

Simultaneous confidence band for stationary covariance function of dense functional data

Wang, Jiangyan^a, Cao, Guanqun^{b,*}, Wang, Li^{c,**}, Yang, Lijian^{d,**}

^aNanjing Audit University, Nanjing, Jiangsu 211815, China

^bAuburn University, Auburn, AL 36849, USA

^cIowa State University, Ames, IA 50011, USA

^dTsinghua University, Beijing 100084, 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.

Keywords: Confidence band, Covariance function, Functional data, Stationary process.

2010 MSC: Primary 62H20, Secondary time series 62G08

1. Introduction

Since [1] 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 [2], [3], 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, [4–8] 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 [9–13]. In this sense, [14] proposed a simultaneous confidence envelope of covariance function for functional data; [15] proposed a consistent estimator for the long-run covariance operator of stationary time series; [16] 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, [9].

*Co-first author.

**Co-Corresponding author.

Email addresses: lilywang@iastate.edu (Wang, Li), yanglijian@tsinghua.edu.cn (Yang, Lijian)

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

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

20 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 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 $L^2(\mathcal{X})$, 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 $\eta(x)$, $x \in \mathcal{X}$, 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 ξ_k , called functional principal component (FPC) scores, are uncorrelated with each other of mean 0 and variance 1. The rescaled eigenfunctions, ϕ_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 $\eta_i(x)$, $x \in \mathcal{X}$, is written as $\eta_i(x) = m(x) + \sum_{k=1}^\infty \xi_{ik} \phi_k(x)$, in which the FPC scores $\{\xi_{ik}\}_{k=1}^\infty$, $i \in \{1, \dots, n\}$, 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

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

where ε_{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 $\mathcal{X} = [0, 1]$, and $X_{ij} = x_j = j/N$, $1 \leq j \leq N$. This type of functional data was considered in [17], [18] and [14], among others. Consequently, our observed data can be written as

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

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 [19, 20]. One may further ask if these random observations at regular grid points would even admit the structure of stationary time series. [14], 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 [21]. 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. The direct smoothing is typically pointwise and nonparametric, while the mixed-effects type usually involves parametric models and hence approximates covariance functions with analytical expressions. For functional data, FPC analysis has become one of the first-line methods; see, for instance, the nonparametric estimation of covariance functions: [8, 14, 17, 22], among others. For longitudinal data, [23] considered reduced rank spline mixed-effects models to describe the modes of the variation; [24] proposed a geometric approach within the framework of marginal maximum likelihood estimation by requiring the trajectories are i.i.d. Gaussian processes.

In this paper, we consider a nonparametric estimation of the stationary 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 observations (a vital feature for us to borrow strength); the second step estimates the covariance function through smoothing using the residuals of the first step. The proposed covariance estimator is smooth and as efficient as the oracle estimator constructed from true $\eta_i(\cdot)$ and the mean $m(\cdot)$.

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 [2]. 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, [25] 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 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, [14] proposed SCBs for the covariance functions, and [7] 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 asymptotically exact SCBs.

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, proofs of the main theoretical results and additional simulation results are presented in Appendices A and B. The full version of the paper can be found in [26].

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

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

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

In such case, a sample-based consistent estimator can be employed, such as the spline smoother proposed in [5]. Denote by $\{t_\ell\}_{\ell=1}^{J_s}$ a sequence of equally-spaced points, $t_\ell = \ell / (J_s + 1)$, $\ell \in \{1, \dots, J_s\}$, $0 < t_1 < \dots < t_{J_s} < 1$, called interior knots, which divide the interval $[0, 1]$ into $(J_s + 1)$ equal subintervals $I_0 = [0, t_1]$, $I_\ell = [t_\ell, t_{\ell+1}]$, $\ell \in \{1, \dots, J_s - 1\}$, $I_{J_s} = [t_{J_s}, 1]$. 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_ℓ , $\ell \in \{0, \dots, J_s\}$, which consists of all $(p - 2)$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $(p - 1)$

on subintervals I_ℓ , $\ell \in \{0, \dots, J_s\}$. Following the notation in [27], 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

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

105 One can then estimate the unknown mean function $m(\cdot)$ as $\widehat{m}(x) = n^{-1} \sum_{i=1}^n \widehat{\eta}_i(x)$, and obtain the covariance estimator

$$\widehat{C}(h) = \frac{1}{1-h} \int_0^{1-h} \frac{1}{n} \sum_{i=1}^n \widehat{Z}_i(x) \widehat{Z}_i(x+h) dx, \quad h \in [0, h_0]. \quad (6)$$

3. Asymptotic Properties

This section studies the asymptotic properties for the proposed estimators.

3.1. Assumptions

110 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 χ , 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 $\langle \phi, \varphi \rangle_N = N^{-1} \sum_{j=1}^N \phi(j/N) \varphi(j/N)$. Correspondingly
115 and respectively, 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 μ -Hölder continuous functions, i.e.,

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

We next introduce some technical assumptions.

- (C1) There exists 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 $p^* = q + \mu$.
- (C2) 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_σ ,
120 $M_0 > 0$, $\sup_{x \in [0,1]} \sigma(x) \leq M_\sigma$, $\sup_{h \in [0, h_0]} |C(h)| \leq M_0$.
- (C3) There exists a constant $\theta > 0$, such that as $N \rightarrow \infty$, $n = n(N) \rightarrow \infty$, $n = O(N^\theta)$.
- (C4) 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 $n \rightarrow \infty$, $\sum_{k_n+1}^\infty \|\phi_k\|_\infty = o(n^{-1/2})$ and $k_n = O(n^\omega)$ for some $\omega > 0$.
- (C5) There are constants $c_1, c_2 \in (0, +\infty)$, $\gamma_1, \gamma_2 \in (1, +\infty)$, $\beta_1, \beta_2 \in (0, 1/2)$, and i.i.d $N(0, 1)$ variables
125 $\{U_{ij,\varepsilon}\}_{i=1,j=1}^{n,N}$, $\{U_{ik,\xi}\}_{i=1,k=1}^{n,k_n}$ such that $\min\{2(3 + (1 + \beta_1)p^*)^{-1}(1 - \beta_2)p^*, 2(1 + \beta_1)^{-1}(\nu - \beta_2)\} > \theta$ for the index ν in Assumption (C2), p^* in Assumption (C1), and

$$\Pr \left\{ \max_{1 \leq k \leq k_n} \max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t U_{ik,\xi} \right| > n^{\beta_1} \right\} < c_1 n^{-\gamma_1}, \quad \Pr \left\{ \max_{1 \leq i \leq n} \max_{1 \leq t \leq N} \left| \sum_{j=1}^t \varepsilon_{ij} - \sum_{j=1}^t U_{ij,\varepsilon} \right| > N^{\beta_2} \right\} < c_2 N^{-\gamma_2}.$$

(C5') The i.i.d 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 ω in Assumption (C4) and θ in Assumption (C3), such that $E|\varepsilon_{11}|^{r_1}$ and $E|\xi_{1k}|^{r_2}$, $k \in \{1, 2, \dots\}$, are finite.

(C6) The spline order $p \geq p^*$, the number of interior knots $J_s \asymp N^\gamma d_N$ for some $\tau > 0$ with $d_N + d_N^{-1} = O(\ln^\tau N)$ as $N \rightarrow \infty$, and for p^* in Assumption (C1), ν in Assumption (C2), θ in Assumption (C3), β_1, β_2 and γ_1 in Assumption (C5)

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

Assumptions (C1)–(C2) are standard in the literature, see [5] for instance. In particular, (C1) and (C4) control the size of the bias of the spline smoother for $m(\cdot)$ and $\phi_k(\cdot)$. Assumption (C2) ensures the variance function is a uniformly bounded function. Assumption (C3) regulates that sample size n grows as a fractional power θ of N , the number of observations per subject. The bounded smoothness of the principal components is guaranteed in Assumption (C4). Assumption (C5) provides a strong approximation of estimation errors and FPC scores. Assumption (C5') is an elementary assumption guaranteeing the high-level Assumption (C5). It is noteworthy that the smoothness of our estimator is controlled by the knots of the splines. Assumption (C6) specifies the number of knots 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), $\gamma = 3/8$, $d_N \asymp \ln \ln N$. 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 (6) 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 $\Delta(h) = \widetilde{C}(h) - C(h)$, $h \in [0, h_0]$.

By the definition of $C(h)$ and $\widetilde{C}(h)$ in (1) and (3), it is easy to show that $C(h) = (1-h)^{-1} \int_0^{1-h} \sum_{k=1}^{\infty} \phi_k(x) \phi_k(x+h) dx$, and $\widetilde{C}(h) = n^{-1} (1-h)^{-1} \sum_{i=1}^n \sum_{k,k'=1}^{\infty} \xi_{ik} \xi_{ik'} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx$. 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. Let

$$\begin{aligned} \Xi(h) &= \sum_{k,k'=1}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \right\}^2 + \sum_{k=1}^{\infty} (E\xi_{1k}^4 - 3) \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_k(x+h) dx \right\}^2 \\ &\quad + \sum_{k,k'=1}^{\infty} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_k(x) \phi_{k'}(x+h) dx \right\} \left\{ \frac{1}{1-h} \int_0^{1-h} \phi_{k'}(x) \phi_k(x+h) dx \right\}. \end{aligned} \quad (7)$$

Theorem 1. Under Assumptions (C1)–(C6), $\sup_{h \in [0, h_0]} |nE\{\Delta(h)\}^2 - \Xi(h)| = o(1)$.

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

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

From (3.2) in [14],

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

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 $\widetilde{C}(\cdot)$ is more efficient than the covariance estimator given in [14].

Proposition 1. *Under Assumptions (C1)–(C6), 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 $h, h' \in [0, h_0]$.

The proof of Proposition 1 is deferred to Section A.3 in Appendix A. Although the oracle smoother $\widetilde{C}(\cdot)$ 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 $\widehat{Z}_i(x) = \widehat{\eta}_i(x) - \widehat{m}(x)$ in place of $Z_i(x)$ in the covariance estimator is asymptotically negligible, that is, two-step estimator $\widehat{C}(\cdot)$ is as efficient as the infeasible estimator $\widetilde{C}(\cdot)$. See Section A.4 in Appendix A for the proof of Proposition 2.

Proposition 2. *Under Assumptions (C1)–(C6), $\sup_{h \in [0, h_0]} |\widehat{C}(h) - \widetilde{C}(h)| = o_p(n^{-1/2})$.*

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

Theorem 2. *Under Assumptions (C1)–(C6), $\sup_{h \in [0, h_0]} |\widehat{C}(h) - C(h) - \Delta(h)| = o_p(n^{-1/2})$.*

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

4. Simultaneous confidence band

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

4.1. Asymptotic SCB

The 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 $C(\cdot)$. It is a direct result of Propositions 1, 2 and Theorem 2, thus the proof is omitted.

Theorem 3. *Under Assumptions (C1)–(C6), for any $\alpha \in (0, 1)$,*

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

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

Corollary 1. *Under Assumptions (C1)–(C6), an asymptotic $100(1-\alpha)\%$ exact SCB for $C(\cdot)$ is given by $\widehat{C}(h) \pm n^{-1/2} Q_{1-\alpha} \Xi^{1/2}(h)$, $h \in [0, h_0]$. While an asymptotic pointwise confidence band for $C(\cdot)$ is given by $\widehat{C}(h) \pm n^{-1/2} z_{1-\alpha/2} \Xi^{1/2}(h)$, $h \in [0, h_0]$.*

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. One is criterion-based selection strategy such as Generalized Cross-Validation (GCV), where the candidate pool for J_s is all the integers between 1 and J_s^* , where $J_s^* = \min\{10, \lfloor n/4 \rfloor\}$. Specifically, the GCV is given as

$$\text{GCV}(J) = \frac{N^{-1} \left\| \bar{\mathbf{Y}} - \mathbf{B}_J (\mathbf{B}_J^\top \mathbf{B}_J)^{-1} \mathbf{B}_J^\top \bar{\mathbf{Y}} \right\|^2}{\left[1 - N^{-1} \text{tr} \left\{ \mathbf{B}_J (\mathbf{B}_J^\top \mathbf{B}_J)^{-1} \mathbf{B}_J^\top \right\} \right]^2},$$

where $\bar{\mathbf{Y}} = (\bar{Y}_{\cdot 1}, \dots, \bar{Y}_{\cdot N})^\top$, where $\bar{Y}_{\cdot j} = n^{-1} \sum_{i=1}^n Y_{ij}$, $\mathbf{B}_J(x) = \{B_{1,p}(x), \dots, B_{J+p,p}(x)\}^\top$, $J = 1, \dots, J_s^*$, and $\mathbf{B}_J = \{\mathbf{B}_J(1/N), \dots, \mathbf{B}_J(N/N)\}^\top$ is the $N \times (J+p)$ design matrix for spline regression. Then the number of knots selected is $J_s = \arg \min_J \text{GCV}(J)$. The second method to choose J_s is the formula based selection strategy as 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, μ) 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 $J_s = \lfloor cN^\gamma \{\ln \ln(N)\}^\gamma \rfloor$ 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

$$\widehat{G}(x, x') = n^{-1} \sum_{i=1}^n \widehat{Z}_i(x) \widehat{Z}_i(x') = \sum_{s=1}^{J_s+p} \sum_{s'=1}^{J_s+p} \widehat{\beta}_{ss'} B_{s,p}(x) B_{s',p}(x'), \quad (8)$$

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 κ to account for the some predetermined proportion of the variance. For example, in our numerical studies below, κ 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 $\mathbf{B} = \{\mathbf{B}(1/N), \dots, \mathbf{B}(N/N)\}^\top$. Then for any $k \in \{1, \dots, \kappa\}$, 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 ψ_k and λ_k can be obtained by solving the eigenequations,

$$\int \widehat{G}(x, x') \widehat{\psi}_k(x') dx' = \widehat{\lambda}_k \widehat{\psi}_k(x), \quad k \in \{1, \dots, \kappa\}. \quad (9)$$

According to (8), solving (9) is equivalent to solving the following: $\mathbf{B}^\top(x) \widehat{\beta} \mathbf{B}^\top \mathbf{B} \widehat{\gamma}_k = \widehat{\lambda}_k \mathbf{B}^\top(x) \widehat{\gamma}_k$, $k \in \{1, \dots, \kappa\}$, where $\widehat{\beta}^\top = (\widehat{\beta}_{s,s'})_{s,s'=1}^{J_s+p}$.

By simple algebra, one needs to solve $\widehat{\beta} \mathbf{B}^\top \mathbf{B} \widehat{\gamma}_k = \widehat{\lambda}_k \widehat{\gamma}_k$, for any $k \in \{1, \dots, \kappa\}$. Consider the following Cholesky decomposition: $\mathbf{B}^\top \mathbf{B} = \mathbf{L}_B \mathbf{L}_B^\top$. Therefore, solving (9) is equivalent to solving $\widehat{\lambda}_k \mathbf{L}_B^\top \widehat{\gamma}_k = \mathbf{L}_B^\top \widehat{\beta} \mathbf{L}_B \mathbf{L}_B^\top \widehat{\gamma}_k$, that is, $\widehat{\lambda}_k$ and $\mathbf{L}_B^\top \widehat{\gamma}_k$, $k \in \{1, \dots, \kappa\}$, 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: $\widehat{\xi}_{ik} = N^{-1} \sum_{j=1}^N \widehat{\lambda}_k^{-1} \{Y_{ij} - \widehat{m}(j/N)\} \widehat{\phi}_k(j/N)$.

4.4. Estimating the variance function $\Xi(h)$ and the percentile $Q_{1-\alpha}$

Notice the fact that (7) inspires 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

$$\widehat{\Xi}(h) = \sum_{k,k'=1}^K \left\{ \frac{1}{1-h} \int_0^{1-h} \widehat{\phi}_k(x) \widehat{\phi}_{k'}(x+h) dx \right\}^2 + \widehat{C}^2(h) + \sum_{k=1}^K (\widehat{\xi}_{1k}^4 - 3) \left\{ \frac{1}{1-h} \int_0^{1-h} \widehat{\phi}_k(x) \widehat{\phi}_k(x+h) dx \right\}^2.$$

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

$$\widehat{\zeta}(h) = \sum_{k \neq k'}^K \frac{1}{1-h} \int_0^{1-h} \epsilon_{kk'} \widehat{\phi}_k(x) \widehat{\phi}_{k'}(x+h) dx + \sum_{k=1}^K \frac{1}{1-h} \int_0^{1-h} \epsilon_k \widehat{\phi}_k(x) \widehat{\phi}_k(x+h) (\widehat{\xi}_{1k}^4 - 1)^{1/2} dx,$$

where $\epsilon_{kk'}$ and ϵ_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

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

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: $Y_{ij} = m(j/N) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij}$, $j \in \{1, \dots, N\}$, $i \in \{1, \dots, n\}$, where $m(x) = \sin\{2\pi(x - 1/2)\}$, ε_{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 [14], 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 (6) 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

$$\begin{aligned} \text{AMSE}(\widehat{G}) &= \frac{1}{500N^2} \sum_{s=1}^{500} \sum_{j,j'=1}^N \left\{ \widehat{G}_s(j/N, j'/N) - G(j/N, j'/N) \right\}^2, \\ \text{AMSE}(\widehat{\lambda}) &= \frac{1}{500K} \sum_{s=1}^{500} \sum_{k=1}^K (\widehat{\lambda}_{ks} - \lambda_k)^2, \quad \text{AMSE}(\widehat{\phi}) = \frac{1}{500N_K} \sum_{s=1}^{500} \sum_{j=1}^N \sum_{k=1}^K \left\{ (\widehat{\phi}_{ks} - \phi_k)(j/N) \right\}^2, \end{aligned}$$

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 (9), 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 σ_ϵ smaller.

Table 1: Average mean squared errors (AMSEs) of $\widehat{C}, \widetilde{C}, \widehat{\lambda}, \widehat{G}, \widehat{\phi}$, coverage rates (CRs) of the proposed SCBs (outside/inside of the parentheses is based on $\widehat{C}, \widetilde{C}$) and average widths (WDs) of SCBs based on $\widehat{C}$. The standard deviation of the errors $\sigma(x) = \sigma_\epsilon = 0.1, 0.5$. The number of knots for the splines are determined by the formula (Formula) and the GCV described in Section 4.2. Results are based on 500 replications.

σ_ϵ	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

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 visualization of actual estimation, Fig. 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 Fig. 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 [14], we consider the following spatial covariance models:

- Spherical model (M1): $C(h; \sigma_s^2, \theta_s) = \sigma_s^2 \{1 - 1.5h/\theta_s + 0.5(h/\theta_s)^3\} I\{h \leq \theta_s\}$;
- 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 Γ is the gamma function, F_ν is the modified Neumann function;
- Gaussian model (M3): $C(h; \sigma_s^2, \theta_s) = \sigma_s^2 \exp(-h^2/\theta_s^2)$.

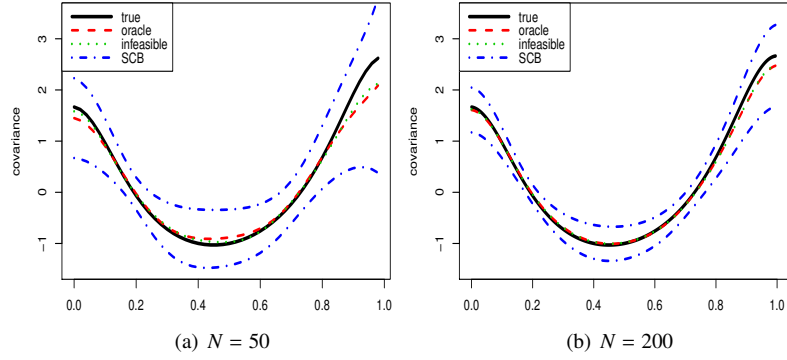


Fig. 1: Plot of true covariance function (thick solid line), oracle estimator $\widehat{C}$ (dashed line) and the 99% SCB (dotted-dashed line), infeasible estimator $\widetilde{C}$ (dotted line) for the covariance function with $\sigma_\epsilon = 0.1$.

Table 2: Average mean squared errors (AMSEs) of $\widehat{C}$, $\widetilde{C}$, $\widehat{\lambda}$, $\widehat{G}$, $\widehat{\phi}$, coverage rates (CRs) of the proposed SCB (outside/inside of the parentheses is based on $\widehat{C}$, $\widetilde{C}$) and average widths (WDs) of SCBs based on $\widehat{C}$. The heteroscedastic errors are based on standard deviation function: $\sigma(x) = \sigma_\epsilon[5 - \exp(x)]/[5 + \exp(x)]$, $\sigma_\epsilon = 0.1, 0.5$. Results are based on 500 replications.

σ_ϵ	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	0.650	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

In the parameterization (following [28] page 29) of the covariance structure, σ_s^2 is the sill and θ_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, $\rho(h) = C(h)/C(0)$, 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 $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$ for M1, and $\sigma(x) = \sigma_\epsilon \{30 + \exp(x/2)\}^{-1} \{30 - \exp(x/2)\}$ for M2 and M3. 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 σ_ϵ is smaller.

Table 3: Average mean squared errors (AMSEs) of $\widehat{C}$, $\widetilde{C}$, $\widehat{\lambda}$, $\widehat{G}$, $\widehat{\phi}$, coverage rates (CRs) of the proposed the SCB (outside/inside of the parentheses is based on $\widehat{C}$, $\widetilde{C}$) and average widths (WDs) of SCBs based on $\widehat{C}$. Results are based on data generated from models M1–M3 in Section 5.2, and 500 replications, in which the standard deviation of the errors $\sigma(x) = \sigma_\epsilon = 0.1, 0.5$.

σ_ϵ	Model	N	AMSE		SCB			
			$\widehat{C}$	$\widetilde{C}$	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

When the covariance structure is not necessarily stationary, [14] 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.

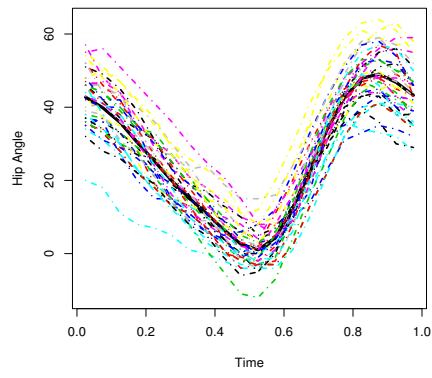
Columns 4–5 of Table 4 present the AMSEs of $\widehat{G}^{\text{PROP}}(x, x')$ and $\widehat{G}^{\text{TPS}}(x, x')$. The results of AMSEs indicate that $\widehat{G}^{\text{PROP}}$ is more accurate than $\widehat{G}^{\text{TPS}}$, while $\widehat{G}^{\text{TPS}}$ usually gives larger AMSE. Columns 6–13 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 4: Average mean squared errors (AMSEs) of $\widehat{G}^{\text{PROP}}(\cdot, \cdot)$, $\widehat{G}^{\text{TPS}}(\cdot, \cdot)$, coverage rates (CRs) and average widths (WDs) of SCE-I and SCE-II. Results are based on data generated from models M1–M3 in Section 5.2 and 500 replications, where the standard deviation of the errors $\sigma(x) = \sigma_\epsilon = 0.1, 0.5$.

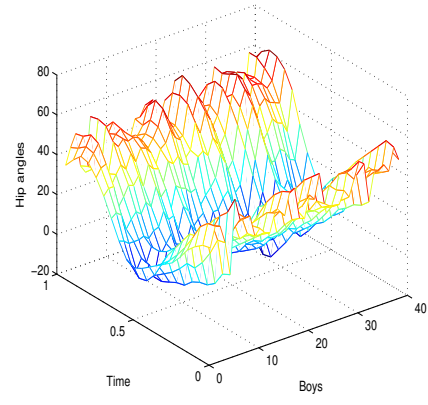
σ_ϵ	Model	N	AMSE		SCE-I				SCE-II			
			$\widehat{G}^{\text{PROP}}$	$\widehat{G}^{\text{TPS}}$	95%		99%		95%		99%	
					CR	WD	CR	WD	CR	WD	CR	WD
0.1	M1	50	0.079	0.123	0.910	1.40	0.960	1.71	0.744	2.08	0.840	2.55
		100	0.039	0.063	0.920	1.01	0.974	1.24	0.852	1.64	0.944	2.02
		200	0.018	0.031	0.946	0.73	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

6. Real data analysis

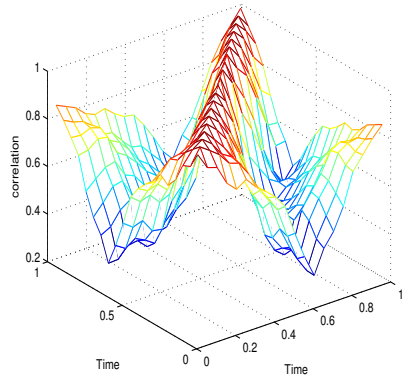
To further illustrate our methodology, we first consider the modeling of the Gait Data collected by the Motion Analysis Laboratory at the Children’s Hospital in San Diego, CA. We focus on the “Hip Angle” functional dataset, which consists of the angles formed by the hip of each boy over his gait cycle. See [29] 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 , $j \in \{1, \dots, N\}$ and $i \in \{1, \dots, n\}$. Fig. 2 (a) shows hip curves together with their estimated mean curve, and Fig. 2 (b) describes the 3D shape of all curves, where “time” is plotted on one axis and sample index on the other.



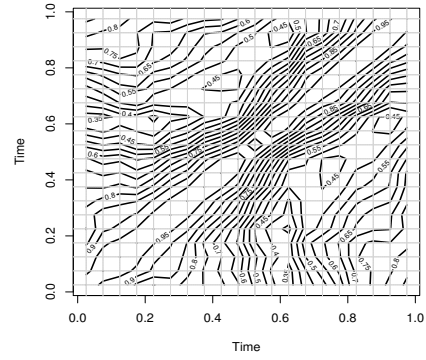
(a)



(b)



(c)



(d)

Fig. 2: (a) Hip angle data (dotted lines) with its mean function estimator (solid line); (b) 3D plot of the data; (c): unsmoothed sample correlation 3D plot for the hip angle data; (d): unsmoothed sample correlation contour plot for the hip angle data.

Fig. 2 (c) and (d) 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. Fig. 3 (a) shows a 3D plot of the proposed covariance matrix estimator $\widehat{G}^{\text{PROP}}(x, x') = \widehat{C}(|x - x'|)$ with its asymptotic SCE. For comparison, the nonstationary covariance function estimator $\widehat{G}^{\text{TPS}}$ and its SCE are also presented; see Fig. 3 (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 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:

$$\text{M1 } H_0: C(h) = C(h; 32, \theta_s) = 32\{1 - 1.5(h/\theta_s) + 0.5(h/\theta_s)^3\}I\{h \leq \theta_s\}, \quad (10)$$

$$\text{M2 } H_0: C(h) = C(h; 32, \theta_s, \nu) = 32\{\Gamma(\nu)\}^{-1}2^{1-\nu}(2\sqrt{\nu}h/\theta_s)^\nu F_\nu(2\sqrt{\nu}h/\theta_s), \quad (11)$$

$$\text{M3 } H_0: C(h) = C(h; 32, \theta_s) = 32\exp(-h^2/\theta_s^2), \quad (12)$$

where $\theta_s = 1.12$ for M1 and M2, $\nu = 1.2$ for M2 and $\theta_s = 2.19$ for M3. In Fig. 4, 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 Fig. 4 (a), one observes that even the 99% SCB cannot contain $C(h; 32, 1.12)$, hence the null hypothesis in (10) is rejected with $p\text{-value} < 0.01$. Fig. 4 (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 (11) and (12) is not rejected with $p\text{-value} > 0.2$.

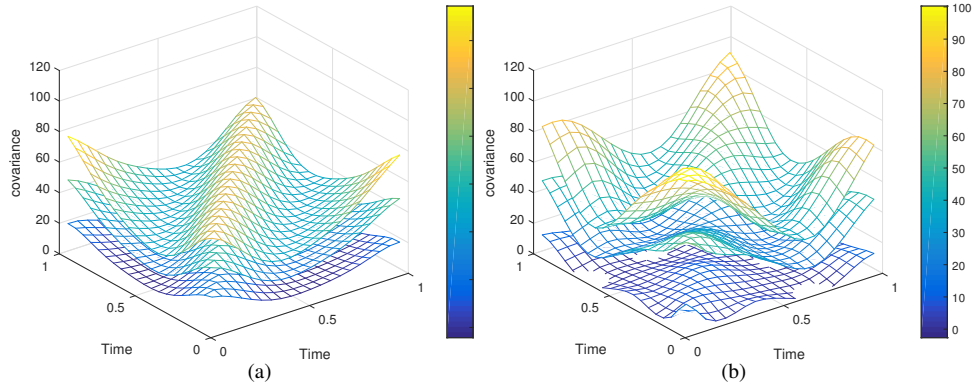


Fig. 3: (a): Covariance matrix estimator based on $\widehat{G}^{\text{PROP}}(x, x')$ (middle) with 95% SCE (up and below); (b): covariance matrix estimator $\widehat{G}^{\text{TPS}}(x, x')$ (middle) of [14] with 95% SCE (up and below).

Acknowledgment

This work is supported in part by SSEC: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 for their constructive comments and suggestions that led to significant improvement of the paper.

Appendix A

This section provides technical lemmas and detailed proofs of the main asymptotic results. Throughout this section, O_p (or o_p) denotes a sequence of random variables of certain order in probability. For instance, $o_p(n^{-1/2})$

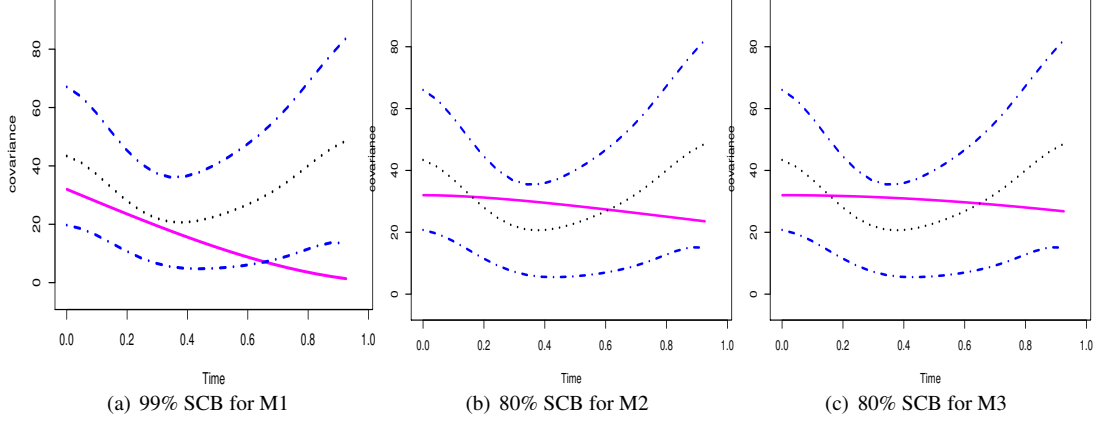


Fig. 4: Covariance function $C(h)$ under H_0 (thick solid line), B-spline covariance estimator $\widehat{C}$ (dotted line), and the SCB based on $\widehat{C}$ (dotted-dashed line) for the hip angle data.

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

For any vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathcal{R}^n$, denote the norm $\|\mathbf{a}\|_r = (|a_1|^r + \dots + |a_n|^r)^{1/r}$, $r \in \{1, 2, \dots\}$, $\|\mathbf{a}\|_\infty = \max(|a_1|, \dots, |a_n|)$. 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. Decomposition

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iN})^\top$, then the spline estimator $\widehat{\eta}_i(x)$ in (5) can be represented as $\widehat{\eta}_i(x) = \mathbf{B}(x)^\top (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{Y}_i$, where $\mathbf{B}$ is the design matrix defined in Section 4.3. Define the empirical inner product matrix of B-spline basis $\{B_{\ell,p}(x)\}_{\ell=1}^{J_s+p}$ as

$$\mathbf{V}_{n,p} = \left\{ \left\langle B_{\ell,p}, B_{\ell',p} \right\rangle_N \right\}_{\ell, \ell'=1}^{J_s+p} = N^{-1} \mathbf{B}^\top \mathbf{B},$$

and, according to Lemma A.3 in [5], for some constant $C_p > 0$, we have $\|\mathbf{V}_{n,p}^{-1}\|_\infty \leq C_p J_s$.

Denote $\boldsymbol{\eta}_i = \{\eta_i(1/N), \dots, \eta_i(N/N)\}^\top$, $\mathbf{m} = \{m(1/N), \dots, m(N/N)\}^\top$, $\mathbf{Z}_i = \{Z_i(1/N), \dots, Z_i(N/N)\}^\top$, $\boldsymbol{\varepsilon}_i = (\sigma(1/N)\varepsilon_{i1}, \dots, \sigma(N/N)\varepsilon_{iN})^\top$. According to model (2), $\boldsymbol{\eta}_i = \mathbf{m} + \mathbf{Z}_i$, then the approximation error $\widehat{\eta}_i(x) - \eta_i(x)$ can be decomposed into the following:

$$\widehat{\eta}_i(x) - \eta_i(x) = \widetilde{\eta}_i(x) - \eta_i(x) + \widetilde{\varepsilon}_i(x), \quad (\text{A1})$$

where

$$\begin{aligned} \widetilde{\eta}_i(x) &= N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \boldsymbol{\eta}_i = \widetilde{m}(x) + \widetilde{Z}_i(x), \\ \widetilde{m}(x) &= N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \mathbf{m}, \quad \widetilde{Z}_i(x) = N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \mathbf{Z}_i, \\ \widetilde{\varepsilon}_i(x) &= N^{-1} \mathbf{B}(x)^\top \mathbf{V}_{n,p}^{-1} \mathbf{B}^\top \boldsymbol{\varepsilon}_i, \end{aligned}$$

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) and (A1), the approximation error of $\widehat{Z}_i(x)$ in (4) to $Z_i(x)$ can be represented by

$$\widehat{Z}_i(x) - Z_i(x) = \widetilde{Z}_i(x) - Z_i(x) + \widetilde{\varepsilon}_i(x) - \frac{1}{n} \sum_{i'=1}^n \left\{ \widetilde{Z}_{i'}(x) + \widetilde{\varepsilon}_{i'}(x) \right\}. \quad (\text{A2})$$

In this section, we provide some technical lemmas. For the sake of saving space, we only state the lemmas and refer to [26] for the proofs of these lemmas.

Lemma A.1. *Under Assumptions (C1)–(C6), as $N \rightarrow \infty$, one has*

$$\begin{aligned} \max_{1 \leq i \leq n} \|\tilde{\eta}_i - \eta_i\|_\infty &= O_{a.s.}\{J_s^{-p^*} (n \ln n)^{2/r_1}\}, \\ \max_{1 \leq i \leq n} \|\tilde{Z}_i - Z_i\|_\infty &= O_{a.s.}\{J_s^{-p^*} (n \ln n)^{2/r_1}\}, \\ \max_{1 \leq i \leq n} \|Z_i\|_\infty &= O_{a.s.}\{(n \ln n)^{2/r_1}\}. \end{aligned}$$

Lemma A.2. *Under Assumptions (C1)–(C6), as $N \rightarrow \infty$, one has*

$$\max_{1 \leq i \leq n} \|\tilde{\varepsilon}_i\|_\infty = O_{a.s.}\{J_s^{1/2} N^{-1/2} (\ln N)^{1/2}\}.$$

Lemma A.3. *Under Assumptions (C1)–(C6), as $N \rightarrow \infty$*

$$\begin{aligned} \max_{1 \leq i \leq n} \|\hat{\eta}_i - \eta_i\|_\infty &= O_p\left\{J_s^{-p^*} (n \ln n)^{2/r_1} + J_s^{1/2} N^{-1/2} (\ln N)^{1/2}\right\}, \\ \max_{1 \leq i \leq n} \|\hat{Z}_i - Z_i\|_\infty &= O_p\left\{J_s^{-p^*} (n \ln n)^{2/r_1} + J_s^{1/2} N^{-1/2} (\ln N)^{1/2}\right\}. \end{aligned} \quad (\text{A3})$$

Lemma A.4. *Assumption (C5) holds under Assumptions (C4) and (C5').*

Lemma A.5. *Under Assumptions (C1)–(C6),*

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

Lemma A.6. *Under Assumptions (C1)–(C6),*

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

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

Lemma A.7. *Under Assumptions (C1)–(C6),*

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

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

Lemma A.8. *Under Assumptions (C1)–(C6), one has*

$$\max_{1 \leq k \leq k_n} \max_{1 \leq \ell \leq J_s+p} \left| (nN)^{-1} \sum_{i=1}^n (\xi_{ik} - U_{ik,\xi}) \sum_{j=1}^N B_{\ell,p}\left(\frac{j}{N}\right) \sigma\left(\frac{j}{N}\right) U_{ij,\varepsilon} \right| = O_{a.s.}(n^{\beta_1-1/2} N^{-1/2} J_s^{-1/2} \ln^{1/2} N),$$

where $0 < \beta_1 < 1/2$.

Lemma A.9. *Under Assumptions (C2)–(C6),*

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

335 **Lemma A.10.** Under Assumptions (C2)–(C6), $\sup_{h \in [0, h_0]} \sup_{x \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n Z_i(x+h) \widetilde{\varepsilon}_i(x) \right| = o_p(n^{-1/2})$.

Lemma A.11. Under Assumptions (C2)–(C6), one has

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

Lemma A.12 (Theorem 1.2 of [30]). Let $\zeta_1, \dots, \zeta_N$ be independent real valued random variables with $E\zeta_j = 0, E\zeta_j^2 = \sigma_j^2 < +\infty, j \in \{1, \dots, N\}$ and let $S_N = \sum_{j=1}^N \zeta_j, V_N^2 = \sum_{j=1}^N \sigma_j^2$. If there exists $c > 0$ such that for $r \geq 3$, $E|\zeta_j|^r \leq c^{r-2} r! \sigma_j^2 < +\infty, j \in \{1, \dots, N\}$, then for each $N > 1, t > 0$,

$$\Pr \left(\left| \frac{S_N}{V_N} \right| \geq t \right) \leq 2 \exp \left\{ -\frac{t^2}{4} \frac{1}{1 + ctV_N^{-1}/2} \right\}.$$

340 **Lemma A.13** (Theorem 2.6.7 of [31]). Suppose that $\xi_i, 1 \leq i \leq n$ are i.i.d with $E(\xi_1) = 0, E(\xi_1^2) = 1$ and $H(x) > 0$ ($x \geq 0$) is an increasing continuous function such that $x^{-2-\gamma}H(x)$ is increasing for some $\gamma > 0$ and $x^{-1} \ln H(x)$ is decreasing with $EH(|\xi_1|) < \infty$. Then there exist constants $C_1, C_2, a > 0$ which depend only on the distribution of ξ_1 and a sequence of Brownian motions $\{W_n(m)\}_{n=1}^\infty$, such that for any $\{x_n\}_{n=1}^\infty$ satisfying $H^{-1}(n) < x_n < C_1 (n \ln n)^{1/2}$ and $S_m = \sum_{i=1}^m \xi_i$, then $\Pr \{ \max_{1 \leq m \leq n} |S_m - W_n(m)| > x_n \} \leq C_2 n \{H(ax_n)\}^{-1}$.

Lemma A.14. Let $W_i \sim N(0, \sigma_i^2), \sigma_i > 0, i \in \{1, \dots, n\}$, for $a > 2$

$$\Pr \left(\max_{1 \leq i \leq n} |W_i/\sigma_i| > a \sqrt{\ln n} \right) < \sqrt{\pi/2} n^{1-a^2/2}.$$

Hence, $(\max_{1 \leq i \leq n} |W_i|) / (\max_{1 \leq i \leq n} \sigma_i) \leq \max_{1 \leq i \leq n} |W_i/\sigma_i| = O_{a.s.}(\sqrt{\ln n})$.

A.3. 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 σ -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 $t \in \{1, \dots, k_n\}$, where k_n satisfies Assumption (C4). 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 ,

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

345 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

$$E(D_t^2(h) | \mathcal{F}_{t-1}) = V_t^{(1)}(h) + V_t^{(2)}(h) + V_t^{(3)}(h), \quad (\text{A4})$$

in which

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

Moreover, one can show that

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

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

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

Note that

$$V_t^{(2)}(h) = (E \xi_{1t}^4 - 1) \left\{ (1-h)^{-1} \int_0^{1-h} \phi_t(x)\phi_t(x+h) dx \right\}^2 < \infty,$$

so one has that

$$\sum_{t=2}^{k_n} V_t^{(2)}(h) \rightarrow \sum_{k=1}^{\infty} (E \xi_{1k}^4 - 1) \left\{ (1-h)^{-1} \int_0^{1-h} \phi_k(x)\phi_k(x+h) dx \right\}^2 < \infty.$$

Similarly,

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

Thus,

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

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

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

Therefore, one has

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

as $n \rightarrow \infty$.

According to (A4), as $n \rightarrow \infty$, one has

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

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

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

350 Applying similar arguments in Lemma 6 of [14], 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 $\epsilon > 0$, $\sup_{h \in [0, h_0]} \sum_{t=2}^{k_n} \mathbb{E}\left\{D_t^3(h)I\left(D_t^2(h) > \epsilon\right)\middle|\mathcal{F}_{t-1}\right\} \rightarrow_p 0$.

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

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

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

A.4. 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

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

Note that by (A2), $\sup_{h \in [0, h_0]} |\text{I}(h)| \leq \max_{1 \leq i \leq n} \|\widehat{Z}_i - Z_i\|_{\infty}^2$. According to (A3),

$$\max_{1 \leq i \leq n} \|\widehat{Z}_i - Z_i\|_{\infty} = O_p \left\{ J_s^{-p^*} n \ln n + J_s^{1/2} N^{-1/2} (\ln N)^{1/2} \right\},$$

By (A2), 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

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

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

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

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

$$\sup_{h \in [0, h_0]} |\widehat{C}(h) - \widetilde{C}(h)| = \sup_{h \in [0, h_0]} |\text{I}(h) + \Pi(h) + \text{III}(h)| = o_p(n^{-1/2}).$$

Appendix B

355 This section presents more results and findings from additional simulation examples.

B.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. The setting of the simulation is the same as in Section 5.1. 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.

Fig. B.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 Fig. B.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. 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 minuses for GCV selected method, respectively.

B.2. More results for spatial covariance models

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

375 The AMSE of the covariance estimators $\widehat{C}$ and $\widetilde{C}$ are reported in columns 3–4 of Table B.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 B.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.

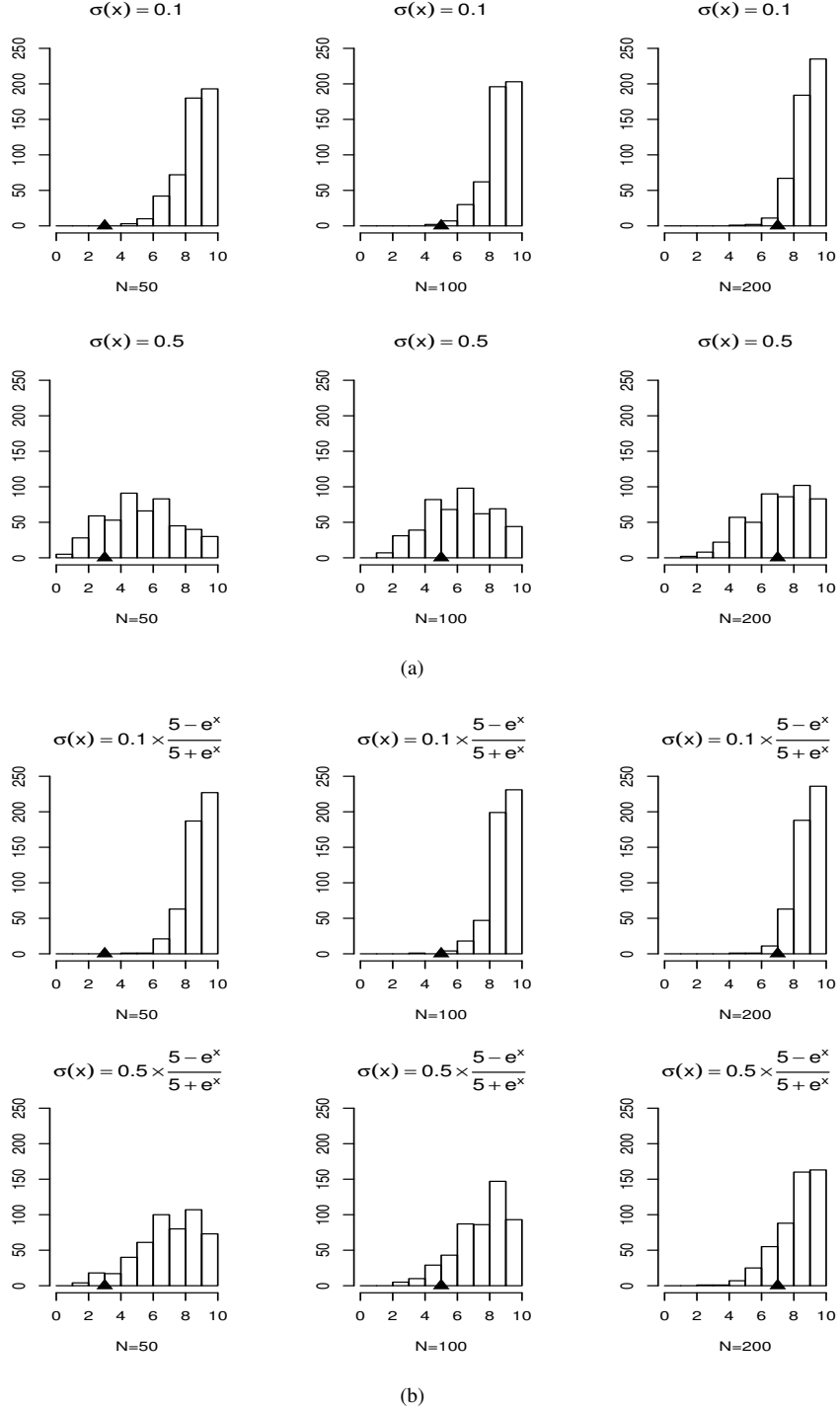


Fig. B.1: The histogram of the GCV selected number of knots in 500 replications with (a) homogeneous and (b) heteroscedastic errors. Black triangles indicate the number of knots suggested by the formula. The GCV selection method is given in Section 4.2.

Table B.1: Average mean squared errors (AMSEs) of $\widehat{C}$, $\widetilde{C}$, coverage rates (CRs) (outside/inside of the parentheses is based on $\widehat{C}$, $\widetilde{C}$), and average widths (WDs) of the asymptotic SCBs based on $\widehat{C}$. The standard deviation function for the heteroscedastic errors $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$ for M1, and $\sigma(x) = \sigma_\epsilon \{30 + \exp(x/2)\}^{-1} \{30 - \exp(x/2)\}$ for M2 and M3. The details of models M1, M2 and M3 are given in Section 5.2. Results are based on 500 replications.

σ_ϵ	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

Table B.2: Average mean squared errors (AMSEs) of $\widehat{G}^{\text{PROP}}(\cdot, \cdot)$, $\widehat{G}^{\text{TPS}}(\cdot, \cdot)$, coverage rates (CRs) and average widths (WDs) of SCE-I and SCE-II. The standard deviation function of the heteroscedastic errors $\sigma(x) = \sigma_\epsilon \{5 + \exp(x)\}^{-1} \{5 - \exp(x)\}$ for M1, and $\sigma(x) = \sigma_\epsilon \{30 + \exp(x/2)\}^{-1} \{30 - \exp(x/2)\}$ for M2 and M3. Details of models M1, M2 and M3 are given in Section 5.2. Results are based on 500 replications.

σ_ϵ	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

References

- [1] J. Ramsay, C. Dalzell, Some tools for functional data analysis, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 53 (1991) 539–572.
- [2] J. Ramsay, B. Silverman, *Functional Data Analysis*, Springer, New York, 2005.
- [3] T. Hsing, R. Eubank, *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear operators*. Wiley Series in Probability and Statistics, Wiley, Chichester, 2015.
- [4] H. Cardot, Nonparametric estimation of smoothed principal components analysis of sampled noisy functions, *J. Nonparametr. Stat.* 12 (2000) 503–538.
- [5] G. Cao, L. Yang, D. Todem, Simultaneous inference for the mean function based on dense functional data, *J. Nonparametr. Stat.* 24 (2012) 359–377.
- [6] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, Springer, Berlin, 2006.
- [7] J. Goldsmith, S. Greven, C. Crainiceanu, Corrected confidence bands for functional data using principal components, *Biometrics* 69 (1) (2013) 41–51.
- [8] P. Hall, H. G. Müller, J. L. Wang, Properties of principal component methods for functional and longitudinal data analysis, *Ann. Statist.* 34 (3) (2006) 1493–1517.
- [9] I. Choi, B. Li, X. Wang, Nonparametric estimation of spatial and space-time covariance function, *J. Agric. Biol. Environ. Stat.* 18 (2013) 611–630.
- [10] P. Diggle, A. Verbyla, Nonparametric estimation of covariance structure in longitudinal data, *Biometrics* 54 (1998) 401–415.
- [11] P. Hall, N. I. Fisher, B. Hoffmann, On the nonparametric estimation of covariance functions, *Ann. Statist.* 22 (4) (1994) 2115–2134.
- [12] J. Yang, H. Zhu, T. Choi, D. D. Cox, Smoothing and mean–covariance estimation of functional data with a bayesian hierarchical model, *Bayesian Anal.* 11 (2016) 649–670.
- [13] J. Yin, Z. Geng, R. Li, H. Wang, Nonparametric covariance model, *Statist. Sinica* 20 (2010) 469–479.
- [14] G. Cao, L. Wang, Y. Li, L. Yang, Oracle-efficient confidence envelopes for covariance functions in dense functional data, *Statist. Sinica* 26 (2016) 359–383.
- [15] L. Horváth, P. Kokoszka, R. Reeder, Estimation of the mean of functional time series and a two-sample problem, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 75 (2013) 103–122.
- [16] U. Pantle, V. Schmidt, E. Spodarev, On the estimation of integrated covariance functions of stationary random fields, *Scand. J. Statist.* 37 (2010) 47–66.
- [17] Y. Li, T. Hsing, Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, *Ann. Statist.* 38 (2010) 3321–3351.
- [18] C. M. Crainiceanu, A. M. Staicu, C. Z. Di, Generalized multilevel functional regression, *J. Amer. Statist. Assoc.* 104 (2009) 1550–1561.
- [19] Fryzlewicz, Ombao, Consistent classification of nonstationary time series using stochastic wavelet representations, *J. Amer. Statist. Assoc.* 104 (2009) 299–312.
- [20] J. Sanderson, P. Fryzlewicz, M. W. Jones, Estimating linear dependence between nonstationary time series using the locally stationary wavelet model, *Biometrika* 97 (2010) 435–446.
- [21] M. Tsyrlunikov, D. Gayfulin, A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications, *Meteorologische Zeitschrift* 5 (2017) 549–566.
- [22] F. Yao, H. G. Müller, J. L. Wang, Functional data analysis for sparse longitudinal data, *J. Amer. Statist. Assoc.* 100 (470) (2005) 577–590.
- [23] G. James, T. Hastie, C. Sugar, Principal component models for sparse functional data, *Biometrika* 87 (3) (2000) 587–602.
- [24] J. Peng, D. Paul, A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data, *Journal of Computational and Graphical Statistics* 18 (4) (2009) 995–1015.
- [25] J. Guo, B. Zhou, J. T. Zhang, Testing the equality of several covariance functions for functional data: A supremum-norm based test, *Comput. Statist. Data Anal.* 124 (2018) 15–26.
- [26] J. Wang, G. Cao, L. Wang, L. Yang, Simultaneous confidence band for stationary covariance function of dense functional data, *arXiv* 1903.05522 (2019).
- [27] C. de Boor, *A Practical Guide to Splines*, Springer, New York, 2001.
- [28] S. Banerjee, B. Carlin, A. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC Press, 2004.
- [29] R. A. Olshen, E. N. Bide, M. P. Wyatt, D. H. Sutherland, Gait analysis and the bootstrap, *Ann. Statist.* 17 (1989) 1419–1440.
- [30] D. Bosq, *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Vol. 110, Springer-Verlag New York, 1998.
- [31] M. Csörgö, P. Révész, *Strong Approximations in Probability and Statistics*, Academic Press, New York-London, 1981.