Niang and Guillas [25] studied parametric estimation of a functional semiparametric partially linear model with autoregressive errors. Zhang et al. [26] proposed a functional polynomial regression with autoregressive errors. Yu et al. [23] considered composite quantile estimation for the partial functional linear regression model with random errors from a short-range dependent and strictly stationary linear process. Xiao and Wang [27] also studied the estimation problem of the partial functional linear model with autoregressive errors. Wang et al. [28] studied the multiple functional linear model with autoregressive errors. Although these functional regression models with autoregressive errors have improved prediction accuracy, the linear assumption on the covariates is sometimes inappropriate. Therefore, it is necessary to develop more general models to fit the more complex data structures. To achieve this goal, by relaxing the linear assumption on the partial functional linear model, a semi-functional linear model with autoregressive errors is developed in this paper. Based on the functional principal component analysis and the spline method, we obtain the estimators of the slope function, nonparametric function, and autoregressive coefficients, respectively.

The rest of this paper is organized as follows. Section 2 describes the semi-functional linear model with autoregressive errors and presents the estimation method and algorithm. In Section 3, we study the asymptotic properties of the proposed estimators under some regular conditions. A Monte Carlo simulation is presented in Section 4 to illustrate the finite sample performance. The proposed method is illustrated by electricity consumption data in Section 5. Sections 6 and 7 present future prospects and summarize the paper, respectively. All technical details and proofs are provided in the appendix.

2. Model and estimation

2.1. Model and estimation

Suppose that $(X_t, Y_t, Z_t), 1 \leq t \leq n$, is a random sample from the semi-functional linear model with autoregressive errors (SFLAR)

$$\begin{aligned} Y_t &= \int_{\mathcal{T}} \beta(s) X_t(s) ds + g(Z_t) + \varepsilon_t, \\ \varepsilon_t &= \sum_{l=1}^{q} a_l \varepsilon_{t-l} + e_t, \end{aligned} \tag{2.1}$$

where Y_t is a real-valued random response variable and Z_t is a real-valued covariate defined on a compact interval $Z = [\underline{z}, \overline{z}]$. The functional covariate $X_t(s)$ defined on an interval $\mathcal{T} \subset R$ is a zero mean, second-order (i.e., $\mathbb{E}|X(s)|^2 < \infty$ for all $s \in \mathcal{T}$) 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 $\langle u, v\rangle = \int_{\mathcal{T}} u(s)v(s)ds$, for any $u, v \in L^2(\mathcal{T})$ and norm $\|u\| = \langle u, u\rangle^{1/2}$. The unknown slope function $\beta(s)$ belongs to $L^2(\mathcal{T})$. For simplicity, throughout this paper, we assume $\mathcal{T} = [0, 1]$. The nonparametric function $g(\cdot)$ is an unknown smooth function. ε_t is a random error with zero mean and finite variance independent of (X_t, Z_t). e_t's are independent and identically distributed with mean 0 and variance σ^2. Note that model (2.1) includes the partial functional linear model with autoregressive errors studied by Xiao and Wang [27].

Since the unknown functions $\beta(\cdot)$ and $g(\cdot)$ are infinite-dimensional parameters, it is impossible to obtain their estimators based on finite observation data. Therefore, we need to apply dimension reduction methods. Specifically, if we denote the covariance function of the functional variable X by $C_X(s, u) = Cov[X(s), X(u)]$, and assume that $C_X(s, u)$ is continuous on the interval $[0, 1]^2$, according to the Mercer theorem, we have

$$C_X(s, u) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s) \phi_i(u),$$

where $\lambda_1 > \lambda_2 > \cdots > 0$ is the ordered eigenvalue sequence and $\phi_1, \phi_2, \cdots$ is the corresponding orthonormal eigenfunction sequence of the covariance operator with kernel $C_X(s,u)$. By the Karhunen-Loève representation, functional variables $X_t(s)$ and the slope function $\beta(s)$ can be respectively expanded as:

$$X_t(s) = \sum_{i=1}^{\infty} \xi_{ti}\phi_i(s), \quad \beta(s) = \sum_{i=1}^{\infty} b_i\phi_i(s), \tag{2.2}$$

where $\xi_{ti} = \int_0^1 X_t(s)\phi_i(s)ds$ are called the principle component scores satisfying $\mathbb{E}\xi_{ti} = 0$ and $\mathbb{E}\xi_{ti}\xi_{tk} = \lambda_i I(i = k)$, and coefficients $b_i = \int_0^1 \beta(s)\phi_i(s)ds$. $\{\phi_i(s)\}_{i=1}^{\infty}$ forms a standard orthogonal basis of $L^2(\mathcal{T})$, commonly referred to as the functional principal component (FPC) basis. Truncate Eqs (2.2) at M, and the model (2.1) can be approximated as

$$Y_t \approx \sum_{i=1}^{M} b_i\xi_{ti} + g(Z_t) + \varepsilon_t. \tag{2.3}$$

Furthermore, we can approximate the nonparametric function $g(\cdot)$ using polynomial splines. Specifically, let G_{p,J_n} be the space of polynomial splines defined on the interval $[\underline{z}, \overline{z}]$ with degree $p-1$ and knot sequence $\underline{z} < z_1 < \cdots < z_{J_n} < \overline{z}$. The space G_{p,J_n} is a J-dimensional linear space, where $J = J(n) = J_n + p$. Following the arguments of de Boor [29], we can conclude that, if it is sufficiently smooth, the nonparametric function $g(Z_t)$ can be approximately expressed as

$$g(Z_t) \approx \sum_{j=1}^{J} c_j B_j(Z_t),$$

where $B_j, j = 1, 2, \cdots, J$, are the B-spline basis functions in G_{p,J_n}. Substituting the approximation of $g(Z_t)$ into the model (2.3) yields

$$Y_t \approx \sum_{i=1}^{M} b_i\xi_{ti} + \sum_{j=1}^{J} c_j B_j(Z_t) + \varepsilon_t. \tag{2.4}$$

However, since the covariance function $C_X(s,u)$ is unknown, the eigenfunctions ϕ_i and variables ξ_{ti} are also unknown and unobservable. To address this issue, we define the empirical version of the covariance function $C_X(s,u)$ by

$$\widehat{C}_X(s,u) = \frac{1}{n}\sum_{t=1}^{n} X_t(s)X_t(u).$$

By Mercer's theorem, we also have

$$\widehat{C}_X(s,u) = \sum_{i=1}^{\infty} \hat{\lambda}_i\hat{\phi}_i(s)\hat{\phi}_i(u),$$

where $(\hat{\lambda}_i, \hat{\phi}_i)$ are the eigenvalue-eigenfunction pairs for the linear operator with kernel $\widehat{C}_X(s,u)$, ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq 0$. We consider $(\hat{\lambda}_i, \hat{\phi}_i)$ and $\widehat{\xi}_{ti} = \int_0^1 X_t(s)\widehat{\phi}_i(s)ds$ as estimators of (λ_i, ϕ_i) and

ξ_{ti}, respectively. Then, the estimators $\hat{\beta}(s) = \sum_{i=1}^{M} \hat{b}_i \hat{\phi}_i(s)$ and $\hat{g}(Z_t) = \sum_{j=1}^{J} \hat{c}_j B_j(Z_t)$ can be obtained by minimizing the following loss function:

$$L_n(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \frac{1}{n-q} \sum_{t=q+1}^{n} [(Y_t - \widehat{\boldsymbol{\xi}}_t^\top \boldsymbol{b} - \boldsymbol{B}_t^\top \boldsymbol{c}) - \sum_{l=1}^{q} a_l (Y_{t-l} - \widehat{\boldsymbol{\xi}}_{t-l}^\top \boldsymbol{b} - \boldsymbol{B}_{t-l}^\top \boldsymbol{c})]^2, \tag{2.5}$$

where $\widehat{\boldsymbol{\xi}_t} = (\widehat{\xi}_{t1}, \widehat{\xi}_{t2}, \cdots, \widehat{\xi}_{tM})^\top$, $\boldsymbol{B_t} = (B_1(Z_t), B_2(Z_t), \cdots, B_J(Z_t))^\top$, $\boldsymbol{a} = (a_1, a_2, \cdots, a_q)^\top$, $\boldsymbol{b} = (b_1, b_2, \cdots, b_M)^\top$, $\boldsymbol{c} = (c_1, c_2, \cdots, c_J)^\top$, and $t = q+1, q+2, \cdots, n$.

To implement the proposed estimation method, we must choose the degrees of the spline functions, the positions, and the number of knots and eigenfunctions. This paper uses a B-spline basis with equally spaced knots and the fixed degree 2. Then, we only need to choose the numbers of B-spline functions (J) and eigenfunctions (M). Various methods, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), can be used to select the truncation parameters J and M. This paper uses the BIC criterion to choose the number of J and M. The selection of J and M is determined by minimizing the following BIC criteria:

$$\mathrm{BIC}(M, J) = \log\left(\frac{\sum_{t=q+1}^{n} [\hat{\varepsilon}_t - \sum_{l=1}^{q} \hat{a}_l \hat{\varepsilon}_{t-l}]^2}{n-q}\right) + \frac{\log(n-q)}{n-q}(M + J + q), \tag{2.6}$$

where $\hat{\varepsilon}_t = Y_t - \sum_{i=1}^{M} \hat{b}_i \hat{\xi}_{ti} - \sum_{j=1}^{J} \hat{c}_j B_j(Z_t)$.

2.2. Algorithm

This subsection mainly introduces the algorithm to estimate the unknown parameters. For ease of description, we rewrite the loss function (2.5) as

$$\begin{aligned} L_n(\boldsymbol{a}, \boldsymbol{\theta}) &= \frac{1}{n-q} \sum_{t=q+1}^{n} [(Y_t - \widehat{\boldsymbol{U}}_{\boldsymbol{t}}^\top \boldsymbol{\theta}) - \sum_{l=1}^{q} a_l (Y_{t-l} - \widehat{\boldsymbol{U}}_{\boldsymbol{t-l}}^\top \boldsymbol{\theta})]^2 \\ &= \frac{1}{n-q} \sum_{t=q+1}^{n} [(Y_t - \sum_{l=1}^{q} a_l Y_{t-l}) - (\widehat{\boldsymbol{U}}_{\boldsymbol{t}} - \sum_{l=1}^{q} a_l \widehat{\boldsymbol{U}}_{\boldsymbol{t-l}})^\top \boldsymbol{\theta}]^2, \end{aligned} \tag{2.7}$$

where $\widehat{\boldsymbol{U}}_{\boldsymbol{t}} = (\widehat{\boldsymbol{\xi}}_{\boldsymbol{t}}^\top, \boldsymbol{B}_{\boldsymbol{t}}^\top)^\top$, $\widehat{\boldsymbol{\xi}}_{\boldsymbol{t}} = (\widehat{\xi}_{t1}, \widehat{\xi}_{t2}, \cdots, \widehat{\xi}_{tM})^\top$. Since there is no closed-form expression of the minimizer of the loss function (2.7), we adopt the two-step iteratively least-squares (TSILS) algorithm to estimate the unknown parameters. Specifically, we first set $\boldsymbol{a}^{(0)} = (0, 0, \cdots, 0)^\top$ and compute $\boldsymbol{\theta}^{(0)}$ by minimizing $L_n(\boldsymbol{a}^{(0)}, \boldsymbol{\theta})$. Afterward, we substitute $\boldsymbol{\theta}^{(0)}$ into the loss function (2.7) to get $\boldsymbol{a}^{(1)}$ by $L_n(\boldsymbol{a}, \boldsymbol{\theta}^{(0)})$, and substitute $\boldsymbol{a}^{(1)}$ into (2.7) to get $\boldsymbol{\theta}^{(1)} = \arg\min L_n(\boldsymbol{a}^{(1)}, \boldsymbol{\theta})$. Repeat this process until convergence. The TSILS algorithm is summarized as follows:

Algorithm 1. The TSILS algorithm initialization: Given the initial value: $\boldsymbol{a}^{(0)} = (0, 0, \cdots, 0)^\top$, compute $\boldsymbol{\theta}^{(0)} = (\widehat{\boldsymbol{U}}^\top \widehat{\boldsymbol{U}})^{-1} \widehat{\boldsymbol{U}}^\top Y$ based on the least squares method, where $Y = (Y_1, Y_2, \cdots, Y_n)^\top$ and $\widehat{\boldsymbol{U}}^\top = (\widehat{\boldsymbol{U}}_{\boldsymbol{1}}, \widehat{\boldsymbol{U}}_{\boldsymbol{2}}, \cdots, \widehat{\boldsymbol{U}}_{\boldsymbol{n}})$.

Step 1. Update $\tilde{\boldsymbol{V}}^{(k)}$, $\tilde{\boldsymbol{H}}^{(k)}$:

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

where $\hat{\varepsilon}_t^{(k)} = Y_t - \widehat{\boldsymbol{U}}_t^\top \hat{\boldsymbol{\theta}}^{(\boldsymbol{k-1})}$, $t = 1, 2, \cdots, n$.
Step 2. Compute $\boldsymbol{a}^{(k)}$:

$$\boldsymbol{a}^{(k)} = (\tilde{\boldsymbol{H}}^{(k)\top} \tilde{\boldsymbol{H}}^{(k)})^{-1} \tilde{\boldsymbol{H}}^{(k)\top} \tilde{\boldsymbol{V}}^{(k)}.$$

Step 3. Update $V_t^{(k)}$, $\boldsymbol{H}_t^{(k)}$:

$$V_t^{(k)} = Y_t - \sum_{l=1}^{q} a_l^{(k)} Y_{t-l}, \quad \boldsymbol{H}_t^{(k)} = \widehat{\boldsymbol{U}}_t - \sum_{l=1}^{q} a_l^{(k)} \widehat{\boldsymbol{U}}_{t-l}.$$

Step 4. Compute $\boldsymbol{\theta}^{(k)}$:

$$\boldsymbol{\theta}^{(k)} = (\boldsymbol{H}^{(k)\top} \boldsymbol{H}^{(\boldsymbol{k})})^{-1} \boldsymbol{H}^{(k)\top} \boldsymbol{V}^{(\boldsymbol{k})},$$

where

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

Step 5. Set $k = k + 1$, and repeat the above steps 1–4 until convergence. Then, the estimation of the parameter vectors $\hat{\boldsymbol{a}} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_q)^\top$, $\hat{\boldsymbol{b}} = (\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_M)^\top$, and $\hat{\boldsymbol{c}} = (\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_J)^\top$ can be obtained.

3. Theoretical properties

In this section, we investigate the theoretical properties of the proposed estimators. First, we assume that the true slope function, nonparametric function, and autoregressive coefficient of the model are $\beta^0(t), g^0(z)$, and $a^0 = \left(a_1^0, a_2^0, \cdots, a_q^0\right)^\top$, respectively. For convenience, the following notation is needed. For two sequences of positive numbers a_n and b_n, $a_n \lesssim b_n$ signifies that a_n/b_n is uniformly bounded and $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. Throughout this paper, the constant C_i may change from line to line. In order to establish the theoretical properties of the estimators, the following assumptions need to be introduced:
Assumption 1. *All roots of the polynomial $A(z) = 1 - \sum_{l=1}^{q} a_l z^l$ lie outside the unit circle in the complex plane.*
Assumption 2. *$\{X_t, Z_t\}_{t=1}^n$ and $\{e_t\}_{t=1}^n$ are independent random sequences and $\{X_t, e_t\}_{t=1}^n$ has the finite fourth-order moment, respectively, i.e. $\mathbb{E}\left(\int X_t^4(s)ds\right) < \infty$, $\mathbb{E}\left(e_t^4\right) < \infty$, for any $1 \leq t \leq n$, and random variable ξ_{ti} satisfy $\mathbb{E}\left(\xi_{ti}^4\right) \leq C_1 \lambda_i^2, i \geq 1$. The density function $f_Z(z)$ of the random variable Z is a continuous bounded function defined on the interval $[\underline{z}, \overline{z}]$. Meanwhile, the random variable e_t is independent of (X_t, Z_t) and $\mathbb{E}(X|Z) = 0$.*
Assumption 3. *The eigenvalues $\lambda_i (i = 1, 2, \cdots)$ of the covariance function C_X and the score coefficients b_i satisfy the following conditions, respectively.*
(1) *There exist some strictly positive constants C_2 and $\alpha > 1$ such that $C_2^{-1} i^{-\alpha} \leq \lambda_i \leq C_2 i^{-\alpha}$ and*

$\lambda_i - \lambda_{i+1} \geq C_2 i^{-\alpha-1}, i \geq 1$.
(**2**) *There exist some strictly positive constants* C_3 *and* $\gamma > \alpha/2 + 3$ *such that* $|b_i| \leq C_3 i^{-\gamma}, i \geq 1$.
Assumption 4. *Let* $0 < \nu \leq 1$. *The nonparametric function* $g(z)$ *is a continuous differentiable function of order* r, *and for* $z_1, z_2 \in [\underline{z}, \overline{z}]$, *there are* $\left|g^{(r)}(z_1) - g^{(r)}(z_2)\right| \leq C_4 |z_1 - z_2|^\nu$. *Define* $p = r + \nu$ *as a smoothness measure of the function* $g(z)$, *which satisfies* $p \geq \gamma + (\alpha - 1)/2$.
Assumption 5. *Define the matrix*

$$\Lambda = \mathbb{E}\left\{(U_t - \sum_{l=1}^{q} a_l^0 U_{t-l})(U_t - \sum_{l=1}^{q} a_l^0 U_{t-l})^\top\right\},$$

$$V = \mathbb{E}(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q})^\top (\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}),$$

where $U_t = (\xi_t^\top, B_t^\top)^\top$, $\xi_t = (\xi_{t1}, \xi_{t2}, \cdots, \xi_{tM})^\top$, $B_t = (B_1(Z_t), B_2(Z_t), \cdots, B_J(Z_t))^\top$, *and the matrix* V *is a positive definite matrix. Denote the minimum eigenvalue of* Λ *by* $\rho_{\min}(\Lambda)$, *which satisfies* $\rho_{\min}(\Lambda) = O_p\left(M^{-\alpha}\right)$.
Assumption 6. *The tuning parameter* M, J *satisfies* $M \asymp J \asymp n^{1/(\alpha+2\gamma)}$.
Remark 1. *Assumption 1 is an invertibility condition on the autoregression process and ensures* ε_t *is a stationary series. Assumption 2 is standard in the literature about the least squares method, which ensures the consistency of* $C_X(t, s)$. *Assumption 3 is established for the consistency of the estimated principal component scores. Assumption 3(1) prevents the spacings among eigenvalues from being too small for identification of the slope function* b_i. *Assumption 3(2) makes the slope function sufficiently smooth relative to the covariance function* $C_X(t, s)$. *Assumption 4 ensures that the nonparametric function* $g(z)$ *is sufficiently smooth so that the spline function can approach it. Assumption 5 is a necessary condition for the model consistency to be established. Assumption 6 gives the order of the truncation parameters* M *and* J *to obtain the convergence speed of* $\hat{\beta}(s)$ *and* $\hat{g}(z)$, *and Assumption 6 requires that the tuning parameters* M *and* J *do not grow too fast.*
Theorem 1. Under the Assumptions 1–6, we have:

$$\|\hat{a} - a^0\|_2^2 = O_p\left(n^{-(\alpha+2\gamma-1)/(\alpha+2\gamma)}\right),$$

$$\|\hat{\beta}(s) - \beta^0(s)\|^2 = O_p\left(n^{-(2\gamma-1)/(\alpha+2\gamma)}\right),$$

$$\|\hat{g}(z) - g^0(z)\|^2 = O_p\left(n^{-(\alpha+2\gamma-1)/(\alpha+2\gamma)}\right).$$

Remark 2. *The theorem shows that the convergence rate of the autoregressive coefficients cannot reach the optimal convergence rate* $O_p\left(n^{-1/2}\right)$ *(such as Wang et al. [30]), which may be due to the existence of functional infinite dimensional explanatory variables. The convergence rates of* $\hat{\beta}(s)$ *and* $\hat{g}(z)$ *are the same as that of the case where the random error is independent and identically distributed (such as Tang [31]), indicating that the autocorrelation of the random error does not affect the convergence rate of the slope function and the nonparametric function.*

4. Simulation study

In this section, we conduct a simulation under different settings to investigate the finite sample performance of our proposed method. The simulation data $\{(X_t, Z_t, Y_t), 1 \leq t \leq n\}$ are generated from the following SFLAR model:

$$Y_t = \int_0^1 \beta(u)X_t(u)du + g(Z_t) + \varepsilon_t, \ \ \varepsilon_t = \sum_{l=1}^{q} a_l\varepsilon_{t-l} + e_t,$$

where Z_t is distributed uniformly on [0,1], and the functional predictor $X_t = \sum_{j=1}^{50} \xi_{tj}\phi_j$, where ξ_{tj} are distributed as independent normal with a mean 0 and variance $\lambda_j = ((j-0.5)\pi)^{-2}$ and $\phi_j = \sqrt{2}sin((j-0.5)\pi t)$. For the unknown functions $\beta(\cdot)$ and $g(\cdot)$, and autoregressive coefficients $(a_1, \cdots, a_q)$, two scenarios have been considered as follows:

Scenario 1. $\beta(u) = \sqrt{2}sin(0.5\pi u) + 3\sqrt{2}sin(1.5\pi u)$, $g(z) = sin(2\pi z)$, $\varepsilon_t = a\varepsilon_{t-1} + e_t$, and the autoregressive coefficient a takes each of the values in the set (–0.5, 0.1, 0.8).

Scenario 2. $\beta(u) = \sqrt{2}sin(0.5\pi u) + 2\sqrt{2}sin(1.5\pi u)$, $g(z) = cos(2\pi z)$, $\varepsilon_t = a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + e_t$, and the autoregressive coefficients (a_1, a_2) take each of the values in the set $\{(0.1, 0.8), (0.5, -0.8)\}$.

The error variables e_t are $N(0, \sigma^2)$. In both scenarios, σ takes 0.5 and 1, corresponding to different signal-to-noise ratios. To implement the proposed method, we apply the BIC criterion of Eq (2.6) to choose the number J of B-spline functions and the number M of eigenfunctions. Specifically, J and M are obtained by selecting the minimum BIC value within the preset range. In this simulation, J is chosen between 3 and 8, and M is selected between 1 and 6. In addition, the errors of the estimators of function $\beta(u)$, $g(z)$ and the model prediction error were evaluated by the integral of square errors $\text{ISE}_1 = \int_0^1 [\hat{\beta}(u) - \beta(u)]^2 du$, $\text{ISE}_2 = \int_0^1 [\hat{g}(z) - g(z)]^2 dz$ and the mean square prediction error (MSPE): $\text{MSPE} = \frac{1}{n}\sum_{t=q+1}^{n}\left(\widehat{Y_t} - g(Z_t) - \int_0^1 \beta(u)X_t(u)\mathrm{d}u\right)^2$, respectively.

In our simulation, we considered sample sizes of $n = 400$, 800, and 1200, with each setting repeated 500 times. Tables 1 and 2 summarize the performance of various estimators regarding bias (Bias), root mean square error (RMSE) for the estimated a and σ, and integrated square errors (ISE) for the estimated $\beta(\cdot)$ and $g(\cdot)$ under normal error conditions. From these tables, we can draw the following observations:

(1) Given the sample size, as the value of σ increases, the bias (Bias) and root mean square error (RMSE) of $\hat{a}$ and $\hat{\sigma}$ increase, and the mean integral square error (ISE) of the functions $\hat{\beta}(u)$ and $\hat{g}(z)$ increase, indicating that with the rise of variance, that is, the signal-to-noise ratio decreases, the estimation accuracy of the model parameters will decrease. The mean square prediction error (MSPE) of the model will increase.

(2) As the sample size n increases from $n = 400$ to 800 and 1200, we see a decrease in the root mean square error (RMSE) and the meawe see n integral square error (ISE), and the performance of the estimation method becomes better and better. When the sample size n increases from $n = 400$ to 800 and 1200, the estimation of the autoregressive coefficient is asymptotically unbiased. The results show that the estimation of the autoregressive coefficient a, and the model parameters $\hat{\beta}(u)$ and $\hat{g}(z)$, are consistent estimators. The estimation curves for the slope and nonparametric function of the model in Scenario 1 ($AR(1)$) with parameters $n = 400$, $a = -0.5$, and $\sigma = 0.5$, and in Scenario 2 ($AR(2)$) with $n = 400$, $a = (0.5, -0.8)$, and $\sigma = 0.5$, were generated from 500 repetitions (see Figures 1 and 2). In the top left of Figures 1 and 2, the true slope function curve $\beta(\cdot)$ (red) and the median estimation curve $\hat{\beta}(\cdot)$ (blue) are displayed for Scenarios 1 and 2, respectively. The top right figure shows a comparison between the true curve (red) and the median estimation curve $\hat{g}(\cdot)$ (blue) of the nonparametric function $g(\cdot)$ in both scenarios. The bottom of Figures 1 and 2 displays the comparison of the estimation curves for the slope and the nonparametric function, repeated 500 times in Scenarios 1 and 2, respectively. Figures 1 and 2 show that when $n = 400$ and $\sigma = 0.5$, the slope $\beta(\cdot)$, the nonparametric function $g(\cdot)$, and the median estimation curve (in blue) closely align with the true curve (in red) for both $AR(1)$ and $AR(2)$. For $\sigma = 0.5$, $a = -0.5$, and $a = (0.5, -0.8)$, the 95% pointwise confidence bands for $\beta(\cdot)$ and $g(\cdot)$ are displayed for sample sizes $n = 400$, 800, and 1200 in Scenarios 1 and 2, respectively.

When $n = 400$, 800 and 1200, and $\sigma = 0.5$, Figures 3 and 4 display the 95% pointwise confidence bands for Scenarios 1 and 2, respectively. Regardless of the changes in the model parameters $\beta(u)$ and $g(z)$ or the errors from $AR(1)$ or $AR(2)$, the proposed estimation method yields estimates for $\beta(u)$ and $g(z)$ that closely align with the true curve. Moreover, as the sample size increases, $\hat{\beta}(u)$ and $\hat{g}(z)$ remain consistent estimates, demonstrating that the proposed model and estimation method perform well in finite samples.

Table 1. The bias and RMSE of $\hat{a}$, $\hat{\sigma}$, the ISE of $\hat{\beta}(u)$, $\hat{g}(z)$, and the MSPE under Scenario 1.

a	n	σ	$\hat{a}$		$\hat{\sigma}$		$\hat{\beta}(u)$	$\hat{g}(z)$	Model
			Bias	RMSE	Bias	RMSE	ISE_1	ISE_2	MSPE
0.1	400	0.5	0.003	0.049	-0.004	0.018	0.053	0.003	0.005
		1	-0.002	0.051	-0.016	0.077	0.118	0.013	0.018
	800	0.5	-0.001	0.036	-0.001	0.013	0.027	0.002	0.003
		1	-0.004	0.033	-0.012	0.050	0.055	0.006	0.009
	1200	0.5	0.000	0.028	0.000	0.010	0.019	0.001	0.002
		1	0.000	0.030	-0.007	0.043	0.039	0.005	0.006
-0.5	400	0.5	0.002	0.043	-0.003	0.018	0.052	0.003	0.004
		1	-0.004	0.044	-0.016	0.072	0.111	0.009	0.013
	800	0.5	0.000	0.032	-0.001	0.012	0.025	0.002	0.002
		1	-0.001	0.031	-0.006	0.048	0.052	0.005	0.007
	1200	0.5	0.000	0.024	0.000	0.010	0.016	0.001	0.002
		1	-0.002	0.025	-0.006	0.042	0.033	0.003	0.005
0.8	400	0.5	-0.006	0.032	-0.003	0.017	0.050	0.018	0.019
		1	-0.007	0.033	-0.020	0.071	0.124	0.066	0.069
	800	0.5	-0.005	0.022	-0.002	0.012	0.024	0.009	0.010
		1	-0.003	0.022	-0.005	0.055	0.044	0.034	0.035
	1200	0.5	-0.003	0.017	0.000	0.010	0.017	0.007	0.007
		1	-0.002	0.017	-0.006	0.040	0.031	0.021	0.022

Table 2. The bias and RMSE of $(\hat{a}_1, \hat{a}_2)$, $\hat{\sigma}$, the ISE of $\hat{\beta}(u)$, $\hat{g}(z)$, and the MSPE under Scenario 2.

(a_1, a_2)	n	σ	$\hat{a}_1$		$\hat{a}_2$		$\hat{\sigma}$		$\hat{\beta}(u)$	$\hat{g}(z)$	Model
			Bias	RMSE	Bias	RMSE	Bias	RMSE	ISE_1	ISE_2	MSPE
	400	0.5	-0.002	0.033	-0.016	0.036	-0.003	0.019	0.027	0.068	0.068
		1	-0.005	0.033	-0.011	0.035	-0.017	0.068	0.085	0.284	0.284
(0.1,0.8)	800	0.5	-0.004	0.022	-0.005	0.022	-0.001	0.013	0.013	0.034	0.034
		1	-0.002	0.022	-0.007	0.022	-0.011	0.054	0.033	0.109	0.110
	1200	0.5	-0.004	0.018	-0.004	0.017	0.001	0.011	0.009	0.022	0.022
		1	-0.001	0.018	-0.003	0.018	-0.008	0.04	0.021	0.085	0.085
	400	0.5	-0.002	0.031	0.000	0.030	-0.005	0.019	0.028	0.003	0.004
		1	-0.002	0.030	-0.001	0.030	-0.025	0.073	0.066	0.011	0.013
(0.5,-0.8)	800	0.5	-0.001	0.022	0.001	0.021	-0.001	0.012	0.013	0.002	0.002
		1	0.002	0.021	-0.001	0.021	-0.010	0.050	0.037	0.005	0.006
	1200	0.5	-0.002	0.018	0.001	0.018	0.001	0.011	0.009	0.002	0.002
		1	0.001	0.017	0.000	0.017	-0.008	0.042	0.020	0.004	0.004

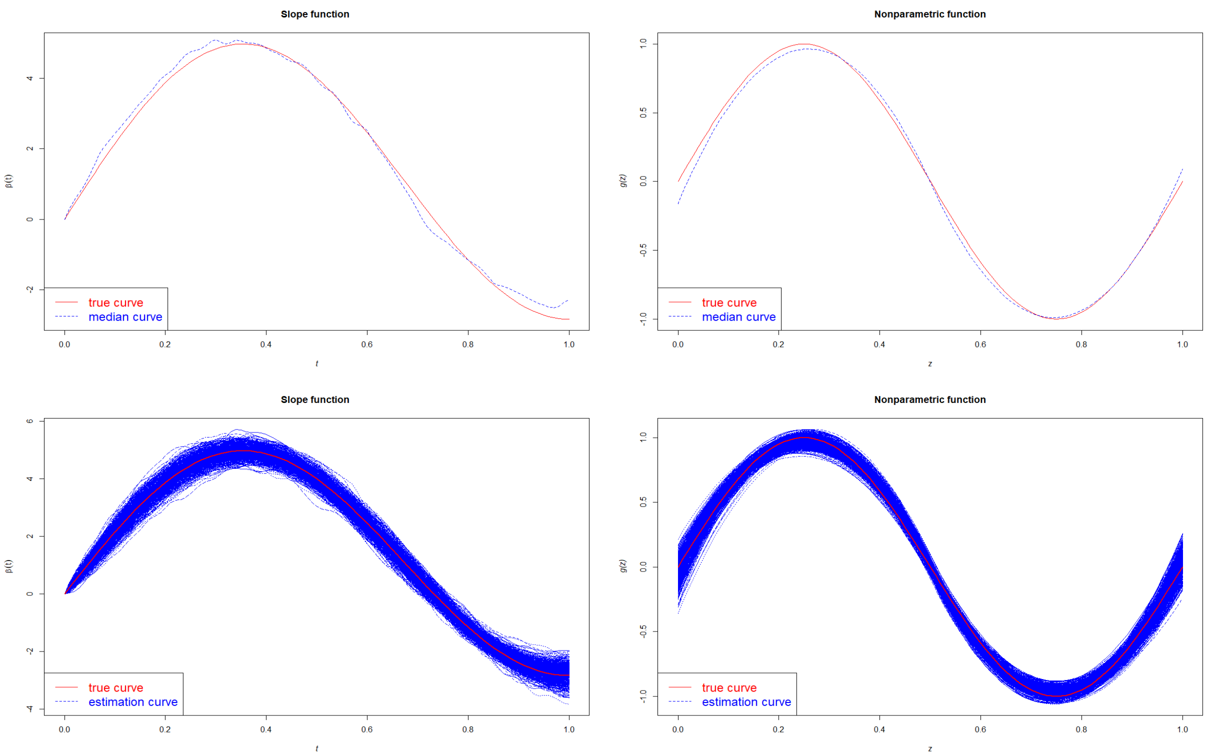

Figure 1. Estimation curves of the slope and nonparametric function under S cenario 1.

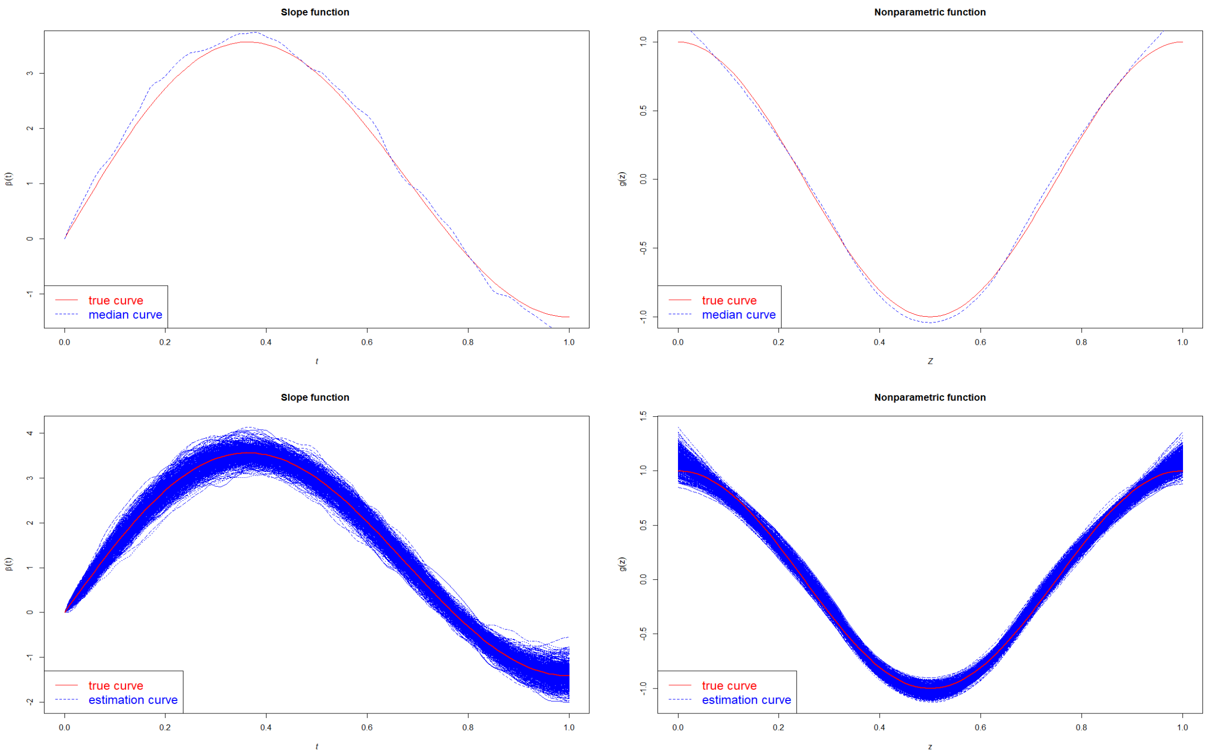

Figure 2. Estimation curves of the slope and nonparametric function under Scenario 2.

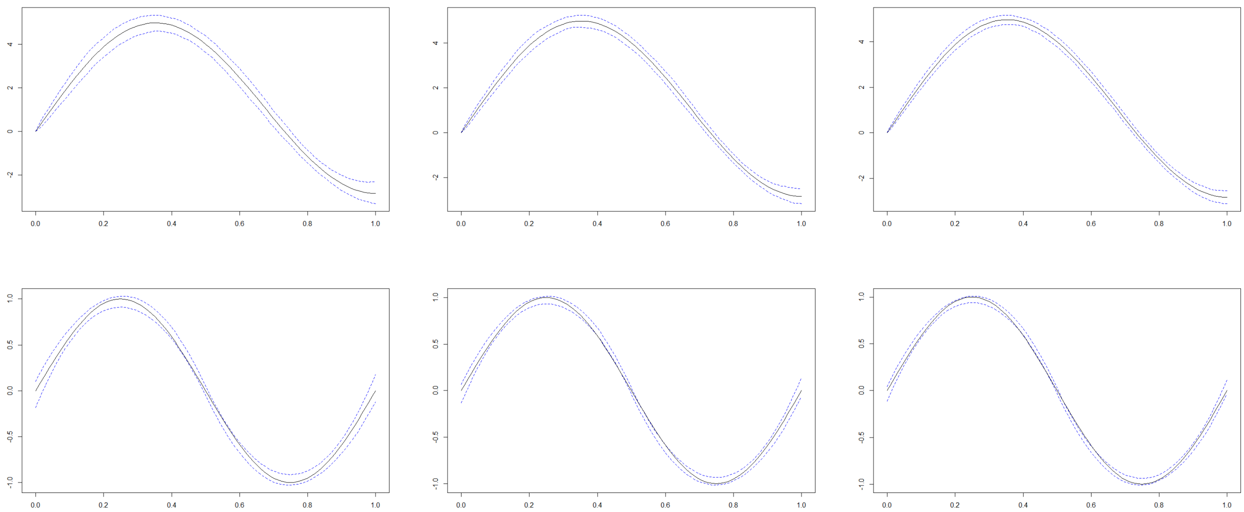

Figure 3. The 95% pointwise confidence band of the slope (top) and nonparametric function (bottom) under Scenario 1.

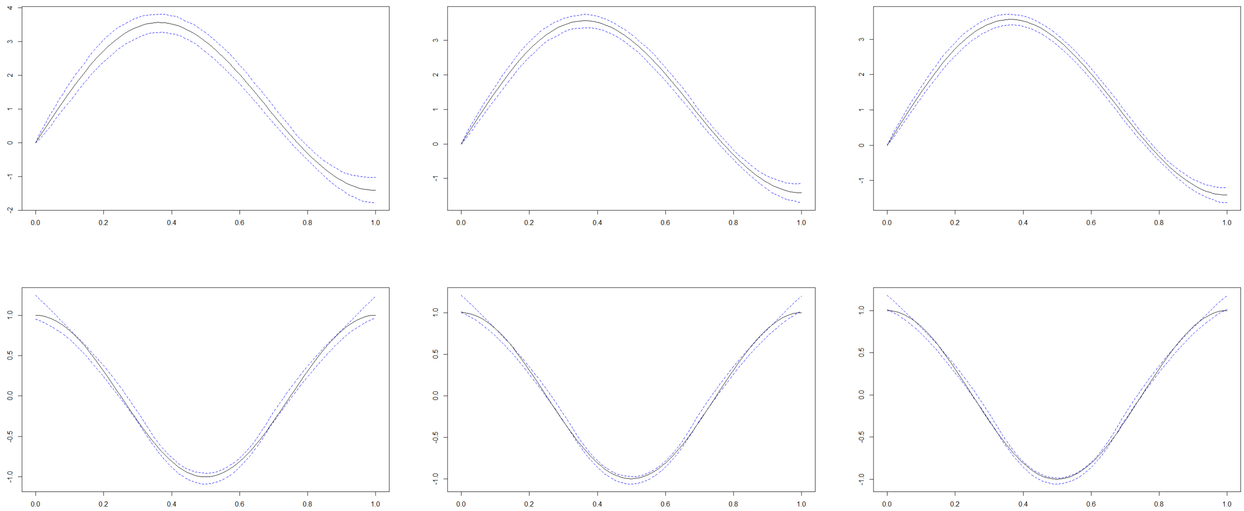

Figure 4. The 95% pointwise confidence band of the slope (top) and nonparametric function (bottom) under Scenario 2.

According to the suggestions of an anonymous reviewer, we run a new simulation to compare the predictive ability of the proposed model with that of existing models. Specifically, the simulation data $\{(X_t, Z_t, Y_t), 1 \leq t \leq n\}$ are generated according to the following two scenarios:
Scenario 3. It is the same as Scenario 1 with autoregressive coefficient $a = 0.5$ and standard deviation $\sigma = 0.5$.
Scenario 4. It is the same as Scenario 3 except for $g(Z) = 2Z$.

For each simulation case, we set the sample size to $n = 400$, and the simulation data is divided into the training data $(X_t, Z_t, Y_t)_{t=1}^{320}$ and the test data $(X_t, Z_t, Y_t)_{t=321}^{400}$. Similar to Aneiros and Vieu [22] and Yu et al. [23], we utilize the following two error criteria: mean quadratic error (MQE):

$$\text{MQE} = \frac{1}{79}\sum_{t=322}^{400}\left(Y_t - \hat{Y}_t\right)^2,$$

and mean relative quadratic error (MRQE):

$$\text{MRQE} = \frac{1}{79}\sum_{s=322}^{400}\frac{\left(Y_t - \hat{Y}_t\right)^2}{\text{Var}(Y)},$$

to evaluate the prediction ability of the model, where $\text{Var}(Y)$ is the empirical variance of $\{Y_t\}_{t=321}^{400}$. To demonstrate the superiority of the proposed model, we compare its prediction results with three standard models: the semi-functional linear model (SFL) from Zhou and Chen [10], the partial functional linear regression model (PFL) by Zhang et al. [26], and the partial functional linear regression model with autoregressive error (PFLAR) by Yu et al. [23]. Each simulation experiment is repeated 200 times, and the results are summarized in Table 3. From Table 3, we can observe that:
(i) When the true model is SFLAR, the proposed model has the best prediction accuracy.
(ii) Using our proposed SFLAR to fit PFLAR data has the same performance as using PFLAR to fit PFLAR data.

Table 3. The MQE and MRQE of different models.

Scenarios	Criterions	Models			
		SFLAR	SFL	PFLAR	PFL
Scenario 3	MQE	0.258	0.346	0.629	0.759
	MRQE	0.160	0.214	0.389	0.469
Scenario 4	MQE	0.256	0.344	0.255	0.341
	MRQE	0.175	0.234	0.174	0.233

5. Applications to real data

In this section, the model and method proposed in this paper will be applied to study the power data set, which can be downloaded from https://www.eia.gov. This paper aims to forecast the monthly retail sales of electricity C_i kilowatt hour (KWH) in the commercial sector. For this purpose, it is considered that the US monthly electricity consumed by commercial sectors which includes the power consumption data from January 1973 to January 2016 (517 months) (Figure 5) and their annual average retail price P_t (per KWH, including tax) (43 years).

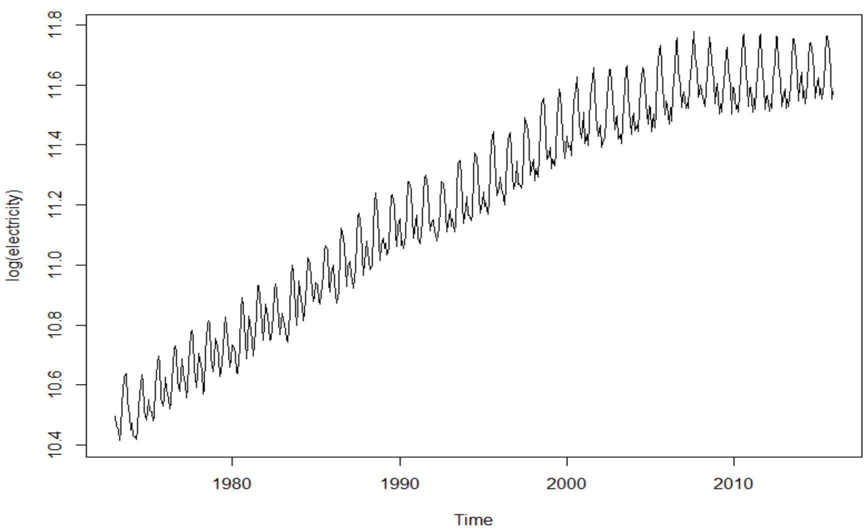

Figure 5. ln (power data) from January 1973 to January 2016.

As shown in Figure 5, the time series shows some linear trends and some heterogeneity in the variance structure. Similar to Aneiros and Vieu [22] and Yu et al. [23], we eliminate the heteroscedasticity and the linear trend of the electricity retail sales data by differencing the ln(data), and obtain the time series (see Figure 6): $\mathcal{X}_{i'} = \ln(C_{i'+1}) - \ln(C_{i'}), i' = 1, 2, \cdots, 516$. Let $X_t(s) = \{\mathcal{X}_{12(t-1)+s}, t = 1, 2, \cdots, 43, s = 1, 2, \cdots, 12\}$ be the monthly ln(difference electricity retail sales data).

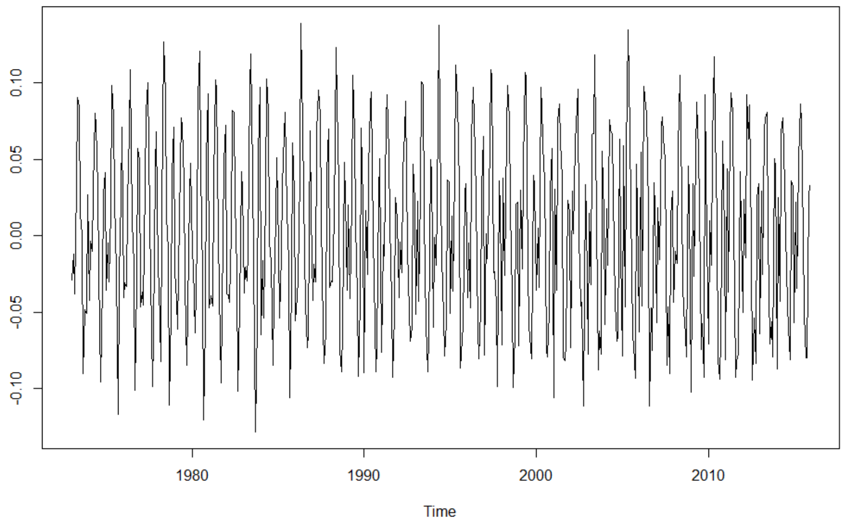

Figure 6. The differenced ln (power data to the commercial sector) from January 1973 to January 2016.

There is only one observation per month, so we use ten cubic B-spline basis functions to convert the discrete monthly power data into functional continuous smooth annual curve predictors $\{X_t(u), u \in [1, 12], t = 1, 2, \cdots, 43\}$ (see Figure 7).

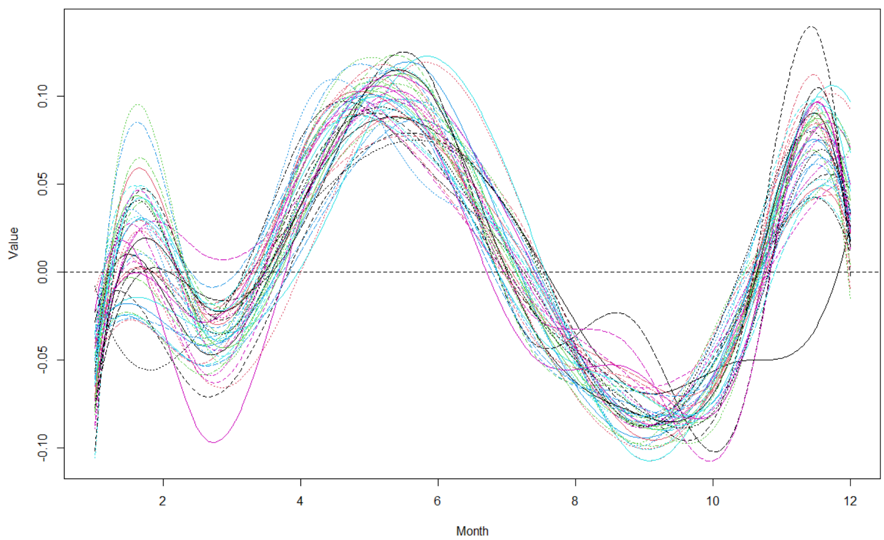

Figure 7. Annual curves of power data to the commercial sector from January 1973 to January 2016.

The response variable is $Y_t(s) = \mathcal{X}_{12t+s}$, which represents the power consumption data in the s-month of the t-year, where $t = 1, 2, \cdots, 42$, $s = 1, 2, \cdots, 12$.

Similar to Aneiros and Vieu [22] and Yu et al. [23], the power data set is divided into the training sample $\{Y_t(s), P_t, X_t(u)\}_{t=1}^{41}$ and the test sample $\{Y_{42}(s), P_{42}, X_{42}(u)\}$. According to Yang et al. [24], we consider the following semi-functional linear model (SFLAR) with autoregressive error to fit the training set:

$$Y_t(s) = \int_1^{12} \beta(u)X_t(u)du + g(P_t) + \varepsilon_t, \ \ \varepsilon_t = \sum_{l=1}^{q} a_l\varepsilon_{t-l} + e_t, \ t = 1, 2, \cdots, 41.$$

Then, the test set verifies the model's predicted value $Y_{42}(s)$. To demonstrate the superiority of the proposed model, we also compare its prediction results with three standard models: The semi-functional linear model (SFL), the partial functional linear regression model (PFL), and the partial functional linear regression model with autoregressive error (PFLAR). In addition, similar to Aneiros and Vieu [22] and Yu et al. [23], we also utilize the mean quadratic error (MQE):

$$\text{MQE} = \frac{1}{12}\sum_{s=1}^{12}\left(Y_{42}(s) - \hat{Y}_{42}(s)\right)^2,$$

and mean relative quadratic error (MRQE):

$$\text{MRQE} = \frac{1}{12}\sum_{s=1}^{12}\frac{\left(Y_{42}(s) - \hat{Y}_{42}(s)\right)^2}{\text{Var}(Y(s))}$$

to evaluate the prediction ability of the model, where $\text{Var}(Y(s))$ is the empirical variance of $\{Y_i(s)\}_{i=1}^{41}$.

The prediction effects of the four models under the test set are summarized in Table 4 and Figure 8. It is observed from Figure 8 that the four models exhibit consistency with the changing trend of the curve in Figure 7. Additionally, the prediction results from the test set in Table 4 and Figure 8 indicate that the SFL is closer to the actual value than the PFL. The MQE and MRQE of the SFL are also smaller than those of the PFL, suggesting that the prediction effect of the SFL is superior to the PFL, regardless of whether the autoregressive error is considered. Furthermore, the model with autoregressive error demonstrates a better prediction effect than the model without, implying that the former is more reasonable. The SFLAR model exhibits the smallest MQE and MRQE among the four models, signifying that the SFLAR model proposed in this paper yields the best prediction effect. Therefore, using the SFLAR model for this data is more reasonable.

Table 4. The MQE and MRQE of different models.

Criterions	Models			
	SFLAR	SFL	PFLAR	PFL
MQE	0.00017	0.00025	0.00027	0.00030
MRQE	0.34483	0.45411	0.47235	0.48765

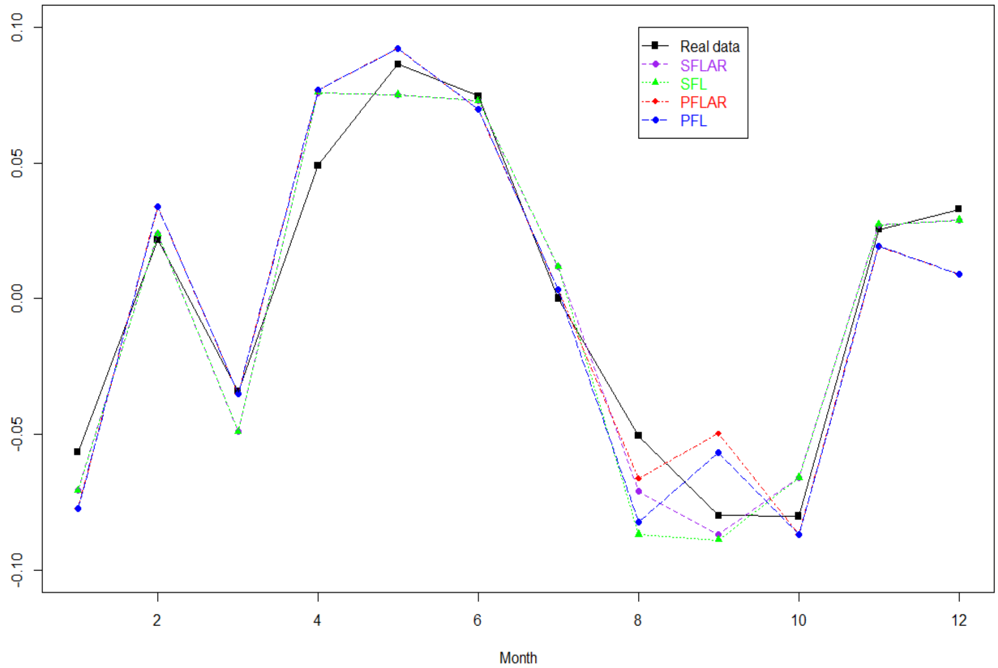

Figure 8. Monthly power data forecast of different models in 2015.

6. Discussion

The results of our study on forecasting monthly retail sales of electricity using an SFLAR model provide important insights into the dynamics of electricity consumption, which captures the influence of both functional and non-functional predictors of electricity sales. Incorporating autoregressive errors allowed the model to effectively capture temporal dependencies within the monthly sales data, thereby enhancing forecasting accuracy. However, our model also has some limitations. For example, we only consider one-dimensional covariate Z in our model. Although our model can be easily generalized to multidimensional situations, it may encounter the curse of dimensionality and computational burden. As one of the anonymous reviewers said, a single index model could provide much more efficient results. Hence, we can extend our model to the semi-functional single index model in future research. At the same time, deep neural networks have been extensively used for nonparametric regression. This also constitutes a topic for our future research.

7. Conclusions

This study focused on the SFLAR model, incorporating a random error sequence with an $AR(p)$ structure. By utilizing Mercer's theorem, the linear component of the function is dimensionally reduced using the method of functional principal component basis expansion. Meanwhile, the nonparametric function is approximated using the B-spline function. The slope function $\beta(\cdot)$, nonparametric function $g(\cdot)$, and autoregressive coefficient a are estimated using a two-step iterative algorithm. Theoretical properties of the estimated parameters are provided under specific regular conditions. Based on the simulation results, the proposed model and estimation method perform well even with a limited sample size. The real data analysis also indicates that the SFLAR model has the best prediction accuracy among these four models when autoregressive errors exist. Therefore, the SFLAR model proposed in this paper is more suitable for this power data.

Author contributions

Bin Yang: Investigation, funding acquisition, formal analysis, validation, software, data curation, writing-original draft, writing-review and editing; Min Chen: Conceptualization, methodology, supervision, project administration, validation; Jianjun Zhou: Conceptualization, funding acquisition, investigation, resources, validation, data curation, writing-original draft, writing-review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors are grateful to the two anonymous referees and the editor for their constructive remarks that greatly improved the manuscript. The research is partially supported by a grant from the National Key R & D Program of China, Grant/Award Number: 2022YFA1003701, National Nature Science Foundation of China (No.11861074, 12261051), Applied Basic Research Project of Yunnan Province (2019FB138), Key Project of Yunnan Provincial Basic Research (Grant No.202401AS070061), and the Yunnan Provincial Department of Education Science Research Fund (Grant No.2024J0087).

Conflict of interest

The authors declare no conflict of interest.

References

1. W. G. Manteiga, A. M. Calvo, Bootstrap in functional linear regression, *J. Stat. Plan. Infer.*, **141** (2011), 453–461. https://doi.org/10.1016/j.jspi.2010.06.027
2. M. F. Bande, P. Galeano, M. G. Manteiga, Functional principal component regression and functional partial least-squares regression: An overview and a comparative study, *Int. Stat. Rev.*, **85** (2017), 61–83. https://doi.org/10.1111/insr.12116
3. F. Ferraty, P. Hall, P. Vieu, Most-predictive design points for functional data predictors, *Biometrika*, **97** (2010), 807–824. http://www.jstor.org/stable/29777138
4. F. Belarbi, S. Chemikh, A. Laksaci, Local linear estimate of the nonparametric robust regression in functional data, *Stat. Probab. Lett.*, **134** (2018), 128–133. https://doi.org/10.1016/j.spl.2017.11.003

5. G. A. Pérez, P. Vieu, Semi-functional partial linear regression, *Stat. Probabil. Lett.*, **76** (2006), 1102–1110. https://doi.org/10.1016/j.spl.2005.12.007

6. H. Shin, Partial functional linear regression, *J. Stat. Plan. Infer.*, **139** (2009), 3405–3418. https://doi.org/10.1016/j.jspi.2009.03.001

7. H. Shin, M. Lee, On prediction rate in partial functional linear regression, *J. Multivariate Anal.*, **103** (2012), 93–106. https://doi.org/10.1016/j.jmva.2011.06.011

8. F. Yao, S. S. Chee, F. Wang, Regularized partially functional quantile regression, *J. Multivariate Anal.*, **156** (2017), 39–56. https://doi.org/10.1016/j.jmva.2017.02.001

9. D. H. Kong, K. J. Xue, F. Yao, H. H. Zhang, Partially functional linear regression in high dimensions, *Biometrika*, **103** (2016), 147–159. https://doi.org/10.1093/biomet/asv062

10. J. J. Zhou, M. Chen, Spline estimators for semi-functional linear model, *Stat. Probabil. Lett.*, **82** (2012), 505–513. https://doi.org/10.1016/j.spl.2011.11.027

11. Q. Y. Peng, J. J. Zhou, N. S. Tang, Varying coefficient partially functional linear regression models, *Stat. Pap.*, **57** (2016), 827–841. https://doi.org/10.1007/s00362-015-0681-3

12. S. Novo, G. Aneiros, P. Vieu, Automatic and location-adaptive estimation in functional single-index regression, *J. Nonparametr. Stat.*, **31** (2019), 364–392. https://doi.org/10.1080/10485252.2019.1567726

13. A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Stat. Plann. Inference.,* **147** (2014), 1–23. https://doi.org/10.1016/j.jspi.2013.04.002

14. A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.*, **146** (2016), 1–6. https://doi.org/10.1016/j.jmva.2015.12.001

15. N. X. Ling, P. Vieu, Nonparametric modelling for functional data: Selected survey and tracks for future, *Statistics*, **52** (2018), 934–949. https://doi.org/10.1080/02331888.2018.1487120

16. N. X. Ling, P. Vieu, On semiparametric regression in functional data analysis, *Wires. Comput. Stat.*, **13** (2020), e1538. https://doi.org/10.1002/wics.1538

17. G. Aneiros, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, *J. Multivariate Anal.*, **170** (2019), 3–9. https://doi.org/10.1016/j.jmva.2018.11.007

18. J. Vilar, R. Cao, G. Aneiros, Forecasting next-day electricity demand and price using nonparametric functional methods, *Int. J. Elec. Power*, **39** (2012), 48–55. https://doi.org/10.1016/j.ijepes.2012.01.004

19. G. Aneiros, J. Vilar, P. Raña, Short-term forecast of daily curves of electricity demand and price, *Int. J. Elec. Power*, **80** (2016), 96–108. https://doi.org/10.1016/j.ijepes.2016.01.034

20. J. Vilar, G. Aneiros, P. Raña Prediction intervals for electricity demand and price using functional data, *Int. J. Elec. Power*, **96** (2018), 457–472. https://doi.org/10.1016/j.ijepes.2017.10.010

21. R. Peláez, G. Aneiros, J. Vilar, Bootstrap prediction regions for daily curves of electricity demand and price using functional data, *Int. J. Elec. Power*, **162** (2024), 110244. https://doi.org/10.1016/j.ijepes.2024.110244

22. G. A. Pérez, P. Vieu, Nonparametric time series prediction: A semi-functional partial linear modeling, *J. Multivariate Anal.*, **99** (2008), 834–857. https://doi.org/10.1016/j.jmva.2007.04.010

23. P. Yu, T. Li, Z. Y. Zhu, Z. Z. Zhang, Composite quantile estimation in partial functional linear regression model with dependent errors, *Metrika*, **82** (2019), 633–656. https://doi.org/10.1007/s00184-018-0699-3

24. B. Yang, M. Chen, J. J. Zhou, Testing for error correlation in semi-functional linear models, *J. Syst. Sci. Complex.*, **36** (2023), 1697–1716. https://doi.org/10.1007/s11424-023-1431-6

25. S. D. Niang, S. Guillas, Functional semiparametric partially linear model with autoregressive errors, *J. Multivariate Anal.*, **101** (2010), 307–315. https://doi.org/10.1016/j.jmva.2008.06.008

26. T. Zhang, P. Dai, Q. Zhang, Joint detection for functional polynomial regression with autoregressive errors, *Commun. Stat.-Theory M.*, **46** (2017), 7837–7854. https://doi.org/10.1080/03610926.2015.1096384

27. P. X. Xiao, G. C. Wang, Partial functional linear regression with autoregressive errors, *Commun. Stat.-Theory M.*, **51** (2022), 4515–4536. https://doi.org/10.1080/03610926.2020.1818097

28. M. Wang, M. L. Shu, J. J. Zhou, S. X. Wu, M. Chen, Least square estimation for multiple functional linear model with autoregressive errors, *Acta Math. Appl. Sin.-E.*, **41** (2022), 84–98. https://doi.org/10.1007/s10255-024-1143-2

29. C. D. Boor, *A practical guide to spline*, New York: Springer, 1978. https://doi.org/10.2307/2006241

30. H. S. Wang, G. D. Li, T. C. Ling, Regression coefficient and autoregressive order shrinkage and selection via the Lasso, *J. R. Stat. Soc. B.*, **69** (2007), 63–78. https://doi.org/10.1111/j.1467-9868.2007.00577.x

31. Q. G. Tang, Estimation for semi-functional linear regression, *Statistics.*, **49** (2015), 1262–1278. https://doi.org/10.1080/02331888.2014.979827

32. L. Schumaker, *Spline functions: Basic theory*, 1 Eds., New York: John Wiley & Sons, 1981. https://doi.org/10.2307/2007301

33. P. Hall, J. L. Horowitz, Methodology and convergence rates for functional linear regression, *Ann. Stat.*, **35** (2007), 70–91. https://doi.org/10.1214/009053606000000957

34. H. Lian, Shrinkage estimation and selection for multiple functional regression, *Stat. Sinica*, **23** (2013), 51–74. https://doi.org/10.5705/ss.2011.160

Appendix

Proof of Theorem 1.

To prove the theorem, we need the following conclusions and lemmas. Let $B_j(z) = \sqrt{J}N_j(z)$, $1 \leq j \leq J$, where $N_j(z)$ is a standard B-spline function. It is known from de Boor [29] that for $z \in [\underline{z}, \overline{z}]$, there are $N_j(z) \geq 0$, $\sum_{j=1}^{J} N_j(z) = 1$, and there exist strictly positive constants M_1 and M_2 such that

$$M_1 \|c^0\|_2^2 \leq \int_{\underline{z}}^{\overline{z}} \left(\sum_{j=1}^{J} c_j^0 B_j(z) \right)^2 dz \leq M_2 \|c^0\|_2^2, \tag{A.1}$$

where $c^0 = (c_1^0, c_2^0, \cdots, c_J^0)^\top$ and $\|\cdot\|_2$ are the Euclidean norm of the vector. It can be seen from Assumption 4 and [32, Deduction 6.21] that there is a positive constant M_3 such that

$$g^0(z) = \sum_{j=1}^{J} c_j^0 B_j(z) + g_n(z), \sup_{z\in[\underline{z},\bar{z}]} |g_n(z)| \le M_3 J^{-p}. \tag{A.2}$$

Lemma 1. *Let* $\hat{\Lambda} = \sum_{t=q+1}^{n} \left\{(\hat{U}_t - \sum_{l=1}^{q} a_l^0 \hat{U}_{t-l})(\hat{U}_t - \sum_{l=1}^{q} a_l^0 \hat{U}_{t-l})^\top\right\}/(n-q)$*, where* $\hat{U}_t = (\hat{\xi}_t^\top, B_t^\top)^\top$, $\hat{\xi}_t = (\hat{\xi}_{t1}, \hat{\xi}_{t2}, \cdots, \hat{\xi}_{tM})^\top$*, and then under the Assumptions 2–6, there is* $\rho_{\min}(\hat{\Lambda}) = O_p(M^{-\alpha})$.

Proof. Note that the matrices Λ and $\hat{\Lambda}$ can be written as:

$$\Lambda = \sum_{l=0}^{q} \sum_{l'=0}^{q} (-a_l^0)(-a_{l'}^0)\mathbb{E}U_{t-l}U_{t-l'}^\top,$$

$$\hat{\Lambda} = \sum_{l=0}^{q} \sum_{l'=0}^{q} (-a_l^0)(-a_{l'}^0)\{\sum_{t=q+1}^{n} \hat{U}_{t-l}\hat{U}_{t-l'}^\top\}/(n-q),$$

where $a_0^0 = -1$, $0 \le l, l' \le q$. For any l, l', we have

$$\mathbb{E}U_{t-l}U_{t-l'}^\top = \begin{pmatrix} \mathbb{E}\xi_{t-l}\xi_{t-l'}^\top & \mathbb{E}\xi_{t-l}B_{t-l'}^\top \\ \mathbb{E}B_{t-l}\xi_{t-l'}^\top & \mathbb{E}B_{t-l}B_{t-l'}^\top \end{pmatrix},$$

$$\sum_{t=q+1}^{n} \hat{U}_{t-l}\hat{U}_{t-l'}^\top/(n-q) = \begin{pmatrix} \sum_{t=q+1}^{n} \hat{\xi}_{t-l}\hat{\xi}_{t-l'}^\top/(n-q) & \sum_{t=q+1}^{n} \hat{\xi}_{t-l}B_{t-l'}^\top/(n-q) \\ \sum_{t=q+1}^{n} B_{t-l}\hat{\xi}_{t-l'}^\top/(n-q) & \sum_{t=q+1}^{n} B_{t-l}B_{t-l'}^\top/(n-q) \end{pmatrix}.$$

Then

$$\begin{aligned} \frac{\sum_{t=q+1}^{n} \hat{U}_{t-l}\hat{U}_{t-l'}^\top}{n-q} - \mathbb{E}U_{t-l}U_{t-l'}^\top &= \frac{\sum_{t=q+1}^{n} \hat{U}_{t-l}\hat{U}_{t-l'}^\top}{n-q} - \frac{\sum_{t=q+1}^{n} U_{t-l}U_{t-l'}^\top}{n-q} \\ &\quad + \frac{\sum_{t=q+1}^{n} U_{t-l}U_{t-l'}^\top}{n-q} - \mathbb{E}U_{t-l}U_{t-l'}^\top \\ &= I_1 + I_2. \end{aligned}$$

For I_1, there is:

$$I_1 = \frac{1}{n-q}\begin{pmatrix} \sum_{t=q+1}^{n}(\hat{\xi}_{t-l}\hat{\xi}_{t-l'}^\top - \xi_{t-l}\xi_{t-l'}^\top) & \sum_{t=q+1}^{n}(\hat{\xi}_{t-l} - \xi_{t-l})B_{t-l'}^\top \\ \sum_{t=q+1}^{n} B_{t-l}(\hat{\xi}_{t-l'}^\top - \xi_{t-l'}^\top) & \mathbf{0} \end{pmatrix}.$$

Since for any $1 \le i, i' \le M$,

$$\begin{aligned} \frac{\sum_{t=q+1}^{n} \hat{\xi}_{(t-l)i}\hat{\xi}_{(t-l')i'}}{n-q} - \frac{\sum_{t=q+1}^{n} \xi_{(t-l)i}\xi_{(t-l')i'}}{n-q} &= \frac{\sum_{t=q+1}^{n}(\hat{\xi}_{(t-l)i} - \xi_{(t-l)i})\hat{\xi}_{(t-l')i'}}{n-q} \\ &\quad + \frac{\sum_{t=q+1}^{n} \xi_{(t-l)i}(\hat{\xi}_{(t-l')i'} - \xi_{(t-l')i'})}{n-q}. \end{aligned}$$

According to condition 2 and Eqs (5.21) and (5.22) in Hall and Horowitz [33], for all $1 \le i, i' \le M$, and $1 \le t \le n$, all have $|\hat{\xi}_{(t-l)i} - \xi_{(t-l)i}|^2 = O_p(n^{-1}i^2) = O_p(n^{-1}M^2)$, $|\hat{\xi}_{(t-l')i'}| = O_p(1)$, and $|\xi_{(t-l)i}| = O_p(1)$, so

$$\frac{\sum_{t=q+1}^{n} \hat{\xi}_{(t-l)i}\hat{\xi}_{(t-l')i'}}{n-q} - \frac{\sum_{t=q+1}^{n} \xi_{(t-l)i}\xi_{(t-l')i'}}{n-q} = O_p(M/\sqrt{n}). \quad (A.3)$$

Similarly, according to condition 2 and the Eqs (5.21) and (5.22) in Hall and Horowitz [33], it can be proved that for all $1 \le i \le M$, $1 \le j' \le J$, $1 \le t \le n$, we have

$$\frac{\sum_{t=q+1}^{n}(\hat{\xi}_{(t-l)i} - \xi_{(t-l)i})B_{j'}(Z_{t-l'})}{n-q} = O_p(M^2/\sqrt{n}). \quad (A.4)$$

According to Assumption 6 and Eqs (A.3) and (A.4), it can be seen that

$$\|I_1\|_1 = O_p(M^3/\sqrt{n}), \quad (A.5)$$

where $\|\cdot\|_1$ represents the row norm of the matrix. Similar to Lian [34], it can be demonstrated that

$$\|I_2\|_1 = O_p(M/\sqrt{n}). \quad (A.6)$$

Therefore, from Assumption 6 and Eqs (A.5) and (A.6), it can be deduced that

$$|\rho_{min}(\hat{\Lambda}) - \rho_{min}(\Lambda)| = O_p(M^3/\sqrt{n}) = o_p(M^{-\alpha}).$$

Combining the above results and Assumption 5, the lemma is proved. The proof of Theorem 1 is given below.

Proof of Theorem 1. Let $\bar{a} = (1, -a_1, \cdots, -a_q)^\top$, $\mathbf{Y}_t = (Y_t, Y_{t-1}, \cdots, Y_{t-q})^\top$, $\mathbf{U}_t = (U_t, U_{t-1}, \cdots, U_{t-q})^\top$, and $\hat{\mathbf{U}}_t = (\hat{U}_t, \hat{U}_{t-1}, \cdots, \hat{U}_{t-q})^\top$. Then $L_n(a, \theta)$ can be written as

$$L_n(\bar{a}, \theta) = \frac{1}{n-q} \sum_{t=q+1}^{n} (\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta)^\top \bar{a}\bar{a}^\top (\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta).$$

Define $r_n = n^{-(\alpha+2\gamma-1)/2(\alpha+2\gamma)}$, $\bar{a} - \bar{a}^0 = r_n u$, and $\theta - \theta^0 = r_n v$, where $u = (0, u_1, \cdots, u_q)^\top$, and $v = (v_{11}, \cdots, v_{1M}, v_{21}, \cdots, v_{2J})^\top = (v_1^\top, v_2^\top)^\top$. Let $\epsilon_t = (\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q})^\top$, and $\mathbf{U}_t^* = (U_t^*, U_{t-1}^*, \cdots, U_{t-q}^*)^\top$, where $U_{t-l}^* = \sum_{i=M+1}^{\infty} b_i^0 \xi_{(t-l)i} + g_n(Z_{t-l})$, $0 \le l \le q$. Then

$$\begin{aligned} L_n(\bar{a}, \theta) - L_n(\bar{a}^0, \theta^0) &= \frac{1}{n-q} \sum_{t=q+1}^{n} [(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta)^\top \bar{a}\bar{a}^\top (\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta) \\ &\quad -(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top \bar{a}^0\bar{a}^{0\top} (\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)] \\ &= \frac{2}{n-q} \sum_{t=q+1}^{n} \epsilon_t^\top \bar{a}^0 W_t(u, v) \\ &\quad + \frac{2}{n-q} \sum_{t=q+1}^{n} \left[(\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 + \mathbf{U}_t^*\right]^\top \bar{a}^0 W_t(u, v) \\ &\quad + \frac{1}{n-q} \sum_{t=q+1}^{n} W_t(u, v)^2, \end{aligned}$$

where $W_t(u,v) = r_n(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top u - r_n(\hat{\mathbf{U}}_t v)^\top \bar{a}^0 - r_n^2(\hat{\mathbf{U}}_t v)^\top u$. From the Cauchy-Schwarz inequality and inequality $xy \le \frac{x^2+y^2}{2}$, we can deduced that

$$\begin{aligned} L_n(\bar{a},\theta) - L_n(\bar{a}^0,\theta^0) &\ge \frac{2}{n-q}\sum_{t=q+1}^{n} \epsilon_t^\top \bar{a}^0 W_t(u,v) \\ &\quad -\frac{2}{n-q}\sum_{t=q+1}^{n}\left\{\left[(\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 + \mathbf{U}_t^*\right]^\top \bar{a}^0\right\}^2 \\ &\quad +\frac{1}{2(n-q)}\sum_{t=q+1}^{n} W_t(u,v)^2 \\ &\triangleq T_1 - T_2 + T_3. \end{aligned}$$

For T_1, there is

$$\begin{aligned} T_1 &= \frac{2r_n}{n-q}\sum_{t=q+1}^{n} \epsilon_t^\top \bar{a}^0 \epsilon_t^\top u \\ &\quad +\frac{2}{n-q}\sum_{t=q+1}^{n} \epsilon_t^\top \bar{a}^0 \left[r_n\left((\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 + \mathbf{U}_t^*\right)^\top u - r_n(\hat{\mathbf{U}}_t v)^\top \bar{a}^0 - r_n^2(\hat{\mathbf{U}}_t v)^\top u\right] \\ &\triangleq T_{11} + T_{12}. \end{aligned}$$

Note that $\epsilon_t^\top \bar{a}^0 = e_t$, and it can be seen from the independence of e_t and $\epsilon_t^\top u$ that

$$T_{11} = O_p\left(\frac{r_n\|u\|_2}{\sqrt{n}}\right) = o_p(r_n^2). \tag{A.7}$$

For T_{12}, it can be seen from Assumptions 2–4 and Assumption 6 that for $0 \le l \le q$, all have

$$\sum_{i=1}^{M} b_i^0(\xi_{(t-l)i} - \hat{\xi}_{(t-l)i}) = O_p(n^{-(\alpha+4\gamma-4)/2(\alpha+2\gamma)}) = o_p(n^{-(\alpha+2\gamma-1)/2(\alpha+2\gamma)}),$$

$$U_{t-l}^* = O_p(n^{-(\alpha+2\gamma-1)/2(\alpha+2\gamma)}),$$

$|\hat{\xi}_{(t-l)i}| = O_p(1)$, and $B_j(Z_{t-l}) = O_p(1)$. From the independence of e_t and (X_t, Z_t), we have

$$T_{12} = O_p\left(\frac{r_n\|v\|_2}{\sqrt{n}}\right) = o_p(r_n^2). \tag{A.8}$$

Combine Eqs (A.6) and (A.7), and we can deduced that

$$T_1 = o_p(r_n^2). \tag{A.9}$$

For T_2, we have

$$T_2 \le \frac{4}{n-q}\sum_{t=q+1}^{n}\left\{\left[(\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0\right]^\top \bar{a}^0\right\}^2 + \frac{4}{n-q}\sum_{t=q+1}^{n}\left(\mathbf{U}_t^{*\top}\bar{a}^0\right)^2 \triangleq T_{21} + T_{22}.$$

From $U^*_{t-l} = O_p(n^{-(\alpha+2\gamma-1)/2(\alpha+2\gamma)})$, we can deduced that $T_{22} = O_p(n^{-(\alpha+2\gamma-1)/(\alpha+2\gamma)})$. For T_{21}, since

$$(\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 = \left((U_t - \hat{U}_t)^\top\theta^0, (U_{t-1} - \hat{U}_{t-1})^\top\theta^0, \cdots, (U_{t-q} - \hat{U}_{t-q})^\top\theta^0\right)^\top,$$

and $(U_{t-l} - \hat{U}_{t-l})^\top\theta^0 = \sum_{i=1}^M b_i^0(\xi_{(t-l)i} - \hat{\xi}_{(t-l)i}) = o_p(n^{-(\alpha+2\gamma-1)/2(\alpha+2\gamma)})$, then

$$T_2 = O_p(n^{-(\alpha+2\gamma-1)/(\alpha+2\gamma)}) = O_p(r_n^2). \qquad (A.10)$$

For T_3, it can be deduced from the inequality $(x-y)^2 \geq x^2/2 - y^2$ and the Cauchy-Schwarz inequality that

$$\begin{aligned}
T_3 &= \frac{r_n^2}{2(n-q)}\sum_{t=q+1}^{n}\left[(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top u - (\hat{\mathbf{U}}_t v)^\top\bar{a}^0 - r_n(\hat{\mathbf{U}}_t v)^\top u\right]^2 \\
&\geq \frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}\left[(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top u - (\hat{\mathbf{U}}_t v)^\top\bar{a}^0\right]^2 - \frac{r_n^4}{2(n-q)}\sum_{t=q+1}^{n}\left[(\hat{\mathbf{U}}_t v)^\top u\right]^2 \\
&\geq \frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}\left[(\hat{\mathbf{U}}_t v)^\top\bar{a}^0\right]^2 + \frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}\left[(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top u\right]^2 \\
&\quad -\frac{r_n^2}{2(n-q)}\sum_{t=q+1}^{n}(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top u(\hat{\mathbf{U}}_t v)^\top\bar{a}^0 - \frac{r_n^4}{2(n-q)}\sum_{t=q+1}^{n}\left[(\hat{\mathbf{U}}_t v)^\top u\right]^2 \\
&\geq \frac{r_n^2}{8(n-q)}\sum_{t=q+1}^{n}\left[(\hat{\mathbf{U}}_t v)^\top\bar{a}^0\right]^2 \\
&\quad +\frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}\left\{\left[(\mathbf{Y}_t - \hat{\mathbf{U}}_t\theta^0)^\top u\right]^2 - 2\left[((\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 + \hat{\mathbf{U}}_t^*)^\top u\right]^2\right\} \\
&\quad -\frac{r_n^2}{2(n-q)}\sum_{t=q+1}^{n}\epsilon_t^\top u(\hat{\mathbf{U}}_t v)^\top\bar{a}^0 - \frac{r_n^4}{2(n-q)}\sum_{t=q+1}^{n}\left[(\hat{\mathbf{U}}_t v)^\top u\right]^2 \\
&\triangleq T_{31} + T_{32} - T_{33} - T_{34}.
\end{aligned}$$

From Lemma 1, we have

$$T_{31} = \frac{r_n^2}{8}v^\top\left[\frac{1}{n-q}\sum_{t=q+1}^{n}\hat{\mathbf{U}}_t^\top\bar{a}^0\bar{a}^{0\top}\hat{\mathbf{U}}_t\right]v \geq \rho_{min}(\hat{\Lambda})\|v\|_2^2/8 = O_p(r_n^2M^{-\alpha})\|v\|_2^2.$$

For T_{32}, there is

$$\begin{aligned}
T_{32} &= \frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}u^\top\epsilon_t\epsilon_t^\top u + \frac{r_n^2}{2(n-q)}\sum_{t=q+1}^{n}u^\top\epsilon_t((\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 + \hat{\mathbf{U}}_t^*)^\top u \\
&\quad -\frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}\left[((\mathbf{U}_t - \hat{\mathbf{U}}_t)\theta^0 + \hat{\mathbf{U}}_t^*)^\top u\right]^2.
\end{aligned}$$

From Assumptions 1 and 2, it can be seen that the sequence $\{\varepsilon_t\}$ is a stationary and ergodic time series, and from the stationary ergodicity theorem, we can deduced that

$$\frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}u^\top\epsilon_t\epsilon_t^\top u = O_p(r_n^2\|u\|_2^2).$$

In addition, from the proof results on T_2 and the Cauchy-Schwarz inequality, we have

$$\frac{r_n^2}{2(n-q)}\sum_{t=q+1}^{n} u^\top \epsilon_t((\mathbf{U}_t-\hat{\mathbf{U}}_t)\theta^0+\hat{\mathbf{U}}_t^*)^\top u = o_p(r_n^2\|u\|_2^2),$$

$$\frac{r_n^2}{4(n-q)}\sum_{t=q+1}^{n}\left[((\mathbf{U}_t-\hat{\mathbf{U}}_t)\theta^0+\hat{\mathbf{U}}_t^*)^\top u\right]^2 = o_p(r_n^2\|u\|_2^2).$$

Therefore

$$T_{32} = O_p(r_n^2\|u\|_2^2).$$

For T_{34}, we have

$$T_{34} = \frac{r_n^4}{2}\sum_{l=1}^{q}\sum_{l'=1}^{q} u_l u_{l'} v^\top\left[\sum_{t=q+1}^{n}\hat{U}_{t-l}\hat{U}_{t-l'}^\top/(n-q)\right]v.$$

Take notice of

$$\mathbb{E}U_{t-l}U_{t-l'}^\top = \begin{pmatrix} \mathbb{E}\xi_{t-l}\xi_{t-l'}^\top & \mathbb{E}\xi_{t-l}B_{t-l'}^\top \\ \mathbb{E}B_{t-l}\xi_{t-l'}^\top & \mathbb{E}(B_{t-l}B_{t-l'}^\top) \end{pmatrix}.$$

From Assumption 2, when $l = l'$, we have

$$\mathbb{E}U_{t-l}U_{t-l'}^\top = \begin{pmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbb{E}(B_{t-l}B_{t-l}^\top) \end{pmatrix},$$

where $\Sigma = diag(\lambda_1, \cdots, \lambda_M)$, and so $v_1^\top\Sigma v_1 = O(\|v_1\|_2^2)$. According to Assumption 2 and Eq (A.1), we know that

$$v_2^\top\mathbb{E}(B_{t-l}B_{t-l}^\top)v_2 = O(\|v_2\|_2^2),$$

and therefore $v^\top\mathbb{E}(U_{t-l}U_{t-l}^\top)v = O(\|v_1\|_2^2+\|v_2\|_2^2)$. When $l \neq l'$, we know that $v^\top\mathbb{E}(U_{t-l}U_{t-l'}^\top)v = O(\|v_2\|_2^2)$. Similar to the proof of Lemma 1, we have

$$\|\sum_{t=q+1}^{n}\hat{U}_{t-l}\hat{U}_{t-l'}^\top/(n-q) - \mathbb{E}(U_{t-l}U_{t-l'}^\top)\|_1 = O_p(M^2/\sqrt{n}) = o_p(1),$$

and then

$$T_{34} = O_p(r_n^4\|u\|_2^2(\|v_1\|_2^2+\|v_2\|_2^2)) = o_p(r_n^2).$$

For T_{33}, there is

$$T_{33} = \frac{r_n^4}{2}\sum_{l=1}^{q}\sum_{l'=0}^{q} u_l(-a_{l'}^0)\sum_{t=q+1}^{n}[\varepsilon_{t-l}U_{t-l'}^\top v + \varepsilon_{t-l}(\hat{U}_{t-l'}-U_{t-l'})^\top v]/(n-q).$$

From the independence of e_t and (X_t, Z_t), it can be seen that ε_t is also independent of (X_t, Z_t), so $\mathbb{E}(\varepsilon_{t-l}\hat{U}_{t-l'}^\top) = 0$. From the stationarity, we know

$$\mathbb{V}ar(\frac{1}{n-q}\sum_{t=q+1}^{n}\varepsilon_{t-l}U_{t-l'}^\top v) = \frac{1}{(n-q)^2}\sum_{t=q+1}^{n}\sum_{t'=q+1}^{n}\mathbb{E}(\varepsilon_{t-l}\varepsilon_{t'-l}v^\top U_{t-l'}U_{t'-l'}^\top v)$$

$$= O(\frac{\|v_1\|_2^2 + \|v_2\|_2^2}{n-q}).$$

Therefore, combining the results of Lemma 1, we can deduced that

$$T_{33} = O_p(r_n^2 \frac{\|u\|_2(\|v_1\|_2^2 + J^{-1}\|v_2\|_2^2)^{1/2}}{\sqrt{n}}) = o_p(r_n^2).$$

From the above results, we have

$$T_3 = O_p(r_n^2\|u\|^2) + O_p(r_n^2 M^{-\alpha})\|v\|^2.$$

Combining the results of T_1, T_2, and T_3, we have

$$L_n(\bar{a}, \theta) - L_n(\bar{a}^0, \theta^0) \geq O_p(r_n^2\|u\|^2) + O_p(r_n^2 M^{-\alpha})\|v\|^2 - O_p(r_n^2).$$

It can be seen that the lower bound of $L_n(\bar{a}, \theta) - L_n(\bar{a}^0, \theta^0)$ is mainly controlled by the positive term T_3 when $\|u\|^2$ and $\|v\|^2$ are large enough. Therefore

$$\|\hat{\theta} - \theta^0\|_2 = O_p(r_n),$$
$$\|\hat{a} - a^0\|_2 = O_p(r_n).$$

Take notice of

$$\begin{aligned}
\|\hat{\beta}(t) - \beta^0(t)\|^2 &= \|\sum_{i=1}^{M} \hat{b}_i\hat{\phi}_i - \sum_{i=1}^{\infty} b_i^0\phi_i\|^2 \\
&\leq 2\|\sum_{i=1}^{M} \hat{b}_i\hat{\phi}_i - \sum_{i=1}^{M} b_i^0\phi_i\|^2 + 2\|\sum_{i=M+1}^{\infty} b_i^0\phi_i\|^2 \\
&\leq 4\|\sum_{i=1}^{M}(\hat{b}_i - b_i^0)\hat{\phi}_i\|^2 + 4\|\sum_{i=1}^{M} b_i^0(\hat{\phi}_i - \phi_i)\|^2 + 2\|\sum_{i=M+1}^{\infty} b_i^0\phi_i\|^2 \\
&\triangleq 4D_1 + 4D_2 + 2D_3.
\end{aligned}$$

It can be seen from $\|\hat{\theta} - \theta^0\|_2 = O_p(r_n)$, and Assumptions 2 and 3, that

$$D_3 = \sum_{i=M+1}^{\infty} b_i^{02} = O(n^{-(2\gamma-1)/(\alpha+2\gamma)}), \quad D_1 = \|\hat{b} - b^0\|_2^2 = O_p(r_n^2) = o_p(n^{-(2\gamma-1)/(\alpha+2\gamma)}),$$

$$D_2 \leq M\sum_{i=1}^{M}\|\hat{\phi}_i - \phi_i\|^2 b_i^{02} = O_p(n^{-(\alpha+4\gamma-4)/(\alpha+2\gamma)}) = o_p(n^{-(2\gamma-1)/(\alpha+2\gamma)}).$$

So

$$\|\hat{\beta}(t) - \beta^0(t)\|^2 = O_p(n^{-(2\gamma-1)/(\alpha+2\gamma)}).$$

From Eqs (A.1) and (A.2), we know that

$$\|\hat{g}(z) - g^0(z)\|^2 \leq 2\int_{\underline{z}}^{\bar{z}}[\sum_{j=1}^{J}(\hat{c}_j - c_j^0)B_j(z)]^2 dz + 2\int_{\underline{z}}^{\bar{z}} g_n^2(z)dz$$

$$\leq \quad 2M_2\|\hat{c} - c^0\|_2^2 + 2M_3J^{-2p}.$$

Combining Assumption 4, Assumption 6, and $\|\hat{\theta} - \theta^0\|_2 = O_p(r_n)$, we have

$$\|\hat{g}(z) - g^0(z)\|^2 = O_p(n^{-(\alpha+2\gamma-1)/(\alpha+2\gamma)}).$$

The proof of the theorem is complete.