

# Simultaneous Confidence Band for Stationary Covariance Function of Dense Functional Data

Jiangyan Wang<sup>a</sup>, Guanqun Cao<sup>b</sup>, Li Wang<sup>c</sup> and Lijian Yang<sup>d</sup>

<sup>a</sup>Nanjing Audit University, China, <sup>b</sup>Auburn University, USA,

<sup>c</sup>Iowa State University, USA, and <sup>d</sup>Tsinghua University, China

*Abstract:* The inference via simultaneous confidence band is studied for stationary covariance function of dense functional data. A two-stage estimation procedure is proposed based on spline approximation, the first stage involving estimation of all the individual trajectories and the second stage involving estimation of the covariance function through smoothing the empirical covariance function. The proposed covariance estimator is smooth and as efficient as the oracle estimator when all individual trajectories are known. An asymptotic simultaneous confidence band (SCB) is developed for the true covariance function, and the coverage probabilities are shown to be asymptotically correct. Intensive simulation experiments are conducted to demonstrate the performance of the proposed estimator and SCB. The proposed method is also illustrated with a real data example.

*Key words and phrases:* Confidence band, Covariance function, Functional data, Stationary.

## 1 Introduction

Since Ramsay and Dalzell (1991) first coined the term “functional data analysis” (FDA), recent years have seen numerous publications emerging in the FDA theory, methods and applications, making it an important area in statistics research. Motivated by specific problems and complex data collected in modern experiments, such as Ramsay and Silverman (2005), Hsing and Eubank (2015), considerable efforts have been made to analyze functional data. The estimation for population mean function and principal component in functional data has been extensively studied, for instance, Cardot (2000), Cao et al. (2012), Ferraty and Vieu (2006), Goldsmith et al. (2013), Hall et al. (2006) and so on.

Related to the smoothness, the second-order structure of random functions can be depicted by the covariance, thus the covariance function is another indispensable ingredient in many areas, such as longitudinal analysis, spatial statistics, and Bayesian hierarchical modeling, see

---

Address for correspondence: Li Wang (lilywang@iastate.edu) and Lijian Yang (yanglijian@tsinghua.edu.cn)

Choi et al. (2013), Diggle and Verbyla (1998), Hall et al. (1994), Yang et al. (2016) and Yin et al. (2010). In this sense, Cao et al. (2016) proposed a simultaneous confidence envelope of covariance function for functional data; Horváth et al. (2013) proposed a consistent estimator for the long-run covariance operator of stationary time series; Pantle et al. (2010) considered the estimation of integrated covariance functions, which is required to construct asymptotic confidence intervals and significance tests for the mean vector in the context of stationary random fields. Since the covariance function measures stronger association among variables that are closer to each other, the employment of covariance function is considerably highlighted in spatial data analysis when the geometric structure of the surface is rough and self-similar. A common situation is that the observations are specified via a Gaussian process whose finite-dimensional joint distributions are determined by a valid covariance function; see, for instance, Choi et al. (2013).

Let  $\{\eta(x), x \in \chi\}$  be a stochastic process defined on a compact interval  $\chi$ , with  $E \int_{\chi} \eta^2(x) dx < +\infty$ . It is covariance stationary if  $G(x, x') = C(|x - x'|)$ , where

$$G(x, x') = \text{Cov} \{ \eta(x), \eta(x') \}, \quad x, x' \in \chi. \quad (1)$$

Consider a collection of  $n$  trajectories  $\{\eta_i(x)\}_{i=1}^n$ , which are i.i.d realizations of  $\eta(x)$ , with mean and covariance functions, say  $m(x) = E\{\eta(x)\}$ ,  $G(x, x') = \text{Cov} \{ \eta(x), \eta(x') \}$ , respectively. The trajectories  $\{\eta_i(x)\}_{i=1}^n$  are decomposed as  $\eta_i(x) = m(x) + Z_i(x)$ , where  $Z_i(x)$  can be viewed as a small-scale variation of  $x$  on the  $i$ th trajectory, and is assumed to be a weakly stationary process with  $E Z_i(x) = 0$  and covariance  $G(x, x') = \text{Cov} \{ Z_i(x), Z_i(x') \}$ .

According to classical functional analysis FDA settings, for  $G(\cdot, \cdot)$ , there exist eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$  and corresponding eigenfunctions  $\{\psi_k\}_{k=1}^{\infty}$ , the latter being an orthonormal basis of  $L^2(\chi)$ , such that  $\sum_{k=1}^{\infty} \lambda_k < \infty$ ,  $G(x, x') = \sum_{k=1}^{\infty} \lambda_k \psi_k(x) \psi_k(x')$ , and  $\int G(x, x') \psi_k(x') dx' = \lambda_k \psi_k(x)$ . The standard process  $\eta(x)$ ,  $x \in \chi$ , then allows the well-known Karhunen-Loève  $L^2$  representation  $\eta(x) = m(x) + \sum_{k=1}^{\infty} \xi_k \phi_k(x)$ , in which the random coefficients  $\xi_k$ , called functional principal component (FPC) scores, are uncorrelated with each other of mean 0 and variance 1. The rescaled eigenfunctions,  $\phi_k$ , called FPC, satisfy that  $\phi_k = \sqrt{\lambda_k} \psi_k$  and  $\int \{ \eta(x) - m(x) \} \phi_k(x) dx = \lambda_k \xi_k$ , for  $k \geq 1$ . The  $i$ th process  $\eta_i(x)$ ,  $x \in \chi$ , is written as  $\eta_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(x)$ , in which the FPC scores  $\{\xi_{ik}\}_{k=1}^{\infty}$ ,  $i = 1, \dots, n$ , are i.i.d copies of  $\{\xi_k\}_{k=1}^{\infty}$ . Although the sequences  $\{\lambda_k\}_{k=1}^{\infty}$ ,  $\{\phi_k(\cdot)\}_{k=1}^{\infty}$  and  $\{\xi_{ik}\}_{k=1}^{\infty}$  exist mathematically, they are either unknown or unobservable.

The actual observed functional data are noisy sampled points from trajectories  $\{\eta_i(x)\}_{i=1}^n$ . Let  $\{(Y_{ij}, X_{ij}), 1 \leq i \leq n, 1 \leq j \leq N\}$  be repeated measurements on a random sample of  $n$  experimental units, where  $Y_{ij}$  is the response observed on the  $i$ th unit at value  $X_{ij}$  of the

variable  $x$ . The observed data can be modeled as

$$Y_{ij} = \eta_i(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij} = m(X_{ij}) + Z_i(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq N,$$

where  $\varepsilon_{ij}$ , independent of  $Z_i(\cdot)$ 's, are i.i.d random errors with mean 0 and variance 1, and  $\sigma^2(\cdot)$  is the variance function of the measurement errors. For the data considered in this paper, without loss of generality,  $\eta_i(\cdot)$  is assumed to be recorded on a regular grid in  $\chi = [0, 1]$ , and  $X_{ij} = x_j = j/N$ ,  $1 \leq j \leq N$ . This type of functional data was considered in Li and Hsing (2010), Crainiceanu et al. (2009) and Cao et al. (2016), among others. Consequently, our observed data can be written as

$$Y_{ij} = m(j/N) + Z_i(j/N) + \sigma(j/N)\varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq N. \quad (2)$$

It would not be a far stretch if the sample points for the  $i$ th subject  $Y_{ij}$  admit the structure of a nonstationary or locally stationary time series, as in Fryzlewicz and Ombao (2009) and Sanderson et al. (2010). One may further ask if these random observations at regular grid points would even admit the structure of stationary time series. Cao et al. (2016), for instance, concluded that the Tecator near-infrared spectra data is nonstationary based on the simultaneous confidence envelope for the covariance function. There are, however, interesting functional data for which the covariance function exhibits stationarity, because a closer relationship between the geometric structures and covariance function relies on the stationary assumption. In particular, the stationary random processes or fields are prominent in the analysis of 1D and 2D signals; see, for instance, the important spatial covariance model studied in Matérn random fields, stationary multivariate time series and the stationary spectral-space statistics studied in physics such as Tsyrlunikov and Gayfulin (2017). As a fundamental issue, the study of covariance structure in stationary stochastic processes can be applied to a wide range of areas such as hydrosciences and geostatistics.

Typically, it is difficult to interpret the covariance function in the case of FDA and longitudinal data analysis. The estimation strategies of the covariance function generally fall into two categories: direct smoothing and mixed-effects type of approaches, depending on whether there is a requirement of covariance reconstruction. In the literature, the mixed-effects type approaches have been considered intensively. For functional data, the functional principal component (FPC) analysis has become one of the first-line methods; see, for instance, the nonparametric estimation of covariance functions: Hall et al. (2006), Li and Hsing (2010) and Cao et al. (2016), among others. For longitudinal data, James et al. (2000) considered reduced rank spline mixed-effects models to describe the modes of variation; Yao et al. (2005) estimated the covariance structure and the FPC score by a conditioning step; Peng and Paul

(2009) proposed a geometric approach within the framework of marginal maximum likelihood estimation by requiring the trajectories are i.i.d. Gaussian processes.

However, to our best knowledge, the direct smoothing study is far from obvious in the FDA setting. This motivates us to develop an approach of direct smoothing for the covariance function. In this paper, we consider a nonparametric estimation of the covariance structure, which is useful either as a guide to the formulation of a parametric model or as the basis for formal inference without imposing parametric assumptions. Our estimation procedure is carried out by spline approximation, where the first step involves the estimation of the  $i$ th trajectory and the mean function, based on dense functional data, which is a vital feature that we can borrow strength; the second step estimates the covariance function through smoothing using the residuals of the first step via direct smoothing. The proposed covariance estimator is smooth and as efficient as the oracle estimator when all trajectories  $\eta_i(\cdot)$  and the mean  $m(\cdot)$  are known.

After estimating the covariance function, our next concern is to provide an inferential tool to further examine the covariance structure. Although a straightforward way is to conduct a hypothesis test, it is not well developed as other FDA methods, due to the difficulty of the infinite-dimensionality of the functional space. The existing methods mainly focus on testing of the mean functions for functional data, such as the pointwise t-test provided by Ramsay and Silverman (2005). However, the hypothesis test for covariance receives relatively little attention even though a global conclusion is often more desirable in real data analysis. In this line, Guo et al. (2018) proposed a supremum-norm based test for the equality of several covariance functions. However, it is a general-purpose smoother that is not designed specifically for covariance operators and it ignores that the smoothness of trajectories in FDA setting, hence the simple averaging of the observations is insufficient to meet the manifold needs in reality.

To surmount these challenges, we develop an asymptotic simultaneous confidence band (SCB), which can be used to test the adequacy and validity of certain covariance models. Specifically, the null hypothesis is  $H_0 : C(h) = C(h; \theta)$  for some  $\theta \in \Theta$ . An SCB is an intuitive and theoretically reliable tool for global inference of functions. For example, in the FDA framework, Cao et al. (2016) proposed SCBs for the covariance functions, and Goldsmith et al. (2013) derived a corrected SCB using principal component. The contribution of this paper is twofold. First, it provides the methodology and asymptotic theory for the estimation of the covariance  $C(\cdot)$  in the framework of stationary dense functional data under mild assumptions; second, the estimator of  $C(\cdot)$  is accompanied by a procedure for constructing asymptotic SCB.

The rest of the paper is organized as follows. In Section 2, we introduce the two-stage B-spline estimation procedure for the covariance function. Section 3 shows that the proposed estimator is as efficient as if all the  $n$  trajectories  $\eta_i(\cdot)$  and the mean function  $m(\cdot)$  are known

over the entire data range. Section 4 presents the asymptotic SCB for the covariance function, and describes the implementation of the SCB. Section 5 carries out intensive simulation studies to evaluate the finite sample performance of the proposed SCB. The methodology is verified by a real data example in Section 6. Technical lemmas and proofs are deferred to Appendices A and B. More simulation studies are carried out in Appendix C. Additional analysis of real data is given in Appendix D.

## 2 B-spline covariance function estimation

In this section, we describe the estimation procedure for the covariance function  $C(\cdot)$ . If the small-scale variation of  $x$ ,  $Z_i(x) = \eta_i(x) - m(x)$ ,  $1 \leq i \leq n$ ,  $x \in [0, 1]$ , on the  $i$ th trajectory could be observed, one would estimate the covariance as

$$\tilde{C}(h) = \frac{1}{1-h} \int_0^{1-h} \frac{1}{n} \sum_{i=1}^n Z_i(x) Z_i(x+h) dx, \quad h \in [0, h_0], \quad (3)$$

where  $h_0 \in (0, 1)$  is a pre-specified upper limit.

Since  $\{Z_i(x)\}_{i=1}^n$ ,  $x \in [0, 1]$ , are unobserved, the above estimator  $\tilde{C}(h)$  is “infeasible” in practice. In this paper, we propose to estimate the covariance function based on the following residuals

$$\hat{Z}_i(x) = \hat{\eta}_i(x) - \hat{m}(x), \quad 1 \leq i \leq n, \quad x \in [0, 1], \quad (4)$$

where  $\hat{\eta}_i(x)$  and  $\hat{m}(x)$  are the estimators of  $\eta_i(x)$  and  $m(x)$ .

In such case, a sample-based consistent estimator can be employed, such as the spline smoother proposed in Cao et al. (2012). Denote by  $\{t_\ell\}_{\ell=1}^{J_s}$  a sequence of equally-spaced points,  $t_\ell = \ell / (J_s + 1)$ ,  $1 \leq \ell \leq J_s$ ,  $0 < t_1 < \dots < t_{J_s} < 1$ , called interior knots, which divide the interval  $[0, 1]$  into  $(J_s + 1)$  equal subintervals  $I_0 = [0, t_1)$ ,  $I_\ell = [t_\ell, t_{\ell+1})$ ,  $\ell = 1, \dots, J_s - 1$ ,  $I_{J_s} = [t_{J_s}, 1]$ . For any positive integer  $p$ , let  $t_{1-p} = \dots = t_0 = 0$  and  $1 = t_{J_s+1} = \dots = t_{J_s+p}$  be auxiliary knots. Let  $\mathcal{S}^{(p-2)} = \mathcal{S}^{(p-2)}[0, 1]$  be the polynomial spline space of order  $p$  on  $I_\ell$ ,  $\ell = 0, \dots, J_s$ , which consists of all  $(p-2)$  times continuously differentiable functions on  $[0, 1]$  that are polynomials of degree  $(p-1)$  on subintervals  $I_\ell$ ,  $\ell = 0, \dots, J_s$ . Following the notation in de Boor (2001), we denote by  $\{B_{\ell,p}(x), 1 \leq \ell \leq J_s + p\}$  the  $p$ th order B-spline basis functions of  $\mathcal{S}^{(p-2)}$ , hence  $\mathcal{S}^{(p-2)} = \left\{ \sum_{\ell=1}^{J_s+p} \lambda_{\ell,p} B_{\ell,p}(x) \mid \lambda_{\ell,p} \in \mathbb{R}, x \in [0, 1] \right\}$ .

The  $i$ th unknown trajectory  $\eta_i(x)$  is estimated by using the following formula

$$\hat{\eta}_i(\cdot) = \arg \min_{g(\cdot) \in \mathcal{S}^{(p-2)}} \sum_{j=1}^N \{Y_{ij} - g(x_j)\}^2. \quad (5)$$

One can then estimate the unknown mean function  $m(\cdot)$  as

$$\widehat{m}(x) = n^{-1} \sum_{i=1}^n \widehat{\eta}_i(x), \quad (6)$$

and obtain the covariance estimator

$$\widehat{C}(h) = \frac{1}{1-h} \int_0^{1-h} \frac{1}{n} \sum_{i=1}^n \widehat{Z}_i(x) \widehat{Z}_i(x+h) dx, \quad h \in [0, h_0]. \quad (7)$$

### 3 Asymptotic Properties

This section studies the asymptotic properties for the proposed estimators.

#### 3.1 Assumptions

To study the asymptotic properties of the two-step spline estimator  $\widehat{C}(\cdot)$ , one needs some assumptions. Throughout the paper, for sequences  $a_n$  and  $b_n$ , denote  $a_n \asymp b_n$  if  $a_n$  and  $b_n$  are asymptotically equivalent. For any function  $\varphi(x)$  defined on a domain  $\chi$ , denote  $\|\varphi\|_\infty = \sup_{x \in \chi} |\varphi(x)|$ , and  $\varphi^{(q)}(x)$  its  $q$ th order derivative with respect to  $x$ . For any  $L^2$  integrable functions  $\phi(x)$  and  $\varphi(x)$ ,  $x \in \chi$ , define their theoretical inner product as  $\langle \phi, \varphi \rangle = \int_\chi \phi(x) \varphi(x) dx$ , and the empirical inner product as  $\langle \phi, \varphi \rangle_N = N^{-1} \sum_{j=1}^N \phi(j/N) \varphi(j/N)$ . The related theoretical and empirical norms are  $\|\phi\|_2^2 = \langle \phi, \phi \rangle$ ,  $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_N$ .

For a non-negative integer  $q$  and a real number  $\mu \in (0, 1]$ , write  $\mathcal{H}^{(q,\mu)}[0, 1]$  as the space of  $\mu$ -Hölder continuous functions, i.e.,

$$\mathcal{H}^{(q,\mu)}[0, 1] = \left\{ \varphi : [0, 1] \rightarrow \mathbb{R} \left| \|\varphi\|_{q,\mu} = \sup_{x,y \in [0,1], x \neq y} \left| \frac{\varphi^{(q)}(x) - \varphi^{(q)}(y)}{|x-y|^\mu} \right| < +\infty \right\}.$$

We next introduce some technical assumptions.

- (A1) There exist an integer  $q > 0$  and a constant  $\mu \in (0, 1]$ , such that the regression function  $m(\cdot) \in \mathcal{H}^{(q,\mu)}[0, 1]$ . In the following, one denotes  $p^* = q + \mu$ .
- (A2) The standard deviation function  $\sigma(\cdot) \in \mathcal{H}^{(0,\nu)}[0, 1]$  for positive index  $\nu \in (0, 1]$  and for some constants  $M_\sigma, M_0 > 0$ ,  $\sup_{x \in [0,1]} \sigma(x) \leq M_\sigma$ ,  $\sup_{h \in [0, h_0]} |C(h)| \leq M_0$ .
- (A3) There exists a constant  $\theta > 0$ , such that as  $N \rightarrow \infty$ ,  $n = n(N) \rightarrow \infty$ ,  $n = \mathcal{O}(N^\theta)$ .
- (A4) The rescaled FPCs  $\phi_k(\cdot) \in \mathcal{H}^{(q,\mu)}[0, 1]$  with  $\sum_{k=1}^\infty \|\phi_k\|_{q,\mu} < +\infty$ ,  $\sum_{k=1}^\infty \|\phi_k\|_\infty < +\infty$ ; for increasing positive integers  $\{k_n\}_{n=1}^\infty$ , as  $n \rightarrow \infty$ ,  $\sum_{k_n+1}^\infty \|\phi_k\|_\infty = \mathcal{O}(n^{-1/2})$  and  $k_n = \mathcal{O}(n^\omega)$  for some  $\omega > 0$ .

(A5) There are positive constants  $c_1, c_2 \in (0, +\infty)$ ,  $\gamma_1, \gamma_2 \in (1, +\infty)$ ,  $\beta_1, \beta_2 \in (0, 1/2)$ , and iid  $N(0, 1)$  variables  $\{U_{ij,\varepsilon}\}_{i=1,j=1}^{n,N}$ ,  $\{U_{ik,\xi}\}_{i=1,k=1}^{n,k_n}$  such that  $\min\left\{\frac{2(1-\beta_2)p^*}{3+(1+\beta_1)p^*}, \frac{2(\nu-\beta_2)}{(1+\beta_1)}\right\} > \theta$  for the index  $\nu$  in Assumption (A2),  $p^*$  in Assumption (A1), and

$$\Pr\left\{\max_{1 \leq k \leq k_n} \max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t U_{ik,\xi} \right| > n^{\beta_1}\right\} < c_1 n^{-\gamma_1},$$

$$\Pr\left\{\max_{1 \leq i \leq n} \max_{1 \leq t \leq N} \left| \sum_{j=1}^t \varepsilon_{ij} - \sum_{j=1}^t U_{ij,\varepsilon} \right| > N^{\beta_2}\right\} < c_2 N^{-\gamma_2}.$$

(A5') The iid variables  $\{\varepsilon_{ij}\}_{i \geq 1, j \geq 1}$  are independent of  $\{\xi_{ik}\}_{i \geq 1, k \geq 1}$ . The number of distinct distributions for all FPC scores  $\{\xi_{ik}\}_{i \geq 1, k \geq 1}$  is finite. There exist constants  $r_1 > 4 + 2\omega$ ,  $r_2 > 4 + 2\theta$ , for  $\omega$  in Assumption (A4) and  $\theta$  in Assumption (A3), such that  $E|\varepsilon_{11}|^{r_2}$  and  $E|\xi_{1k}|^{r_1}, k = 1, 2, \dots$  are finite.

(A6) The spline order  $p \geq p^*$ , the number of interior knots  $J_s \asymp N^\tau d_N$  for some  $\tau > 0$  with  $d_N + d_N^{-1} = \mathcal{O}(\log^\tau N)$  as  $N \rightarrow \infty$ , and for  $p^*$  in Assumption (A1),  $\nu$  in Assumption (A2),  $\theta$  in Assumption (A3),  $\beta_1, \beta_2$  and  $\gamma_1$  in Assumption (A5)

$$\max\left\{\frac{5\theta}{4p^*}, \frac{\theta + (\gamma_1 + 1 + \omega)^{-1} 8\theta\beta_1}{2p^*}, 1 - \nu\right\} < \gamma < 1 - \frac{\theta}{2} - \beta_2 - \frac{\theta}{2}\beta_1.$$

Assumptions (A1)–(A2) are standard in the literature, see Cao et al. (2012) and Song and Yang (2009) for instance. In particular, (A1) and (A4) control the size of the bias of the spline smoother for  $m(\cdot)$  and  $\phi_k(\cdot)$ . Assumption (A2) ensures the variance function is a uniformly bounded function. Assumption (A3) regulates that sample size  $n$  grows as a fractional power  $\theta$  of  $N$ , the number of observations per subject. The bounded smoothness of the principal components is guaranteed in Assumption (A4). Assumption (A5) provides a strong approximation of estimation errors and FPC scores. Assumption (A5') is an elementary assumption to guarantee the high level Assumption (A5). It is noteworthy that the smoothness of our estimator is controlled by the knots of the splines. Assumption (A6) specifies the requirement that the number of knots has to meet for the B-spline smoothing.

**REMARK 1.** *These assumptions are mild conditions that can be satisfied in many practical situations. One simple and reasonable setup for the above parameters  $q, \mu, \theta, p, \gamma$  can be as follows:  $q + \mu = p^* = 4$ ,  $\nu = 1$ ,  $\theta = 1$ ,  $p = 4$  (cubic spline),  $\gamma = 3/8$ ,  $d_N \asymp \log \log N$ . These constants are used as defaults in implementing the method; see Section 4.*

### 3.2 Oracle efficiency

We now show that the proposed two-step estimator  $\widehat{C}(\cdot)$  defined in (7) is oracle-efficient, i.e., it is as efficient as if all trajectories  $\eta_i(\cdot)$  are known over the entire data range. To begin with, we first investigate the asymptotic property of the infeasible covariance estimator  $\widetilde{C}(h)$ . Denote by  $\Delta(h) = \widetilde{C}(h) - C(h)$ ,  $h \in [0, h_0]$ .

According to the definition of  $C(h)$  and  $\widetilde{C}(h)$  in (1) and (3), one has

$$\begin{aligned} C(h) &= \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \mathbb{E}(\xi_{ik}\xi_{ik'}) \frac{1}{1-h} \int_0^{1-h} \phi_k(x)\phi_{k'}(x+h) dx \\ &= \frac{1}{1-h} \int_0^{1-h} \sum_{k=1}^{\infty} \phi_k(x)\phi_k(x+h) dx, \\ \widetilde{C}(h) &= \frac{1}{n(1-h)} \sum_{i=1}^n \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \xi_{ik}\xi_{ik'} \int_0^{1-h} \phi_k(x)\phi_{k'}(x+h) dx. \end{aligned}$$

Thus,

$$\Delta(h) = \frac{1}{1-h} \sum_{k,k'=1}^{\infty} (\bar{\xi}_{\cdot kk'} - \delta_{kk'}) \int_0^{1-h} \phi_k(x)\phi_{k'}(x+h) dx,$$

where  $\bar{\xi}_{\cdot kk'} = n^{-1} \sum_{i=1}^n \xi_{ik}\xi_{ik'}$ , and  $\delta_{kk'} = 1$  for  $k = k'$  and 0 otherwise.

Then the asymptotic mean squared error of the infeasible covariance estimator  $\widetilde{C}(\cdot)$  is provided in Theorem 1 below.

**THEOREM 1.** *Under Assumptions (A1)–(A6),  $\sup_{h \in [0, h_0]} |n\mathbb{E}\{\Delta(h)\}^2 - \Xi(h)| = o(1)$ , in which*

$$\begin{aligned} \Xi(h) &= \sum_{k,k'=1}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x)\phi_{k'}(x+h) dx \right\}^2 \\ &\quad + \sum_{k,k'=1}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x)\phi_{k'}(x+h) dx \right\} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_{k'}(x)\phi_k(x+h) dx \right\} \\ &\quad + \sum_{k=1}^{\infty} (\mathbb{E}\xi_{1k}^4 - 3) \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x)\phi_k(x+h) dx \right\}^2. \end{aligned} \tag{8}$$

**REMARK 2.** *By rewriting  $\Xi(h)$ , one has*

$$\begin{aligned} \Xi(h) &= \sum_{k=1}^{\infty} (\mathbb{E}\xi_{1k}^4 - 1) \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x)\phi_k(x+h) dx \right\}^2 \\ &\quad + \sum_{k < k'}^{\infty} \left[ \frac{1}{1-h} \left\{ \int_0^{1-h} \phi_k(x)\phi_{k'}(x+h) dx + \int_0^{1-h} \phi_{k'}(x)\phi_k(x+h) dx \right\} \right]^2. \end{aligned}$$



Following from (3.2) in Cao et al. (2016),

$$V(x, x+h) = \sum_{k < k'}^{\infty} \{ \phi_k(x) \phi_{k'}(x+h) + \phi_{k'}(x) \phi_k(x+h) \}^2 + \sum_{k=1}^{\infty} \phi_k^2(x) \phi_k^2(x+h) (\mathbb{E} \xi_{1k}^4 - 1),$$

thus,  $(1-h)^{-1} \int_0^{1-h} V(x, x+h) dx \geq \Xi(h)$ ,  $h \in [0, h_0]$ . Therefore, if the covariance function is stationary, the infeasible estimator  $\tilde{C}(\cdot)$  is more efficient than the covariance estimator given in Cao et al. (2016).

**PROPOSITION 1.** Under Assumptions (A1)–(A6), as  $N \rightarrow \infty$ ,  $\sqrt{n} \Delta(\cdot) \rightarrow_D \zeta(\cdot)$ , where  $\zeta(\cdot)$  is a Gaussian process defined on  $[0, h_0]$  such that  $\mathbb{E} \zeta(h) = 0$ ,  $\mathbb{E} \zeta^2(h) = \Xi(h)$ , with covariance function

$$\begin{aligned} \Omega(h, h') &= \text{Cov}(\zeta(h), \zeta(h')) = (1-h)^{-1} (1-h')^{-1} \\ &\times \left\{ \int_0^{1-h} \int_0^{1-h'} \sum_{k, k'=1}^{\infty} \phi_k(x) \phi_k(x') \phi_{k'}(x+h) \phi_{k'}(x'+h') dx dx' \right. \\ &+ \int_0^{1-h} \int_0^{1-h'} \sum_{k, k'=1}^{\infty} \phi_k(x) \phi_k(x'+h') \phi_{k'}(x+h) \phi_{k'}(x') dx dx' \\ &\left. + \int_0^{1-h} \int_0^{1-h'} \sum_{k=1}^{\infty} (\mathbb{E} \xi_{1k}^4 - 3) \phi_k(x) \phi_k(x+h) \phi_k(x') \phi_k(x'+h') dx dx' \right\}, \end{aligned}$$

for any  $h, h' \in [0, h_0]$ .

The proof is deferred to the Appendix. Although the oracle smoother  $\tilde{C}(\cdot)$  enjoys the desirable theoretical property, it is not a statistic since  $Z_i(x) = \eta_i(x) - m(x)$  is unknown. According to Proposition 2 below, the price for using  $\hat{Z}_i(x) = \hat{\eta}_i(x) - \hat{m}(x)$  in place of  $Z_i(x)$  in the covariance estimator is asymptotically negligible, that is, two-step estimator  $\hat{C}(\cdot)$  is as efficient as the infeasible estimator  $\tilde{C}(\cdot)$ .

**PROPOSITION 2.** Under Assumptions (A1)–(A6),  $\sup_{h \in [0, h_0]} |\hat{C}(h) - \tilde{C}(h)| = \mathcal{O}_p(n^{-1/2})$ .

Combining the above two propositions, we obtain the following result.

**THEOREM 2.** Under Assumptions (A1)–(A6),  $\sup_{h \in [0, h_0]} |\hat{C}(h) - C(h) - \Delta(h)| = \mathcal{O}_p(n^{-1/2})$ .

Theorem 2 indicates that  $\Delta(h)$  is the leading term of  $\hat{C}(h) - C(h)$ .

## 4 Simultaneous confidence band

In this section, we construct the SCB for the covariance function  $C(\cdot)$ .

## 4.1 Asymptotic SCB

Next theorem presents the asymptotic behavior of the maximum of the normalized deviation of the covariance estimator  $\widehat{C}(\cdot)$ , which sheds the lights on how to construct the asymptotic SCB for  $C(\cdot)$ .

**THEOREM 3.** *Under Assumptions (A1)–(A6), for any  $\alpha \in (0, 1)$ ,*

$$\lim_{N \rightarrow \infty} \Pr \left\{ \sup_{h \in [0, h_0]} n^{1/2} \left| \widehat{C}(h) - C(h) \right| \Xi(h)^{-1/2} \leq Q_{1-\alpha} \right\} = 1 - \alpha,$$

$$\lim_{N \rightarrow \infty} \Pr \left\{ n^{1/2} \left| \widehat{C}(h) - C(h) \right| \Xi(h)^{-1/2} \leq z_{1-\alpha/2} \right\} = 1 - \alpha, \quad \forall h \in [0, h_0],$$

where  $Q_{1-\alpha}$  is the 100  $(1 - \alpha)^{th}$  percentile of the absolute maxima distribution of  $\zeta(h)\Xi^{-1/2}(h)$ , while  $z_{1-\alpha/2}$  is denoted as the 100  $(1 - \alpha/2)^{th}$  percentile of the standard normal distribution, and  $\zeta(h)$  is the mean zero Gaussian process defined in Proposition 1.

Theorem 3 is a direct result of Propositions 1, 2 and Theorem 2, thus the proof is omitted.

**COROLLARY 1.** *Under Assumptions (A1)–(A6), an asymptotic 100  $(1 - \alpha)$  % exact SCB for  $C(\cdot)$  is  $\widehat{C}(h) \pm n^{-1/2} Q_{1-\alpha} \Xi^{1/2}(h)$ ,  $h \in [0, h_0]$ . While an asymptotic pointwise confidence band for  $C(\cdot)$  is given by  $\widehat{C}(h) \pm n^{-1/2} z_{1-\alpha/2} \Xi^{1/2}(h)$ ,  $h \in [0, h_0]$ .*

Note that the percentile  $Q_{1-\alpha}$  and the variance function  $\Omega(h, h')$  have to be estimated from the data. These issues are addressed in Section 4.4.

## 4.2 Knots selection

In spline smoothing, the number of knots is often treated as an unknown tuning parameters, and the fitting results can be sensitive to it. Though in the literature there is no optimal method to choose  $J_s$ , we recommend the following two ways: (a) criterion-based selection strategies such as Generalized Cross Validation (GCV) and Bayesian Information Criterion (BIC), and the candidate pool for  $J_s$  is all the integers between 1 and  $J_s^*$ , where  $J_s^* = \min\{10, \lfloor n/4 \rfloor\}$ ; (b) formula based selection strategies stated in Remark 1, specifically, we seek  $J_s$  that satisfies Assumption (C6) such that  $J_s \asymp N^\gamma d_N$ . In practice, the smoothness order  $(q, \mu)$  of  $m(\cdot)$  and  $\phi_k(\cdot)$  are taken as default  $(3, 1)$  or  $(4, 0)$  with a matching spline order  $p = 4$  (cubic spline). Therefore, we suggest  $J_s = \lfloor cN^\gamma \{\log \log(N)\}^\gamma \rfloor$  for some positive constant  $c$ . Note that the default of parameter  $\gamma = 3/8$  satisfies the condition given in Assumption (C6). In our extensive simulation studies, we find that  $c = 0.8$  is a good choice for the tuning parameter. Both methods give very similar estimators and SCBs in our numerical studies.

### 4.3 FPC analysis

We now describe how to obtain the covariance function  $\widehat{G}(\cdot, \cdot)$ , and its eigenfunctions  $\widehat{\phi}_k(\cdot)$  and eigenvalues  $\widehat{\lambda}_k$  in the FPC analysis.

We estimate  $G(\cdot, \cdot)$  by

$$\widehat{G}(x, x') = n^{-1} \sum_{i=1}^n \widehat{Z}_i(x) \widehat{Z}_i(x') = \sum_{s=1}^{J_s+p} \sum_{s'=1}^{J_s+p} \widehat{\beta}_{ss'} B_{s,p}(x) B_{s',p}(x'), \quad (9)$$

where  $\widehat{Z}_i$  is defined in (4) and  $\widehat{\beta}_{ss'}$ 's are the coefficients.

In FPC applications, it is typical to truncate the spectral decomposition at an integer  $\kappa$  to account for the some predetermined proportion of the variance. For example, in our numerical studies below,  $\kappa$  is selected as the number of eigenvalues that can explain 95% of the variation in the data. Next, let  $\mathbf{B}(x) = \{B_{1,p}(x), \dots, B_{J_s+p,p}(x)\}^\top$ , and the  $N \times (J_s + p)$  design matrix  $\mathbf{B}$  for spline regression is

$$\mathbf{B} = \{\mathbf{B}(1/N), \dots, \mathbf{B}(N/N)\}^\top = \begin{pmatrix} B_{1,p}(1/N) & \cdots & B_{J_s+p,p}(1/N) \\ \vdots & \cdots & \vdots \\ B_{1,p}(N/N) & \cdots & B_{J_s+p,p}(N/N) \end{pmatrix}. \quad (10)$$

Then for any  $k = 1, \dots, \kappa$ , we consider the following spline approximation for  $\psi_k(\cdot)$ :  $\widehat{\psi}_k(x') = \sum_{\ell=1}^{J_s+p} \widehat{\gamma}_{\ell k} B_{\ell,p}(x')$ , where  $\widehat{\gamma}_{\ell k}$ 's are coefficients of B-spline estimator subject to  $\widehat{\gamma}_k^\top \mathbf{B}^\top \mathbf{B} \widehat{\gamma}_k = 1$  with  $\widehat{\gamma}_k = (\widehat{\gamma}_{1,k}, \dots, \widehat{\gamma}_{J_s+p,k})^\top$ . The estimates of eigenfunctions and eigenvalues correspond  $\psi_k$  and  $\lambda_k$  can be obtained by solving the eigenequations,

$$\int \widehat{G}(x, x') \widehat{\psi}_k(x') dx' = \widehat{\lambda}_k \widehat{\psi}_k(x), \quad k = 1, \dots, \kappa. \quad (11)$$

According to (9), solving (11) is equivalent to solve the following:  $\mathbf{B}^\top(x) \widehat{\beta} \mathbf{B}^\top \mathbf{B} \widehat{\gamma}_k = \widehat{\lambda}_k \mathbf{B}^\top(x) \widehat{\gamma}_k$ ,  $k = 1, \dots, \kappa$ , where  $\widehat{\beta}^\top = (\widehat{\beta}_{s,s'})_{s,s'=1}^{J_s+p}$ .

By simple algebra and Lemma 3.1 in Wang and Yang (2009), one needs to solve  $\widehat{\beta} \mathbf{B}^\top \mathbf{B} \widehat{\gamma}_k = \widehat{\lambda}_k \widehat{\gamma}_k$ , for any  $k = 1, \dots, \kappa$ . Consider the following Cholesky decomposition:  $\mathbf{B}^\top \mathbf{B} = \mathbf{L}_B \mathbf{L}_B^\top$ . Therefore, solving (11) is equivalent to solve  $\widehat{\lambda}_k \mathbf{L}_B^\top \widehat{\gamma}_k = \mathbf{L}_B^\top \widehat{\beta} \mathbf{L}_B \mathbf{L}_B^\top \widehat{\gamma}_k$ , that is,  $\widehat{\lambda}_k$  and  $\mathbf{L}_B^\top \widehat{\gamma}_k$ ,  $k = 1, \dots, \kappa$ , are the eigenvalues and unit eigenvectors of  $\mathbf{L}_B^\top \widehat{\beta} \mathbf{L}_B$ . In other words,  $\widehat{\gamma}_k$  is obtained by multiplying  $(\mathbf{L}_B^\top)^{-1}$  immediately after the unit eigenvectors of  $\mathbf{L}_B^\top \widehat{\beta} \mathbf{L}_B$ , hence  $\widehat{\psi}_k(\cdot)$  is obtained. Consequently,  $\widehat{\phi}_k(x') = \widehat{\lambda}_k^{1/2} \widehat{\psi}_k(x')$ . Then, the  $k$ th FPC score of the  $i$ th curve can be estimated by a numerical integration

$$\widehat{\xi}_{ik} = \frac{1}{N} \sum_{j=1}^N \widehat{\lambda}_k^{-1} \left\{ Y_{ij} - \widehat{m} \left( \frac{j}{N} \right) \right\} \widehat{\phi}_k \left( \frac{j}{N} \right).$$

#### 4.4 Estimating the variance function $\Xi$ and the percentile $Q_{1-\alpha}$

Notice the fact that (8) entails us to estimate the variance function  $\Xi(\cdot)$  by merely computing  $\widehat{\xi}_{1k}^4$ ,  $\widehat{C}(\cdot)$  and  $\widehat{\phi}_k$ . In practice, the following estimator is employed

$$\begin{aligned}\widehat{\Xi}(h) &= \sum_{k,k'=1}^{\kappa} \left( \frac{1}{1-h} \int_0^{1-h} \widehat{\phi}_k(x) \widehat{\phi}_{k'}(x+h) dx \right)^2 \\ &\quad + \widehat{C}^2(h) + \sum_{k=1}^{\kappa} \left( \mathbb{E} \widehat{\xi}_{1k}^4 - 3 \right) \left\{ \frac{1}{1-h} \int_0^{1-h} \widehat{\phi}_k(x) \widehat{\phi}_k(x+h) dx \right\}^2.\end{aligned}$$

Next, to derive the percentile  $Q_{1-\alpha}$ , the Gaussian process is simulated as follows

$$\begin{aligned}\widehat{\zeta}(h) &= \sum_{k \neq k'}^{\kappa} \frac{1}{1-h} \int_0^{1-h} \epsilon_{kk'} \widehat{\phi}_k(x) \widehat{\phi}_{k'}(x+h) dx \\ &\quad + \sum_{k=1}^{\kappa} \frac{1}{1-h} \int_0^{1-h} \epsilon_k \widehat{\phi}_k(x) \widehat{\phi}_k(x+h) \left( \mathbb{E} \widehat{\xi}_{1k}^4 - 1 \right)^{1/2} dx,\end{aligned}$$

where  $\epsilon_{kk'}$  and  $\epsilon_k$  are independent standard Gaussian random variables. Hence,  $\widehat{\zeta}(h)$  is a zero mean Gaussian process with variance function  $\widehat{\Xi}(h)$  and covariance function

$$\begin{aligned}\widehat{\Omega}(h, h') &= \text{Cov} \left\{ \widehat{\zeta}(h), \widehat{\zeta}(h') \right\} \\ &= \frac{1}{1-h} \frac{1}{1-h'} \int_0^{1-h} \int_0^{1-h'} \left\{ \sum_{k,k'=1}^{\kappa} \widehat{\phi}_k(x) \widehat{\phi}_k(x') \widehat{\phi}_{k'}(x+h) \widehat{\phi}_{k'}(x'+h') \right. \\ &\quad \left. + \sum_{k=1}^{\kappa} \left( \mathbb{E} \widehat{\xi}_{1k}^4 - 3 \right) \widehat{\phi}_k(x) \widehat{\phi}_k(x+h) \widehat{\phi}_k(x') \widehat{\phi}_k(x'+h') \right\} dx dx' + \widehat{C}(h) \widehat{C}(h'),\end{aligned}$$

for any  $h, h' \in [0, h_0]$ . A large number of independent realizations of  $\widehat{\zeta}(h)$  are simulated, then the maximal absolute deviation for each copy of  $\widehat{\zeta}(h) \widehat{\Xi}^{-1/2}(h)$  is taken. Eventually,  $Q_{1-\alpha}$  is estimated by the empirical percentiles of these maximum values.

## 5 Simulation Studies

To illustrate the finite-sample behavior of our confidence bands, we conduct simulation studies to illustrate the finite-sample performance of the proposed method.

### 5.1 General study

The data are generated from the following model:  $Y_{ij} = m(j/N) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij}$ ,  $1 \leq j \leq N$ ,  $1 \leq i \leq n$ , where  $m(x) = \sin\{2\pi(x-1/2)\}$ ,  $\varepsilon_{ij}$  are i.i.d standard normal variables,  $\phi_k(x) = \sqrt{\lambda_k} \psi_k(x)$  with  $\lambda_k = (1/4)^{\lfloor k/2 \rfloor}$ ,  $\psi_{2k-1}(x) = \sqrt{2} \cos(2k\pi x)$ ,

$\psi_{2k}(x) = \sqrt{2} \sin(2k\pi x)$ ,  $k \geq 1$ . We consider both homogenous errors with  $\sigma(x) = \sigma_\epsilon$  and strongly heteroscedastic errors with  $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$ , where the noise level  $\sigma_\epsilon = 0.1, 0.5$ . Since Assumption (C5) is satisfied, following Cao et al. (2016), we truncate  $\sum_{k=1}^{\infty} \xi_{ik} \phi_k(j/N)$  at 1000. The number of curves  $n = \lfloor cN^\theta \rfloor$  with  $c = 0.8$  and  $\theta = 1$ , and the number of observations per curve  $N$  is taken to be 50, 100 and 200, respectively. Each simulation is repeated 500 times. Throughout this section, the mean function is estimated by cubic splines, i.e.,  $p = 4$ , with the number of knots selected using the formula and GCV given in Section 4.2.

First, we examine the accuracy of the proposed two-stage estimation procedure. The average mean squared error (AMSE) is computed to assess the performance of the covariance estimators  $\widehat{C}(\cdot)$  and  $\widetilde{C}(\cdot)$  defined in (7) and (3), respectively. The AMSE of  $\widehat{G}(\cdot, \cdot)$ , the eigenvalue  $\widehat{\lambda}_k$ 's and the eigenfunction  $\widehat{\phi}_k$ 's are defined as

$$\text{AMSE}(\widehat{G}) = \frac{1}{500N^2} \sum_{s=1}^{500} \sum_{j,j'=1}^N \left\{ \widehat{G}_s \left( \frac{j}{N}, \frac{j'}{N} \right) - G \left( \frac{j}{N}, \frac{j'}{N} \right) \right\}^2,$$

$$\text{AMSE}(\widehat{\lambda}) = \frac{1}{500\kappa} \sum_{s=1}^{500} \sum_{k=1}^{\kappa} (\widehat{\lambda}_{ks} - \lambda_k)^2, \text{AMSE}(\widehat{\phi}) = \frac{1}{500N\kappa} \sum_{s=1}^{500} \sum_{j=1}^N \sum_{k=1}^{\kappa} \left\{ (\widehat{\phi}_{ks} - \phi_k) \left( \frac{j}{N} \right) \right\}^2,$$

where  $\widehat{G}_s, \widehat{\lambda}_{ks}, \widehat{\phi}_{ks}$  represent the values of the  $s$ -th replication of  $\widehat{G}(\cdot, \cdot), \widehat{\lambda}_k, \widehat{\phi}_k$  in (11), respectively. Our simulation results based on homogeneous and heteroscedastic variance functions are listed in Tables 1 and 2, respectively. One concludes that a larger noise level leads to a higher AMSE over all, and the AMSEs of the two estimators  $\widehat{C}(\cdot)$  and  $\widetilde{C}(\cdot)$  are very similar in each scenario. Moreover, the  $\text{AMSE}(\widehat{\lambda})$  is getting smaller when  $N$  is increasing in each scenario. The GCV method has smaller  $\text{AMSE}(\widehat{\lambda})$  and  $\text{AMSE}(\widehat{G})$  than the formula method does. The value of the AMSE for  $\widehat{\phi}$  varies in each of the scenarios. When  $N = 50$ , the  $\text{AMSE}(\widehat{\phi})$  based on the formula method is smaller than that based on the GCV method, while a converse phenomenon is observed when  $N = 100, 200$ .

Tables 1 and 2 also present the empirical coverage rate (CR), i.e., the percentage of the event that the true curve  $C(\cdot)$  is entirely covered by the SCB among all 500 replications, respectively. As the sample size increases, the CR of the SCB becomes closer to the nominal confidence level, which shows a positive confirmation of Theorem 3. In addition, the average widths (WD) of the bands are calculated and presented in columns 9 and 11 of Tables 1–2. It is obvious that the width tends to be narrower when the sample size becomes larger and noise level  $\sigma_\epsilon$  smaller.

Overall, the performance of the SCB based on estimator  $\widehat{C}$  is indistinguishable from the infeasible SCB based on estimator  $\widetilde{C}$ ; and they approximate the nominal level as  $N$  increases. The knots number selected by the GCV yield similar results as those of the formula. For visu-

Table 1: Simulation results based on homogeneous errors with  $\sigma(x) = \sigma_\epsilon$ : AMSE of estimators  $\widehat{C}$ ,  $\widetilde{C}$ ,  $\widehat{\lambda}$ ,  $\widehat{G}$ ,  $\widehat{\phi}$ ; CR (outside/inside of the parentheses is based on  $\widehat{C}$ ,  $\widetilde{C}$ ) and WD of SCBs based on  $\widehat{C}$ .

$\sigma_\epsilon$	$N$	AMSE					SCB			
		$\widehat{C}$	$\widetilde{C}$	$\widehat{\lambda}$	$\widehat{G}$	$\widehat{\phi}$	95%		99%	
							CR	WD	CR	WD
0.1 Formula	50	0.068	0.065	0.014	0.130	0.661	0.866(0.892)	1.25	0.926(0.942)	1.51
	100	0.034	0.035	0.006	0.051	0.871	0.910(0.922)	0.92	0.970(0.970)	1.11
	200	0.016	0.016	0.003	0.025	0.845	0.958(0.962)	0.67	0.992(0.992)	0.80
0.1 GCV	50	0.065	0.065	0.009	0.095	0.806	0.868(0.894)	1.25	0.940(0.948)	1.51
	100	0.035	0.034	0.005	0.048	0.790	0.914(0.920)	0.92	0.968(0.970)	1.11
	200	0.016	0.016	0.002	0.025	0.769	0.960(0.960)	0.67	0.992(0.994)	0.80
0.5 Formula	50	0.070	0.065	0.014	0.133	0.651	0.864(0.892)	1.26	0.920(0.938)	1.53
	100	0.035	0.035	0.006	0.052	0.876	0.910(0.922)	0.93	0.966(0.970)	1.11
	200	0.016	0.016	0.003	0.025	0.845	0.956(0.960)	0.67	0.992(0.992)	0.80
0.5 GCV	50	0.070	0.065	0.012	0.107	0.871	0.858(0.896)	1.27	0.918(0.940)	1.53
	100	0.036	0.035	0.006	0.052	0.826	0.898(0.920)	0.92	0.958(0.964)	1.11
	200	0.017	0.016	0.003	0.026	0.798	0.946(0.960)	0.67	0.988(0.992)	0.80

alization of actual estimation, Figure 1 depicts the true covariance  $C(\cdot)$ , the spline covariance estimators  $\widehat{C}(\cdot)$ , as well as the 99% SCB for  $C(\cdot)$ . They are all based on a typical run under the setting  $N = 50$ ,  $N = 200$  and  $\sigma_\epsilon = 0.1$ . It is clear from Figure 1 that the estimator  $\widehat{C}(\cdot)$  is very close to the true covariance function  $C(\cdot)$  and the true covariance function is entirely covered by the SCB.

## 5.2 Spatial covariance models

In order to compare the finite-sample performance of the proposed estimator to that of Cao et al. (2016), we consider the following spatial covariance models:

- Spherical model (M1):  $C(h; \sigma_s^2, \theta_s) = \sigma_s^2 \{1 - \frac{3}{2} \frac{h}{\theta_s} + \frac{1}{2} (\frac{h}{\theta_s})^3\} I \{h \leq \theta_s\}$ ;
- Matérn model (M2):  $C(h; \sigma_s^2, \theta_s, \nu) = \sigma_s^2 \{\Gamma(\nu)\}^{-1} 2^{1-\nu} (2\sqrt{\nu}h/\theta_s)^\nu F_\nu(2\sqrt{\nu}h/\theta_s)$ , where  $\Gamma$  is the gamma function,  $F_\nu$  is the modified Neumann function;
- Gaussian Model (M3):  $C(h; \sigma_s^2, \theta_s) = \sigma_s^2 \exp(-h^2/\theta_s^2)$ .

In the parameterization (following Banerjee et al. (2004) page 29) of the covariance structure,  $\sigma_s^2$  is the sill and  $\theta_s$  is the range parameter. In the simulation, we set  $\sigma_s^2 = 2$  for M1, M2 and M3, and choose  $\theta_s = 1$  for M1 and M2,  $\theta_s = 3$  for M3, while for M2,  $\nu = 3$ . Since  $C(h) \rightarrow 0$  as  $h \rightarrow \infty$ , in practice, we only numerically evaluate the covariance  $C(h)$  over the “effective range” defined as the distance beyond which the correlation between observations,

Table 2: Simulation results based on heteroscedastic errors with  $\sigma(x) = \sigma_\epsilon \frac{5-\exp(x)}{5+\exp(x)}$ : AMSE of estimators  $\hat{C}$ ,  $\tilde{C}$ ,  $\hat{\lambda}$ ,  $\hat{G}$ ,  $\hat{\phi}$ ; CR (outside/inside of the parentheses is based on  $\hat{C}$ ,  $\tilde{C}$ ), and WD of SCBs based on  $\hat{C}$ .

$\sigma_\epsilon$	$N$	AMSE					SCB			
		$\hat{C}$	$\tilde{C}$	$\hat{\lambda}$	$\hat{G}$	$\hat{\phi}$	95%		99%	
							CR	WD	CR	WD
0.1 Formula	50	0.068	0.065	0.014	0.130	0.650	0.862(0.890)	1.25	0.930(0.942)	1.51
	100	0.034	0.035	0.006	0.051	0.870	0.912(0.922)	0.92	0.966(0.970)	1.11
	200	0.016	0.016	0.003	0.025	0.859	0.954(0.960)	0.67	0.988(0.990)	0.80
0.1 GCV	50	0.065	0.065	0.009	0.095	0.808	0.868(0.896)	1.25	0.938(0.948)	1.51
	100	0.034	0.035	0.005	0.048	0.772	0.918(0.920)	0.92	0.970(0.970)	1.11
	200	0.016	0.016	0.002	0.025	0.772	0.960(0.960)	0.67	0.992(0.994)	0.80
0.5 Formula	50	0.070	0.065	0.014	0.133	0.651	0.864(0.890)	1.26	0.926(0.940)	1.52
	100	0.035	0.035	0.006	0.051	0.879	0.916(0.920)	0.92	0.966(0.970)	1.11
	200	0.016	0.016	0.003	0.025	0.847	0.960(0.964)	0.67	0.988(0.990)	0.80
0.5 GCV	50	0.067	0.065	0.010	0.098	0.857	0.868(0.898)	1.26	0.930(0.942)	1.51
	100	0.035	0.034	0.005	0.049	0.817	0.912(0.922)	0.92	0.962(0.968)	1.11
	200	0.016	0.016	0.003	0.025	0.783	0.954(0.962)	0.67	0.992(0.994)	0.80

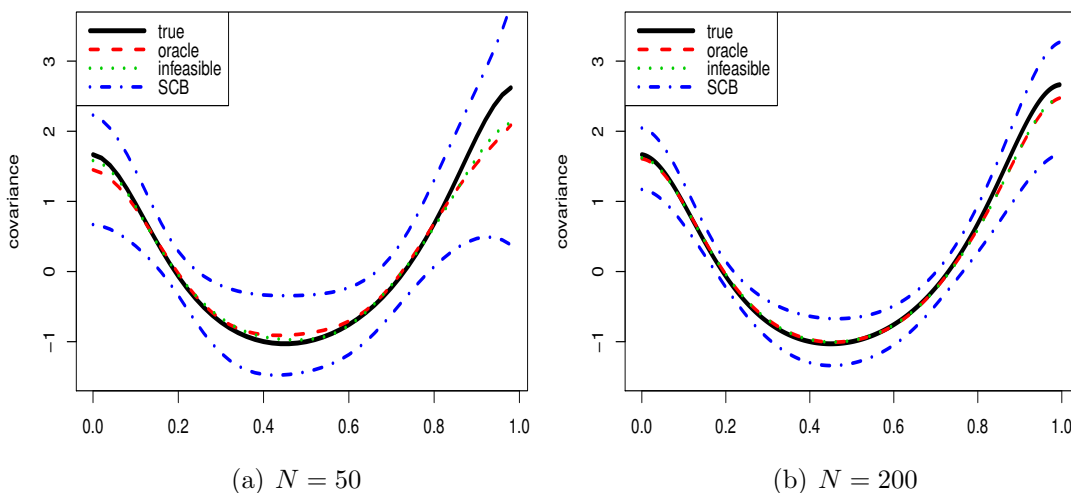


Figure 1: Plot of true covariance function (thick solid line), oracle estimator  $\hat{C}$  (dashed line) and the 99% SCB (dotted-dashed line), infeasible estimator  $\tilde{C}$  (dotted line) for the covariance function with  $\sigma_\epsilon = 0.1$ .

$\rho(h) = C(h)/C(0)$ , is less than or equal to 0.05. In such sense, we choose the compact interval  $[0, s]$  to represent the “effective range”, where  $s$  is the largest  $h$  satisfying  $\rho(h) \leq 0.05$ . An exception of this phenomenon is the spherical model that has an exact range  $[0, \theta_s]$ , i.e.,  $C(h) = 0$  when  $h = \theta_s$ . To be consistent in our evaluation of the methods, we apply the “effective range” to the spherical model as well.

Our data are generated from  $Y_{ij} = m(x_j) + Z_i(x_j) + \sigma(x_j)\varepsilon_{ij}$ , where  $m(x) = \sin\{2\pi(x - 1/2)\}$ ,  $\{x_j\}_{j=1}^N$  are equally spaced grid points over “effective range”  $[0, s]$ ,  $\varepsilon_{ij} \sim N(0, 1)$  are i.i.d variables, and the process  $Z_i(\cdot)$  is generated from a zero mean Gaussian process. We examine the performance of models containing homogeneous errors with  $\sigma(x) = \sigma_\epsilon$  and heteroscedastic errors with  $\sigma(x) = \sigma_\epsilon \frac{5 - \exp(x/2)}{5 + \exp(x/2)}$  for M1, and  $\sigma(x) = \sigma_\epsilon \frac{30 - \exp(x/2)}{30 + \exp(x/2)}$  for M2 and M3. The results are similar to each other, so we only present the results with homogeneous errors. The number of curves  $n = \lfloor 0.8N \rfloor$  with  $N = 50, 100$  and  $200$ , and the noise levels are  $\sigma_\epsilon = 0.1, 0.5$ . The mean function is estimated by cubic splines, i.e.,  $p = 4$ , with the number of knots selected using the formula given in Section 4.2. The GCV selected knots yield similar results but it is more time consuming, hence they are not summarized here.

The AMSE of the covariance estimators  $\widehat{C}$  and  $\widetilde{C}$  are reported in columns 4–5 of Table 3. The performance of the two estimators is very similar. Columns 6 and 8 present the empirical coverage rate CR, i.e., the percentage of the true curve  $C(\cdot)$  entirely covered by the SCB, based on 95% and 99% confidence levels, respectively. As the sample size increases, the coverage probability of the SCB becomes closer to the nominal confidence level. In addition, the WDs of the bands are calculated and presented in columns 7 and 9 in Table 3. It is obvious that the width tends to be narrower when the sample size becomes larger or  $\sigma_\epsilon$  is smaller.

When the covariance structure is not necessarily stationary, Cao et al. (2016) proposed a tensor-product bivariate B-spline estimator  $\widehat{G}^{\text{TPS}}(x, x')$  and a SCB for the covariance function  $G(x, x') = \text{Cov}\{Z_1(x), Z_1(x')\}$ . Following the suggestion of one referee, to assess the accuracy of recovering  $G(\cdot, \cdot)$ , the covariance function estimators  $\widehat{C}$  is also presented in 2D to make a comparison, say,  $\widehat{G}^{\text{PROP}}(x, x') = \widehat{C}(|x - x'|)$ . In addition, the simultaneous confidence envelopes (SCE) is constructed by using  $\widehat{G}^{\text{PROP}}(x, x')$  and  $\widehat{G}^{\text{TPS}}(x, x')$  are compared, named SCE-I and SCE-II, respectively.

Columns 4–5 of Table 4 present the AMSEs of  $\widehat{G}^{\text{PROP}}(x, x')$  and  $\widehat{G}^{\text{TPS}}(x, x')$ . The results of AMSEs indicate that  $\widehat{G}^{\text{PROP}}$  is more accurate than  $\widehat{G}^{\text{TPS}}$ , while  $\widehat{G}^{\text{TPS}}$  usually gives larger AMSE. Columns 6–13 of Table 4 report the CR and WD of SCE-I and SCE-II. One sees that the CRs of SCE-I are much closer to the nominal levels than those of SCE-II, and increasing the sample size helps to improve the CR of the SCEs to their nominal levels. One also observes the widths of the SCE-I are much narrower than those of the SCE-II. These findings indicate our proposed SCE-I is more efficient than SCE-II when the true covariance function is stationary.



Table 3: Simulation results based on homogeneous errors with  $\sigma(x) = \sigma_\epsilon$ : AMSE of estimators  $\hat{C}$ ,  $\tilde{C}$ ; CR (outside/inside of the parentheses is based on  $\hat{C}$  and  $\tilde{C}$ ), and WD of SCBs based on  $\hat{C}$ .

$\sigma_\epsilon$	Model	$N$	AMSE		SCB					
			$\hat{C}$	$\tilde{C}$	95%		99%			
					CR	WD	CR	WD		
0.1	M1	50	0.082	0.081	0.910(0.918)		1.37	0.960(0.966)		1.68
		100	0.040	0.040	0.920(0.926)		0.99	0.974(0.978)		1.21
		200	0.019	0.018	0.946(0.952)		0.72	0.980(0.986)		0.87
	M2	50	0.096	0.095	0.904(0.908)		1.44	0.950(0.954)		1.78
		100	0.048	0.049	0.926(0.924)		1.05	0.978(0.978)		1.30
		200	0.022	0.022	0.958(0.958)		0.76	0.992(0.994)		0.94
	M3	50	0.109	0.109	0.906(0.910)		1.50	0.954(0.958)		1.86
		100	0.055	0.055	0.922(0.928)		1.09	0.976(0.978)		1.35
		200	0.025	0.025	0.960(0.958)		0.79	0.988(0.990)		0.98
0.5	M1	50	0.082	0.080	0.896(0.912)		1.38	0.952(0.964)		1.70
		100	0.040	0.040	0.920(0.928)		0.99	0.980(0.980)		1.21
		200	0.019	0.018	0.938(0.946)		0.72	0.988(0.988)		0.88
	M2	50	0.097	0.096	0.896(0.908)		1.46	0.946(0.958)		1.80
		100	0.048	0.049	0.914(0.930)		1.06	0.978(0.980)		1.30
		200	0.022	0.022	0.954(0.958)		0.77	0.990(0.994)		0.94
	M3	50	0.111	0.111	0.908(0.916)		1.51	0.952(0.962)		1.88
		100	0.055	0.055	0.912(0.924)		1.10	0.974(0.978)		1.36
		200	0.025	0.025	0.958(0.956)		0.79	0.988(0.990)		0.98

Table 4: Simulation results based on homogeneous errors with  $\sigma(x) = \sigma_\epsilon$ : AMSE of estimators  $\hat{G}^{\text{PROP}}(\cdot, \cdot)$ ,  $\hat{G}^{\text{TPS}}(\cdot, \cdot)$ ; CR and WD of SCE-I and SCE-II.

$\sigma_\epsilon$	Model	$N$	AMSE		SCE-I				SCE-II			
			$\hat{G}^{\text{PROP}}$	$\hat{G}^{\text{TPS}}$	95%		99%		95%		99%	
					CR	WD	CR	WD	CR	WD	CR	WD
0.1	M1	50	0.079	0.123	0.910	1.40	0.960	1.71	0.744	2.08	0.840	2.55
		100	0.039	0.063	0.920	1.01	0.974	1.24	0.852	1.64	0.944	2.02
		200	0.018	0.031	0.946	0.73	0.980	0.90	0.904	1.19	0.964	1.47
	M2	50	0.096	0.148	0.904	1.50	0.950	1.86	0.682	2.09	0.816	2.56
		100	0.048	0.072	0.926	1.10	0.978	1.35	0.780	1.62	0.898	2.00
		200	0.022	0.036	0.958	0.79	0.992	0.98	0.926	1.16	0.976	1.43
	M3	50	0.114	0.153	0.906	1.57	0.954	1.95	0.724	2.15	0.824	2.65
		100	0.057	0.074	0.922	1.15	0.976	1.42	0.852	1.52	0.942	1.86
		200	0.026	0.039	0.960	0.83	0.988	1.21	0.882	1.09	0.956	1.33
0.5	M1	50	0.079	0.133	0.896	1.41	0.952	1.73	0.740	2.10	0.858	2.58
		100	0.039	0.064	0.920	1.02	0.980	1.25	0.814	1.64	0.924	2.03
		200	0.018	0.032	0.938	0.74	0.988	0.90	0.896	1.19	0.968	1.46
	M2	50	0.097	0.146	0.898	1.51	0.946	1.87	0.666	2.13	0.780	2.60
		100	0.048	0.071	0.914	1.10	0.978	1.35	0.768	1.62	0.906	2.00
		200	0.022	0.036	0.954	0.79	0.990	0.98	0.928	1.16	0.970	1.43
	M3	50	0.115	0.161	0.908	1.58	0.952	1.97	0.694	2.16	0.792	2.65
		100	0.057	0.075	0.912	1.15	0.974	1.42	0.814	1.53	0.928	1.87
		200	0.026	0.037	0.958	0.83	0.988	1.03	0.864	1.09	0.956	1.33

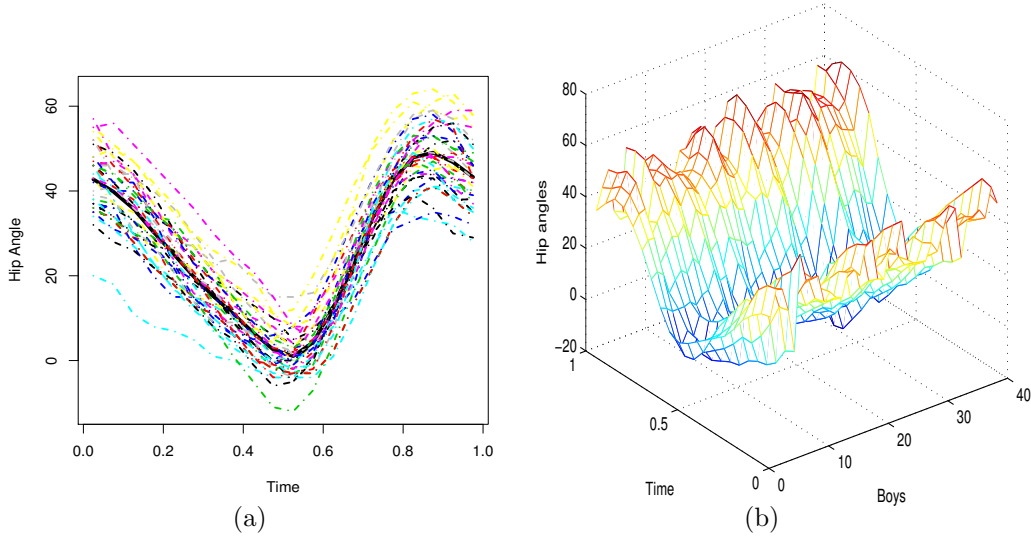


Figure 2: (a) Hip angle data (dotted lines) with its mean function estimator (solid line); (b) 3D plot.

## 6 Real data analysis

To further illustrate our methodology, we first consider the modeling of the Gait Data collected by the Motion Analysis Laboratory at the Children’s Hospital in San Diego, CA. We focus on the “Hip Angle” functional dataset, which consists of the angles formed by the hip of each boy over his gait cycle. See Olshen et al. (1989) for the details. In the study, the cycle begins and ends at the point where the heel of the limb under observation strikes the ground, which has been translated into values over  $[0, 1]$ . There are measurements on  $n = 39$  samples (boys), where for each sample  $N = 20$  hip angles were recorded every 0.05 second with time being measured on  $[0, 1]$ . Denote by  $Y_{ij}$  the hip angle of the  $i$ th boy at the time  $x_j$ ,  $j = 1, \dots, N$  and  $i = 1, \dots, n$ . Figure 2 (a) shows hip curves together with their estimated mean curve, and Figure 2 (b) describes the 3D shape of all curves, where “time” is plotted on one axis and sample index on the other.

Figure 3 (a) and (b) display the 3D and contour plots of the sample correlation of the hip data. From the plot, the contours are almost parallel to the main diagonal, indicating that the variation of the hip angles can be considered as an approximately stationary process. Figure 4 (a) shows a 3D plot of the proposed covariance matrix estimator  $\hat{G}^{\text{PROP}}(x, x') = \hat{C}(|x - x'|)$  with its asymptotic SCE. For comparison, the nonstationary covariance function estimator  $\hat{G}^{\text{TPS}}$  and its SCE are also presented; see Figure 4 (b).

As mentioned in Section 1, SCB is a very insightful and useful tool to examine the adequacy of certain parametric specifications of a covariance function. Now we make use of the proposed

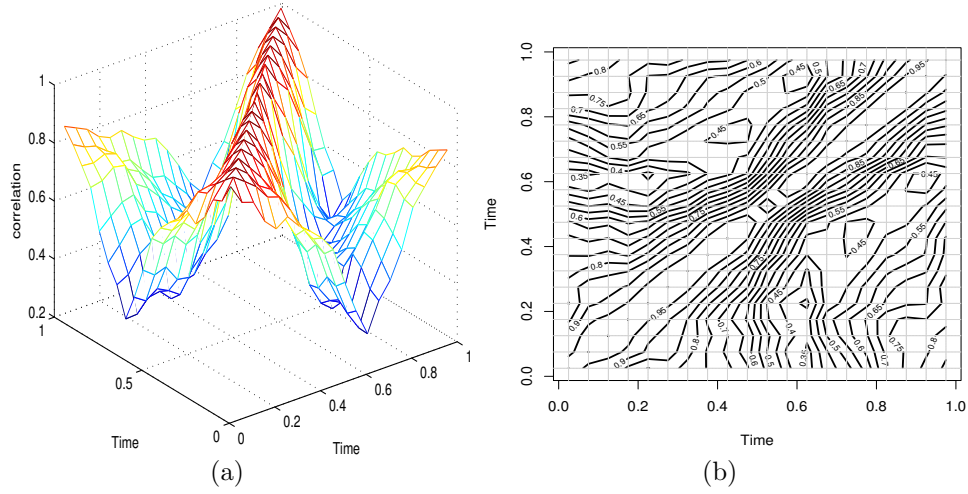


Figure 3: (a): Unsmoothed sample correlation 3D plot for the hip angle data; (b): Unsmoothed sample correlation contour plot for the hip angle data.

SCB to test if this hip data has a parametric covariance form like M1, M2 or M3. We set the null hypothesis  $H_0$  for M1, M2 and M3 in the following:

$$\text{M1 } H_0: C(h) = C(h; 32, \theta_s) = 32\{1 - 1.5(h/\theta_s) + 0.5(h/\theta_s)^3\}I\{h \leq \theta_s\}, \quad (12)$$

$$\text{M2 } H_0: C(h) = C(h; 32, \theta_s, \nu) = 32\{\Gamma(\nu)\}^{-1}2^{1-\nu}\{2\sqrt{\nu}h/\theta_s\}^\nu F_\nu(2\sqrt{\nu}h/\theta_s), \quad (13)$$

$$\text{M3 } H_0: C(h) = C(h; 32, \theta_s) = 32\exp(-h^2/\theta_s^2), \quad (14)$$

where  $\theta_s = 1.12$  for M1 and M2,  $\nu = 1.2$  for M2 and  $\theta_s = 2.19$  for M3. In Figure 5, the thick solid line is the covariance function  $C(h)$  under  $H_0$ , the center dashed line is the B-spline estimator, and the dotted-dashed lines are the SCBs. From Figure 5 (a), one observes that even the 99% SCB cannot contain  $C(h; 32, 1.12)$ , hence the null hypothesis in (12) is rejected with  $p$ -value  $< 0.01$ . Figure 5 (b) and (c) indicate that the 80% SCB contains  $C(h; 32, 1.2, 1.12)$  and  $C(h; 32, 2.19)$ , the null hypothesis in (13) and (14) is not rejected with  $p$ -value  $> 0.2$ .

## Acknowledgment

This work is supported in part by National Natural Science Foundation of China awards NSFC 11771240, 11801272, Natural Science Foundation of Jiangsu BK20180820; Natural Science Foundation of the Higher Education Institutions of Jiangsu Province 17KJB110005, 19KJA180002, China Scholarship Council, the National Science Foundation grants DMS 1542332, DMS 1736470 and DMS 1916204. The authors are truly grateful to the editor, the associate editor, two reviewers, and Mr. Jie Li from Tsinghua University Center for Statistical Science

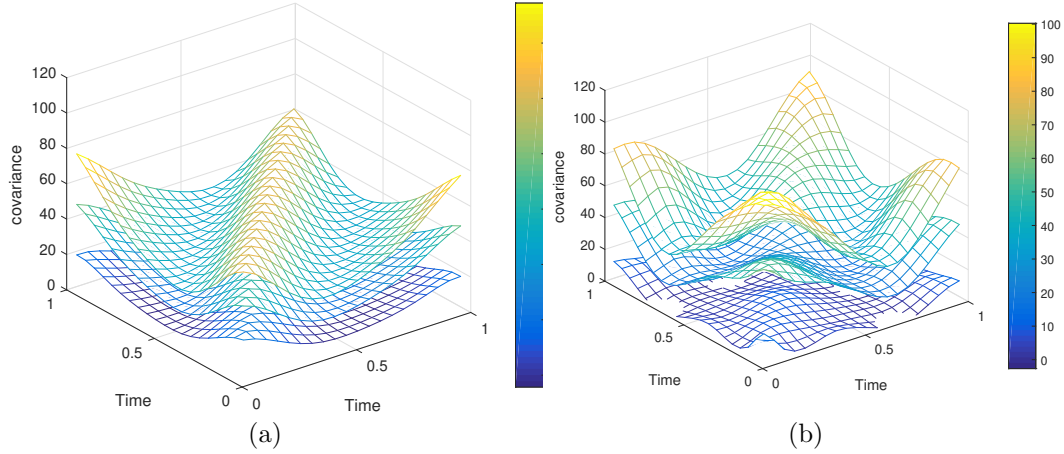


Figure 4: (a): Covariance matrix estimator based on  $\widehat{G}^{\text{PROP}}(x, x')$  (middle) with 95% SCE (up and below); (b): covariance matrix estimator  $\widehat{G}^{\text{TPS}}(x, x')$  (middle) of Cao et al. (2016) with 95% SCE (up and below).

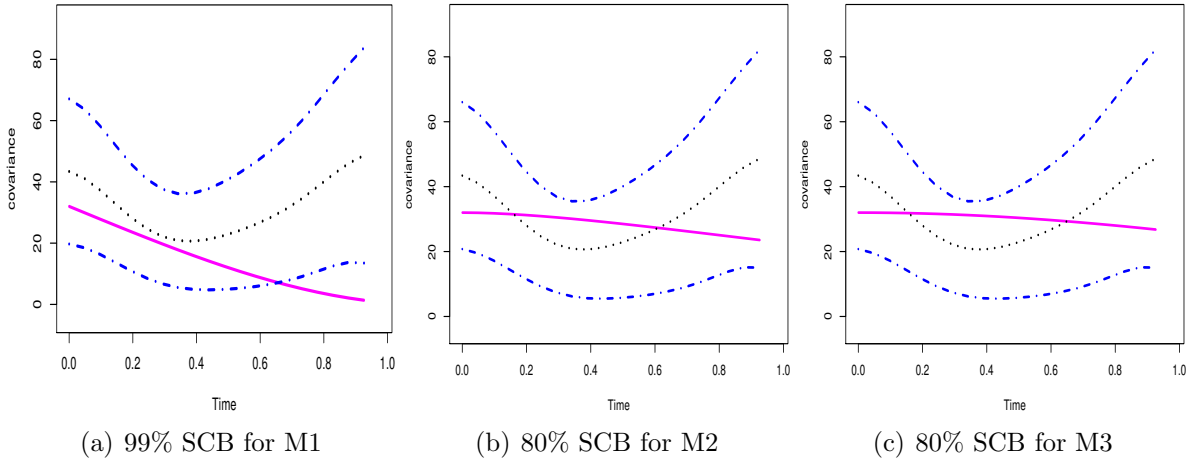


Figure 5: Covariance function  $C(h)$  under  $H_0$  (thick solid line), B-spline covariance estimator  $\widehat{C}$  (dotted line), and the SCB based on  $\widehat{C}$  (dotted-dashed line) for the hip angle data.

for their constructive comments and suggestions that led to significant improvement of the paper.

## Appendices

### A. Technical Lemmas and Proofs of Propositions 1 and 2

Throughout this section,  $\mathcal{O}_p$  (or  $\mathcal{o}_p$ ) denotes a sequence of random variables of certain order in probability. For instance,  $\mathcal{O}_p(n^{-1/2})$  means a smaller order than  $n^{-1/2}$  in probability, and by  $\mathcal{O}_{a.s.}$  (or  $\mathcal{o}_{a.s.}$ ) almost surely  $\mathcal{O}$  (or  $\mathcal{o}$ ). In addition,  $\mathcal{U}_p$  denotes a sequence of random functions which are  $\mathcal{O}_p$  uniformly defined in the domain.

For any vector  $\mathbf{a} = (a_1, \dots, a_n) \in \mathcal{R}^n$ , denote the norm  $\|\mathbf{a}\|_r = (|a_1|^r + \dots + |a_n|^r)^{1/r}$ ,  $1 \leq r < +\infty$ ,  $\|\mathbf{a}\|_\infty = \max(|a_1|, \dots, |a_n|)$ . For any matrix  $\mathbf{A} = (a_{ij})_{i=1, j=1}^{m, n}$ , denote its  $L_r$  norm as  $\|\mathbf{A}\|_r = \max_{\mathbf{a} \in \mathcal{R}^n, \mathbf{a} \neq \mathbf{0}} \|\mathbf{A}\mathbf{a}\|_r \|\mathbf{a}\|_r^{-1}$ , for  $r < +\infty$  and  $\|\mathbf{A}\|_r = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$ , for  $r = \infty$ .

#### A.1 Lemmas

Let  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iN})^\top$ , then the spline estimator  $\hat{\eta}_i(x)$  in (5) can be represented as  $\hat{\eta}_i(x) = \mathbf{B}(x)^\top (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{Y}_i$ , where  $\mathbf{B}$  is given in (10). Define the empirical inner product matrix of B-spline basis  $\{B_{\ell,p}(x)\}_{\ell=1}^{J_s+p}$  as  $\mathbf{V}_{n,p} = \{\langle B_{\ell,p}, B_{\ell',p} \rangle_N\}_{\ell, \ell'=1}^{J_s+p} = N^{-1} \mathbf{B}^\top \mathbf{B}$ , and according to Lemma A.3 in Cao et al. (2012), for some constant  $C_p > 0$

$$\|\mathbf{V}_{n,p}^{-1}\|_\infty \leq C_p J_s. \quad (\text{A.1})$$

According to model (2),  $\eta_i = \mathbf{m} + \mathbf{Z}_i$ , where  $\eta_i = \{\eta_i(1/N), \dots, \eta_i(N/N)\}^\top$ ,  $\mathbf{Z}_i = \{Z_i(1/N), \dots, Z_i(N/N)\}^\top$ ,  $\mathbf{m} = \{m(1/N), \dots, m(N/N)\}^\top$ , then the approximation error  $\hat{\eta}_i(x) - \eta_i(x)$  can be decomposed into the following:

$$\hat{\eta}_i(x) - \eta_i(x) = \tilde{\eta}_i(x) - \eta_i(x) + \tilde{\varepsilon}_i(x), \quad (\text{A.2})$$

where  $\tilde{\varepsilon}_i(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \varepsilon_i$ , and

$$\tilde{\eta}_i(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \eta_i = \tilde{m}(x) + \tilde{Z}_i(x), \quad (\text{A.3})$$

$$\tilde{m}(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \mathbf{m}, \quad \tilde{Z}_i(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \mathbf{Z}_i, \quad (\text{A.4})$$

where  $\varepsilon_i = (\sigma(1/N)\varepsilon_{i1}, \dots, \sigma(N/N)\varepsilon_{iN})^\top$ . Thus, one has  $\widehat{\eta}_i(x) - \eta_i(x) = \widetilde{Z}_i(x) - Z_i(x) + \widetilde{m}(x) - m(x) + \widetilde{\varepsilon}_i(x)$ . Therefore, by (4), (6) and (A.2), the approximation error of  $\widehat{Z}_i(x)$  in (4) to  $Z_i(x)$  can be represented by

$$\widehat{Z}_i(x) - Z_i(x) = \widetilde{Z}_i(x) - Z_i(x) + \widetilde{\varepsilon}_i(x) - \frac{1}{n} \sum_{i'=1}^n \left\{ \widetilde{Z}_{i'}(x) + \widetilde{\varepsilon}_{i'}(x) \right\}. \quad (\text{A.5})$$

**LEMMA A.1.** *Under Assumptions (A1)–(A6), as  $N \rightarrow \infty$ , one has*

$$\begin{aligned} \max_{1 \leq i \leq n} \|\widetilde{\eta}_i - \eta_i\|_\infty &= \mathcal{O}_{a.s.}\{J_s^{-p^*} (n \log n)^{2/r_1}\}, \\ \max_{1 \leq i \leq n} \|\widetilde{Z}_i - Z_i\|_\infty &= \mathcal{O}_{a.s.}\{J_s^{-p^*} (n \log n)^{2/r_1}\}, \quad \max_{1 \leq i \leq n} \|Z_i\|_\infty = \mathcal{O}_{a.s.}\{(n \log n)^{2/r_1}\}. \end{aligned} \quad (\text{A.6})$$

**LEMMA A.2.** *Under Assumptions (A1)–(A6), as  $N \rightarrow \infty$ , one has*

$$\max_{1 \leq i \leq n} \|\widetilde{\varepsilon}_i\|_\infty = \mathcal{O}_{a.s.}\{J_s^{1/2} N^{-1/2} (\log N)^{1/2}\}.$$

The next lemma follows from Lemmas A.1, A.2, (A.2) and (A.5).

**LEMMA A.3.** *Under Assumptions (A1)–(A6), as  $N \rightarrow \infty$ ,*

$$\begin{aligned} \max_{1 \leq i \leq n} \|\widehat{\eta}_i - \eta_i\|_\infty &= \mathcal{O}_p \left\{ J_s^{-p^*} (n \log n)^{2/r_1} + J_s^{1/2} N^{-1/2} (\log N)^{1/2} \right\}, \\ \max_{1 \leq i \leq n} \|\widehat{Z}_i - Z_i\|_\infty &= \mathcal{O}_p \left\{ J_s^{-p^*} (n \log n)^{2/r_1} + J_s^{1/2} N^{-1/2} (\log N)^{1/2} \right\}. \end{aligned} \quad (\text{A.7})$$

**LEMMA A.4.** *Assumption (A5) holds under Assumptions (A4) and (A5').*

**LEMMA A.5.** *Under Assumptions (A1)–(A6),*

$$\sup_{h \in [0, h_0]} \left| \frac{1}{n(1-h)} \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \left\{ \widetilde{Z}_i(x) - Z_i(x) \right\} dx \right| = \mathcal{O}_p(n^{-1/2}).$$

**LEMMA A.6.** *Under Assumptions (A1)–(A6),*

$$\max_{1 \leq i \leq n} \max_{1 \leq \ell \leq J_s + p} \left| N^{-1} \sum_{j=1}^N B_{\ell,p}(j/N) \sigma(j/N) U_{ij,\varepsilon} \right| = \mathcal{O}_{a.s.}(N^{-1/2} J_s^{-1/2} \log^{1/2} N),$$

where  $U_{ij,\varepsilon}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq N$ , are iid standard normal random variables.

LEMMA A.7. Under Assumptions (A1)–(A6),

$$\max_{\substack{1 \leq k \leq kn \\ 1 \leq \ell \leq J_s + p}} \left| \frac{1}{nN} \sum_{i=1}^n U_{ik,\xi} \left\{ \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma(j/N) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \right| = \mathcal{O}_{a.s.} \left( n^{-1/2} N^{\beta_2 - 1} \log^{1/2} N \right),$$

where  $0 < \beta_2 < 1/2$ .

LEMMA A.8. Under Assumptions (A1)–(A6), one has

$$\begin{aligned} & \max_{\substack{1 \leq k \leq kn \\ 1 \leq \ell \leq J_s + p}} \left| (nN)^{-1} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{ij,\varepsilon} \right| \\ &= \mathcal{O}_{a.s.} (n^{\beta_1 - 1/2} N^{-1/2} J_s^{-1/2} \log^{1/2} N), \quad \text{for } 0 < \beta_1 < 1/2. \end{aligned}$$

LEMMA A.9. Under Assumptions (A2)–(A6),

$$\max_{\substack{1 \leq k \leq kn \\ 1 \leq \ell \leq J_s + p}} \left| \frac{1}{nN} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \left\{ \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \right| = \mathcal{O}_{a.s.} \left( n^{\beta_1} N^{\beta_2 - 1} \right).$$

LEMMA A.10. Under Assumptions (A2)–(A6),

$$\sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n Z_i(x+h) \tilde{\varepsilon}_i(x) \right| = \mathcal{O}_p(n^{-1/2}).$$

LEMMA A.11. Under Assumptions (A2)–(A6), one has

$$\sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n Z_i(x+h) \frac{1}{n} \sum_{i'=1}^n \tilde{Z}_{i'}(x) \right| = \mathcal{O}_p(n^{-1/2}).$$

LEMMA A.12. Let  $W_i \sim N(0, \sigma_i^2)$ ,  $\sigma_i > 0$ ,  $i = 1, \dots, n$ , then for  $a > 2$

$$\Pr \left( \max_{1 \leq i \leq n} |W_i/\sigma_i| > a\sqrt{\log n} \right) < 2n^{1-a^2/2}. \quad (\text{A.8})$$

Hence,  $(\max_{1 \leq i \leq n} |W_i|) / (\max_{1 \leq i \leq n} \sigma_i) \leq \max_{1 \leq i \leq n} |W_i/\sigma_i| = \mathcal{O}_{a.s.}(\sqrt{\log n})$ .

## A.2 Proof of Proposition 1

Let  $\mathcal{F}_t = \sigma(\bar{\xi}_{\cdot 11}, \bar{\xi}_{\cdot 12}, \dots, \bar{\xi}_{\cdot 1t}, \bar{\xi}_{\cdot 22}, \dots, \bar{\xi}_{\cdot t-1, t}, \bar{\xi}_{\cdot tt})$ , so that  $\mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \mathcal{F}_4 \subseteq \dots$  is an increasing sequence of  $\sigma$ -fields. Denote

$$\begin{aligned} S_t(h) &= \sqrt{n}\Delta(\cdot) = \sqrt{n} \sum_{1 \leq k \neq k' \leq t} \bar{\xi}_{\cdot kk'} \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \\ &\quad + \sqrt{n} \sum_{1 \leq k \leq t} (\bar{\xi}_{\cdot kk} - 1) \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_k(x+h) dx, \end{aligned}$$

for  $t = 1, \dots, k_n$ , where  $k_n$  satisfies Assumption (A4). We show that  $S_t(h)$  is a martingale process in  $h \in [0, h_0]$ .

Define  $D_t(h) = S_t(h) - S_{t-1}(h)$ , thus,

$$D_t(h) = \frac{\sqrt{n}}{1-h} \left\{ \sum_{k=1}^{t-1} \bar{\xi}_{\cdot kt} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx + (\bar{\xi}_{\cdot tt} - 1) \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \right\},$$

which is  $\mathcal{F}_t$ -measurable. While notice that for any  $t$ ,

$$\begin{aligned} \mathbb{E}(D_t(h) | \mathcal{F}_{t-1}) &= \frac{\sqrt{n}}{1-h} \mathbb{E} \left\{ \sum_{k=1}^{t-1} \bar{\xi}_{\cdot kt} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \right. \\ &\quad \left. + (\bar{\xi}_{\cdot tt} - 1) \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \middle| \mathcal{F}_{t-1} \right\} \\ &= \frac{\sqrt{n}}{1-h} \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n \xi_{it} \sum_{k=1}^{t-1} \xi_{ik} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \middle| \mathcal{F}_{t-1} \right\} \\ &\quad + \sqrt{n} \mathbb{E} \left\{ (\bar{\xi}_{\cdot tt} - 1) \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \middle| \mathcal{F}_{t-1} \right\} = 0, \end{aligned}$$

which implies that  $\{D_t(h), t = 2, 3, \dots\}$  is a martingale difference process with respect to  $\{\mathcal{F}_{t-1}, t = 2, 3, \dots\}$ .

Next denote

$$\mathbb{E}(D_t^2(h) | \mathcal{F}_{t-1}) = V_t^{(1)}(h) + V_t^{(2)}(h) + V_t^{(3)}(h), \quad (\text{A.9})$$

in which

$$\begin{aligned} V_t^{(1)}(h) &= n \mathbb{E} \left[ \left\{ n^{-1} \sum_{i=1}^n \xi_{it} \sum_{k=1}^{t-1} \xi_{ik} \frac{1}{1-h} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \right\}^2 \middle| \mathcal{F}_{t-1} \right], \\ V_t^{(2)}(h) &= n \mathbb{E} \left[ \left\{ (\bar{\xi}_{\cdot tt} - 1) \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \right\}^2 \middle| \mathcal{F}_{t-1} \right], \\ V_t^{(3)}(h) &= 2n \mathbb{E} \left[ \left\{ n^{-1} \sum_{i=1}^n \xi_{it} \sum_{k=1}^{t-1} \xi_{ik} \frac{1}{1-h} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \right\} \right. \\ &\quad \left. \times (\bar{\xi}_{\cdot tt} - 1) \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \middle| \mathcal{F}_{t-1} \right]. \end{aligned}$$



Moreover, one can show that

$$\begin{aligned}
V_t^{(1)}(h) &= \mathbb{E} \left[ \left\{ \frac{1}{n} \sum_{i=1}^n \xi_{it} \sum_{k=1}^{t-1} \xi_{ik} \frac{1}{1-h} \int_0^{1-h} \{ \phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h) \} dx \right\}^2 \middle| \mathcal{F}_{t-1} \right] \\
&= \mathbb{E} \sum_{k=1}^{t-1} \left[ \frac{1}{n} \sum_{i=1}^n \xi_{it}^2 \xi_{ik}^2 \left( \frac{1}{1-h} \int_0^{1-h} \{ \phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h) \} dx \right)^2 \middle| \mathcal{F}_{t-1} \right] \\
&= \mathbb{E} \xi_{1t}^2 \sum_{k=1}^{t-1} \bar{\xi}_{\cdot kk} \left[ \frac{1}{1-h} \int_0^{1-h} \{ \phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h) \} dx \right]^2,
\end{aligned}$$

therefore, one has when  $n \rightarrow \infty$ ,

$$\begin{aligned}
\sum_{t=2}^{k_n} V_t^{(1)}(h) &\rightarrow \sum_{k \neq k'}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \right\}^2 \\
&+ \sum_{k \neq k'}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \right\} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x+h) \phi_{k'}(x) dx \right\} < \infty.
\end{aligned}$$

Note that  $V_t^{(2)}(h) = (\mathbb{E} \xi_{1t}^4 - 1) \left\{ (1-h)^{-1} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \right\}^2 < \infty$ , so one has that  $\sum_{t=2}^{k_n} V_t^{(2)}(h) \rightarrow \sum_{k=1}^{\infty} (\mathbb{E} \xi_{1k}^4 - 1) \left\{ (1-h)^{-1} \int_0^{1-h} \phi_k(x) \phi_k(x+h) dx \right\}^2 < \infty$ . Similarly,

$$\begin{aligned}
V_t^{(3)}(h) &= 2n \mathbb{E} \left[ \left\{ \sum_{k=1}^{t-1} \frac{1}{n} \sum_{i=1}^n \xi_{it} \xi_{ik} \frac{1}{1-h} \int_0^{1-h} \{ \phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h) \} dx \right\} \right. \\
&\quad \times \left. \left( \frac{1}{n} \sum_{i=1}^n \xi_{it}^2 - 1 \right) \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \mathcal{F}_{t-1} \right] \\
&= 2 (\mathbb{E} \xi_{1t}^3 - 1) \mathbb{E} \left[ \sum_{k=1}^{t-1} \bar{\xi}_{\cdot k} \frac{1}{1-h} \int_0^{1-h} \{ \phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h) \} dx \right. \\
&\quad \times \left. \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \mathcal{F}_{t-1} \right],
\end{aligned}$$

where  $\bar{\xi}_{\cdot k} = n^{-1} \sum_{i=1}^n \xi_{ik}$ . Next, notice that

$$\sup_{h \in [0, h_0]} \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \frac{1}{1-h} \int_0^{1-h} \{ \phi_k(x) \phi_{k'}(x+h) + \phi_{k'}(x) \phi_k(x+h) \} dx < \infty.$$

Therefore, one has

$$\begin{aligned}
\sum_{t=2}^{k_n} V_t^{(3)}(h) &\rightarrow 2 \sum_{t=2}^{k_n} (\mathbb{E} \xi_{1t}^3 - 1) \sum_{k=1}^{\infty} \mathbb{E} (\bar{\xi}_{\cdot k} | \mathcal{F}_{t-1}) \times \frac{1}{(1-h)^2} \\
&\quad \times \int_0^{1-h} \{ \phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h) \} dx \times \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \rightarrow_p 0,
\end{aligned}$$

as  $n \rightarrow \infty$ .

According to (A.9), as  $n \rightarrow \infty$ , one has

$$\begin{aligned} \sum_{t=2}^{k_n} \mathbb{E} (D_t^2(h) | \mathcal{F}_{t-1}) &\rightarrow_p \sum_{k \neq k'}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \right\}^2 \\ &+ \sum_{k \neq k'}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \right\} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x+h) \phi_{k'}(x) dx \right\} \\ &+ (\mathbb{E} \xi_{1t}^4 - 1) \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \right\}^2. \end{aligned}$$

Denote by  $\mathbb{E} (D_t^3(h) | \mathcal{F}_{t-1}) = d_t^{(1)}(h) + 3d_t^{(2)}(h) + 3d_t^{(3)}(h) + d_t^{(4)}(h)$ , where

$$\begin{aligned} d_t^{(1)}(h) &= n^{3/2} \mathbb{E} \left[ \left\{ \sum_{k=1}^{t-1} \bar{\xi}_{\cdot kt} \frac{1}{1-h} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \right\}^3 \middle| \mathcal{F}_{t-1} \right], \\ d_t^{(2)}(h) &= n^{3/2} \mathbb{E} \left[ \left\{ \sum_{k=1}^{t-1} \bar{\xi}_{\cdot kt} \frac{1}{1-h} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \right\}^2 \right. \\ &\quad \times (\bar{\xi}_{\cdot tt} - 1) \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \middle| \mathcal{F}_{t-1} \left. \right], \\ d_t^{(3)}(h) &= n^{3/2} \mathbb{E} \left[ \left\{ \sum_{k=1}^{t-1} \bar{\xi}_{\cdot kt} \frac{1}{1-h} \int_0^{1-h} \{\phi_k(x) \phi_t(x+h) + \phi_t(x) \phi_k(x+h)\} dx \right\} \right. \\ &\quad \times (\bar{\xi}_{\cdot tt} - 1)^2 \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \right\}^2 \middle| \mathcal{F}_{t-1} \left. \right], \\ d_t^{(4)}(h) &= n^{3/2} \mathbb{E} \left[ \left\{ (\bar{\xi}_{\cdot tt} - 1) \frac{1}{1-h} \int_0^{1-h} \phi_t(x) \phi_t(x+h) dx \right\}^3 \middle| \mathcal{F}_{t-1} \right]. \end{aligned}$$

Applying similar arguments in Lemma 6 of Cao et al. (2016), one has  $\sum_{t=2}^{k_n} \mathbb{E} \{d_t^{(i)}(h) | \mathcal{F}_{t-1}\} \rightarrow_p 0$ , for  $i = 1, 2, 3, 4$ . Hence, for any  $\epsilon > 0$ ,  $\sup_{h \in [0, h_0]} \sum_{t=2}^{k_n} \mathbb{E} \{D_t^3(h) I(D_t^2(h) > \epsilon) | \mathcal{F}_{t-1}\} \rightarrow_p 0$ .

By the uniform central limit theorem, one has  $\sqrt{n} \Delta(\cdot) = S_t(h) \rightarrow_D \zeta(\cdot)$ , as  $n \rightarrow \infty$ , where  $\zeta(h)$  is a Gaussian process such that  $\mathbb{E} \zeta(h) = 0$ ,

$$\begin{aligned} \Xi(h) = \mathbb{E} \zeta^2(h) &= \sum_{k=1}^{\infty} (\mathbb{E} \xi_{1k}^4 - 1) \left( \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_k(x+h) dx \right)^2 \\ &+ \sum_{k < k'}^{\infty} \left\{ \frac{1}{1-h} \left( \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx + \int_0^{1-h} \phi_{k'}(x) \phi_k(x+h) dx \right) \right\}^2, \end{aligned}$$

and covariance function

$$\begin{aligned} \Omega(h, h') &= \text{Cov}(\zeta(h), \zeta(h')) = \frac{1}{1-h} \frac{1}{1-h'} \\ &\times \left\{ \int_0^{1-h} \int_0^{1-h'} \sum_{k, k'=1}^{\infty} \phi_k(x) \phi_k(x') \phi_{k'}(x+h) \phi_{k'}(x'+h') dx dx' \right. \\ &+ \int_0^{1-h} \int_0^{1-h'} \sum_{k, k'=1}^{\infty} \phi_k(x) \phi_k(x'+h') \phi_{k'}(x+h) \phi_{k'}(x') dx dx' \\ &\left. + \int_0^{1-h} \int_0^{1-h'} \sum_{k=1}^{\infty} (\mathbb{E}\xi_{1k}^4 - 3) \phi_k(x) \phi_k(x+h) \phi_k(x') \phi_k(x'+h') dx dx' \right\}, \end{aligned}$$

for any  $h, h' \in [0, h_0]$ . The proposition is proved.

### A.3 Proof of Proposition 2

We decompose the difference between  $\widehat{C}(h)$  and  $\widetilde{C}(h)$  into the following three terms:  $\widehat{C}(h) - \widetilde{C}(h) = \text{I}(h) + \text{II}(h) + \text{III}(h)$ , where

$$\begin{aligned} \text{I}(h) &= \frac{1}{n(1-h)} \int_0^{1-h} \sum_{i=1}^n \left\{ \widehat{Z}_i(x) - Z_i(x) \right\} \left\{ \widehat{Z}_i(x+h) - Z_i(x+h) \right\} dx, \quad (\text{A.10}) \\ \text{II}(h) &= \frac{1}{n(1-h)} \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \left\{ \widehat{Z}_i(x) - Z_i(x) \right\} dx, \\ \text{III}(h) &= \frac{1}{n(1-h)} \int_0^{1-h} \sum_{i=1}^n Z_i(x) \left\{ \widehat{Z}_i(x+h) - Z_i(x+h) \right\} dx. \end{aligned}$$

Note that by (A.5),  $\sup_{h \in [0, h_0]} |\text{I}(h)| \leq \max_{1 \leq i \leq n} \|\widehat{Z}_i - Z_i\|_{\infty}^2$ . According to (A.7),

$$\max_{1 \leq i \leq n} \|\widehat{Z}_i - Z_i\|_{\infty} = \mathcal{O}_p \left\{ J_s^{-p^*} n \log n + J_s^{1/2} N^{-1/2} (\log N)^{1/2} \right\}.$$

By (A.5), one has

$$\begin{aligned} \text{II}(h) &= \frac{1}{n(1-h)} \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \left\{ \widehat{Z}_i(x) - Z_i(x) \right\} dx \\ &= \frac{1}{n(1-h)} \left[ \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \left\{ \widetilde{Z}_i(x) - Z_i(x) \right\} dx + \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \widetilde{\varepsilon}_i(x) dx \right] \\ &\quad - \frac{1}{n^2(1-h)} \left[ \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \sum_{i'=1}^n \widetilde{Z}_{i'}(x) dx + \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \sum_{i'=1}^n \widetilde{\varepsilon}_{i'}(x) dx \right]. \end{aligned}$$

Similar to the proof of Lemma A.10, it is easy to see

$$\sup_{h \in [0, h_0]} \frac{1}{n^2(1-h)} \left| \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \sum_{i'=1}^n \tilde{\varepsilon}_{i'}(x) dx \right| = \mathcal{O}_p(n^{-1/2}).$$

Consequently, by Lemmas A.5, A.10 and A.11, one has

$$\sup_{h \in [0, h_0]} |\mathbb{II}(h)| = \sup_{h \in [0, h_0]} \frac{1}{n(1-h)} \left| \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \left\{ \widehat{Z}_i(x) - Z_i(x) \right\} dx \right| = \mathcal{O}_p(n^{-1/2}).$$

Similarly, one can show that  $\sup_{h \in [0, h_0]} |\mathbb{III}(h)| = \sup_{h \in [0, h_0]} |\mathbb{II}(h)|$ . Consequently,

$$\sup_{h \in [0, h_0]} |\widehat{C}(h) - \widetilde{C}(h)| = \sup_{h \in [0, h_0]} |\mathbb{I}(h) + \mathbb{II}(h) + \mathbb{III}(h)| = \mathcal{O}_p(n^{-1/2}).$$

## B. Proofs of Technical Lemmas

In this section, we provide the proofs of technical lemmas introduced in Appendix A.

PROOF OF LEMMA A.1. For any  $k = 1, 2, \dots$ , let  $\phi_k = (\phi_k(1/N), \dots, \phi_k(N/N))^\top$ , and denote  $\widetilde{\phi}_k(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \phi_k$ . According to (A.3),  $\widetilde{\eta}_i(x) = \widetilde{m}(x) + \sum_{k=1}^{\infty} \xi_{ik} \widetilde{\phi}_k(x)$ , therefore,

$$\widetilde{\eta}_i(x) - \eta_i(x) = \widetilde{m}(x) - m(x) + \sum_{k=1}^{\infty} \xi_{ik} \left\{ \widetilde{\phi}_k(x) - \phi_k(x) \right\}.$$

By Lemma A.4 of Cao et al. (2012), there exists a constant  $C_{q,\mu} > 0$ , such that

$$\|\widetilde{m} - m\|_\infty \leq C_{q,\mu} \|m\|_{q,\mu} J_s^{-p^*}, \quad \|\widetilde{\phi}_k - \phi_k\|_\infty \leq C_{q,\mu} \|\phi_k\|_{q,\mu} J_s^{-p^*}, \quad k \geq 1.$$

Thus, by Assumption (A4), one obtains

$$\|\widetilde{\eta}_i - \eta_i\|_\infty \leq \|\widetilde{m} - m\|_\infty + \sum_{k=1}^{\infty} |\xi_{ik}| \|\widetilde{\phi}_k - \phi_k\|_\infty \leq C_{q,\mu} W_i J_s^{-p^*},$$

where  $W_i = \|m\|_{q,\mu} + \sum_{k=1}^{\infty} |\xi_{ik}| \|\phi_k\|_{q,\mu}$ ,  $i = 1, \dots, n$ , are iid nonnegative random variables with finite absolute moment. According to Assumption (A6), one has

$$\Pr \left\{ \max_{1 \leq i \leq n} W_i > (n \log n)^{2/r_1} \right\} \leq n \frac{\mathbb{E} W_i^{r_1}}{(n \log n)^2} = \mathbb{E} W_i^{r_1} n^{-1} (\log n)^{-2},$$

thus,  $\sum_{n=1}^{\infty} \Pr \{ \max_{1 \leq i \leq n} W_i > n \log n \} \leq \mathbb{E} W_i^{r_1} \sum_{n=1}^{\infty} n^{-1} (\log n)^{-2} < +\infty$ , so  $\max_{1 \leq i \leq n} W_i = \mathcal{O}_{a.s.} \left\{ (n \log n)^{2/r_1} \right\}$  and (A.6) is proved. Similarly, one obtains that  $\max_{1 \leq i \leq n} \|Z_i\|_\infty =$

$\mathcal{O}_{a.s.} \{(n \log n)^{2/r_1}\}$  and  $\max_{1 \leq i \leq n} \|\tilde{Z} - Z_i\|_\infty = \mathcal{O}_{a.s.} \{J_s^{-p^*} (n \log n)^{2/r_1}\}$ . Lemma A.1 holds consequently. ■

PROOF OF LEMMA A.2. Note that  $\{\varepsilon_{ij}\}$  are iid variables with  $\mathbb{E}(\varepsilon_{ij}^2) = 1$  and Lemma 2 of Wang (2012) implies that  $\|B_{\ell,p}\|_{2,N}^2 \asymp J_s^{-1}$  uniformly for all  $1 - p \leq \ell \leq J_s$ . Thus, one has uniformly for all  $1 - p \leq \ell \leq J_s$ ,

$$\mathbb{E} \left\{ \frac{1}{N} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \varepsilon_{ij} \right\}^2 = \frac{1}{N^2} \sum_{j=1}^N B_{\ell,p}^2 \left( \frac{j}{N} \right) \sigma^2 \left( \frac{j}{N} \right) = \frac{1}{N} \|B_{\ell,p}\sigma\|_{2,N}^2 \asymp J_s^{-1} N^{-1}.$$

Applying Bernstein inequality of Theorem 1.2 of Bosq (1998), similar to the proof of Lemma A.6, one has  $\max_{1 \leq i \leq n} \|N^{-1} \mathbf{B}^T \varepsilon_i\|_\infty = \mathcal{O}_{a.s.} \{J_s^{-1/2} N^{-1/2} (\log N)^{1/2}\}$ . Therefore, by recalling (A.1) and  $\tilde{\varepsilon}_i(x)$  in (A.2), one obtains

$$\max_{1 \leq i \leq n} \|\tilde{\varepsilon}_i\|_\infty = \max_{1 \leq i \leq n} \left\| N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^T \varepsilon_i \right\|_\infty = \mathcal{O}_{a.s.} \left\{ J_s^{1/2} N^{-1/2} (\log N)^{1/2} \right\}.$$

The lemma follows. ■

PROOF OF LEMMA A.4. Under Assumption (A5'),  $\mathbb{E}|\xi_{ik}|^{r_1} < +\infty$ ,  $r_1 > 4 + 2\omega$ ,  $\mathbb{E}|\varepsilon_{ij}|^{r_2} < +\infty$ ,  $r_2 > 4 + 2\theta$ , where  $\omega$  is defined in Assumption (A4) and  $\theta$  is defined in Assumption (A3), so there exists some  $\beta_1, \beta_2 \in (0, 1/2)$ , such that  $r_1 > (2 + \omega) / \beta_1$ ,  $r_2 > (2 + \theta) / \beta_2$ .

Let  $H(x) = x^{r_1}$ . Theorem 2.6.7 of Csörgo and Révész (1981) entails that there exist constants  $c_{1k}$  and  $a_k$  depending on the distribution of  $\xi_{ik}$ , such that for  $x_n = n^{\beta_1}$ ,  $n/H(a_k x_n) = a_k^{-r_1} n^{1-r_1\beta_1}$  and iid  $\mathcal{N}(0, 1)$  variables  $U_{ik,\xi}$ ,

$$\Pr \left\{ \max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t U_{ik,\xi} \right| > n^{\beta_1} \right\} < c_{1k} a_k^{-r_1} n^{1-r_1\beta_1},$$

by noticing that  $r_1 > (2 + \omega) / \beta_1$ ,  $\gamma_1 = r_1\beta_1 - 1 - \omega > 1$ . Since Assumption (A5') ensures that the number of distinct distributions for  $\xi_{ik}$  is finite, there is a common  $c_1 > 0$ , such that  $\max_{1 \leq k \leq k_n} \Pr \left\{ \max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t U_{ik,\xi} \right| > n^{\beta_1} \right\} < c_1 n^{-\gamma_1}$ , and consequently, there is a  $C_1 > 0$  such that

$$\Pr \left\{ \max_{\substack{1 \leq k \leq k_n \\ 1 \leq t \leq n}} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t U_{ik,\xi} \right| > n^{\beta_1} \right\} < k_n c_1 n^{1-r_1\beta_1} \leq C_1 n^{-\gamma_1}.$$

Likewise, under Assumption (A5), taking  $H(x) = x^{r_2}$ , Theorem 2.6.7 of Csörgo and Révész (1981) implies that there exists constants  $c_2$  and  $b$  depending on the distribution of  $\varepsilon_{ij}$ , such that for  $x_N = N^{\beta_2}$ ,  $N/H(ax_N) = b^{-r_2} c_2^{-r_2} N^{1-r_2\beta_2}$  and iid standard normal random variables  $U_{ij,\varepsilon}$  such that

$$\max_{1 \leq i \leq n} \Pr \left\{ \max_{1 \leq t \leq N} \left| \sum_{j=1}^t \varepsilon_{it} - \sum_{j=1}^t U_{it,\varepsilon} \right| > N^{\beta_2} \right\} < c_2 b^{-r_2} N^{1-\gamma_2\beta_2},$$

and consequently there is a  $C_2 > 0$  such that

$$\Pr \left\{ \max_{\substack{1 \leq i \leq n \\ 1 \leq t \leq N}} \left| \sum_{j=1}^t \varepsilon_{it} - \sum_{j=1}^t U_{it,\varepsilon} \right| > N^{\beta_2} \right\} < c_2 b^{-r_2} n \times N^{1-\gamma_2 \beta_2} \leq C_2 N^{\theta+1-\gamma_2 \beta_2}.$$

Since  $r_2 \beta_2 > (2 + \theta)$ , there is  $\gamma_2 = r_2 \beta_2 - 1 - \theta > 1$  and Assumption (A5) follows. The lemma holds consequently. ■

PROOF OF LEMMA A.5. According to Lemma A.1, one has

$$\begin{aligned} \sup_{h \in [0, h_0]} \left| \frac{1}{n(1-h)} \int_0^{1-h} \sum_{i=1}^n Z_i(x+h) \left\{ \tilde{Z}_i(x) - Z_i(x) \right\} dx \right| &\leq \max_{1 \leq i \leq n} \sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} |Z_i(x+h)| \\ &\times \max_{1 \leq i \leq n} \|\tilde{Z}_i - Z_i\|_\infty = \mathcal{O}_{a.s.} \left\{ J_s^{-p^*} (n \log n)^{4/r_1} \right\} = \mathcal{O}_p(n^{-1/2}). \end{aligned}$$

The proof is completed. ■

PROOF OF LEMMA A.6. We apply Lemma A.12 to obtain the uniform bound for the zero mean Gaussian variables  $N^{-1} \sum_{j=1}^N B_{\ell,p}(j/N) \sigma(j/N) U_{ij,\varepsilon}$ ,  $1 \leq i \leq n, 1 \leq \ell \leq J_s + p$  with variance  $N^{-1} \|B_{\ell,p} \sigma\|_{2,N}^2 \leq CN^{-1} J_s^{-1}$ . It follows from Lemma A.12 that

$$\begin{aligned} \max_{\substack{1 \leq i \leq n \\ 1 \leq \ell \leq J_s + p}} \left| N^{-1} \sum_{j=1}^N B_{\ell,p}(j/N) \sigma(j/N) U_{ij,\varepsilon} \right| &= \mathcal{O}_{a.s.} \left\{ N^{-1/2} J_s^{-1/2} \log^{1/2}(J_s + p) n \right\} \\ &= \mathcal{O}_{a.s.} \left( N^{-1/2} J_s^{-1/2} \log^{1/2} N \right), \end{aligned} \quad (\text{B.1})$$

where the last step follows from Assumptions (A4) and (A3) on the order of  $J_s$  and  $n$  relative to  $N$ . The lemma is proved. ■

PROOF OF LEMMA A.7. Applying Lemma A.4, one has

$$\max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq N}} \left| N^{-1} \sum_{t=1}^j (\varepsilon_{it} - U_{it,\varepsilon}) \right| = \mathcal{O}_{a.s.}(N^{\beta_2-1}).$$

Next, we denote the following sequences:

$$\begin{aligned} Q_{1,lk} &= (nN)^{-1} \sum_{i=1}^n U_{ik,\xi} \left[ \sum_{j=1}^{N-1} \left\{ B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) - B_{\ell,p} \left( \frac{j+1}{N} \right) \sigma \left( \frac{j+1}{N} \right) \right\} \sum_{t=1}^j (\varepsilon_{it} - U_{it,\varepsilon}) \right], \\ Q_{2,lk} &= (nN)^{-1} \sum_{i=1}^n U_{ik,\xi} \left\{ B_{\ell,p}(1) \sigma(1) \sum_{t=1}^N (\varepsilon_{it} - U_{it,\varepsilon}) \right\}. \end{aligned}$$

Further denote the  $\sigma$ -field  $\mathcal{F}_\varepsilon = \sigma\{\varepsilon_{ij}, i, j = 1, 2, \dots\}$ , then for  $1 \leq k \leq k_n, 1 \leq \ell \leq J_s + p$ ,  $Q_{1,\ell k} | \mathcal{F}_\varepsilon =_D N(0, \sigma_{\ell k}^2)$ , in which

$$\sigma_{\ell k}^2 = n^{-2} \sum_{i=1}^n \left[ \sum_{j=1}^{N-1} \left\{ B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) - B_{\ell,p} \left( \frac{j+1}{N} \right) \sigma \left( \frac{j+1}{N} \right) \right\} N^{-1} \sum_{t=1}^j (\varepsilon_{it} - U_{it,\varepsilon}) \right]^2.$$

Applying (A.8), for any  $a > 2$

$$\Pr \left[ \max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |Q_{1,\ell k}| / \sigma_{\ell k} > a \{\log k_n (J_s + p)\}^{1/2} \middle| \mathcal{F}_\varepsilon \right] < 2 \{k_n (J_s + p)\}^{1-a^2/2},$$

and hence

$$\Pr \left[ \max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |Q_{1,\ell k}| / \sigma_{\ell k} > a \{\log k_n (J_s + p)\}^{1/2} \right] < 2 \{k_n (J_s + p)\}^{1-a^2/2}.$$

Taking large enough  $a$ , while noting Assumptions (A6) and (A4) on the order of  $J_s$  and  $k_n$  relative to  $N$ , one concludes with Borel-Cantelli Lemma that

$$\max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |Q_{1,\ell k}| / \sigma_{\ell k} = \mathcal{O}_{a.s.} \left( \{\log k_n (J_s + p)\}^{1/2} \right) = \mathcal{O}_{a.s.} (\log^{1/2} N). \quad (\text{B.2})$$

Next, the B spline basis satisfies

$$\left| B_{\ell,p} \left( \frac{j}{N} \right) - B_{\ell,p} \left( \frac{j+1}{N} \right) \right| \leq N^{-1} \|B_{\ell,p}\|_{0,1} \leq C J_s N^{-1}$$

uniformly over  $1 \leq j \leq N$  and  $1 \leq \ell \leq J_s + p$ , while Assumptions (A2) and (A6) imply that  $J_s N^{-1} \sim N^\gamma d_N N^{-1} \sim N^{\gamma-1} d_N \gg N^{-\nu}$ , hence

$$\left| \sigma \left( \frac{j}{N} \right) - \sigma \left( \frac{j+1}{N} \right) \right| \leq N^{-\nu} \|\sigma\|_{0,\nu} \leq C J_s N^{-1}$$

uniformly over  $1 \leq j \leq N$ . it then follows that

$$\begin{aligned} \sigma_{\ell k}^2 &\leq n^{-1} \left\{ \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq N}} \left| N^{-1} \sum_{t=1}^j (\varepsilon_{it} - U_{it,\varepsilon}) \right| \right\}^2 \left\{ C N J_s^{-1} \times \|B_{\ell,p}\|_{0,1} N^{-1} \right\}^2 \\ &\leq C n^{-1} \left\{ \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq N}} \left| N^{-1} \sum_{t=1}^j (\varepsilon_{it} - U_{it,\varepsilon}) \right| \right\}^2 = \mathcal{U}_{a.s.} (n^{-1} N^{2\beta_2 - 2}). \end{aligned} \quad (\text{B.3})$$

Putting together the bounds in (B.2) and (B.3), one obtains that

$$\max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |Q_{1,\ell k}| = \mathcal{O}_{a.s.} (n^{-1/2} N^{\beta_2 - 1} \log^{1/2} N).$$

Similarly,  $\max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s + p} |\mathbb{Q}_{2, \ell k}| = \mathcal{O}_{a.s.}(n^{-1/2} N^{\beta_2 - 1} \log^{1/2} N)$ . Finally, the Lemma is proved by noticing that

$$\begin{aligned} & \left| n^{-1} \sum_{i=1}^n U_{ik, \xi} \left\{ N^{-1} \sum_{j=1}^N B_{\ell, p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij, \varepsilon}) \right\} \right| \\ &= \left| (nN)^{-1} \sum_{i=1}^n U_{ik, \xi} \left[ \sum_{j=1}^{N-1} \left\{ B_{\ell, p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) - B_{\ell, p} \left( \frac{j+1}{N} \right) \sigma \left( \frac{j+1}{N} \right) \right\} \sum_{t=1}^j (\varepsilon_{it} - U_{it, \varepsilon}) \right] \right. \\ & \quad \left. + (nN)^{-1} \sum_{i=1}^n U_{ik, \xi} \left\{ B_{\ell, p}(1) \sigma(1) \sum_{t=1}^N (\varepsilon_{it} - U_{it, \varepsilon}) \right\} \right| \leq |\mathbb{Q}_{1, \ell k}| + |\mathbb{Q}_{2, \ell k}|. \blacksquare \end{aligned}$$

PROOF OF LEMMA A.8. According to Lemma A.5 in Cao et al. (2012), under Assumptions (A4)–(A5),  $\max_{1 \leq k \leq k_n} \max_{1 \leq t \leq n} |\sum_{i=1}^t (\xi_{ik} - U_{ik, \xi})| = \mathcal{O}_{a.s.}(n^{\beta_1})$ . Next, denote

$$\begin{aligned} \mathbb{Q}_{3, \ell k} &= (nN)^{-1} \sum_{t=1}^{n-1} \sum_{i=1}^t (\xi_{ik} - U_{ik, \xi}) \sum_{j=1}^N B_{\ell, p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{tj, \varepsilon}, \\ \mathbb{Q}_{4, \ell k} &= -(nN)^{-1} \sum_{t=1}^{n-1} \sum_{i=1}^t (\xi_{ik} - U_{ik, \xi}) \sum_{j=1}^N B_{\ell, p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{(t+1)j, \varepsilon}, \\ \mathbb{Q}_{5, \ell k} &= (nN)^{-1} \sum_{t=1}^n (\xi_{tk} - U_{tk, \xi}) \sum_{j=1}^N B_{\ell, p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{nj, \varepsilon}. \end{aligned}$$

Denote the  $\sigma$ -field  $\mathcal{F}_\xi = \sigma \{ \xi_{ij}, i, j = 1, 2, \dots \}$ , then for  $1 \leq k \leq k_n, 1 \leq \ell \leq J_s + p$ , one has  $\mathbb{Q}_{3, \ell k} | \mathcal{F}_\xi =_D N \left( 0, \sigma_{lk, 3}^2 \right)$ , where

$$\sigma_{lk, 3}^2 = 2(nN)^{-2} \sum_{t=1}^{n-1} \left\{ \sum_{i=1}^t (\xi_{ik} - U_{ik, \xi}) \right\}^2 \sum_{j=1}^N B_{\ell, p}^2 \left( \frac{j}{N} \right) \sigma^2 \left( \frac{j}{N} \right).$$

Similar to Lemma A.7, applying (A.8), for any  $a > 2$

$$\Pr \left[ \max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |\mathbb{Q}_{3, \ell k}| / \sigma_{lk, 3} > a \{ \log k_n (J_s + p) \}^{1/2} \middle| \mathcal{F}_\xi \right] \leq 2 \{ k_n (J_s + p) \}^{1-a^2/2},$$

and hence

$$\Pr \left[ \max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |\mathbb{Q}_{3, \ell k}| / \sigma_{lk, 3} > a \{ \log k_n (J_s + p) \}^{1/2} \right] \leq 2 \{ k_n (J_s + p) \}^{1-a^2/2}.$$

Taking large enough  $a$ , according to Assumptions (A4) and (A6), one concludes with Borel-Cantelli Lemma that

$$\max_{\substack{1 \leq k \leq k_n \\ 1 \leq \ell \leq J_s + p}} |\mathbb{Q}_{3, \ell k}| / \sigma_{lk, 3} = \mathcal{O}_{a.s.} \left( \{ \log k_n (J_s + p) \}^{1/2} \right) = \mathcal{O}_{a.s.}(\log^{1/2} N). \quad (\text{B.4})$$



Noticing that  $\max_{1 \leq \ell \leq J_s+p} \|B_{\ell,p}\sigma\|_{2,N}^2 = \mathcal{O}(J_s^{-1})$ , one has

$$\begin{aligned} \sigma_{lk,3}^2 &= 2n^{-2} \left\{ \sum_{t=1}^{n-1} \sum_{i=1}^t (\xi_{ik} - U_{ik,\xi}) \right\}^2 N^{-2} \sum_{j=1}^N B_{\ell,p}^2 \left( \frac{j}{N} \right) \sigma^2 \left( \frac{j}{N} \right) \\ &\leq 2N^{-1} \|B_{\ell,p}\sigma\|_{2,N}^2 \left\{ (n-1) \times \max_{1 \leq t \leq n} \left| n^{-1} \sum_{i=1}^t (\xi_{ik} - U_{ik,\xi}) \right| \right\}^2 \leq cJ_s^{-1} N^{-1} n^{2\beta_1-1}. \end{aligned} \quad (\text{B.5})$$

Putting together the bounds in (B.4) and (B.5), one obtains that

$$\max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} |Q_{3,lk}| = \mathcal{O}_{a.s.} \left( n^{\beta_1-1/2} J_s^{-1/2} N^{-1/2} \log^{1/2} N \right).$$

Thus, one can show the following similarly,

$$\max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} |Q_{4,lk}| = \mathcal{O}_{a.s.} \left( n^{\beta_1-1/2} N^{-1/2} J_s^{-1/2} \log^{1/2} N \right).$$

$$\max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} |Q_{5,lk}| = \mathcal{O}_{a.s.} \left( n^{\beta_1-1/2} N^{-1/2} J_s^{-1/2} \log^{1/2} N \right).$$

Therefore, the lemma holds by noticing that

$$\begin{aligned} &\max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} \left| n^{-1} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) N^{-1} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{ij,\varepsilon} \right| \\ &\leq \max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} |Q_{3,lk}| + \max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} |Q_{4,lk}| + \max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} |Q_{5,lk}|. \blacksquare \end{aligned}$$

PROOF OF LEMMA A.9. Simple algebra provides that

$$\begin{aligned} &\sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \left\{ \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \\ &= \sum_{t=1}^{n-1} \sum_{i=1}^t (\xi_{ik} - U_{ik,\xi}) \left\{ \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{tj} - U_{tj,\varepsilon}) \right\} \\ &\quad - \sum_{t=1}^{n-1} \sum_{i=1}^t (\xi_{ik} - U_{ik,\xi}) \left\{ \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{(t+1)j} - U_{(t+1)j,\varepsilon}) \right\} \\ &\quad - \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \left\{ \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{nj} - U_{nj,\varepsilon}) \right\}. \end{aligned}$$

By noticing that

$$\begin{aligned} &\max_{1 \leq t \leq n} \max_{1 \leq k \leq k_n} \left| n^{-1} \sum_{i=1}^t (\xi_{ik} - U_{ik,\xi}) \right| = \mathcal{O}_{a.s.} \left( n^{\beta_1-1} \right), \\ &\max_{1 \leq t \leq n} \max_{1 \leq \ell \leq J_s+p} \left| N^{-1} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{tj} - U_{tj,\varepsilon}) \right| = \mathcal{O}_{a.s.} (N^{\beta_2-1}), \end{aligned}$$

one has

$$\begin{aligned}
& \max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s + p} \left| n^{-1} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \left\{ N^{-1} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \right| \\
& \leq n \max_{1 \leq k \leq k_n} \left| n^{-1} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \right| \max_{1 \leq \ell \leq J_s + p} \max_{1 \leq i \leq n} \left| N^{-1} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right| \\
& = \mathcal{O}_{a.s.}(n^{\beta_1} N^{\beta_2 - 1}).
\end{aligned}$$

The lemma holds. ■

PROOF OF LEMMA A.10. By the definition of  $\tilde{\varepsilon}_i(x)$  in (A.2), one has

$$Z_i(x+h)\tilde{\varepsilon}_i(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \left\{ Z_i(x+h) \sigma \left( \frac{j}{N} \right) \varepsilon_{ij} \right\}_{j=1}^N,$$

which implies that

$$\begin{aligned}
\sum_{i=1}^n Z_i(x+h)\tilde{\varepsilon}_i(x) &= \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{\infty} \phi_k(x+h) \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right\} \\
&= \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \left( \sum_{k=1}^{k_n} + \sum_{k=k_n+1}^{\infty} \right) \phi_k(x+h) \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right\}.
\end{aligned}$$

First, uniformly for  $1 \leq \ell \leq J_s + p, h \in [0, h_0]$ , one has

$$\left| \frac{1}{nN} \sum_{k=k_n+1}^{\infty} \phi_k(x+h) \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right| \leq D,$$

where  $D = \max_{1 \leq \ell \leq J_s + p} \sum_{k=k_n+1}^{\infty} \|\phi_k\|_\infty (nN)^{-1} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p}(j/N) \sigma \left( \frac{j}{N} \right) |\xi_{ik}| |\varepsilon_{ij}|$ , and

$$\text{ED} \leq \max_{1 \leq \ell \leq J_s + p} \sum_{k=k_n+1}^{\infty} \|\phi_k\|_\infty (nN)^{-1} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \mathbb{E} |\xi_{ik}| \mathbb{E} |\varepsilon_{ij}| \leq c J_s^{-1} \sum_{k=k_n+1}^{\infty} \|\phi_k\|_\infty.$$

As Assumption (A4) guarantees that  $\sum_{k=k_n+1}^{\infty} \|\phi_k\|_\infty \ll n^{-1/2}$ , while  $\|\mathbf{V}_{n,p}^{-1}\| \leq C J_s$  for large  $N$ , one has

$$\sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=k_n+1}^{\infty} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right\} \right| = \mathcal{O}_p(n^{-1/2}). \quad (\text{B.6})$$

Next, one bounds the sum  $\sum_{k=1}^{k_n}$ . According to Assumption (A4), (A.3) and Lemma A.7, one has

$$\begin{aligned} & \sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n \xi_{ik} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \right| \\ &= \mathcal{O}_p(J_s n^{-1/2} N^{\beta_2 - 1}). \end{aligned}$$

In addition, by Lemma A.8, one obtains that

$$\begin{aligned} & \sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{ij,\varepsilon} \right\} \right| \\ &= \mathcal{O}_p(n^{\beta_1 - 1/2} N^{-1/2} J_s^{1/2}). \end{aligned}$$

Note that  $U_{ik,\xi}$  and  $U_{ij,\varepsilon}$  are independent standard normal random variables. Similar to (B.1), one obtains that

$$\max_{1 \leq \ell \leq J_s + p} \left| \frac{1}{n} \sum_{i=1}^n U_{ik,\xi} \left\{ \frac{1}{N} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{ij,\varepsilon} \right\} \right| = \mathcal{O}_{a.s.} \left( N^{-1/2} J_s^{-1/2} n^{-1/2} \sqrt{\log N} \right).$$

It is easy to see that

$$\begin{aligned} & \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right\} \right| \\ & \leq \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \right| \\ & \quad + \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n U_{ik,\xi} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) (\varepsilon_{ij} - U_{ij,\varepsilon}) \right\} \right| \\ & \quad + \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{n} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) N^{-1} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{ij,\varepsilon} \right\} \right| \\ & \quad + \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{k_n} \phi_k(x+h) \frac{1}{n} \sum_{i=1}^n U_{ik,\xi} \left\{ N^{-1} \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) U_{ij,\varepsilon} \right\} \right\} \right|. \end{aligned}$$

Therefore, combining Lemmas A.6–A.9, one has

$$\begin{aligned} & \sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=k_n+1}^{\infty} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right\} \right| \\ &= \mathcal{O}_{a.s.} \left( J_s n^{-1/2} N^{\beta_2 - 1} \log^{1/2} N \right) + \mathcal{O}_{a.s.} \left( n^{\beta_1 - 1/2} N^{-1/2} J_s^{1/2} \log^{1/2} N \right) \\ & \quad + \mathcal{O}_{a.s.} \left( J_s n^{\beta_1} N^{\beta_2 - 1} \right) + \mathcal{O}_{a.s.} \left( n^{-1/2} N^{-1/2} J_s^{1/2} \log^{1/2} N \right). \end{aligned}$$

Hence, the proof is completed by noticing that

$$\begin{aligned}
& \sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n Z_i(x+h) \tilde{\varepsilon}_i(x) \right| \\
&= \sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{\infty} \phi_k(x+h) \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N B_{\ell,p} \left( \frac{j}{N} \right) \sigma \left( \frac{j}{N} \right) \xi_{ik} \varepsilon_{ij} \right\} \right| \\
&= \mathcal{O}_p(n^{-1/2}) + \mathcal{O}_{a.s.} \left( J_s n^{-1/2} N^{\beta_2-1} \log^{1/2} N \right) + \mathcal{O}_{a.s.} \left( n^{\beta_1-1/2} N^{-1/2} J_s^{1/2} \log^{1/2} N \right) \\
&\quad + \mathcal{O}_{a.s.} \left( J_s n^{\beta_1} N^{\beta_2-1} \right) + \mathcal{O}_{a.s.} \left( n^{-1/2} N^{-1/2} J_s^{1/2} \log^{1/2} N \right) = \mathcal{O}_p(n^{-1/2}). \blacksquare
\end{aligned}$$

PROOF OF LEMMA A.11. Notice that

$$\begin{aligned}
Z_i(x+h) n^{-1} \sum_{i'=1}^n \tilde{Z}_{i'}(x) &= N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \left\{ Z_i(x+h) n^{-1} \sum_{i'=1}^n \sum_{k=1}^{\infty} \xi_{i'k} \phi_k(j/N) \right\}_{j=1}^N \\
&= \frac{1}{N} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \left\{ \sum_{k=1}^{\infty} \xi_{ik} \phi_k(x+h) \frac{1}{n} \sum_{i'=1}^n \sum_{k'=1}^{\infty} \xi_{i'k'} \phi_{k'}(j/N) \right\}_{j=1}^N. \quad (\text{B.7})
\end{aligned}$$

Let  $\bar{\xi}_{\cdot k} = n^{-1} \sum_{i=1}^n \xi_{ik}$ , then

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n Z_i(x+h) \frac{1}{n} \sum_{i'=1}^n \tilde{Z}_{i'}(x) &= \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{\infty} \phi_k(x+h) \bar{\xi}_{\cdot k} \sum_{k'=1}^{\infty} \langle B_{\ell,p}, \phi_{k'} \rangle_N \bar{\xi}_{\cdot k'} \right\}_{\ell=1}^{J_s+p} \\
&= \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{\infty} \phi_k(x+h) \bar{\xi}_{\cdot k} \sum_{k'=1}^{\infty} \langle B_{\ell,p}, \phi_{k'} \rangle_N \bar{\xi}_{\cdot k'} \right\}_{\ell=1}^{J_s+p}.
\end{aligned}$$

Now uniformly for  $1 \leq \ell \leq J_s + p$ ,  $h \in [0, h_0]$ , one has

$$\left| \sum_{k=1}^{\infty} \phi_k(x+h) \bar{\xi}_{\cdot k} \sum_{k'=1}^{\infty} \langle B_{\ell,p}, \phi_{k'} \rangle_N \bar{\xi}_{\cdot k'} \right| \leq C J_s^{-1} \mathbf{S},$$

where  $\mathbf{S} = \sum_{k=1}^{\infty} \|\phi_k\|_{\infty} \sum_{k'=1}^{\infty} \|\phi_{k'}\|_{\infty} |\bar{\xi}_{\cdot k} \bar{\xi}_{\cdot k'}|$ , and according to (A5) in Cao et al. (2012),

$$\mathbf{E} \mathbf{S} \leq \sum_{k=1}^{\infty} \|\phi_k\|_{\infty} \sum_{k'=1}^{\infty} \|\phi_{k'}\|_{\infty} \max_{1 \leq k \leq \infty} \mathbf{E} |\bar{\xi}_{\cdot k}|^2 \leq cn^{-1} \sum_{k=1}^{\infty} \|\phi_k\|_{\infty} \sum_{k'=1}^{\infty} \|\phi_{k'}\|_{\infty}.$$

As Assumption (A4) guarantees that  $\sum_{k=1}^{\infty} \|\phi_k\|_{\infty} < \infty$ , while  $\|\mathbf{V}_{n,p}^{-1}\| \leq C J_s$  for large  $N$ , one obtains that

$$\sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \left\{ \sum_{k=1}^{\infty} \phi_k(x+h) \bar{\xi}_{\cdot k} \sum_{k'=1}^{\infty} \langle B_{\ell,p}, \phi_{k'} \rangle_N \bar{\xi}_{\cdot k'} \right\}_{\ell=1}^{J_s+p} \right| = \mathcal{O}_p(n^{-1/2}). \quad (\text{B.8})$$

The lemma holds. ■

PROOF OF LEMMA A.12. Note that

$$\begin{aligned} \Pr \left( \max_{1 \leq i \leq n} \left| \frac{W_i}{\sigma_i} \right| > a\sqrt{\log n} \right) &\leq \sum_{i=1}^n \Pr \left( \left| \frac{W_i}{\sigma_i} \right| > a\sqrt{\log n} \right) \leq 2n \left\{ 1 - \Phi \left( a\sqrt{\log n} \right) \right\} \\ &< 2n \frac{\phi \left( a\sqrt{\log n} \right)}{a\sqrt{\log n}} \leq 2n\phi \left( a\sqrt{\log n} \right) = \sqrt{2/\pi} n^{1-a^2/2}, \end{aligned}$$

for  $n \rightarrow \infty$ ,  $a > 2$ , which proves (A.8). The lemma follows by applying Borel-Cantelli Lemma with choice of  $a > 2$ . ■

## C. More Simulation Results and Findings

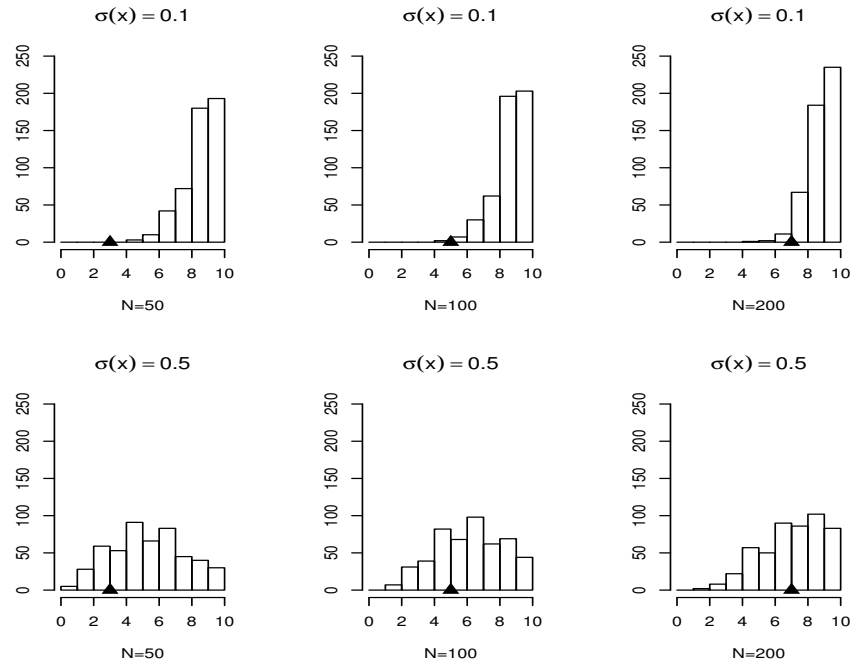
### C.1 A simulation study to evaluate the knots selection methods

In this section, we conduct a simulation study to evaluate the performance of the knots selection methods proposed in Section 4.2 in the main paper. The setting of the simulation is the same as in Section 5.1 in the main paper. For model fitting, the mean function is estimated by cubic splines, and the number of knots of the splines,  $J_s$ , is selected using either the formula-based method (Formula) and the GCV method (GCV) described in Section 4.2. Each simulation is repeated 500 times.

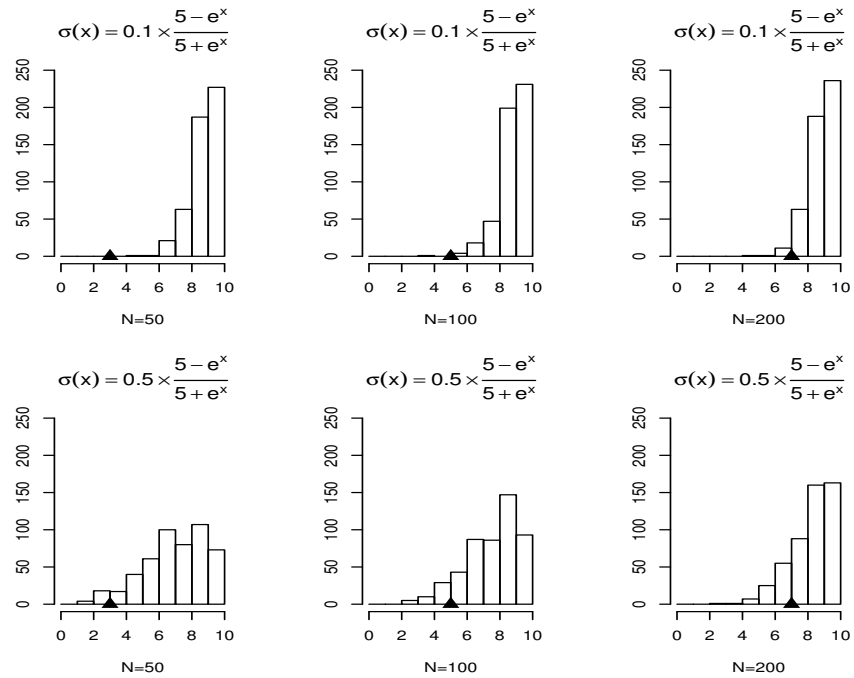
Figure C.1 below shows the frequency bar plot of the GCV-selected  $J_s$  over 500 replications, where the black triangles indicate the number of knots suggested using the formula given in Section 4.2. From Figure C.1, one sees that on average the GCV method tends to select a slightly larger number of knots than the formula method does, but both methods provide similar results as shown in Tables 1 and 2 in the main paper. The GCV method is indeed more time-consuming than the formula method. For example, in scenario  $N = 50$  and  $\sigma_\epsilon = 0.1$  of Table 1, it takes 50 seconds for the formula and 9 minutes for GCV selected method, respectively.

### C.2 More results for spatial covariance models

Tables C.1–C.2 report some simulation results based on the spatial covariance model presented in Section 5.2 in the main paper. Specifically, we report the simulation results based on the data generated from the model with the heteroscedastic errors:  $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$



(a)



(b)

Figure C.1: Bar graphs of the GCV selected number of knots in 500 replications with (a) homogeneous and (b) heteroscedastic errors. Black triangles indicate the number of knots suggested by the formula.

Table C.1: Simulation results based on heteroscedastic errors with  $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$  for M1, and  $\sigma(x) = \sigma_\epsilon \{30 + \exp(x/2)\}^{-1} \{30 - \exp(x/2)\}$  for M2 and M3: AMSE of the estimators  $\widehat{C}$ ,  $\widetilde{C}$ ; CR (outside/inside of the parentheses is based on  $\widehat{C}$ ,  $\widetilde{C}$ ), and the WD of the asymptotic SCBs based on  $\widehat{C}$ .

$\sigma_\epsilon$	Model	$N$	AMSE		SCB			
			$\widehat{C}$	$\widetilde{C}$	95%		99%	
					CR	WD	CR	WD
0.1	M1	50	0.082	0.081	0.916(0.920)	1.37	0.958(0.966)	1.68
		100	0.040	0.040	0.922(0.926)	0.99	0.974(0.978)	1.21
		200	0.019	0.018	0.940(0.952)	0.72	0.980(0.986)	0.87
	M2	50	0.095	0.096	0.906(0.904)	1.44	0.950(0.954)	1.78
		100	0.048	0.049	0.922(0.926)	1.05	0.980(0.976)	1.30
		200	0.022	0.022	0.958(0.958)	0.76	0.992(0.994)	0.94
	M3	50	0.109	0.109	0.904(0.908)	1.50	0.954(0.956)	1.86
		100	0.055	0.055	0.922(0.928)	1.09	0.976(0.978)	1.35
		200	0.025	0.025	0.960(0.958)	0.79	0.988(0.990)	0.98
0.5	M1	50	0.082	0.081	0.898(0.918)	1.38	0.962(0.970)	1.69
		100	0.040	0.040	0.918(0.924)	0.99	0.976(0.980)	1.21
		200	0.019	0.018	0.948(0.952)	0.72	0.982(0.986)	0.88
	M2	50	0.096	0.096	0.904(0.908)	1.45	0.948(0.958)	1.79
		100	0.048	0.049	0.916(0.926)	1.06	0.974(0.980)	1.30
		200	0.022	0.022	0.960(0.956)	0.77	0.990(0.994)	0.94
	M3	50	0.110	0.109	0.904(0.908)	1.51	0.948(0.958)	1.87
		100	0.055	0.055	0.910(0.924)	1.10	0.976(0.976)	1.36
		200	0.025	0.025	0.960(0.958)	0.79	0.988(0.990)	0.98

for M1, and  $\sigma(x) = \sigma_\epsilon \{30 + \exp(x/2)\}^{-1} \{30 - \exp(x/2)\}$  for M2 and M3. The number of curves  $n = \lfloor 0.8N \rfloor$  with  $N = 50, 100$  and  $200$ , and the noise levels are  $\sigma_\epsilon = 0.1, 0.5$ . The mean function is estimated by cubic splines, i.e.,  $p = 4$ , with the number of knots selected using the formula method.

The AMSE of the covariance estimators  $\widehat{C}$  and  $\widetilde{C}$  are reported in columns 3–4 of Table C.1. The performance of the two estimators is very similar. Columns 5 and 7 present the empirical coverage rate CR, i.e., the percentage of the true curve  $C(\cdot)$  entirely covered by the SCB, based on 95% and 99% confidence levels, respectively. As the sample size increases, the coverage probability of the SCB becomes closer to the nominal level. Columns 3–4 of Table C.2 present the AMSEs of  $\widehat{G}^{\text{PROP}}(x, x')$  and  $\widehat{G}^{\text{TPS}}(x, x')$ . The results of AMSEs indicate that  $\widehat{G}^{\text{PROP}}$  is more accurate than  $\widehat{G}^{\text{TPS}}$ , while  $\widehat{G}^{\text{TPS}}$  usually gives larger AMSE. Columns 5–12 of Table 4 report the CR and WD of SCE-I and SCE-II. One sees that the CRs of SCE-I are much closer to the nominal levels than those of SCE-II, and increasing the sample size helps to improve the CR of the SCEs to their nominal levels. One also observes the widths of the SCE-I are much narrower than those of the SCE-II.

Table C.2: Simulation results based on heteroscedastic errors with  $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$  for M1, and  $\sigma(x) = \sigma_\epsilon \{30 + \exp(x/2)\}^{-1} \{30 - \exp(x/2)\}$  for M2 and M3: AMSE of the estimators  $\widehat{G}^{\text{PROP}}(\cdot, \cdot)$ ,  $\widehat{G}^{\text{TPS}}(\cdot, \cdot)$ ; CR and WD of SCE-I and SCE-II.

$\sigma_\epsilon$	Model	$N$	AMSE		SCE-I				SCE-II			
			$\widehat{G}^{\text{PROP}}$	$\widehat{G}^{\text{TPS}}$	95%		99%		95%		99%	
					CR	WD	CR	WD	CR	WD	CR	WD
0.1	M1	50	0.079	0.121	0.916	1.40	0.958	1.71	0.722	2.07	0.838	2.55
		100	0.039	0.061	0.922	1.01	0.974	1.24	0.850	1.64	0.952	2.03
		200	0.018	0.032	0.940	0.73	0.980	0.90	0.880	1.19	0.964	1.46
	M2	50	0.097	0.150	0.906	1.50	0.950	1.86	0.710	2.08	0.830	2.56
		100	0.048	0.071	0.922	1.10	0.980	1.35	0.786	1.63	0.900	2.00
		200	0.022	0.036	0.958	0.79	0.992	0.98	0.930	1.16	0.976	1.43
	M3	50	0.114	0.153	0.904	1.57	0.954	1.95	0.720	2.15	0.816	2.64
		100	0.057	0.074	0.922	1.15	0.976	1.42	0.856	1.52	0.944	1.86
		200	0.026	0.039	0.960	0.83	0.988	1.03	0.878	1.09	0.956	1.33
0.5	M1	50	0.079	0.129	0.902	1.41	0.964	1.73	0.730	2.10	0.834	2.57
		100	0.039	0.064	0.920	1.02	0.980	1.25	0.810	1.65	0.918	2.03
		200	0.018	0.035	0.942	0.74	0.984	0.90	0.914	1.21	0.980	1.49
	M2	50	0.097	0.143	0.904	1.51	0.948	1.86	0.674	2.10	0.802	2.57
		100	0.048	0.071	0.916	1.10	0.972	1.35	0.774	1.63	0.894	2.01
		200	0.022	0.036	0.960	0.79	0.990	0.98	0.924	1.16	0.970	1.43
	M3	50	0.114	0.149	0.900	1.58	0.948	1.96	0.696	2.12	0.820	2.60
		100	0.057	0.075	0.910	1.15	0.974	1.42	0.830	1.52	0.930	1.87
		200	0.026	0.038	0.960	0.83	0.988	1.03	0.876	1.09	0.944	1.33

## D. Additional Real Data Analysis

### D.1 Biscuit Dough Piece Data

In the following, we apply the methodology to a “biscuit dough piece data”, which is an experiment involved varying the composition of biscuit dough pieces by measuring the quantitative near infrared reflectance (NIR) spectroscopy. Quantitative NIR spectroscopy plays an important role in analyzing such diverse materials as food and drink, pharmaceutical products, and petrochemicals. The NIR spectrum of a sample of, say, wheat flour is a continuous curve measured by modern scanning instruments at hundreds of equally spaced wavelengths. Then the information contained in the curve can be used to predict the chemical composition of the sample. For a full description of the experiment, see Osborne et al. (1984). This dataset is available in the R package “fds” (<https://cran.r-project.org/web/packages/fds/fds.pdf>), which contains several subsets such as “nirc” (calibration) and “nirp” (prediction).

We focus on the calibration set “nirc”, which contains 40 doughs and 700 point NIR spectra for each dough. According to Brown et al. (2001), the observation number 23 in the calibration set appears as an outlier. Thus we remove it and take the other samples (doughs). Hence, there



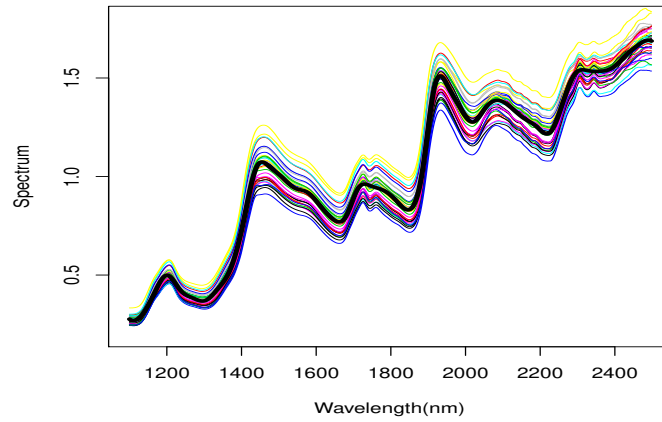


Figure D.1: Biscuit dough pieces data curves (thin solid lines) with their mean function estimator (thick solid line).

are measurements on  $n = 39$  samples, where for each sample  $N = 700$  spectral was recorded every 2 nanometre (nm) with wavelength being measured on  $[1100, 2498]$ . Denote by  $Y_{ij}$  the spectral of the  $i$ th sample at the wavelength  $x_j$ ,  $j = 1, \dots, N$  and  $i = 1, \dots, n$ .

Figure D.1 displays NIR spectra curves together with their estimated mean curve, and “Wavelength (nm)” is plotted on  $x$ -axis and “Spectrum” on the  $y$ -axis. An illustrative 3D plot of SCE of the Biscuit dough piece data is depicted is in Figure D.2: (a) based on  $\hat{G}^{\text{PROP}}$  (middle) and its 95% SCE (upper and lower); (b) based on  $\hat{G}^{\text{TPS}}$  (middle) with its 95% asymptotic confidence envelop (up and below).

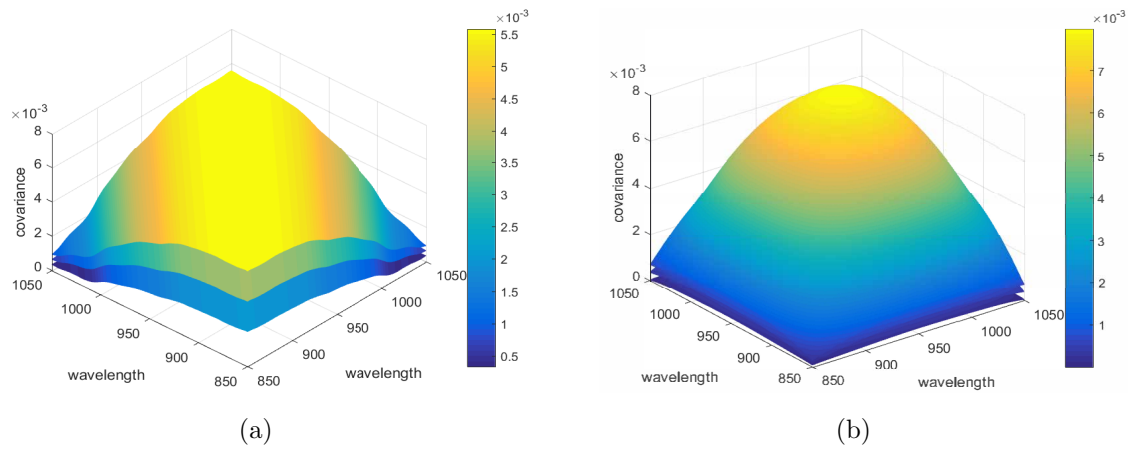


Figure D.2: Biscuit dough piece data plots: (a) covariance estimator  $\hat{G}^{\text{PROP}}$  and its 95% SCE; (b) covariance estimator  $\hat{G}^{\text{TPS}}$  and its 95% SCE.

# Bibliography

- Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC Press.
- Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, vol. 110, Springer-Verlag New York.
- Brown, P., Fearn, T., and Vannucci, M. (2001), “Bayesian wavelet regression on curves with application to a spectroscopic calibration problem,” *J. Amer. Statist. Assoc.*, 96, 398–408.
- Cao, G., Wang, L., Li, Y., and Yang, L. (2016), “Oracle-efficient confidence envelopes for covariance functions in dense functional data,” *Statist. Sinica*, 26, 359–383.
- Cao, G., Yang, L., and Todem, D. (2012), “Simultaneous inference for the mean function based on dense functional data,” *J. Nonparametr. Stat.*, 24, 359–377.
- Cardot, H. (2000), “Nonparametric estimation of smoothed principal components analysis of sampled noisy functions,” *J. Nonparametr. Stat.*, 12, 503–538.
- Choi, I., Li, B., and Wang, X. (2013), “Nonparametric estimation of spatial and space-time covariance function,” *J. Agric. Biol. Environ. Stat.*, 18, 611–630.
- Crainiceanu, C. M., Staicu, A. M., and Di, C. Z. (2009), “Generalized multilevel functional regression,” *J. Amer. Statist. Assoc.*, 104, 1550–1561.
- Csörgö, M. and Révész, P. (1981), *Strong Approximations in Probability and Statistics*, Academic Press, New York-London.
- de Boor, C. (2001), *A Practical Guide to Splines*, Springer, New York.
- Diggle, P. and Verbyla, A. (1998), “Nonparametric estimation of covariance structure in longitudinal data,” *Biometrics*, 54, 401–415.

- Ferraty, F. and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, Springer, Berlin.
- Fryzlewicz and Ombao (2009), “Consistent classification of nonstationary time series using stochastic wavelet representations,” *J. Amer. Statist. Assoc.*, 104, 299–312.
- Goldsmith, J., Greven, S., and Crainiceanu, C. (2013), “Corrected confidence bands for functional data using principal components,” *Biometrics*, 69, 41–51.
- Guo, J., Zhou, B., and Zhang, J. T. (2018), “Testing the equality of several covariance functions for functional data: A supremum-norm based test,” *Comput. Statist. Data Anal.*, 124, 15–26.
- Hall, P., Fisher, N. I., and Hoffmann, B. (1994), “On the nonparametric estimation of covariance functions,” *Ann. Statist.*, 22, 2115–2134.
- Hall, P., Müller, H. G., and Wang, J. L. (2006), “Properties of Principal Component Methods for Functional and Longitudinal Data Analysis,” *Ann. Statist.*, 34, 1493–1517.
- Horváth, L., Kokoszka, P., and Reeder, R. (2013), “Estimation of the mean of functional time series and a two-sample problem,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 75, 103–122.
- Hsing, T. and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear operators*. Wiley Series in Probability and Statistics, Wiley, Chichester.
- James, G., Hastie, T., and Sugar, C. (2000), “Principal component models for sparse functional data,” *Biometrika*, 87, 587–602.
- Li, Y. and Hsing, T. (2010), “Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data,” *Ann. Statist.*, 38, 3321–3351.
- Olshen, R. A., Biden, E. N., Wyatt, M. P., and Sutherland, D. H. (1989), “Gait Analysis and the Bootstrap,” *Ann. Statist.*, 17, 1419–1440.
- Osborne, B. G., Fearn, T., Miller, A. R., and Douglas, S. (1984), “Application of near infrared reflectance spectroscopy to the compositional analysis of biscuits and biscuit doughs,” *J. Sci. Food Agric.*, 35, 99–105.
- Pantle, U., Schmidt, V., and Spodarev, E. (2010), “On the estimation of integrated covariance functions of stationary random fields,” *Scand. J. Statist.*, 37, 47–66.

- Peng, J. and Paul, D. (2009), “A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data,” *Journal of Computational and Graphical Statistics*, 18, 995–1015.
- Ramsay, J. and Dalzell, C. (1991), “Some tools for functional data analysis,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 53, 539–572.
- Ramsay, J. and Silverman, B. (2005), *Functional Data Analysis*, Springer, New York.
- Sanderson, J., Fryzlewicz, P., and Jones, M. W. (2010), “Estimating linear dependence between nonstationary time series using the locally stationary wavelet model,” *Biometrika*, 97, 435–446.
- Song, Q. and Yang, L. (2009), “Spline confidence bands for variance functions,” *J. Nonparametr. Stat.*, 21, 589–609.
- Tsyrlunikov, M. and Gayfulin, D. (2017), “A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications,” *Meteorologische Zeitschrift*, 5, 549–566.
- Wang, J. (2012), “Modelling time trend via spline confidence band,” *Ann. Inst. Statist. Math.*, 64, 275–301.
- Wang, J. and Yang, L. (2009), “Polynomial spline confidence bands for regression curves,” *Statist. Sinica*, 19, 325–342.
- Yang, J., Zhu, H., Choi, T., and Cox, D. D. (2016), “Smoothing and mean–covariance estimation of functional data with a Bayesian hierarchical model,” *Bayesian Anal.*, 11, 649–670.
- Yao, F., Müller, H. G., and Wang, J. L. (2005), “Functional data analysis for sparse longitudinal data,” *J. Amer. Statist. Assoc.*, 100, 577–590.
- Yin, J., Geng, Z., Li, R., and Wang, H. (2010), “Nonparametric covariance model,” *Statist. Sinica*, 20, 469–479.