Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 53 (4) (2024), 1196 – 1217
DOI : 10.15672/hujms.1324888

Research Article

Estimation in functional partially linear spatial autoregressive model

Yuping Hu, Siyu Wu, Sanying Feng*

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

Abstract

Functional regression has been a hot topic in statistical research. However, not much work has been done when response variables are cross-sectionally dependent variables and explanatory variables contain a real-valued scalar variable and a functional-valued random variable. In this paper, we consider a new functional partially linear spatial autoregressive model. Based on the functional principal components analysis and basis function approximation, we obtain the estimators of the unknown parameter and functions through the instrumental variables estimation method. The asymptotic normality and convergence rates of estimators are proved under some mild conditions. In addition, we illustrate the finite sample performance of the proposed estimation method through simulation study and a real data analysis.

Mathematics Subject Classification (2020). 62G05, 62G07

Keywords. Functional partially linear spatial autoregressive model, spatial autoregression, functional principal component analysis, instrument variable

1. Introduction

Functional data analysis has received widespread attention due to its application in multiple disciplines, including chemistry, biology, medicine, and economics, etc. There have been amounts of significant work on functional data analysis [3, 7, 9, 10, 24]. With the development of nonparametric technology, partially linear models are widely used. A good deal of researches have been done on the combination of partially linear models and functional data. For example, Shin [26] proposed a partially functional linear model. Feng and Xue [8] studied a partially functional linear varying coefficient model. Kong et al. [17] studied a new partially functional linear model and established the consistency and oracle properties of the proposed method under some mild conditions. Yu et al. [32] discussed the estimators of a single-index partially functional linear regression model by B-spline approximation. Further, functional partially linear models have been proposed because of their flexibility and interpretability. Based on the functional principal components analysis (FPCA) and kernel estimation method, Lian [20] discussed the parametric and nonparametric estimators of the functional partially linear model, and proved the asymptotic properties of the estimators. Tang [30] studied the estimators of functional partially

*Corresponding Author.
Email addresses: hyp@zzu.edu.cn (Y. Hu), 2499226134@qq.com (S. Wu), fsy5801@zzu.edu.cn (S. Feng)
Received: 09.07.2023; Accepted: 27.07.2024

linear models through FPCA and B-spline approximation method. Zhou and Chen [34] approximated the slope function and nonparametric function with polynomial splines for a semi-functional linear model. All of the aforementioned functional partially linear models have a limitation that the response variables are cross-sectionally independent.

However, the response variables may be cross-sectionally dependent, which are frequently encountered in economic, finance, and environment, etc. There has been considerable work about cross-sectionally dependent response variables [1, 4, 5]. Su and Jin [29] applied the quasi-maximum likelihood for partially linear spatial autoregressive models. Malikov and Sun [23] considered the case where spatial autoregressive parameters can vary across units, and proposed several nonparametric GMM estimators for a flexible semiparametric spatial autoregressive model. Du et al. [6] developed a partially linear additive spatial autoregressive model and studied the asymptotic properties of the proposed estimators.

When the response variables are cross-sectionally dependent and explanatory variables contain a real-valued scalar variable and a functional-valued random variable, typical methods developed for the aforementioned functional partially linear models will be invalid. Therefore, it is necessary to develop a new method to deal with the spatial dependency. In this paper, we consider the combination of the functional partially linear model and spatial autoregressive model, and propose a new functional partially linear spatial autoregressive model with the following form:

$$Y_i = \lambda \sum_{j=1}^{N} w_{ij} Y_j + \int_{\mathscr{T}} \gamma(t) X_i(t) \mathrm{d}t + g(Z_i) + V_i. \tag{1.1}$$

Y_i is the ith observation of the real-valued dependent variable, λ is a scalar parameter, w_{ij} is the (i,j)th element of the matrix $\boldsymbol{W}_N$ which is a known $N \times N$ spatial weight matrix with zero diagonal elements. The spatial weight matrix $\boldsymbol{W}_N$ is defined according to the distance between individuals. The distance here is generalized and not just the distance between geographical locations, such as economic distance and so on. $X_i(t)$ be zero mean stochastic process belonging to $L^2(\mathscr{T})$, for the sake of simplicity, we suppose throughout that $\mathscr{T}$=$[0,1]$. $\gamma(t)$ is an unknown square integrable slope function on $[0,1]$. Z_i is the ith observation of covariate Z, and for simplicity, Z is assumed to distribute on the compact interval $[c,d]$. Without loss of generality, the interval $[c,d]$ can be assumed to be $[0,1]$. $g(z)$ is an unknown smooth function on $[0,1]$ with the assumption that $E[g(Z)] = 0$, in order to ensure the identifiability of the nonparametric function. V_i are independent and identically distributed random errors with zero mean and finite variance σ^2.

To the best of our knowledge, the above functional partially linear spatial autoregressive model has not yet been studied in the scientific literature. The model is flexible in practice and can deal with functional data and cross-sectionally dependent data. In this paper, our purpose is to develop the theories and methods for estimating the parameter λ, the nonparametric function $g(z)$, and the slope function $\gamma(t)$ of model (1.1). Specifically, $\gamma(t)$ is processed based on the FPCA, $g(z)$ is approximated by the B-spline basis function, and then the estimators of parametric and nonparametric components are obtained utilizing the instrumental variable methods. Under some regularity conditions, the asymptotic properties of the estimators are established.

The remainder of the paper is structured as follows. In Section 2, we introduce the estimation method. In Section 3, we investigate the asymptotic properties of the estimators. The results of the simulation study are presented in Section 4. Section 5 gives a real data analysis. We conclude the paper and propose some interesting directions for future research in Section 6. Lastly, technical proofs are given in Appendix.

2. Estimation procedures

In order to fit the functional data, we consider FPCA. Denote the covariance function of $X(t)$ by $K_X(s,t) = \mathrm{Cov}(X(s), X(t))$. If $K_X(s,t)$ is continuous on $\mathscr{T} \times \mathscr{T}$, Mercer's theorem implies that $K_X(s,t) = \sum_{k=1}^{\infty} \tau_k \phi_k(s)\phi_k(t)$ with $\sum_{k=1}^{\infty} \tau_k < \infty$, where $\{\phi_k(t)\}$ is a complete orthogonal basis sequence in $L^2(\mathscr{T})$ and $\{\tau_k\}$ is a non-increasing sequence of non-negative eigenvalues.

By the Karhunen-Loève expansion, $X_i(t)$ can be represented as

$$X_i(t) = \sum_{k=1}^{\infty} U_{ik}\phi_k(t),$$

where $U_{ik} = \langle X_i(t), \phi_k(t)\rangle$ are uncorrelated random variables with mean zero and variances $E(U_{ik}^2) = \tau_k$, and $\langle \cdot, \cdot \rangle$ represents the $L^2(\mathscr{T})$ inner product. Expanded on the orthogonal eigenbasis $\{\phi_k(t)\}$, the slope function can be represented as

$$\gamma(t) = \sum_{k=1}^{\infty} \gamma_k \phi_k(t),$$

where $\gamma_k = \langle \gamma(t), \phi_k(t)\rangle$.

In practice, ϕ_k are unknown. Therefore, it is necessary to find the estimators. For this purpose, we consider the empirical version of $K_X(s,t)$ as follows

$$\widehat{K}_X(s,t) = \frac{1}{N}\sum_{i=1}^{N}(X_i(s) - \bar{X}(s))(X_i(t) - \bar{X}(t)) = \sum_{k=1}^{\infty} \widehat{\tau}_k \widehat{\phi}_k(s)\widehat{\phi}_k(t),$$

where the $(\widehat{\tau}_k,\ \widehat{\phi}_k)$ are pairs of eigenvalue and eigenfunction for the covariance operator associated with $\widehat{K}_X$ and $\widehat{\tau}_1 \geq \widehat{\tau}_2 \geq \cdots \geq 0$. We use $(\widehat{\tau}_k,\ \widehat{\phi}_k)$ as the estimators of $(\tau_k,\ \phi_k)$.

Thus, Model (1.1) can be well-approximated by

$$Y_i \approx \lambda \sum_{j=1}^{N} w_{ij} Y_j + \sum_{k=1}^{m} \gamma_k \langle X_i(t), \widehat{\phi}_k(t)\rangle + g(Z_i) + V_i, \tag{2.1}$$

where m is sufficiently large.

Let $\boldsymbol{Y}_N = (Y_1, \cdots, Y_N)^T$, $\boldsymbol{Z}_N = (Z_1, \cdots, Z_N)^T$, $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_m)^T$, $\boldsymbol{V}_N = (V_1, \cdots, V_N)^T$, and $\boldsymbol{U}_N = (\langle X_i, \widehat{\phi}_k\rangle)_{N\times m}$ is an $N \times m$ matrix with the (i,k)th element is $\langle X_i, \widehat{\phi}_k\rangle$. Then, Model (2.1) can be written as matrix notation

$$\boldsymbol{Y}_N \approx \lambda \boldsymbol{W}_N \boldsymbol{Y}_N + \boldsymbol{U}_N \boldsymbol{\gamma} + g(\boldsymbol{Z}_N) + \boldsymbol{V}_N. \tag{2.2}$$

The estimator of nonparametric function $g(z)$ is obtained by the method of spline approximation. Let $\boldsymbol{B}(z) = (B_1(z), \cdots, B_{K_N+l+1}(z))$ be a set of B-spline basis functions with order $l+1$, and $0 = z_0 < z_1 < \cdots < z_{K_N-1} < z_{K_N} < 1$ are the quasi-uniform internal knots. Therefore, $g(z)$ can be approximated by a linear combination of normalized B-spline basis functions $g(z) \approx \sum_{j=1}^{K_N+l+1} B_j(z)\alpha_j = \boldsymbol{B}(z)\boldsymbol{\alpha}$ where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{K_N+l+1})^T$ is called the spline coefficient vector. Furthermore, invoking $E[g(Z)] = 0$, we let $\overline{B}_j = \frac{1}{N}\sum_{i=1}^{N} B_j(Z_i)$, $\pi_j(z) = B_j(z) - \overline{B}_j$, and $g(z) \approx g_N(z) = \sum_{j=1}^{K_N+l+1} \pi_j(z)\alpha_j = \boldsymbol{\Pi}(z)\boldsymbol{\alpha}$, where $\boldsymbol{\Pi}(z) = (\pi_1(z), \cdots, \pi_{K_N+l+1}(z))$. It is obvious that $\sum_{i=1}^{N} g_N(Z_i) = 0$.

With the B-spline approximation, Model (2.2) can be rewritten as follows:

$$\begin{aligned}\boldsymbol{Y}_N &\approx \lambda \boldsymbol{W}_N \boldsymbol{Y}_N + \boldsymbol{U}_N \boldsymbol{\gamma} + \boldsymbol{\Pi}(\boldsymbol{Z}_N)\boldsymbol{\alpha} + \boldsymbol{V}_N \\ &= \lambda \boldsymbol{W}_N \boldsymbol{Y}_N + \boldsymbol{U}_N \boldsymbol{\gamma} + \boldsymbol{\Pi}\boldsymbol{\alpha} + \boldsymbol{V}_N,\end{aligned} \tag{2.3}$$

where $\boldsymbol{\Pi} = (\pi_1(\boldsymbol{Z_N}), \cdots, \pi_{K_N+l+1}(\boldsymbol{Z_N}))$ and $\pi_j(\boldsymbol{Z_N}) = (\pi_j(Z_1), \cdots, \pi_j(Z_N))^T$, $j = 1, \cdots, K_N + l + 1$.

Let $\boldsymbol{P} = \boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T$ denote the projection matrix onto the space spanned by $\boldsymbol{U}_N$, we have

$$(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}_N \approx (\boldsymbol{I}-\boldsymbol{P})\lambda\boldsymbol{W}_N\boldsymbol{Y}_N + (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\Pi}\boldsymbol{\alpha} + (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{V}_N. \tag{2.4}$$

Let $\boldsymbol{Q} = (\boldsymbol{W}_N\boldsymbol{Y}_N, \boldsymbol{\Pi})$, and $\boldsymbol{\theta} = (\lambda, \boldsymbol{\alpha}^T)^T$. Applying the two stage least squares procedure proposed by [15], we proposed the estimator as shown below:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q})^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}_N, \tag{2.5}$$

where $\boldsymbol{M} = \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T$ and $\boldsymbol{H}$ is a matrix of instrumental variables. Therefore, we have $\widehat{\lambda} = \boldsymbol{\xi}^T\widehat{\boldsymbol{\theta}}$, where $\boldsymbol{\xi} = (1, 0, \cdots, 0)^T$ is a $(K_N + l + 2)$-dimensional column vector. We also can obtain an estimator of $g(z)$ by $\widehat{g}(z) = \boldsymbol{\Pi}(z)\widehat{\boldsymbol{\alpha}}$. Formula (2.5) leads to an estimator of $\boldsymbol{\gamma}$ as

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \cdots, \widehat{\gamma}_m)^T = (\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T(\boldsymbol{Y}_N - \boldsymbol{Q}\widehat{\boldsymbol{\theta}}).$$

Consequently, we can use $\widehat{\gamma}(t) = \sum_{k=1}^m \widehat{\gamma}_k\widehat{\phi}_k(t)$ as the estimator of $\gamma(t)$.

We often assume that the instrumental variables are given in the estimation procedure. In practice, we need to discuss that how to construct an appropriate instrumental variables matrix $\boldsymbol{H}$. According to the result of [18], the best instrumental variables matrix should be constructed as

$$\boldsymbol{H}^{\text{best}} = E(\boldsymbol{Q}|\boldsymbol{\Pi}, \boldsymbol{U}_N) = \Big(\boldsymbol{W}_N(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}(\boldsymbol{U}_N\boldsymbol{\gamma}_0 + \boldsymbol{\Pi}\boldsymbol{\alpha}_0), \boldsymbol{\Pi}\Big),$$

where $\lambda_0, \boldsymbol{\gamma}_0$ and $\boldsymbol{\alpha}_0$ are the corresponding true values. However, the true values $\lambda_0, \boldsymbol{\gamma}_0$ and $\boldsymbol{\alpha}_0$ are unknown. Therefore, we refer to the two-step iterative method suggested by [33] to obtain the initial estimators. In the first step, the following instrumental variables are obtained

$$\tilde{\boldsymbol{H}} = \Big(\boldsymbol{W}_N(\boldsymbol{I}-\tilde{\lambda}\boldsymbol{W}_N)^{-1}(\boldsymbol{U_N}\tilde{\boldsymbol{\gamma}}, \boldsymbol{\Pi}), \boldsymbol{\Pi}\Big),$$

where $\tilde{\lambda}$ and $\tilde{\boldsymbol{\gamma}}$ are obtained by simply regressing $\boldsymbol{Y}_N$ on pseudo regressor variables $\boldsymbol{W}_N\boldsymbol{Y}_N, \boldsymbol{U}_N$ and $\boldsymbol{\Pi}$. In the second step, we obtain the estimators $\bar{\lambda}$, $\bar{\boldsymbol{\gamma}}$ and $\bar{\boldsymbol{\theta}}$ by $\tilde{\boldsymbol{H}}$, and then construct the final instrumental variables as follows:

$$\boldsymbol{H} = \left(\boldsymbol{W}_N\left(\boldsymbol{I}-\bar{\lambda}\boldsymbol{W}_N\right)^{-1}(\boldsymbol{U_N}\bar{\boldsymbol{\gamma}} + \boldsymbol{\Pi}\bar{\boldsymbol{\alpha}}), \boldsymbol{\Pi}\right).$$

3. Asymptotic properties

In this section we discuss the asymptotic normality of $\widehat{\lambda}$ and the convergence rates of $\widehat{\gamma}(t)$ and $\widehat{g}(z)$ based on the following assumptions. Denote λ_0, $\gamma_0(\cdot)$ and $g_0(\cdot)$ be the true values of λ, $\gamma(\cdot)$ and $g(\cdot)$ respectively. Accordingly, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\alpha}_0$ are the true values of $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ respectively. The Euclidean norm is represented by $\|\cdot\|$ and $\|f(t)\|_2^2 = \int_{\mathscr{T}} f^2(t)\mathrm{d}t$ for all $f(t) \in L^2(\mathscr{T})$. It's worth noting that C denotes a positive constant that may be different at each appearance throughout this paper.

Assumption 3.1. *The matrix $\boldsymbol{I}-\lambda\boldsymbol{W}_N$ is nonsingular for any $\lambda \in (-\underline{d}_n, \overline{d}_n)$, where $0 < \underline{d}_n, \overline{d}_n < \infty$.*

Assumption 3.2. *The row and column sums of the matrices $\boldsymbol{W}_N$ and $(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}$ are bounded uniformly in absolute value.*

Assumption 3.3. *It is assumed that random function $X(t)$ and random variables U_k satisfy $E\|X(t)\|_2^4 < \infty$ and $E\left(U_k^4\right) \le C\tau_k^2$, $k \ge 1$.*

Assumption 3.4. *There exists some constants $a > 1$ and $b > a/2+1$ such that $C^{-1}j^{-a} \le \tau_j \le Cj^{-a}$, $\tau_j - \tau_{j+1} \ge Cj^{-a-1}$ and $|\gamma_{0j}| \le Cj^{-b}$, $j \ge 1$, where $\gamma_{0j} = \langle\gamma_0(t), \phi_j(t)\rangle$.*

Assumption 3.5. *$g(\cdot)$ is r times continuously differentiable on $[0, 1]$ for some $r \ge 2$.*

Assumption 3.6. *trace* $\left(\frac{K_N}{N}\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\right)^{-1}$ *is bounded in probability, where* $\tilde{\boldsymbol{Q}} = (\boldsymbol{W}_N(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)), \boldsymbol{\Pi}) = (\boldsymbol{S}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)), \boldsymbol{\Pi})$ *and*

$$\boldsymbol{\eta}_0 = \left(\int_0^1 \gamma_0(t)X_1(t)\mathrm{d}t, \cdots, \int_0^1 \gamma_0(t)X_N(t)\mathrm{d}t\right)^T.$$

Assumption 3.7. $\boldsymbol{\xi}^T\left(\frac{\Lambda^T\Lambda}{N}\right)^{-1}\boldsymbol{\xi}$ *converges to a positive constant in probability, where* $\Lambda^T = \tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}$.

Assumption 3.8. *For matrix* $\tilde{\boldsymbol{Q}} = (\boldsymbol{S}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)), \boldsymbol{\Pi})$, *there exists a constant* λ_c *such that* $\lambda_c\boldsymbol{I} - \frac{K_N}{N}\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}^T$ *is positive semidefinite matrix.*

Assumption 3.9. *The distribution of* Z *is absolutely continuous and its density is bounded away from zero and infinity on* $[0,1]$.

Remark 3.10. Assumption 3.1 defines the parameter space of λ as a general interval $(-\underline{d}_n, \overline{d}_n)$ around zero such that $\boldsymbol{I}-\lambda\boldsymbol{W}_N$ is nonsingular. In practical applications, $\boldsymbol{W}_N$ is often row-normalized such that $\sum_{j=1}^N \omega_{ij} = 1$ for $i = 1, \cdots, N$. In that case, it can be guaranteed that $\boldsymbol{I}-\lambda\boldsymbol{W}_N$ is nonsingular for any $\lambda \in (-1,1)$. For a general $\boldsymbol{W}_N$, we define the interval $(-\underline{d}_n, \overline{d}_n)$ to be a subset of $(-1/\mu_n, 1/\mu_n)$, where $\mu_n = \min\{\max_{1\le i\le N}\sum_{j=1}^N|\omega_{ij}|, \max_{1\le j\le N}\sum_{i=1}^N|\omega_{ij}|\}$. According to Lemmas 1 and 2 of [16], $\boldsymbol{I}-\lambda\boldsymbol{W}_N$ is nonsingular for any $\lambda \in (-1/\mu_n, 1/\mu_n)$. If Assumption 3.2 holds, there exists a constant C independent of N such that $C\boldsymbol{I}-\boldsymbol{S}\boldsymbol{S}^T$ is positive semi-definite. Assumptions 3.3-3.4 consist of regularity assumptions for functional data [10–12, 26]. For example, Assumption 3.4 ensures that the slope function $\gamma_0(t)$ is identifiable and it is sufficiently smooth relative to the covariance function. Assumptions 3.5-3.6 are required to realize the optimal convergence rate of $g(\cdot)$. Assumption 3.7 is used to represent the asymptotic variance of $\widehat{\lambda}$. Assumption 3.8 is required to ensure the identifiability of parameter $\boldsymbol{\theta}_0$. Assumption 3.9 requires a boundedness condition on the covariate, which is often assumed in asymptotic analysis of nonparametric regression problems (see Condition 1 of [28]).

The following Theorem 3.11 states the convergence rates of the estimators of the slope function and nonparametric function.

Theorem 3.11. *Suppose that Assumptions 3.1-3.9 hold,* $m \sim N^{1/(a+2b)}$, $K_N \sim N^{1/(2r+1)}$, *then*

$$\|\widehat{\gamma}(t)-\gamma_0(t)\|_2^2 = O_p\left(N^{-\frac{2b-1}{a+2b}}\right) + O_p\left(N^{-\frac{2r-1}{2r+1}} + N^{-\frac{2r}{2r+1}+\frac{1}{a+2b}}\right),$$
$$\|\widehat{g}(z)-g_0(z)\|_2^2 = O_p\left(N^{-\frac{2r}{2r+1}}\right) + O_p\left(N^{-\frac{a+2b-1}{a+2b}}\right).$$

Remark 3.12. If we assume that Assumptions 3.1-3.9 hold, and take $m \sim K_N \sim N^{1/(a+2b)} \sim N^{1/(2r+1)}$, then

$$\|\widehat{\gamma}(t)-\gamma_0(t)\|_2^2 = O_p\left(N^{-\frac{2b-1}{a+2b}}\right),$$

$$\|\widehat{g}(z)-g_0(z)\|_2^2 = O_p\left(N^{-\frac{2r}{2r+1}}\right).$$

We can see that the convergence rate of $\widehat{\gamma}(t)$ is the same as the rate established by [10], and it is optimal in the sense of minimax. Nonparametric function estimator $\widehat{g}(z)$ has the same optimal convergence rate established by [27].

Theorem 3.13. *Suppose that Assumptions 3.1-3.9 hold, $m \sim N^{1/(a+2b)} = o(K_N)$, $N/K_N^{2r+1} = o(1)$ and $K_N/N = o(1)$, then*

$$\sqrt{N}(\widehat{\lambda} - \lambda_0) \xrightarrow{D} N(0, \sigma^2 \varsigma),$$

where $\varsigma = \mathrm{plim}_{N\to\infty} \boldsymbol{\xi}^T \left(\frac{\Lambda^T\Lambda}{N}\right)^{-1} \frac{\Lambda^T(I-P)\Lambda}{N} \left(\frac{\Lambda^T\Lambda}{N}\right)^{-1} \boldsymbol{\xi}$ *and "$\xrightarrow{D}$" denotes convergence in distribution.*

4. Simulation study

4.1. Choosing the smoothing parameters

We know that the selection of the truncation parameter m and the knot number K_N is crucial. In general, we choose smoothing parameters based on some data-driven approaches, for example, Yao et al. [31] used CV and AIC criterions to determine the truncation parameter m, Ma [22] selected the knot number K_N for the B-spline basis according to BIC criterion. In this paper, we consider cubic splines (i.e., l=3) and employ the BIC and AIC criterions to choose the smoothing parameters as follows:

Specifically, we define

$$RSS = \sum_{i=1}^{N} \left(Y_i - \widehat{\lambda} \sum_{j=1}^{N} w_{ij} Y_j - \sum_{k=1}^{m} \widehat{\gamma}_k \langle X_i(t), \widehat{\phi}_k(t) \rangle - \sum_{j=1}^{K_N+l+1} \pi_j(Z_i) \widehat{\alpha}_j \right)^2,$$

where $\widehat{\lambda}$, $\widehat{\gamma}_k$ and $\widehat{\alpha}_j$ are the estimators in Section 2.

Method I : the truncation parameter m and the knot number K_N can be selected by the BIC criterion. Specifically, we minimize

$$\mathrm{BIC}(m, K_N) = \log \frac{RSS}{N} + \frac{\log(N)}{N} \{m + (K_N + l + 1)\}.$$

Method II : We choose the truncation parameter m and the knot number K_N in two steps:

Step 1. m is selected by the cumulative proportion of the principal component analysis. The cumulative proportion is defined as follows:

$$Q = \frac{\sum_{j=1}^{m} \widehat{\tau}_j}{\sum_{k=1}^{\infty} \widehat{\tau}_k}.$$

We choose m with a cumulative proportion of over 0.9.

Step 2. K_N is selected according to the BIC criterion. Specifically, we minimize

$$\mathrm{BIC}(K_N) = \log \frac{RSS}{N} + \frac{\log(N)}{N}(K_N + l + 1).$$

Method III : Similar to **Method** II, we choose m based on the cumulative proportion. The knot number K_N is selected by the AIC criterion. Specifically, we minimize

$$\mathrm{AIC}(K_N) = \log \frac{RSS}{N} + \frac{2}{N}(K_N + l + 1).$$

We will compare the performances of the proposed estimators based on the above three methods in simulations.

4.2. Analysis of simulation results

In this subsection, we will compare the differences in the selection of smoothing parameters by the above three methods. Further, we will evaluate the finite sample performances of the estimators given in the previous section by conducting 500 Monte Carlo simulation studies. The data $\{Y_i\}$ are generated from the following model

$$\boldsymbol{Y}_N = \lambda \boldsymbol{W}_N \boldsymbol{Y}_N + \int_0^1 \gamma(t) \boldsymbol{X}_N(t) \mathrm{d}t + g(\boldsymbol{Z}_N) + \boldsymbol{V}_N, \tag{4.1}$$

where $\boldsymbol{Z}_N = (Z_1, Z_2, \cdots, Z_N)^T$, and Z_i following the uniform distributions on $[0,1]$, $\gamma(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$ and $g(z) = 8(z-1/3)^2 - 1$, $V_N \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$. Similar to [2] and [19], we take $\boldsymbol{W}_N = \boldsymbol{I}_R \otimes \boldsymbol{B}_p$, where $\boldsymbol{B}_p = (\boldsymbol{l}_p \boldsymbol{l}_p^T - \boldsymbol{I}_p)/(p-1)$, $\boldsymbol{l}_p$ is the p-dimensional vector with all elements being 1 and $\otimes$ is a Kronecker product. That means we focus on the spatial scenario with R number of districts, where in each district, there are p members with each neighbor of a member giving equal weight. The sample size is $N = Rp$.

We assume the functional predictors can be expressed as $X_i(t) = \sum_{j=1}^{50} U_{ij}\phi_j(t)$, where $\phi_j(t) = \sqrt{2}\sin((j-0.5)\pi t)$ and U_{ij} are independently distributed as the normal with mean zero and variance $\tau_j = ((j-0.5)\pi)^{-2}$. In addition, we assume that the actual observation values are realized by $\{X_i(\cdot)\}$ at an equally spaced grid of 100 points in $[0,1]$.

The simulation studies are realized by different numerical values of R for 40 and 70, p for 3, 5, and 8, and σ^2 for 0.25 and 1. To compare the effects of different spatial dependence, we consider $\lambda = 0.2, 0.5, 0.8$ for the simulation studies, where $\lambda = 0.2$ represents the relatively weak spatial dependence, $\lambda = 0.5$ indicates a moderate degree of spatial dependence, and the relatively strong spatial dependence is represented by $\lambda = 0.8$.

Throughout the simulations, for scalar parameter λ, we evaluate the accuracy of the parameter estimator by average Bias and standard deviation (SD). We evaluate the performances of the estimators of the slope function $\gamma(t)$ and the nonparametric function by the square root of average squared errors (RASE). The RASE of $\widehat{\gamma}(t)$ is defined as

$$\mathrm{RASE}(\widehat{\gamma}(\mathrm{t})) = \left\{ \frac{1}{N_1} \sum_{q=1}^{N_1} [\widehat{\gamma}(t_q) - \gamma(t_q)]^2 \right\}^{1/2}.$$

In our simulation, $N_1 = 200$ and $\{t_q, q = 1, \cdots, N_1\}$ are the regular grid points at which the function $\widehat{\gamma}(t)$ is evaluated. The RASE of $\widehat{g}(z)$ is defined as

$$\mathrm{RASE}(\widehat{g}(z)) = \left\{ \frac{1}{N_2} \sum_{l=1}^{N_2} [\widehat{g}(z_l) - g(z_l)]^2 \right\}^{1/2},$$

where $\{z_l, l = 1, \cdots, N_2\}$ are grid points which are chosen to be equally spaced in $[0,1]$ and $N_2 = 200$ is used. To further evaluate the performances of the estimators of $\gamma(t)$ and $g(z)$, we consider the SDs of $\widehat{\gamma}(t)^*$ and $\widehat{g}(z)^*$, where $\widehat{\gamma}(t)^*$ and $\widehat{g}(z)^*$ represent square root of average squared errors of $\widehat{\gamma}(t)$ and $\widehat{g}(z)$ respectively.

The simulation results based the three methods of smoothing parameters choosing are listed in Tables 1–3. Tables 1–3 listed the average Bias, average RASEs of $\widehat{\gamma}(t)$ and $\widehat{g}(z)$, and SDs of $\widehat{\lambda}$, $\widehat{\gamma}(t)^*$ and $\widehat{g}(z)^*$. Comparing the simulation results of the three model selection methods, we find that the performances of the proposed estimators are similar.

The following three conclusions can be drawn from the simulation results: (1) The average Bias of $\widehat{\lambda}$ is small in all simulations, in other words, the parameter estimator is nearly unbiased. (2) The average RASEs of $\widehat{\gamma}(t)$ and $\widehat{g}(z)$ are small for all cases and decrease as N increases or σ^2 decreases, and it can be concluded that the estimate curves fit well to the corresponding true curves. (3) The SD of $\widehat{\lambda}$ decreases as R increases or σ^2 decreases, and the SDs of $\widehat{\gamma}(t)^*$ and $\widehat{g}(z)^*$ decrease as N increases or σ^2 decreases. Figure 1 shows the simulation effect when the spatial effect coefficient and sample size are the smallest but the variance is the largest in the case of the smooth parameters selected by **Method** II. It can be observed that the estimate curves approximate the true curves. Therefore, the simulation results show that the proposed estimation method is effective.

Table 1. Simulation results for $\lambda = 0.2$.

			$\sigma^2 = 0.25$						$\sigma^2 = 1$					
			Bias	RASE		SD			Bias	RASE		SD		
Method	R	p	$\hat{\lambda}$	$\hat{\gamma}(t)$	$\hat{g}(z)$	$\hat{\lambda}$	$\hat{\gamma}(t)^*$	$\hat{g}(z)^*$	$\hat{\lambda}$	$\hat{\gamma}(t)$	$\hat{g}(z)$	$\hat{\lambda}$	$\hat{\gamma}(t)^*$	$\hat{g}(z)^*$
I	40	3	-8.8×10^{-4}	0.372	0.127	0.028	0.151	0.060	-0.010	0.630	0.240	0.094	0.421	0.073
		5	-8.5×10^{-4}	0.276	0.117	0.029	0.087	0.056	-0.008	0.462	0.197	0.092	0.290	0.061
		8	7.7×10^{-5}	0.216	0.113	0.033	0.073	0.047	-0.007	0.363	0.170	0.100	0.255	0.046
	70	3	-1.6×10^{-4}	0.269	0.119	0.018	0.094	0.054	-0.003	0.441	0.197	0.066	0.256	0.059
		5	8.4×10^{-4}	0.210	0.112	0.022	0.071	0.046	-9.0×10^{-4}	0.339	0.164	0.067	0.178	0.043
		8	-6.9×10^{-4}	0.166	0.110	0.023	0.049	0.040	-0.008	0.262	0.144	0.069	0.123	0.038
II	40	3	-1.2×10^{-3}	0.361	0.127	0.028	0.122	0.060	-0.010	0.617	0.239	0.094	0.292	0.074
		5	-9.0×10^{-4}	0.277	0.117	0.029	0.088	0.056	-0.008	0.476	0.197	0.092	0.238	0.060
		8	-9.3×10^{-5}	0.218	0.113	0.033	0.069	0.047	-0.008	0.377	0.170	0.100	0.171	0.046
	70	3	-2.1×10^{-4}	0.266	0.119	0.018	0.084	0.054	-0.004	0.467	0.197	0.066	0.221	0.059
		5	7.6×10^{-4}	0.209	0.112	0.022	0.066	0.046	-1.1×10^{-3}	0.351	0.164	0.067	0.166	0.043
		8	-7.8×10^{-4}	0.168	0.110	0.023	0.050	0.040	-0.008	0.276	0.145	0.069	0.120	0.038
III	40	3	-1.1×10^{-3}	0.361	0.129	0.028	0.123	0.059	-0.008	0.619	0.255	0.093	0.296	0.080
		5	-7.6×10^{-4}	0.277	0.119	0.029	0.088	0.055	-0.007	0.477	0.210	0.091	0.238	0.064
		8	6.5×10^{-5}	0.218	0.115	0.033	0.069	0.047	-0.007	0.377	0.184	0.099	0.170	0.047
	70	3	-1.6×10^{-4}	0.266	0.121	0.018	0.084	0.053	-0.004	0.467	0.210	0.066	0.221	0.062
		5	8.7×10^{-4}	0.209	0.113	0.022	0.066	0.045	-6.8×10^{-4}	0.352	0.174	0.067	0.166	0.046
		8	-6.4×10^{-4}	0.168	0.111	0.023	0.050	0.040	-0.008	0.276	0.152	0.069	0.120	0.041

Table 2. Simulation results for $\lambda = 0.5$.

			$\sigma^2 = 0.25$						$\sigma^2 = 1$					
			Bias	RASE		SD			Bias	RASE		SD		
Method	R	p	$\hat{\lambda}$	$\hat{\gamma}(t)$	$\hat{g}(z)$	$\hat{\lambda}$	$\hat{\gamma}(t)^*$	$\hat{g}(z)^*$	$\hat{\lambda}$	$\hat{\gamma}(t)$	$\hat{g}(z)$	$\hat{\lambda}$	$\hat{\gamma}(t)^*$	$\hat{g}(z)^*$
I	40	3	-5.7×10^{-4}	0.373	0.127	0.020	0.152	0.060	-0.007	0.656	0.241	0.067	0.468	0.074
		5	-6.1×10^{-4}	0.275	0.117	0.019	0.086	0.056	-0.006	0.461	0.198	0.062	0.283	0.061
		8	2.4×10^{-5}	0.216	0.113	0.021	0.073	0.047	-0.005	0.368	0.170	0.065	0.263	0.046
	70	3	-1.3×10^{-4}	0.269	0.119	0.013	0.093	0.053	-0.003	0.450	0.198	0.047	0.268	0.060
		5	5.5×10^{-4}	0.210	0.112	0.015	0.071	0.046	-7.8×10^{-4}	0.343	0.165	0.045	0.185	0.043
		8	-4.7×10^{-4}	0.166	0.110	0.015	0.049	0.040	-0.006	0.262	0.145	0.045	0.123	0.039
II	40	3	-9.4×10^{-4}	0.361	0.127	0.020	0.122	0.060	-0.008	0.621	0.241	0.068	0.297	0.074
		5	-6.5×10^{-4}	0.277	0.117	0.019	0.088	0.056	-0.006	0.478	0.197	0.062	0.238	0.061
		8	-9.7×10^{-5}	0.218	0.113	0.021	0.069	0.047	-0.005	0.377	0.170	0.065	0.172	0.046
	70	3	-1.9×10^{-4}	0.266	0.119	0.013	0.084	0.054	-0.003	0.469	0.198	0.047	0.221	0.060
		5	4.8×10^{-4}	0.209	0.112	0.015	0.066	0.046	-9.3×10^{-4}	0.352	0.165	0.045	0.167	0.043
		8	-5.2×10^{-4}	0.168	0.110	0.015	0.050	0.040	-0.006	0.276	0.145	0.045	0.120	0.039
III	40	3	-7.9×10^{-4}	0.360	0.129	0.020	0.123	0.059	-0.006	0.624	0.255	0.067	0.301	0.079
		5	-5.3×10^{-4}	0.277	0.119	0.019	0.088	0.055	-0.005	0.479	0.211	0.061	0.239	0.064
		8	2.1×10^{-5}	0.218	0.115	0.022	0.069	0.047	-0.004	0.377	0.184	0.064	0.171	0.047
	70	3	-1.0×10^{-4}	0.266	0.121	0.013	0.084	0.053	-0.002	0.470	0.210	0.047	0.221	0.062
		5	5.7×10^{-4}	0.209	0.113	0.015	0.066	0.045	-3.2×10^{-4}	0.353	0.174	0.045	0.168	0.046
		8	-4.3×10^{-4}	0.168	0.111	0.015	0.050	0.040	-0.005	0.276	0.153	0.045	0.120	0.041

Table 3. Simulation results for $\lambda = 0.8$.

			$\sigma^2 = 0.25$						$\sigma^2 = 1$					
			Bias	RASE		SD			Bias	RASE		SD		
Method	R	p	$\hat{\lambda}$	$\hat{\gamma}(t)$	$\hat{g}(z)$	$\hat{\lambda}$	$\hat{\gamma}(t)^*$	$\hat{g}(z)^*$	$\hat{\lambda}$	$\hat{\gamma}(t)$	$\hat{g}(z)$	$\hat{\lambda}$	$\hat{\gamma}(t)^*$	$\hat{g}(z)^*$
I	40	3	-2.7×10^{-4}	0.373	0.127	0.009	0.148	0.060	-0.003	0.667	0.243	0.030	0.468	0.074
		5	-2.7×10^{-4}	0.276	0.118	0.008	0.087	0.056	-0.002	0.461	0.198	0.027	0.245	0.061
		8	-4.0×10^{-6}	0.216	0.114	0.009	0.073	0.047	-0.002	0.370	0.170	0.027	0.265	0.045
	70	3	-6.9×10^{-5}	0.269	0.119	0.006	0.093	0.053	-0.001	0.471	0.199	0.021	0.305	0.060
		5	2.2×10^{-4}	0.210	0.112	0.006	0.071	0.046	-3.8×10^{-4}	0.348	0.165	0.019	0.187	0.044
		8	-2.0×10^{-4}	0.166	0.110	0.006	0.049	0.040	-0.002	0.264	0.145	0.019	0.124	0.039
II	40	3	-4.6×10^{-4}	0.361	0.127	0.009	0.123	0.060	-0.004	0.629	0.243	0.031	0.305	0.075
		5	-2.9×10^{-4}	0.277	0.118	0.008	0.088	0.056	-0.003	0.481	0.198	0.027	0.237	0.061
		8	-5.5×10^{-5}	0.218	0.113	0.009	0.069	0.047	-0.002	0.379	0.171	0.027	0.173	0.046
	70	3	-1.0×10^{-4}	0.266	0.119	0.006	0.084	0.053	-0.002	0.474	0.199	0.021	0.222	0.060
		5	1.9×10^{-4}	0.209	0.112	0.006	0.066	0.046	-5.0×10^{-4}	0.354	0.165	0.019	0.169	0.044
		8	-2.2×10^{-4}	0.168	0.110	0.006	0.050	0.040	-0.002	0.277	0.145	0.019	0.121	0.039
III	40	3	-3.6×10^{-4}	0.361	0.129	0.009	0.123	0.059	-0.003	0.632	0.259	0.030	0.308	0.080
		5	-2.4×10^{-4}	0.277	0.119	0.008	0.088	0.056	-0.002	0.483	0.212	0.026	0.239	0.063
		8	-6.2×10^{-6}	0.218	0.115	0.009	0.069	0.047	-0.002	0.379	0.184	0.027	0.172	0.047
	70	3	-5.5×10^{-5}	0.266	0.121	0.006	0.084	0.053	-0.001	0.474	0.212	0.021	0.223	0.062
		5	2.3×10^{-4}	0.209	0.114	0.006	0.066	0.045	-1.6×10^{-4}	0.355	0.174	0.019	0.169	0.046
		8	-1.8×10^{-4}	0.168	0.111	0.006	0.050	0.040	-0.002	0.277	0.153	0.019	0.121	0.040

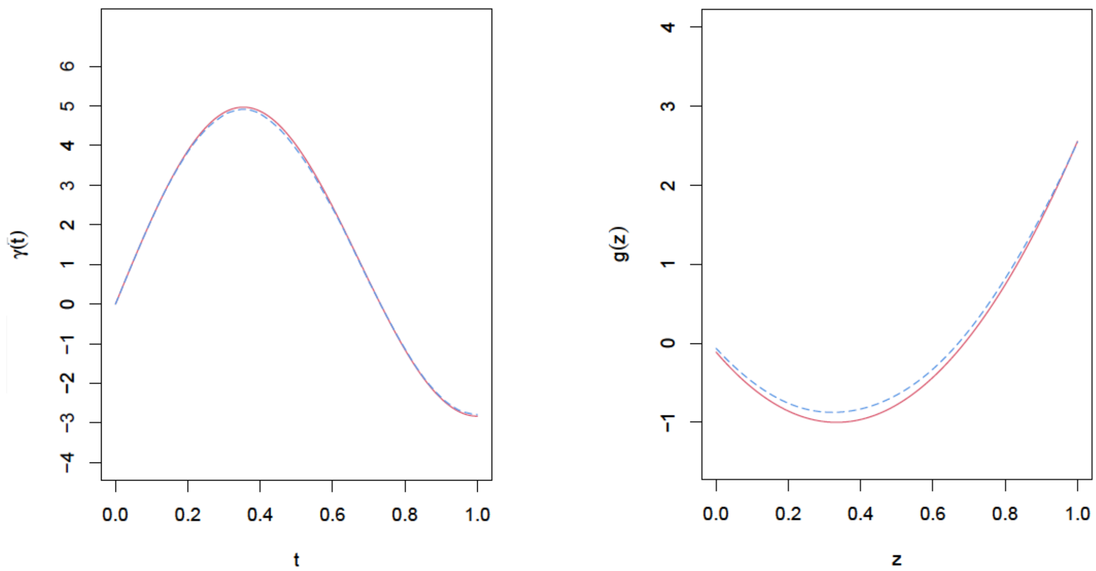

Figure 1. Simulation result of $\widehat{\gamma}(t)$ and $\widehat{g}(z)$ when $\lambda = 0.2, R = 40, p = 3, \sigma^2 = 1$. The solid curve denotes the true curve, the dash curve denotes its estimate.

4.3. Comparative study of spatial data processing with spatial model and non-spatial model

We will study the results of deliberately ignoring the inherent spatial structure of the spatially dependent data. Specifically, the data is still generated as described in Section 4.2 but the following model is considered

$$\boldsymbol{Y}_i = \int_0^1 \gamma(t)\boldsymbol{X}_i(t)\mathrm{d}t + g(\boldsymbol{Z}_i) + \boldsymbol{V}_i, \tag{4.2}$$

which is a functional partially linear model.

Same as the previous estimation process, we use FPCA to deal with functional data, and for nonparametric function, we approximate it by linear combinations of B-spline basis. Then, we have

$$\boldsymbol{Y}_N \approx \boldsymbol{U}_N\boldsymbol{\gamma} + \boldsymbol{\Pi}\boldsymbol{\alpha} + \boldsymbol{V}_N.$$

Because there is no endogeneity, ordinary least squares method can be used directly. Then, we have $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{\Pi}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}_N$ and $\widehat{\boldsymbol{\gamma}} = (\boldsymbol{U_N}^T\boldsymbol{U_N})^{-1}\boldsymbol{U_N}^T(\boldsymbol{Y}_N - \boldsymbol{\Pi}\widehat{\boldsymbol{\alpha}})$. As we discussed before, the results of the three methods of smoothing parameters choosing are not significantly different. Therefore, we select the smoothing parameters according to **Method** II directly.

Table 4 shows the results of 500 Monte Carlo simulations. Compared with Tables 1–3 we can see that, when the spatial structure is ignored, the RASEs of $\widehat{\gamma}(t)$ and $\widehat{g}(z)$ become larger and the SDs of $\widehat{\gamma}(t)^*$ and $\widehat{g}(z)^*$ increase, which also coincides with what was discovered from Figure 2 and Figure 3. Moreover, as the spatial dependence increases, the difference between the two models in terms of RASEs and SDs increases rapidly. These results demonstrate that, compared with the traditional econometric model, the spatial autoregressive model can effectively solve the spatial dependence.

Table 4. Simulation results of the functional partially linear model (4.2).

			$\lambda = 0.2$				$\lambda = 0.5$				$\lambda = 0.8$			
			RASE		SD		RASE		SD		RASE		SD	
σ^2	R	p	$\widehat{\gamma}(t)$	$\widehat{g}(z)$	$\widehat{\gamma}(t)^*$	$\widehat{g}(z)^*$	$\widehat{\gamma}(t)$	$\widehat{g}(z)$	$\widehat{\gamma}(t)^*$	$\widehat{g}(z)^*$	$\widehat{\gamma}(t)$	$\widehat{g}(z)$	$\widehat{\gamma}(t)^*$	$\widehat{g}(z)^*$
0.25	40	3	0.397	0.135	0.136	0.058	0.839	0.265	0.343	0.080	3.823	1.190	1.351	0.299
		5	0.292	0.121	0.090	0.055	0.507	0.184	0.188	0.060	2.246	0.725	0.767	0.195
		8	0.228	0.115	0.072	0.047	0.357	0.145	0.132	0.045	1.454	0.463	0.521	0.126
	70	3	0.298	0.127	0.096	0.052	0.760	0.252	0.277	0.067	3.793	1.168	1.028	0.236
		5	0.223	0.114	0.068	0.046	0.450	0.171	0.158	0.046	2.187	0.701	0.647	0.153
		8	0.176	0.111	0.052	0.040	0.292	0.135	0.101	0.038	1.337	0.435	0.400	0.095
1	40	3	0.643	0.247	0.315	0.075	1.042	0.367	0.550	0.119	4.050	1.300	1.709	0.390
		5	0.487	0.200	0.245	0.060	0.671	0.259	0.342	0.079	2.381	0.807	0.995	0.236
		8	0.387	0.171	0.178	0.046	0.494	0.201	0.255	0.054	1.560	0.520	0.707	0.152
	70	3	0.493	0.204	0.240	0.061	0.897	0.319	0.433	0.099	3.914	1.241	1.347	0.318
		5	0.361	0.166	0.175	0.044	0.559	0.218	0.268	0.061	2.285	0.745	0.824	0.196
		8	0.283	0.145	0.125	0.038	0.377	0.166	0.181	0.045	1.381	0.458	0.525	0.120

5. Real data analysis

In this section, we will validate the procedure proposed in this paper by analyzing an econometric dataset. The data originate from the National Bureau of Statistics of the People's Republic of China and are collected according to the 30 provincial administrative regions in hinterland of China, excluding Tibet. The used variables are described in Table 5. Specifically, DOP, which is defined as the ratio of total import and export of goods to GDP, is observed from 1995 to 2015, whereas all other variables are observed in 2015. ICP is defined as the ratio of industrial added value to carbon dioxide emissions. ER is the ratio of pollutant charge to GDP. The proposed functional partially linear spatial autoregressive model is used to analyze the data, where ICP is the response variable, DOP is the functional variable and ER is the real-valued explanatory variable.

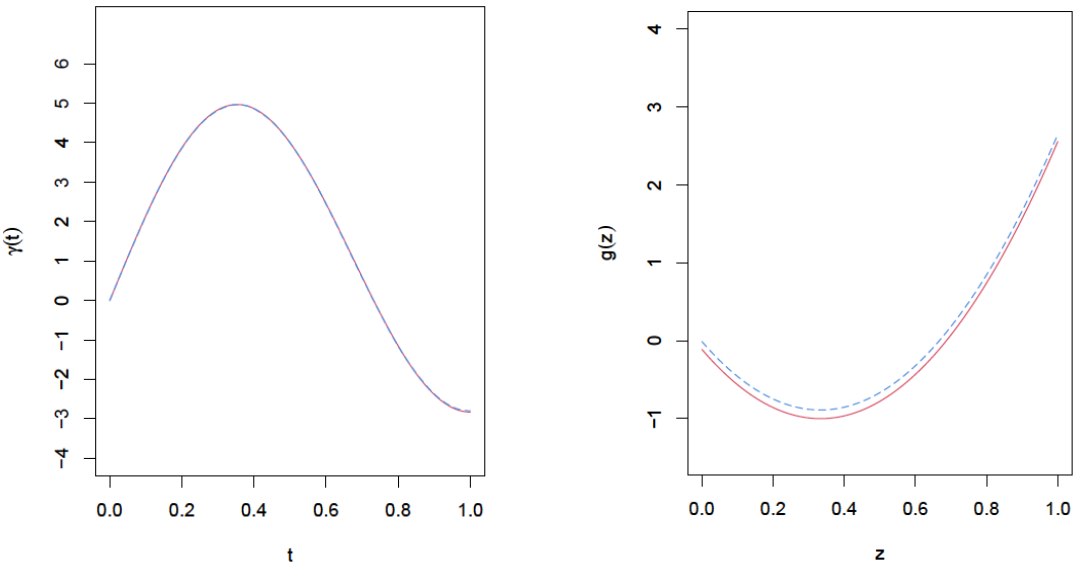

Figure 2. Simulation result of the functional partially linear spatial autoregressive model when $\lambda = 0.8, R = 70, p = 8, \sigma^2 = 0.25$.

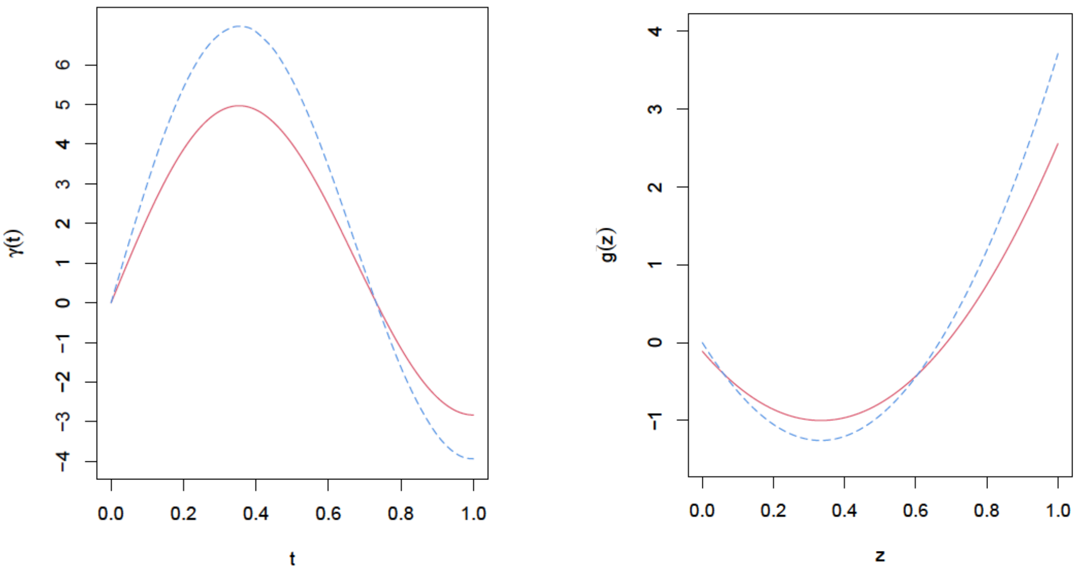

Figure 3. Simulation result of the functional partially linear model (4.2) when $\lambda = 0.8, R = 70, p = 8, \sigma^2 = 0.25$.

Table 5. The variables used in the data analysis.

Variable	Description
ICP	Industrial carbon productivity
DOP	Degree of opening to the outside worldP
ER	Environmental regulation

Similar to [21], we use Moran test to verify whether the dependent variable has spatial correlation. The Moran's I statistic is defined as follows:

$$Moran's\ I = \frac{N \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij}) \sum_{j=1}^{N} (Y_j - \bar{Y})^2},$$

where Y_i and Y_j represent the index value of province i and j respectively, N is the number of provinces, ω_{ij} is the (i, j) element of the spatial weight matrix $\boldsymbol{W}_N$, representing the connection relation between provinces. Here, we use geographical distance to construct the spatial weight matrix $\boldsymbol{W}_N$, and the row of $\boldsymbol{W}_N$ is standardized. After the calculation, Moran's I statistic is 0.196 and p-value is 0.009. Therefore, it is necessary to construct a spatial autoregressive model.

Initially, we use B-splines basis to extend the discrete data of DOP to smooth curves, as shown in Figure 4.

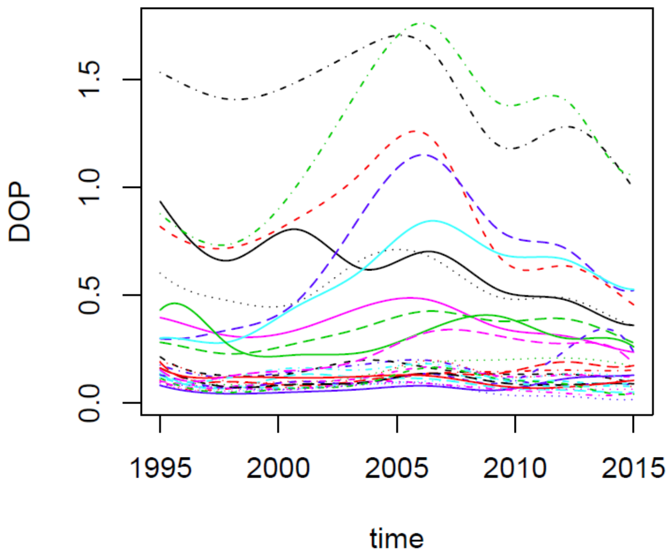

Figure 4. Curves of DOP for 30 provinces in mainland China.

According to the estimation approach proposed in this paper, we obtain the estimated value of the spatial coefficient λ is 0.62, which is consistent with the result of Moran test. And the estimated curves of slope function $\gamma(t)$ and non-parametric function $g(z)$ are shown in Figure 5.

From the left panel of Figure 5, we found that the impact of DOP on ICP gradually decreases over time. In the mid-1990s, due to the relatively low productivity level in China, the impact of DOP on ICP was very significant. With the continuous improvement of China's productivity level, the impact of DOP on ICP is gradually decreasing, which is also a foreseeable result. Due to the global economic crisis around 2010, we can also observe that the impact of DOP on ICP reached its lowest point at that time. From the right panel of Figure 5, we can find that ER and ICP do not always show an reverse relationship. Appropriate environmental regulations can promote industrial enterprises to optimize resource allocation, thereby improving production efficiency. Therefore, how to formulate reasonable environmental regulatory rules is an important issue.

Next, we will compare some models to prove the effectiveness of our proposed model. We use the average residual sum of squares (ARSS) to evaluate the performances of the models, which is defined as

$$ARSS = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2,$$

where $\hat{Y}_i$ is corresponding fitting values. The four models and their corresponding ARSS values are shown in Table 6.

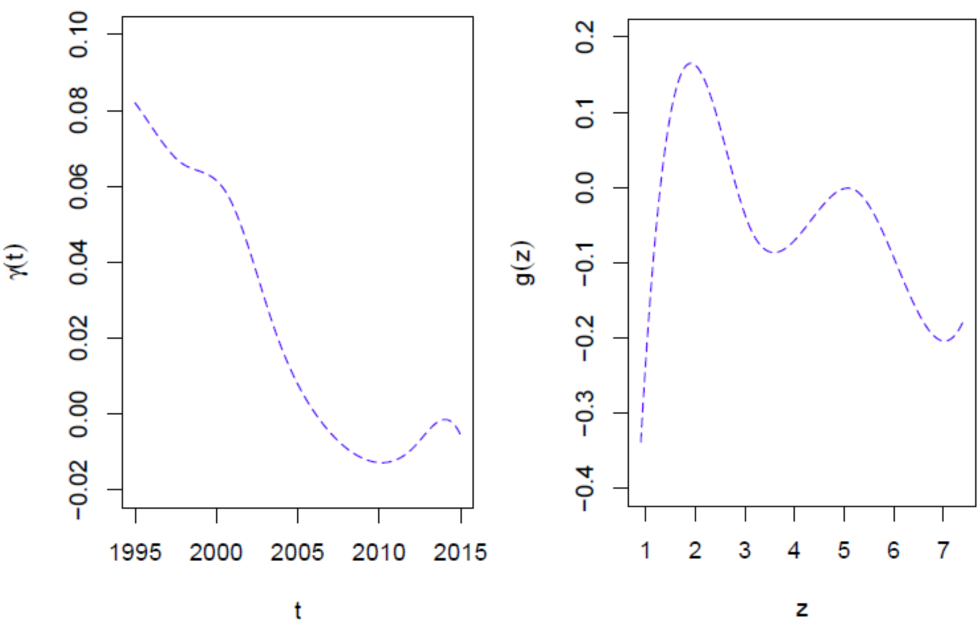

Figure 5. The estimated curves of $\gamma(t)$ and $g(z)$.

Table 6. The ARSSs for different models.

Model	ARSS
$Y_i = \lambda \sum_{j=1}^{30} w_{ij} Y_j + g(Z_i) + V_i$	0.023
$Y_i = \int_{1995}^{2015} \gamma(t) X_i(t) \mathrm{d}t + g(Z_i) + V_i$	0.037
$Y_i = \lambda \sum_{j=1}^{30} w_{ij} Y_j + \int_{1995}^{2015} \gamma(t) X_i(t) \mathrm{d}t + V_i$	0.025
$Y_i = \lambda \sum_{j=1}^{30} w_{ij} Y_j + \int_{1995}^{2015} \gamma(t) X_i(t) \mathrm{d}t + g(Z_i) + V_i$	0.013

From Table 6, we can verify that it is useful to analyze industrial carbon productivity by considering the spatially dependent structures and the nonlinear effects of environmental regulation.

6. Conclusions

In this paper, we propose the estimators of a functional partially linear spatial autoregressive model, where the estimation method mainly relies on instrumental variables and two-stage least squares method. The slope function and nonparametric function are approximated by functional principal component analysis and B-spline basis respectively, and the theoretical properties of the resulting estimators are established under some mild conditions. The simulation and real data analysis show that the proposed estimation method is effective.

There are many interesting directions for future research. In this paper, we only consider the estimators of the spatial coefficient, slope function, and nonparametric function, but do not take into account another important aspect of statistical analysis, which is testing the effects of predictors. In the future, we hope to be able to identify model structures by testing the major effects of the scalar predictor and functional predictor. Another interesting question is to find some robust estimators for the proposed model.

Acknowledgements

We would like to thank the editor, the associate editor and the two anonymous referees for their constructive comments and suggestions that significantly improved the paper.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. This work is supported by the National Social Science Foundation of China (23BTJ061) and the Humanities and Social Science Project of Ministry of Education of China (21YJC910003).

Data availability. The data used in this paper are from the National Bureau of Statistics of the People's Republic of China.

References

[1] L. Anselin, *Spatial Econometrics: Methods and Models*, Kluwer, Dordrecht, NL, 1988.
[2] A.C. Case, *Spatial patterns in household demand*, Econometrica **59** (4), 953-965, 1991.
[3] J.M. Chiou and H.G. Müller, *Linear manifold modelling of multivariate functional data*, J. R. Stat. Soc. B **76** (3), 605-626, 2014.
[4] A. Cliff and J.K. Ord, *Spatial Autocorrelation*, Pion, London, UK, 1973.
[5] N.A. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, USA, 1993.
[6] J. Du, X.Q. Sun, R.Y. Cao and Z.Z. Zhang, *Statistical inference for partially linear additive spatial autoregressive models*, Spat. Stat. **25**, 52-67, 2018.
[7] Y.Y. Fan, G.M. James and P. Radchenko, *Functional additive regression*, Ann. Stat. **43** (5), 2296-2325, 2015.
[8] S.Y. Feng and L.G. Xue, *Partially functional linear varying coefficient model*, Statistics **50** (4), 717-732, 2016.
[9] P. Hall and G. Hooker, *Truncated linear models for functional data*, J. R. Stat. Soc. B **78** (3), 637-653, 2016.
[10] P. Hall and J.L. Horowitz, *Methodology and convergence rates for functional linear regression*, Ann. Stat. **35** (1), 70-91, 2007.
[11] Y.P. Hu, S.Y. Wu, S.Y. Feng and J.L. Jin, *Estimation in partial functional linear spatial autoregressive model*, Mathematics **8** (10), Doi: 10.3390/math8101680, 2020.
[12] Y.P. Hu, L.G. Xue, J. Zhao and L.Y. Zhang, *Skew-normal partial functional linear model and homogeneity test*, J. Stat. Plan. Infer. **204**, 116-127, 2020.
[13] J. Huang, J.L. Horowitz and F. R. Wei, *Variable selection in nonparametric additive models*, Ann. Stat. **38** (4), 2282-2313, 2010.
[14] J.Z. Huang and H. Shen, *Functional coefficient regression models for non-linear time series: a polynomial spline approach*, Scand. J. Stat. **31** (4), 515-534, 2004.
[15] H.H. Kelejian and I.R. Prucha, *A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances*, J. Real. Estate. Finance **17** (1), 99-121, 1998.

[16] H.H. Kelejian and I.R. Prucha, *Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances*, J. Econometrics **157** (1), 53-67, 2010.
[17] D. Kong, K.J. Xue, F. Yao and H.H. Zhang, *Partially functional linear regression in high dimensions*, Biometrika **103** (1), 147-159, 2016.
[18] L.F. Lee, *Best spatial two-tage least squares estimators for a spatial autoregressive model with autoregressive disturbances*, Economet. Rev. **22** (4), 307-335, 2003.
[19] L.F. Lee, *Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models*, Econometrica **72** (6), 1899-1925, 2004.
[20] H. Lian, *Functional partial linear model*, J. Nonparametr. Stat. **23** (1), 115-128, 2011.
[21] R.Y. Long, T.X. Shao and H. Chen, *Spatial econometric analysis of China's province-level industrial carbon productivity and its influencing factors*, Appl. Energ. **166**, 210-219, 2016.
[22] S.J. Ma, *Estimation and inference in functional single-index models*, Ann. I. Stat. Math. **68** (1), 181-208, 2016.
[23] E. Malikov and Y. Sun, *Semiparametric estimation and testing of smooth coefficient spatial autoregressive models*, J. Econometrics **199** (1), 12-34, 2017.
[24] J.O. Ramsay and B.W. Silverman, *Functional Data Analysis*, Springer, New York, USA, 1997.
[25] L.L. Schumaker, *Spline Function: Basic theory*, Wiley, New York, USA, 1981.
[26] H. Shin, *Partial functional linear regression*, J. Stat. Plan. Infer. **139** (10), 3405-3418, 2009.
[27] C.J. Stone, *Optimal rates of convergence for nonparametric estimators*, Ann. Stat. **8** (6), 1348-1360, 1980.
[28] C.J. Stone, *Additive regression and other nonparametric models*, Ann. Stat. **13** (2), 689-705, 1985.
[29] L.J. Su and S.N. Jin, *Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models*, J. Econometrics **157** (1), 18-33, 2010.
[30] Q.G. Tang, *Estimation for semi-functional linear regression*, Statistics **49** (6), 1262-1278, 2015.
[31] F. Yao, H.G. Müller and J.L. Wang, *Functional data analysis for sparse longitudinal data*, J. Am. Stat. Assoc. **100** (470), 577-590, 2005.
[32] P. Yu, J. Du and Z.Z. Zhang, *Single-index partially functional linear regression model*, Stat. Pap. **61** (3), 1107-1123, 2020.
[33] Y.Q. Zhang and D.M. Shen, *Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects*, J. Stat. Plan. Infer. **159**, 64-80, 2015.
[34] J.J. Zhou and M. Chen, *Spline estimators for semi-functional linear model*, Stat. Probabil. Lett. **82** (3), 505-513, 2012.

Appendix

Lemma .1. *Under the conditions of Theorem 3.11, one has*

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 = O_p\left(\frac{K_N(K_N+m)}{N}\right).$$

Proof. Let $\boldsymbol{e}_N = \boldsymbol{\eta}_0 - \boldsymbol{U}_N\boldsymbol{\gamma}_0$ and $\boldsymbol{\varepsilon}_N = g_0(\boldsymbol{Z}_N) - \boldsymbol{\Pi}\boldsymbol{\alpha}_0$, then

$$\begin{aligned}
\boldsymbol{Y}_N &= \lambda_0\boldsymbol{W}_N\boldsymbol{Y}_N + \boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N) + \boldsymbol{V}_N \\
&= \lambda_0\boldsymbol{W}_N\boldsymbol{Y}_N + \boldsymbol{e}_N + \boldsymbol{U}_N\boldsymbol{\gamma}_0 + \boldsymbol{\varepsilon}_N + \boldsymbol{\Pi}\boldsymbol{\alpha}_0 + \boldsymbol{V}_N \\
&= \boldsymbol{Q}\boldsymbol{\theta}_0 + \boldsymbol{U}_N\boldsymbol{\gamma}_0 + \boldsymbol{e}_N + \boldsymbol{\varepsilon}_N + \boldsymbol{V}_N.
\end{aligned}$$

By the definition of $\widehat{\boldsymbol{\theta}}$, we have

$$\begin{aligned}\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0 &= (\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q})^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}_N-\boldsymbol{\theta}_0\\ &= (\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q})^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\left[(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}_N\right]-\boldsymbol{\theta}_0\\ &= (\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q})^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})(\boldsymbol{Q}\boldsymbol{\theta}_0+\boldsymbol{e}_N+\boldsymbol{\varepsilon}_N+\boldsymbol{V}_N)-\boldsymbol{\theta}_0\\ &= (\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q})^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})(\boldsymbol{e}_N+\boldsymbol{\varepsilon}_N+\boldsymbol{V}_N).\end{aligned}$$

First, consider $\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q}$. Recall that $\boldsymbol{Y}_N=(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)+\boldsymbol{V}_N)$, it has

$$\begin{aligned}\boldsymbol{Q} &= (\boldsymbol{W}_N\boldsymbol{Y}_N,\boldsymbol{\Pi})\\ &= (\boldsymbol{W}_N(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)+\boldsymbol{V}_N),\boldsymbol{\Pi})\\ &= (\boldsymbol{W}_N(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)),\boldsymbol{\Pi})+(\boldsymbol{W}_N(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}\boldsymbol{V}_N,\boldsymbol{0})\\ &\triangleq \tilde{\boldsymbol{Q}}+\tilde{\boldsymbol{e}},\end{aligned}$$

where $\tilde{\boldsymbol{Q}}=(\boldsymbol{S}(\boldsymbol{\eta}_0+g_0(\boldsymbol{Z}_N)),\boldsymbol{\Pi})$, $\tilde{\boldsymbol{e}}=(\boldsymbol{S}\boldsymbol{V}_N,\boldsymbol{0})$, $\boldsymbol{S}=\boldsymbol{W}_N(\boldsymbol{I}-\lambda_0\boldsymbol{W}_N)^{-1}$. Hence, one has

$$\begin{aligned}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q} &= \tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}+\tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{e}}\\ &\quad+\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{e}}+\tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\\ &\triangleq R_{11}+R_{12}+R_{13}+R_{14},\end{aligned}$$

where

$$\begin{aligned}R_{11} &= \tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}},\\ R_{12} &= \tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{e}},\\ R_{13} &= \tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{e}},\\ R_{14} &= \tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}.\end{aligned}$$

By the properties of projection matrix and Assumption 3.2, we have

$$\begin{aligned}&E\left[\boldsymbol{V}_N{}^T\boldsymbol{S}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N\right]\\ =&E\left[trace\left\{\boldsymbol{V}_N{}^T\boldsymbol{S}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N\right\}\right]\\ =&E\left[trace\left\{\boldsymbol{V}_N{}^T\boldsymbol{S}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N\right\}\right]\\ =&E\left[trace\left\{(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\boldsymbol{H}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N\boldsymbol{V}_N{}^T\boldsymbol{S}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\right\}\right]\\ \leq&C\sigma^2E\left[trace\left\{(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\boldsymbol{H}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\right\}\right]\\ \leq&C\sigma^2E\left[trace\left\{(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\boldsymbol{H}^T\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\right\}\right]\\ =&O(K_N).\end{aligned}$$

Hence, we have

$$\boldsymbol{V}_N{}^T\boldsymbol{S}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N=O_p(K_N).$$

Thus, $R_{12} = O_p(K_N)$. By straightforward algebra, one has $E(R_{13}) = 0$. In addition, based on Assumption 3.2 and Assumption 3.8, we have

$$\begin{aligned}
&E\left(\left\|\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N\right\|^2\right)\\
=&E\left[trace\left\{\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{S}\boldsymbol{V}_N\boldsymbol{V}_N{}^T\boldsymbol{S}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\right\}\right]\\
\le&C\sigma^2E\left[trace\left\{\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\right\}\right]\\
\le&C\sigma^2E\left[trace\left\{\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\right\}\right]\\
\le&C\sigma^2E\left[trace\left\{(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\boldsymbol{H}^T(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\right\}\right]\\
\le&C\lambda_c\sigma^2\frac{N}{K_N}E\left[trace\left\{(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\boldsymbol{H}^T\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-\frac{1}{2}}\right\}\right]\\
=&O(N).
\end{aligned}$$

Therefore, we have $R_{13} = O_p(\sqrt{N})$. Similarly, we have $R_{14} = O_p(\sqrt{N})$. Combining the convergence rates of R_{12}, R_{13} and R_{14}, we have

$$\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q} = R_{11} + O_p(\sqrt{N}).$$

Now, we consider $\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N$. Obviously,

$$\begin{aligned}
\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N =&\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N + \tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N\\
\triangleq&R_{21}+R_{22},
\end{aligned}$$

where

$$\begin{aligned}
R_{21} &= \tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N,\\
R_{22} &= \tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N.
\end{aligned}$$

We have

$$\begin{aligned}
\|\boldsymbol{e}_N\| &= \left\|\sum_{j=1}^{\infty}\gamma_{0j}\langle\boldsymbol{X}_N,\phi_j\rangle - \sum_{j=1}^{m}\gamma_{0j}\langle\boldsymbol{X}_N,\widehat{\phi}_j\rangle\right\|\\
&= \left\|\sum_{j=1}^{m}\gamma_{0j}\langle\boldsymbol{X}_N,\phi_j-\widehat{\phi}_j\rangle + \sum_{j=m+1}^{\infty}\gamma_{0j}\langle\boldsymbol{X}_N,\phi_j\rangle\right\|\\
&\le \left\|\sum_{j=1}^{m}\gamma_{0j}\langle\boldsymbol{X}_N,\phi_j-\widehat{\phi}_j\rangle\right\| + \left\|\sum_{j=m+1}^{\infty}\gamma_{0j}\langle\boldsymbol{X}_N,\phi_j\rangle\right\|.
\end{aligned}$$

By Lemma 1(b) of [17] with the help of Assumption 3.3 and Assumption 3.4, we have

$$\left\|\widehat{\phi}_j - \phi_j\right\| = O_p(jN^{-\frac{1}{2}}).$$

By Assumption 3.3 and Assumption 3.4, one has

$$\begin{aligned}
\left\|\sum_{j=1}^{m}\gamma_{0j}\langle X_i,\phi_j-\widehat{\phi}_j\rangle\right\|^2 &\le \|X_i(t)\|_2^2\left\|\sum_{j=1}^{m}(\phi_j-\widehat{\phi}_j)\gamma_{0j}\right\|^2\\
&= O_p(\sum_{j=1}^{m}j^{-b}jN^{-\frac{1}{2}})^2\\
&= O_p(N^{-1}m^{4-2b}).
\end{aligned}$$

By Assumption 3.5, we have

$$E\left[\sum_{j=m+1}^{\infty}\gamma_{0j}\langle X_i,\phi_j\rangle\right]=0,$$

$$\mathrm{Var}\left[\sum_{j=m+1}^{\infty}\gamma_{0j}\langle X_i,\phi_j\rangle\right]=\sum_{j=m+1}^{\infty}\gamma_{0j}^2\tau_j\le C\sum_{j=m+1}^{\infty}j^{-(a+2b)}=O(N^{-1}m).$$

Therefore, we have

$$\left\|\sum_{j=m+1}^{\infty}\gamma_{0j}\langle X_i,\phi_j\rangle\right\|^2=O_p(N^{-1}m).$$

Further, we have

$$\begin{aligned}\|\boldsymbol{e}_N\|^2&\le N\cdot\left\|\sum_{j=1}^{m}\gamma_{0j}\langle X_i,\phi_j-\widehat{\phi}_j\rangle\right\|^2+N\cdot\left\|\sum_{j=m+1}^{\infty}\gamma_{0j}\langle X_i,\phi_j\rangle\right\|^2\\&=O_p(m^{4-2b})+O_p(m)\\&=O_p(m).\end{aligned}$$

Combining this with Assumption 3.9, we have

$$\begin{aligned}E(\|R_{21}\|^2)&=E\left[trace\left\{\boldsymbol{e}_N{}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N\right\}\right]\\&\le\lambda_c\frac{N}{K_N}E\left[trace\left\{\boldsymbol{e}_N{}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N\right\}\right]\\&\le\lambda_c\frac{N}{K_N}E\left[trace\left\{\boldsymbol{e}_N{}^T\boldsymbol{e}_N\right\}\right]\\&=O(\frac{Nm}{K_N}).\end{aligned}$$

Thus, we can get $R_{21}=O_p(\sqrt{\frac{Nm}{K_N}})$. Similarly, we have $R_{22}=o_p(\sqrt{N})$.

Then we consider $\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N$. Obviously,

$$\begin{aligned}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N=&\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N+\tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N\\\triangleq&R_{31}+R_{32},\end{aligned}$$

where

$$\begin{aligned}R_{31}&=\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N,\\R_{32}&=\tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N.\end{aligned}$$

By Assumption 3.5 and Lemma 1 of [13], we have $\|\boldsymbol{\varepsilon}_N\|=O_p(\sqrt{N}K_N^{-r})$. Combining this with $K_N=O(N^{\frac{1}{2r+1}})$, we have

$$\begin{aligned}E(\|R_{31}\|^2)&=E\left[trace\left\{\boldsymbol{\varepsilon}_N{}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N\right\}\right]\\&\le\lambda_c\frac{N}{K_N}E\left[trace\left\{\boldsymbol{\varepsilon}_N{}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N\right\}\right]\\&\le\lambda_c\frac{N}{K_N}E\left[trace\left\{\boldsymbol{\varepsilon}_N{}^T\boldsymbol{\varepsilon}_N\right\}\right]\\&=O(N).\end{aligned}$$

Thus, we can get $R_{31}=O_p(\sqrt{N})$. Similarly, we have $R_{32}=o_p(\sqrt{N})$.

Then, one has

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)&=\sqrt{N}(\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Q})^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})(\boldsymbol{e}_N+\boldsymbol{\varepsilon}_N+\boldsymbol{V}_N)\\&=\left(\frac{R_{11}}{N}+o_p(1)\right)^{-1}\left(\frac{\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})}{\sqrt{N}}(\boldsymbol{e}_N+\boldsymbol{\varepsilon}_N+\boldsymbol{V}_N)\right).\end{aligned}$$

Next, we consider

$$\begin{aligned}&\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N}{\sqrt{N}}\right)\\&=\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N}{\sqrt{N}}\right)+\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\tilde{\boldsymbol{e}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N}{\sqrt{N}}\right)\\&=\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{R_{31}}{\sqrt{N}}\right)+\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{R_{32}}{\sqrt{N}}\right)\\&\triangleq A_{11}+A_{12}.\end{aligned}$$

By Assumption 3.6, we have

$$\begin{aligned}E\left(\|A_{11}\|^2\right)=&E\left[trace\left\{\left(\frac{R_{11}}{N}\right)^{-1}\frac{R_{31}^TR_{31}}{N}\left(\frac{R_{11}}{N}\right)^{-1}\right\}\right]\\\le&CE\left[trace\left\{\left(\frac{R_{11}}{N}\right)^{-2}\right\}\right]\\\le&CE\left[trace\left\{\left(\frac{R_{11}}{N}\right)^{-1}\right\}\right]^2\\=&O(K_N^2).\end{aligned}$$

Then, we have $A_{11}=O_p(K_N)$. Similarly, combining Assumption 3.6 and $\|R_{32}\|^2=o_p(N)$, we have $E\left(\|A_{12}\|^2\right)=o(K_N^2)$. Hence, we have $A_{12}=o_p(K_N)$. Further, we have

$$\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\varepsilon}_N}{\sqrt{N}}\right)=O_p(K_N).$$

Similarly, we have

$$\begin{aligned}&\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}_N}{\sqrt{N}}\right)=O_p(\sqrt{K_Nm})+o_p(K_N),\\&\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{V}_N}{\sqrt{N}}\right)=o_p(K_N).\end{aligned}$$

Thus, we have

$$\left\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|^2=O_p\left(\frac{K_N(K_N+m)}{N}\right).$$

□

Proof of Theorem 3.11. We have

$$
\begin{aligned}
\|\widehat{\gamma}(t)-\gamma_0(t)\|^2 &= \left\|\sum_{j=1}^{m}\widehat{\gamma}_j\widehat{\phi}_j-\sum_{j=1}^{\infty}\gamma_{0j}\phi_j\right\|^2\\
&\le 2\left\|\sum_{j=1}^{m}\widehat{\gamma}_j\widehat{\phi}_j-\sum_{j=1}^{m}\gamma_{0j}\phi_j\right\|^2+2\left\|\sum_{j=m+1}^{\infty}\gamma_{0j}\phi_j\right\|^2\\
&\le 4\left\|\sum_{j=1}^{m}(\widehat{\gamma}_j-\gamma_{0j})\widehat{\phi}_j\right\|^2+4\left\|\sum_{j=1}^{m}\gamma_{0j}(\widehat{\phi}_j-\phi_j)\right\|^2+2\sum_{j=m+1}^{\infty}\gamma_{0j}^2\\
&\triangleq 4B_1+4B_2+2B_3.
\end{aligned}
$$

According to the $\left\|\widehat{\phi}_j-\phi_j\right\|^2=O_p(N^{-1}j^2)$ and orthogonality of $\widehat{\phi}_j$, we have

$$
\begin{aligned}
B_2 &= \left\|\sum_{j=1}^{m}\gamma_{0j}(\widehat{\phi}_j-\phi_j)\right\|^2 \le m\sum_{j=1}^{m}\left\|\widehat{\phi}_j-\phi_j\right\|^2\gamma_{0j}^2 \le \frac{m}{N}O_p\left(\sum_{j=1}^{m}j^2\gamma_{0j}^2\right)\\
&= O_p\left(N^{-1}m\sum_{j=1}^{m}j^{2-2b}\right)=O_p\left(N^{-1}m\right)=o_p\left(N^{-\frac{2b-1}{a+2b}}\right),
\end{aligned}
$$

and

$$
B_3=\sum_{j=m+1}^{\infty}\gamma_{0j}^2\le C\sum_{j=m+1}^{\infty}j^{-2b}=O\left(m^{-(2b-1)}\right)=O\left(N^{-\frac{2b-1}{a+2b}}\right).
$$

For B_1, a simple calculation yields

$$
B_1=\left\|\sum_{j=1}^{m}(\widehat{\gamma}_j-\gamma_{0j})\widehat{\phi}_j\right\|^2\le\sum_{j=1}^{m}|\widehat{\gamma}_j-\gamma_{0j}|^2=\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2.
$$

According to the estimate process, we have

$$
\begin{aligned}
\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0 &= (\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T(\boldsymbol{Y}_N-\boldsymbol{Q}\widehat{\boldsymbol{\theta}})-\boldsymbol{\gamma}_0\\
&= (\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T(\boldsymbol{Q}\boldsymbol{\theta}_0+\boldsymbol{U}_N\boldsymbol{\gamma}_0+\boldsymbol{e}_N+\boldsymbol{\varepsilon}_N+\boldsymbol{V}_N-\boldsymbol{Q}\widehat{\boldsymbol{\theta}})-\boldsymbol{\gamma}_0\\
&= (\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{V}_N+(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{e}_N+(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{\varepsilon}_N\\
&\quad+(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{Q}(\boldsymbol{\theta}_0-\widehat{\boldsymbol{\theta}})\\
&\triangleq B_{11}+B_{12}+B_{13}+B_{14},
\end{aligned}
$$

where

$$
\begin{aligned}
&B_{11}=(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{V}_N,\quad B_{12}=(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{e}_N,\\
&B_{13}=(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{\varepsilon}_N,\quad B_{14}=(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{Q}(\boldsymbol{\theta}_0-\widehat{\boldsymbol{\theta}}).
\end{aligned}
$$

First, we consider B_{11}. Because the eigenvalues of $\frac{1}{N}\boldsymbol{U}_N{}^T\boldsymbol{U}_N$ are bounded in probability, we have

$$\begin{aligned}
E(\|B_{11}\|^2) &= E\left[\boldsymbol{V}_N{}^T\boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{V}_N\right] \\
&\leq \frac{1}{N^2}E\left[trace\left\{\boldsymbol{V}_N{}^T\boldsymbol{U}_N\boldsymbol{U}_N{}^T\boldsymbol{V}_N\right\}\right] \\
&\leq \frac{1}{N}E\left[trace\left\{(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-\frac{1}{2}}\boldsymbol{U}_N{}^T\boldsymbol{V}_N\boldsymbol{V}_N{}^T\boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-\frac{1}{2}}\right\}\right] \\
&= \frac{\sigma^2}{N}E\left[trace\left\{(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-\frac{1}{2}}\boldsymbol{U}_N{}^T\boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-\frac{1}{2}}\right\}\right] \\
&= O\left(N^{-1}m\right).
\end{aligned}$$

Hence, we have $B_{11} = O_p\left(\sqrt{N^{-1}m}\right)$. Next, based on the $\|\boldsymbol{e}_N\|^2 = O_p(m)$, we consider B_{12},

$$\begin{aligned}
E(\|B_{12}\|^2) &= E\left[\boldsymbol{e}_N{}^T\boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{e}_N\right] \\
&\leq \frac{1}{N}E\left[\boldsymbol{e}_N{}^T\boldsymbol{e}_N\right] \\
&= O\left(N^{-1}m\right).
\end{aligned}$$

Hence, we have $B_{12} = O_p\left(\sqrt{N^{-1}m}\right)$. Now, based on the $\|\boldsymbol{\varepsilon}_N\|^2 = O_p(NK_N^{-2r})$, we consider B_{13},

$$\begin{aligned}
E(\|B_{13}\|^2) &= E\left[\boldsymbol{\varepsilon}_N{}^T\boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{\varepsilon}_N\right] \\
&\leq \frac{1}{N}E\left[\boldsymbol{\varepsilon}_N{}^T\boldsymbol{\varepsilon}_N\right] \\
&= O\left(K_N^{-2r}\right).
\end{aligned}$$

Hence, we have $B_{13} = O_p\left(K_N^{-r}\right)$. Based on Lemma 1, Assumption 3.2 and Assumption 3.8, we can get

$$\begin{aligned}
&(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\boldsymbol{Q}^T\boldsymbol{Q}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) \\
=&(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T(\boldsymbol{W}_N\boldsymbol{Y}_N, \boldsymbol{\Pi})^T(\boldsymbol{W}_N\boldsymbol{Y}_N, \boldsymbol{\Pi})(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) \\
=&(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\begin{pmatrix}\boldsymbol{Y}_N{}^T\boldsymbol{W}_N{}^T\boldsymbol{W}_N\boldsymbol{Y}_N & \boldsymbol{Y}_N{}^T\boldsymbol{W}_N{}^T\boldsymbol{\Pi} \\ \boldsymbol{\Pi}^T\boldsymbol{W}_N\boldsymbol{Y}_N & \boldsymbol{\Pi}^T\boldsymbol{\Pi}\end{pmatrix}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) \\
=&(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\begin{pmatrix}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N) + \boldsymbol{V}_N)^T\boldsymbol{S}^T\boldsymbol{S}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N) + \boldsymbol{V}_N) & (\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N) + \boldsymbol{V}_N)^T\boldsymbol{S}^T\boldsymbol{\Pi} \\ \boldsymbol{\Pi}^T\boldsymbol{S}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N) + \boldsymbol{V}_N) & \boldsymbol{\Pi}^T\boldsymbol{\Pi}\end{pmatrix}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) \\
=&(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\begin{pmatrix}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N))^T\boldsymbol{S}^T\boldsymbol{S}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N)) & (\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N))^T\boldsymbol{S}^T\boldsymbol{\Pi} \\ \boldsymbol{\Pi}^T\boldsymbol{S}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N)) & \boldsymbol{\Pi}^T\boldsymbol{\Pi}\end{pmatrix}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) \\
+&(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\begin{pmatrix}\boldsymbol{V}_N{}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{V}_N & \boldsymbol{V}_N{}^T\boldsymbol{S}^T\boldsymbol{\Pi} \\ \boldsymbol{\Pi}^T\boldsymbol{S}\boldsymbol{V}_N & 0\end{pmatrix}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) + (\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\begin{pmatrix}2\boldsymbol{V}_N{}^T\boldsymbol{S}^T\boldsymbol{S}(\boldsymbol{\eta}_0 + g_0(\boldsymbol{Z}_N)) & 0 \\ 0 & 0\end{pmatrix}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}) \\
=&O_p\left(K_N\left(K_N + m\right)\right).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
E(\|B_{14}\|^2) &= E\left[(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\boldsymbol{Q}^T\boldsymbol{U}_N(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}(\boldsymbol{U}_N{}^T\boldsymbol{U}_N)^{-1}\boldsymbol{U}_N{}^T\boldsymbol{Q}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})\right] \\
&\leq \frac{1}{N}E\left[(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^T\boldsymbol{Q}^T\boldsymbol{Q}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})\right] \\
&= O\left(\frac{K_N\left(K_N + m\right)}{N}\right).
\end{aligned}$$

This implies that $B_{14} = O_p\left(\sqrt{\frac{K_N(K_N+m)}{N}}\right)$. Then, we have

$$\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 = O_p\left(N^{-1}m + K_N^{-2r} + N^{-1}K_N^2 + N^{-1}K_N m\right) = O_p\left(N^{-\frac{2r-1}{2r+1}} + N^{-\frac{2r}{2r+1}+\frac{1}{a+2b}}\right).$$

Thus, we have $B_1 = O_p\left(N^{-\frac{2r-1}{2r+1}} + N^{-\frac{2r}{2r+1}+\frac{1}{a+2b}}\right)$. In combination with the convergence rates of B_2 and B_3, we have

$$\|\widehat{\gamma}(t) - \gamma_0(t)\|_2^2 = O_p\left(N^{-\frac{2b-1}{a+2b}}\right) + O_p\left(N^{-\frac{2r-1}{2r+1}} + N^{-\frac{2r}{2r+1}+\frac{1}{a+2b}}\right).$$

Finally, we consider the convergence rate of $\widehat{g}(z)$. Note that

$$\|\widehat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}\|^2 \le \left\|\widehat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}\right\|^2 = O_p\left(\frac{K_N(K_N+m)}{N}\right).$$

By the Corollary 6.21 of [25], we have

$$\|\boldsymbol{\Pi}(z)\boldsymbol{\alpha}_0 - g_0(z)\|_2^2 = O_p(K_N^{-2r}).$$

Invoking formula (10) in [14], one has

$$\begin{aligned}
\|\widehat{g}(z) - g_0(z)\|_2^2 &= \|\boldsymbol{\Pi}(z)\widehat{\boldsymbol{\alpha}} - g_0(z)\|_2^2 \\
&\le 2\|\boldsymbol{\Pi}(z)\widehat{\boldsymbol{\alpha}} - \boldsymbol{\Pi}(z)\boldsymbol{\alpha}_0\|_2^2 + 2\|\boldsymbol{\Pi}(z)\boldsymbol{\alpha}_0 - g_0(z)\|_2^2 \\
&= O(K_N^{-1})\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|^2 + O_p(K_N^{-2r}) \\
&= O_p\left(\frac{K_N+m}{N}\right) + O_p(K_N^{-2r}) \\
&= O_p\left(N^{-\frac{a+2b-1}{a+2b}}\right) + O_p\left(N^{-\frac{2r}{2r+1}}\right).
\end{aligned}$$

□

Proof of Theorem 3.13. Similar to the proof of Lemma .1, it is easy to show that $R_{21} = R_{31} = o_p(\sqrt{N})$. Then, by $K_N/N = o(1)$, we have

$$\begin{aligned}
\sqrt{N}(\widehat{\lambda} - \lambda_0) =& \sqrt{N}\boldsymbol{\xi}^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
=& \sqrt{N}\boldsymbol{\xi}^T\left(R_{11} + O_p(N^{1/2}) + O_p(K_N)\right)^{-1}\boldsymbol{Q}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})(\boldsymbol{e}_N + \boldsymbol{\varepsilon}_N + \boldsymbol{V}_N) \\
=& \boldsymbol{\xi}^T\left(\frac{R_{11}}{N}\right)^{-1}\left(\frac{\tilde{\boldsymbol{Q}}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{P})}{\sqrt{N}}\right)\boldsymbol{V}_N + o_p(1) \\
=& \boldsymbol{\xi}^T\left(\frac{\Lambda^T\Lambda}{N}\right)^{-1}\frac{\Lambda^T(I-P)}{\sqrt{N}}\boldsymbol{V}_N + o_p(1).
\end{aligned}$$

Invoking the central limit theorem and Assumption 3.7, we have

$$\sqrt{N}(\widehat{\lambda} - \lambda_0) \xrightarrow{D} N(0, \sigma^2\varsigma).$$

□